

Editor-in-Chief

**Luciano Martini**

University of Milan  
Milan, Italy

## Associate Editors

**Eli Y. Adashi**

University of Utah School of Medicine  
Salt Lake City, Utah, United States

**John P. Bilezikian**

Columbia University  
New York, New York, United States

**George P. Chrousos**

National Institute of Child Health  
and Human Development  
Bethesda, Maryland, United States

**Junichi Fukata**

Kochi Women's University  
Kochi City, Kochi, Japan

**Julianne Imperato-McGinley**

Weill Medical College of Cornell University  
New York, New York, United States

**Hiroo Imura**

Prime Minister's Office,  
Government of Japan  
Tokyo, Japan

**Antonio Liuzzi**

Istituto Auxologico Italiano  
Verbania, Italy

**Frank L. Moore**

Oregon State University  
Corvallis, Oregon, United States

**Daniel Porte**

VA San Diego Health Care System  
San Diego, California, United States

**Jens F. Rehfeld**

Copenhagen University Hospital  
Copenhagen, Denmark

**Ellis Samols**

University of Nevada,  
Las Vegas School of Medicine  
Las Vegas, Nevada, United States

**Gustav Schonfeld**

Washington University,  
St. Louis School of Medicine  
St. Louis, Missouri, United States

**Pierre C. Sizonenko**

University of Geneva, Hôpital la Tour  
Geneva, Switzerland

**Wilmar M. Wiersinga**

University of Amsterdam  
Amsterdam, The Netherlands



The field of endocrinology is inextricably linked to physiology. The specialty was initially founded when it became clear that various glands produced hormones that exerted characteristic physiologic effects on growth, reproduction, and metabolism. Early descriptions of hormone deficiency syndromes such as Addison's disease and myxedema were soon followed by hormone replacement strategies, often resulting in dramatic clinical effects. These observations unleashed intensive efforts to isolate and characterize the steroid and peptide hormones produced by the adrenal, thyroid, parathyroid, pituitary, and pancreatic islets, and other glands. The success of this era was epitomized by the isolation of insulin and the successful treatment of children with type 1 diabetes mellitus in 1922. The development of radioimmunoassays (RIAs) was a monumental advance that allowed hormones to be measured in various physiologic conditions. RIAs transformed endocrinology more than any other field. The ability to measure hormone levels during stimulation and suppression tests firmly established the principles of feedback regulation and formed the basis for many current diagnostic algorithms. RIAs also revealed the natural patterns of hormone secretion, including circadian rhythms and reproductive cycles, as well as hormonal responses to sleep, meals, stress, exercise, and other daily life events. Our understanding of hormone action has been accelerated by studies of their membrane and nuclear receptors, which convey hormone specificity in target tissues. The signaling pathways elicited by these receptors constitute intricate and complex networks that inform the cell about its external environment. Recombinant DNA technology has been essential for cloning the genes and cDNAs that encode large families of hormones and receptors. Growth hormone, chorionic gonadotropin, and

somatostatin were among the first mammalian cDNAs to be cloned. With completion of the human genome project, all of the genes that encode hormones and their receptors have, in principle, been identified. However, many of these receptors remain "orphans" with still-unknown ligands and incompletely defined functions. Not surprisingly, genetic advances have revealed remarkable insight into inherited endocrine disorders.

Most physicians are attracted to the field of endocrinology because it so beautifully integrates physiology, biochemistry, and cell signaling with patient care. Clinical manifestations of endocrine disorders can usually be explained by understanding the physiologic role of hormones—whether deficient or excessive. The conceptual framework for understanding hormone secretion, hormone action, and principles of feedback control provides the clinician with a logical diagnostic approach that typically employs appropriate laboratory testing and/or imaging studies. The fact that many endocrine disorders are amenable to cure or effective treatment also makes the practice of endocrinology especially satisfying. Because most glands are inaccessible to physical examination, endocrinologists are trained to detect key features of the medical history and subtle physical signs that point toward true endocrine disease. Increasingly, the challenge is to identify endocrine disorders at their earliest stages rather than when the clinical manifestations are obvious. Terms such as subclinical hypothyroidism, impaired glucose tolerance, and incidental adrenal or pituitary adenoma have crept into our vocabulary and have changed our approach to patients. Laboratory testing takes on added importance as we attempt to diagnose more subtle forms of disease.

Building on this strong foundation of basic science, the knowledge base in endocrinology continues to change rapidly. In addition to the dramatic advances

generated from genetics and molecular biology, the field has benefited from the introduction of an unprecedented number of new drugs, particularly for the management of diabetes and osteoporosis. Common diseases such as diabetes, hypertension, obesity, and osteoporosis have also been the subject of numerous large-scale clinical trials that provide a powerful evidence base for medical decision-making.

The rapid changes in medicine mandate that physicians continuously update their knowledge base and clinical skills. The *Encyclopedia of Endocrine Diseases* recognizes this challenge and provides a remarkable compilation of current knowledge in basic and clinical endocrinology. This ambitious four-volume set provides nearly 500 articles on basic and clinical endocrinology. The topics range from classic endocrine subjects such as hypothyroidism and acromegaly to new dimensions of the field including adipocytokines, ghrelin, and the role of the aldosterone receptor in cardiovascular disease. The international group of authors are experts in their topics, which have been subdivided to provide in-depth coverage. Thus, acromegaly is separated into articles on clinical features, diagnosis, and therapy to provide the level of detail needed to manage the most challenging cases. The standardized format and clear illustrations help to quickly offer answers. It is difficult to imagine an endocrine topic not covered in this encyclopedia, which provides a new, comprehensive reference for the daunting body of knowledge in endocrinology.

With a four-volume *Encyclopedia of Endocrine Diseases* in hand, it is interesting to speculate about

the remaining big questions and future discoveries in endocrinology. What are the genetic and adaptive elements that cause such a broad normal range for hormone values? It is humbling to recognize that we still have an incomplete understanding of the hormonal control of fundamental processes such as the onset of puberty, appetite control, gonadal differentiation into testes or ovaries, islet cell regeneration, insulin resistance, and causes of autoimmune endocrine disease. We still have much to learn about the optimal way to deliver many hormone therapies to mimic normal physiology. This topic is prominent in our efforts to provide intensive insulin replacement, or to replace growth hormone or cortisol, such that the beneficial effects outweigh complications. New therapies, such as intermittent PTH for osteoporosis, will be subjected to additional clinical trials, and one can easily imagine emerging questions about how to cycle the therapy and how to use it in relation to other treatments that alter osteoblast and osteoclast function. Gene transfer and stem cell strategies provide promising treatments for disorders such as diabetes and osteoporosis. The *Encyclopedia of Endocrine Diseases* can help to foster these discoveries by keeping researchers and clinicians at the cutting edge of endocrinology.

J. Larry Jameson, M.D., Ph.D.  
Irving S. Cutter Professor and Chairman  
Department of Medicine  
Feinberg School of Medicine  
Northwestern University  
Chicago, Illinois, United States



Among the most complex constructs in the body, the endocrine system comprises a group of glands that secrete hormones directly into the bloodstream, together with the receptors for these hormones and the intracellular signaling pathways they invoke. The endocrine system maintains and regulates stable functioning by using hormones to control metabolism, temperature, biological cycles, internal fluid volume, reproduction, growth, and development. Fortunately, the system is a marvel when functioning optimally. However, the ways in which its processes, actions, and functions may go awry are myriad.

The *Encyclopedia of Endocrine Diseases* is not meant as a primer on the subject of endocrinology, but instead is intended to provide a comprehensive reference work on the extensive spectrum of diseases and disorders that can occur within the endocrine system. This groundbreaking encyclopedia is especially timely, as there have been dramatic discoveries in the field of endocrinology over the past 10 to 20 years, particularly with respect to diagnosis techniques and treatment methods. Indeed, during the time since the encyclopedia was conceived, new hormones have been named.

To bring a major reference work of such broad scope from initial conception to final publication involved a great deal of planning, staging, and organization, together with the efforts of innumerable individuals. At the start, the broadest possible list of topics was compiled and a distinguished multinational panel of 14 associate editors was assembled. Throughout the editorial process, the editors supervised their subject area of expertise, recommended and corresponded with article contributors, reviewed the subsequent manuscripts, and continuously helped to refine the topics list.

The encyclopedia is intended to serve as a useful and comprehensive source of information spanning

the many and varied aspects of the endocrine system. It consists of nearly 500 topics explored by some 800 eminent clinicians and scientists from around the world, a veritable who's who of endocrine research. Here the interested reader can find articles on newly discovered hormones such as ghrelin and leptin; articles about such maladies as hypertension, hypoglycemia, diabetes, cancer, osteoporosis, kidney stones, Graves' disease, Paget's disease, Alzheimer's disease, Noonan syndrome, Langerhans cell disease, Cushing's syndrome, thyroid and pituitary disorders; and articles dealing with subjects ranging from the evolution of the endocrine systems, the mechanisms of hormone action, and the endocrine failure in aging to the integration between the nervous and the endocrine systems.

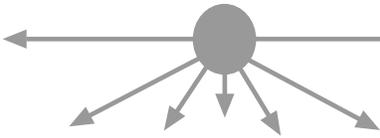
Written to be accessible to both the clinical and nonclinical reader, all of the articles are formatted in similar fashion and each is intended as a stand-alone presentation. Beginning each article is a glossary list defining key terms that may be unfamiliar to the reader and are important to an understanding of the article. The body of the article begins with a brief introduction to the subject under discussion, bold headings lead the reader through the text, and figures and tables explain and illuminate most articles. Following the article are reference citations to provide the reader with access to further in-depth considerations of the topic and cross-references to related entries in the encyclopedia. A compilation of all glossary terms appearing in the complete four-volume work is presented in the final volume as a dictionary of subject matter relevant to the endocrine system and its disorders.

It is my hope that the *Encyclopedia of Endocrine Diseases* proves to be a valuable resource to a deservedly diverse readership, and particularly to students, many of whom it may well attract to the rewarding field of

endocrinology. The project would not have been possible without cooperation, coordination, and reliance on e-mail among the key people, who were located in Japan, The Netherlands, Denmark, Switzerland, Italy, and the United States. I am greatly indebted to the dedicated and unstinting efforts of my associate editors, as well as the diligence and generosity of spirit of the Elsevier/Academic Press personnel who shepherded the project: Tari Paschall, Chris Morris, Carolan Gladden, and Joanna Dinsmore. To all of our

contributors go profound thanks for investing time and energy to produce their articles, which together have made the encyclopedia.

*Luciano Martini  
Editor-in-Chief  
Professor of Endocrinology  
University of Milan  
Milan, Italy*



# A

## ABCA1 Defects

Gerd Schmitz and Wolfgang Drobnik  
University of Regensburg, Regensburg, Germany

### Glossary

**ATP-binding cassette (ABC) transporters** A family of multispan transmembrane proteins that mediate the active uptake or efflux of specific substrates across biological membrane systems.

**high-density lipoproteins** Plasma lipoprotein subclass mediating reverse cholesterol transport.

**lipid microdomain** Plasma membrane compartments with a specific lipid composition characterized by a distinct physical state and insolubility in certain nonionic detergents.

**ABCA1** belongs to the family of ATP-binding cassette transporters, which represent one of the biggest multispan membrane protein families described by the early 21st century.

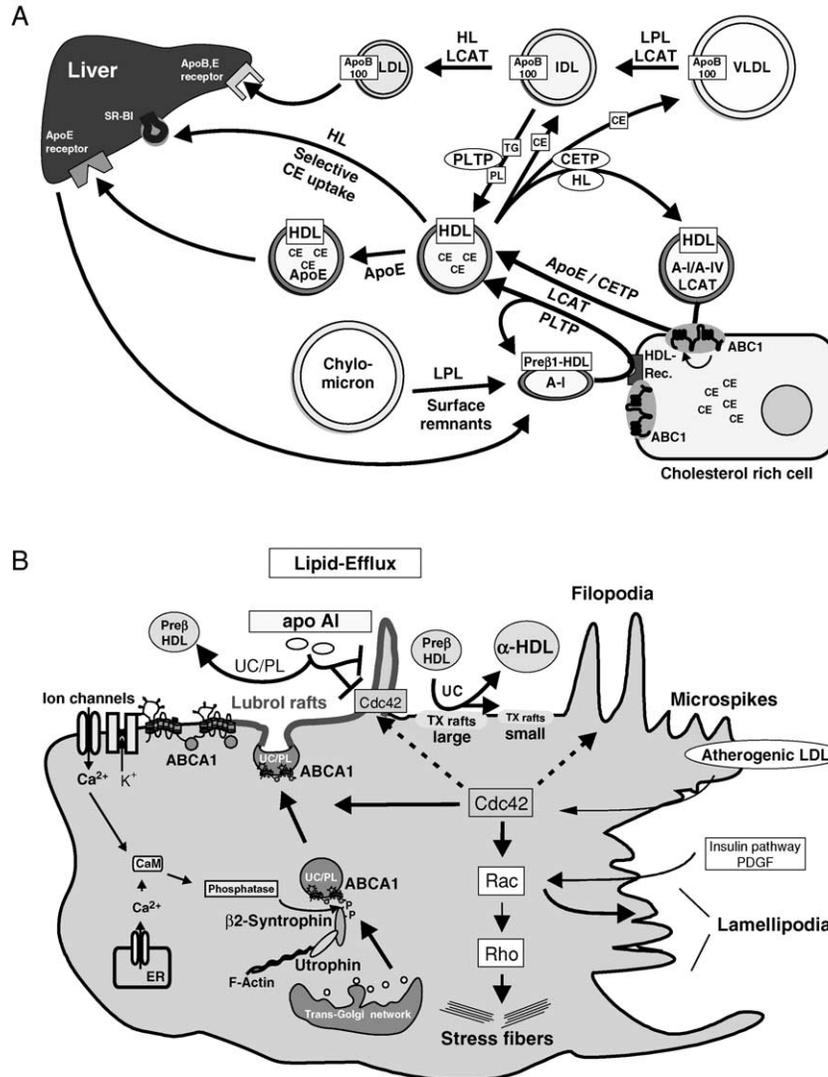
### INTRODUCTION

ATP-binding cassette (ABC) transporters have attracted much attention because mutations in these molecules are the cause of various human inherited diseases. Functional ABC transporters usually consist of two transmembrane domains (TMDs) and two nucleotide-binding domains (NBDs) or ABCs that are present either in one polypeptide chain (full-size transporter) or in two polypeptides (half-size transporter). A signature motif located between both ABCs is characteristic of each of the seven ABC subfamilies (ABCA–ABCG) described by the early 21st century. The majority of these proteins mediate the active uptake or efflux of specific substrates across various biological membrane systems, whereby two different groups of ABC proteins can be distinguished by their mode of action. The one group of ABC transporters (e.g., ABCB1 (MDR1), ABCC1 (MRP1)) has strong ATPase activity, and the resulting free energy is directly coupled to the movement of molecules across membranes, whereas the other group (e.g., ABCC7 (CFTR), ABCC9 (SUR2), ABCA1) is characterized by very low

ATP hydrolysis but conformational change subsequent to ATP binding, which is linked to regulatory processes rather than direct transport-pump activity. A multitude of substrates is transported by the various ABC family members, including glutathione, glucuronate, or sulphate conjugates; xenobiotics; peptides; nucleotides; ions; and various lipid species. Members of the ABCA family especially are involved in the transport of steroids and various phospholipid and sphingolipid species, and the transcriptional control of at least seven ABCA members is controlled or influenced by lipids. These data indicate an important role of the whole ABCA subfamily in cellular lipid transport processes.

### ABCA1 DEFECTS

ABC defects cause high-density lipoprotein (HDL) deficiency syndromes. Lipid-rich  $\alpha$ -HDL originates from lipid-poor discoidal apolipoprotein–phospholipid complexes, with apo AI being the main apolipoprotein (Fig. 1A). Apo AI binds to cells and promotes vesicle transport and exocytosis of cholesterol and phospholipids. Cholesterol acquisition followed by lecithin:cholesterol-acyltransferase (LCAT)-mediated esterification results in the formation of lipid-rich spherical  $\alpha$ -HDL. The cholesteryl esters associated with these mature HDL particles could be removed from the circulation by a scavenger receptor BI-mediated selective uptake into hepatocytes. Disorders of HDL metabolism could result from mutations in various genes along this metabolic pathway, including LCAT, apo AI, and ABCA1. ABCA1 is a 2261 amino acid protein with a molecular weight of 220 kDa that is expressed in a multitude of human organs, including liver, adrenal tissues, placenta, and spleen. Mutations in the ABCA1 gene have been identified as a cause for Tangier disease and other HDL deficiencies. Tangier disease is a rare inherited disorder of HDL metabolism. It was initially described by D. S. Fredrickson in 1961 and has since been diagnosed in at least 70 patients from 60 families. The patients are characterized by a complete deficiency of  $\alpha$ -HDL and a severe



**Figure 1** (A) High-density lipoprotein metabolism. (B) ABCA1: A regulator of lipid rafts, vesicular transport, and filipodia formation.

reduction of apo AI to 1 to 3% of normal, accompanied by low plasma cholesterol and normal or elevated triglycerides. The main clinical signs include the accumulation of cholesteryl esters in various tissues, hyperplastic orange tonsils, splenomegaly, and relapsing neuropathy. In addition, some Tangier patients have premature coronary artery disease (CAD), whereas others, even those over 60 years of age, are without any clinical symptoms of CAD. The clinical phenotype of Tangier disease and the biochemical features (e.g., low HDL) are inherited in an autosomal recessive mode and a co-dominant mode, respectively. In 1998, the genetic defect in Tangier disease was confined to chromosome 9q31, followed by the

demonstration that ABCA1 that is contained within this candidate region is subject to sterol-dependent regulation. Subsequent studies have shown that homozygous mutations in ABCA1 are the underlying defect in Tangier disease, whereas heterozygous mutations are found in patients with the more frequent and less severe familial HDL deficiency, which is inherited in a dominant mode and lacks clinical features of Tangier disease (i.e., at least a subgroup of familial HDL deficiency patients are Tangier heterozygotes). The essential role of ABCA1 in the regulation of HDL metabolism was further supported by demonstrating that targeted disruption of the ABCA1 gene in mice produces a phenotype similar to human Tangier

disease, whereas overexpression of ABCA1 increased HDL cholesterol levels. ABCA1-deficient cells are characterized by a complete loss of apo AI-mediated cholesterol and phospholipid efflux, indicating that ABCA1 is critically involved in the first step of reverse cholesterol transport. Without the initial ABCA1-dependent lipidation, apo AI undergoes rapid degradation, explaining the extremely low apo AI levels in Tangier patients and the increased catabolism of apo AI and HDL in Tangier patients infused with radiolabeled normal HDL. In BAC transgenic mice overexpressing ABCA1, a highly significant correlation was observed between the increase in cholesterol efflux from various tissues and the elevation of plasma HDL. Similarly, in patients with heterozygote ABCA1 mutations, the reductions in cholesterol efflux and plasma HDL were closely correlated. Together, these data provided direct evidence that ABCA1-dependent cholesterol efflux is a major determinant of plasma HDL cholesterol levels.

### ABCA1: A REGULATOR OF APO AI-MEDIATED LIPID EFFLUX

The exact cellular processes facilitating and regulating ABCA1-dependent lipid efflux are the subject of intense investigation. Data suggest that apo AI interacts either directly with ABCA1 or with lipid domains in close proximity to ABCA1. This interaction stimulates a Golgi- and energy-dependent vesicular transport process, resulting in the translocation of intracellular cholesterol and phospholipids to sites accessible to the apolipoprotein. Data also suggest that this apo AI-mediated lipid efflux is a two-step mechanism, with an initial ABCA1-dependent efflux of phospholipids and a subsequent ABCA1-independent efflux of cholesterol to the newly formed apo AI-phospholipid complex. Moreover, it has been shown that ABCA1 rapidly recycles between the cell surface and the intracellular compartments, although it is currently unclear whether this recycling is involved in the lipid transport process or instead regulates synthesis and degradation of ABCA1. Regarding the direct function of ABCA1, it was initially assumed that ABCA1 functions as an active pump translocating cholesterol from the inner leaflet of the plasma membrane to the outer leaflet, where it is accessible to the uptake by apo AI. However, in contrast to ABC transporters that exert bona fide pump function (e.g., MDR-1), ABCA1, similar to the ABC regulator proteins CFTR and SUR, shows only marginal intrinsic ATPase activity. Thus, ABCA1 may act as a transport

facilitator rather than as an active pump. Interestingly, both the C terminus of CFTR and ABCA1 contain a PDZ domain-binding sequence. By using the yeast-two-hybrid system, we demonstrated a direct interaction of ABCA1 with the PDZ domain-containing protein  $\beta$ 2-syntrophin. Immunoprecipitation confirmed these results and identified utrophin as an ABCA1 interaction partner. In analogy to the function of  $\beta$ 2-syntrophin-utrophin complexes in anchoring insulin-containing secretory granules, it is tempting to speculate that the interaction of ABCA1 with  $\beta$ 2-syntrophin-utrophin regulates the availability of ABCA1 at the cell surface. Additional ABCA1 interacting proteins include components of t-SNARE complexes, which are involved in targeted vesicle transport and are also known to interact with the N terminus of CFTR.

### REGULATION OF ABCA1 EXPRESSION AND FUNCTION

Several factors have been shown to control the expression of ABCA1. Since the initial finding that cholesterol influx into the cell potentially induces ABCA1 expression, a number of transcriptional control elements have been characterized. Tissue-specific regulation of ABCA1 is controlled by the transcription factors Sp1/3, USF1/2, and HNF-1 $\alpha$ , and considerable attention has been paid to nuclear liver X receptors (LXR) as inducers of ABCA1 expression in response to lipid loading. In addition, the zinc finger protein ZNF202 appears to function as a major repressor of ABCA1 transcription, and the oncostatin M-induced ABCA1 transcription provides a new concept for how members of the IL6 family of cytokines may regulate lipid transport proteins. Additional regulators of the ABCA1-dependent lipid efflux pathway include cAMP, phospholipase C, phospholipase D, and bioactive sphingolipids such as ceramide, sphingosine, and sphingosine-1-phosphate. The effects of these signaling pathways are probably cell type dependent, and further work is necessary to determine the exact mechanisms by which they control ABCA1-dependent cell function.

### ABCA1 AND SUSCEPTIBILITY TO ATHEROSCLEROSIS

Considering the known reverse relationship between HDL cholesterol levels and the risk of premature CAD, several groups have investigated the role of ABCA1 in atherogenesis. Thus, it has been reported

that homozygote and heterozygote mutations in ABCA1 are associated with an increased prevalence of premature CAD that correlates to the reduction in HDL cholesterol. Furthermore, it has been demonstrated that diet-induced development of atherosclerotic lesions is significantly reduced in transgenic mice overexpressing ABCA1. In contrast, complete inactivation of ABCA1 in apo E<sup>-/-</sup> and LDL<sup>-/-</sup> mice had no effect on the development of atherosclerotic lesions, although it markedly reduced HDL cholesterol. It was suggested that the proposed atherogenic effect of complete ABCA1 deficiency may be compensated by a less atherogenic lipid profile, a hypothesis that may also partially explain the lack of premature atherosclerosis in a significant number of Tangier patients. Importantly, two independent studies showed that targeted disruption of ABCA1 in leukocytes of LDL<sup>-/-</sup> or apo E<sup>-/-</sup> mice resulted in the development of more advanced atherosclerotic lesions without significantly affecting HDL levels. Together, these data indicate that ABCA1 clearly serves an anti-atherogenic function, although this may involve properties of ABCA1 that are independent of plasma lipids and HDL levels.

Several factors may account for the protective effect of ABCA1 in atherogenesis. First, ABCA1-mediated cholesterol efflux may significantly compensate excessive cholesterol uptake by macrophages in the vessel wall without significantly influencing plasma HDL levels. Second, ABCA1 has been implicated in the engulfment of apoptotic cells by macrophages. Thus, it is conceivable that the ABCA1-mediated phagocytic activity of lesion macrophages may counteract excessive accumulation of apoptotic material that, in return, may stimulate the inflammatory response within the vascular wall. Finally, we have previously hypothesized that ABCA1 function regulates the differentiation, lineage commitment (phagocytic vs dendritic cells), and targeting of monocytes into the vascular wall or the RES. This concept has been substantiated by recent work from our laboratory demonstrating accumulation of macrophages in liver and spleen in LDL receptor-deficient mouse chimeras that selectively lack ABCA1 in their blood cells. The fact that the absence of ABCA1 from leukocytes is sufficient to induce aberrant monocyte recruitment into specific tissues identifies ABCA1 as a critical leukocyte factor in the control of monocyte targeting. An interesting clue as to how ABCA1 may be implicated in the control of monocyte/macrophage trafficking at the cellular level comes from the observation that apo AI-mediated lipid efflux is paralleled by the down-regulation of the protein Cdc42 and filipodia formation, which may mitigate monocyte recruitment to the artery wall.

## ABCA1: A REGULATOR OF MEMBRANE PROTRUSIONS AND LIPID MICRODOMAINS

CDC42 is a member of the family of small GTP-binding proteins that controls a wide range of cellular functions, including cytoskeletal modulation, formation of filipodia, and vesicular processing. Similar to ABCA1, the protein expression of CDC42 is increased by cholesterol loading of monocytes, whereas deloading by apo AI and HDL has the opposite effect. These changes in CDC42 expression are paralleled by alterations in M-CSF-induced filipodia formation and fMLP-induced chemotaxis, with an increase on E-LDL-mediated cholesterol loading and a decrease in response to apo AI and HDL. ABCA1-deficient monocytes from Tangier patients showed reduced filipodia formation and decreased CDC42 expression. Thus, the ABCA1 pathway is linked to the formation of membrane protrusions, which may be of significant relevance for the anti-atherogenic effects of apo AI and HDL. Further evidence for this functional link was provided by Matsuzawa and colleagues, who demonstrated that overexpression of ABCA1 in HEK293 cells induces formation of filipodia and long membrane protrusions. These ABCA1 effects and the apo AI-mediated lipid efflux in MDCK cells were significantly reduced by a dominant negative form of CDC42. Together with the finding that ABCA1 co-immunoprecipitated with CDC42, the data suggest a role for CDC42 as a downstream mediator of ABCA1 function. Moreover, it could be imagined that the principal function of ABCA1 is to facilitate the supply of choline-phospholipids and cholesterol for newly emerging plasma membrane extensions, which in the presence of apo AI are transferred to the extracellular acceptor rather than being used for the formation of new membrane areas (Fig. 1B). This hypothesis would be in accordance with the slightly impaired intestinal cholesterol absorption observed in *Abca1*<sup>-/-</sup> mice, a process that involves the microvilli of the enterocyte brush boarder membrane. Further support is derived from findings that apo AI preferentially depletes cholesterol and phospholipids from a novel type of cholesterol-based microdomain called Lubrol raft. Röper and colleagues showed that these lipid microdomains are building units for different forms of plasma membrane protrusions and that CDC42 and ABCA1 were partially localized to these domains. In fibroblasts, apo AI-induced lipid efflux also involved classical Triton X-100 rafts, and independent of the cell type, these domains were also modified by spherical HDL<sub>3</sub> (Fig. 1B). Because Triton rafts are recognized as

platforms of signal transduction and cell regulation, the observed effect may be of major importance for the role of HDL in atherogenesis.

### See Also the Following Articles

Atherogenesis • Hypercholesterolemias, Familial Defective ApoB (FDB) and LDL Receptor Defects • Low HDL/High HDL Syndromes

### Further Reading

- Aiello, R. J., Brees, D., Bourassa, P. A., Royer, L., Lindsey, S., Coskran, T., Haghpassand, M., and Francone, O. L. (2002). Increased atherosclerosis in hyperlipidemic mice with inactivation of *abca1* in macrophages. *Arterioscler. Thromb. Vasc. Biol.* **22**, 630–637.
- Bodzioch, M., Orso, E., Klucken, J., Langmann, T., Bottcher, A., Diederich, W., Drobnik, W., Barlage, S., Buchler, C., Porsch-Ozcuremez, M., Kaminski, W. E., Hahmann, H. W., Oette, K., Rothe, G., Aslanidis, C., Lackner, K. J., and Schmitz, G. (1999). The gene encoding ATP-binding cassette transporter 1 is mutated in Tangier disease. *Nat. Genet.* **22**, 347–351.
- Brooks-Wilson, A., Marcil, M., Clee, S. M., Zhang, L. H., Roomp, K., van Dam, M., Yu, L., Brewer, C., Collins, J. A., Molhuizen, H. O., Loubser, O., Ouellette, B. F., Fichter, K., Ashbourne-Excoffon, K. J., Sensen, C. W., Scherer, S., Mott, S., Denis, M., Martindale, D., Frohlich, J., Morgan, K., Koop, B., Pimstone, S., Kastelein, J. J., and Hayden, M. R. (1999). Mutations in ABC1 in Tangier disease and familial high-density lipoprotein deficiency. *Nat. Genet.* **22**, 336–345.
- Costet, P., Luo, Y., Wang, N., and Tall, A. R. (2000). Sterol-dependent transactivation of the ABC1 promoter by the liver X receptor/retinoid X receptor. *J. Biol. Chem.* **275**, 28240–28245.
- Dean, M., Hamon, Y., and Chimini, G. (2001). The human ATP-binding cassette (ABC) transporter superfamily. *J. Lipid Res.* **42**, 1007–1017.
- Diederich, W., Orso, E., Drobnik, W., and Schmitz, G. (2001). Apolipoprotein A1 and HDL(3) inhibit spreading of primary human monocytes through a mechanism that involves cholesterol depletion and regulation of CDC42. *Atherosclerosis* **159**, 313–324.
- Drobnik, W., Borsukova, H., Bottcher, A., Pfeiffer, A., Liebisch, G., Schutz, G. J., Schindler, H., and Schmitz, G. (2002). ATP-binding cassette transporter A1 (ABCA1) affects total body sterol metabolism. *Traffic* **3**, 268–278.
- Drobnik, W., Lindenthal, B., Lieser, B., Ritter, M., Christiansen, W. T., Liebisch, G., Giesa, U., Igel, M., Borsukova, H., Buchler, C., Fung-Leung, W. P., Von Bergmann, K., and Schmitz, G. (2001). Apo A1/ABCA1-dependent and HDL3-mediated lipid efflux from compositionally distinct cholesterol-based microdomains. *Gastroenterology* **120**, 1203–1211.
- Joyce, C. W., Amar, M. J., Lambert, G., Vaisman, B. L., Paigen, B., Najib-Fruchart, J., Hoyt, R. F., Jr., Neufeld, E. D., Remaley, A. T., Fredrickson, D. S., Brewer, H. B., Jr., and Santamarina-Fojo, S. (2002). The ATP binding cassette transporter A1 (ABCA1) modulates the development of aortic atherosclerosis in C57BL/6 and apoE-knockout mice. *Proc. Natl. Acad. Sci. USA* **99**, 407–412.
- Klucken, J., Buchler, C., Orso, E., Kaminski, W. E., Porsch-Ozcuremez, M., Liebisch, G., Kapinsky, M., Diederich, W., Drobnik, W., Dean, M., Allikmets, R., and Schmitz, G. (2000). ABCG1 (ABC8), the human homolog of the *Drosophila* white gene, is a regulator of macrophage cholesterol and phospholipid transport. *Proc. Natl. Acad. Sci. USA* **97**, 817–822.
- Langmann, T., Porsch-Ozcuremez, M., Heimerl, S., Probst, M., Moehle, C., Taher, M., Borsukova, H., Kielar, D., Kaminski, W. E., Dittrich-Wengenroth, E., and Schmitz, G. (2002). Identification of sterol-independent regulatory elements in the human ATP-binding cassette transporter A1 promoter. *J. Biol. Chem.* **277**, 14443–14450.
- Orso, E., Broccardo, C., Kaminski, W. E., Bottcher, A., Liebisch, G., Drobnik, W., Gotz, A., Chambenoit, O., Diederich, W., Langmann, T., Spruss, T., Luciani, M. F., Rothe, G., Lackner, K. J., Chimini, G., and Schmitz, G. (2000). Transport of lipids from golgi to plasma membrane is defective in tangier disease patients and Abc1-deficient mice. *Nat. Genet.* **24**, 192–196.
- Porsch-Ozcuremez, M., Langmann, T., Heimerl, S., Borsukova, H., Kaminski, W. E., Drobnik, W., Honer, C., Schumacher, C., and Schmitz, G. (2001). The zinc finger protein 202 (ZNF202) is a transcriptional repressor of ATP binding cassette transporter A1 (ABCA1) and ABCG1 gene expression and a modulator of cellular lipid efflux. *J. Biol. Chem.* **276**, 12427–12433.
- Roper, K., Corbeil, D., and Huttner, W. B. (2000). Retention of prominin in microvilli reveals distinct cholesterol-based lipid micro-domains in the apical plasma membrane. *Nat. Cell Biol.* **2**, 582–592.
- Rust, S., Rosier, M., Funke, H., Real, J., Amoura, Z., Piette, J. C., Deleuze, J. F., Brewer, H. B., Duverger, N., Deneffe, P., and Assmann, G. (1999). Tangier disease is caused by mutations in the gene encoding ATP-binding cassette transporter 1. *Nat. Genet.* **22**, 352–355.
- Szakacs, G., Langmann, T., Ozvegy, C., Orso, E., Schmitz, G., Varadi, A., and Sarkadi, B. (2001). Characterization of the ATPase cycle of human ABCA1: implications for its function as a regulator rather than an active transporter. *Biochem. Biophys. Res. Commun.* **288**, 1258–1264.
- Tsukamoto, K., Hirano, K., Tsujii, K., Ikegami, C., Zhongyan, Z., Nishida, Y., Ohama, T., Matsuura, F., Yamashita, S., and Matsuzawa, Y. (2001). ATP-binding cassette transporter-1 induces rearrangement of actin cytoskeletons possibly through Cdc42/N-WASP. *Biochem. Biophys. Res. Commun.* **287**, 757–765.
- Wang, N., Silver, D. L., Thiele, C., and Tall, A. R. (2001). ATP-binding cassette transporter A1 (ABCA1) functions as a cholesterol efflux regulatory protein. *J. Biol. Chem.* **276**, 23742–23747.



# Abetalipoproteinemia

John R. Wetterau and Richard E. Gregg  
Bristol-Myers Squibb, Princeton, New Jersey, United States

## Glossary

**chylomicron** Triglyceride-rich lipoprotein made in the intestine that transports dietary lipid to various tissues throughout the body.

**microsomal triglyceride transfer protein (MTP)** Heterodimeric lipid transfer protein found within the lumen of the endoplasmic reticulum of enterocytes and hepatocytes that promotes the transport of triglyceride and cholesteryl ester between phospholipid membranes in *in vitro* lipid transfer assays.

**protein disulfide isomerase** Multifunctional redox chaperone protein found within the lumen of the endoplasmic reticulum that promotes the proper folding of newly synthesized secretory proteins that contain disulfide bonds.

**very low-density lipoprotein (VLDL)** Triglyceride-rich lipoprotein made in the liver and intestine that transports lipid to peripheral tissues.

Abetalipoproteinemia is an autosomal recessive disease caused by mutations in the gene encoding the microsomal triglyceride transfer protein. Affected individuals have defects in the production of plasma lipoproteins that contain apolipoprotein B: chylomicrons, very low-density lipoproteins, and low-density lipoproteins. As a result of the defect, subjects have plasma cholesterol and triglyceride levels of approximately 40 and 10mg/dl, respectively. Neuromuscular and retinal degeneration occurs due to a secondary deficiency in vitamin E, a fat-soluble vitamin that depends on lipoproteins for its absorption and transport throughout the body.

## INTRODUCTION

In 1950, F. A. Bassen and A. L. Kornzweig diagnosed a patient with retinitis pigmentosa, malformed erythrocytes, celiac disease, and ataxia. This disease was later named abetalipoproteinemia when H. B. Salt and colleagues associated a similar syndrome with an absence of plasma lipoproteins with beta electrophoretic

mobility. Subsequently, it was found that many of the pathological consequences of the disease, particularly the neurological findings, were related to vitamin E deficiency. In the 1990s, the molecular basis of the disease was elucidated when a link between mutations in the gene encoding the microsomal triglyceride transfer protein [MTP; a lipid transfer protein located at the sites of chylomicron and very low-density lipoprotein (VLDL) assembly], a defect in lipoprotein assembly, and abetalipoproteinemia was established, thus demonstrating that the proximal cause of abetalipoproteinemia is a mutation in the MTP gene.

## ABETALIPOPROTEINEMIA IS CAUSED BY MUTATIONS IN THE MTP GENE

### Overview of Lipoprotein Metabolism

Plasma lipids are transported throughout the body in lipid-protein complexes. Chylomicrons and VLDL are triglyceride-rich emulsions surrounded by a monolayer of phospholipid, free cholesterol, and protein. They transport dietary and endogenously synthesized triglyceride, respectively, to peripheral tissues, where it is hydrolyzed to produce free fatty acids that can be used as an energy source or stored as fat in adipocytes. Following lipolysis, chylomicron remnants are rapidly cleared from plasma. Some VLDL remnants are cleared directly from plasma, but a portion are converted to cholesterol and cholesteryl ester-rich LDLs, which are subsequently removed from plasma by a receptor-mediated process. The primary protein component of VLDL is apolipoprotein B (apoB)-100 (4536 amino acids). The primary protein component of chylomicrons is apoB-48 (2152 amino acids), which is encoded by an edited version of the mRNA that encodes apoB-100.

High-density lipoproteins (HDLs) are small, dense particles secreted directly from the liver or made from excess surface components of chylomicrons and VLDL following hydrolysis of their triglyceride

core. HDLs play an important role in the transport of cholesterol from peripheral tissues to the liver, where it can be processed and transported out of the body.

### Genetic Link between a Defect in MTP, a Defect in Lipoprotein Assembly, and Abetalipoproteinemia

The MTP is a heterodimeric lipid transfer protein found in the lumen of the endoplasmic reticulum and Golgi apparatus. One subunit is protein disulfide isomerase, a 58-kDa multifunctional redox chaperone protein that catalyzes the proper folding of newly synthesized proteins that contain disulfide bonds within the endoplasmic reticulum. The second subunit is a novel 97-kDa subunit that confers the lipid transfer activity to the protein complex. In lipid transfer assays utilizing synthetic membrane substrates, MTP accelerates the transfer of triglyceride, cholesteryl ester, and, to a lesser extent, phospholipid between membranes.

Intestinal biopsies from abetalipoproteinemic subjects were found to be devoid of MTP activity and the unique large subunit of MTP. Following the cloning of the MTP large subunit, various missense, nonsense, frameshift, and splice site mutations were identified in the gene encoding the MTP large subunit in the patients studied. All mutations were either homozygous or compound heterozygous and explained the complete absence of functional MTP protein, consistent with the autosomal recessive transmission of the disease. These and subsequent studies of specific inhibitors of MTP lipid transfer activity established that MTP is required for the assembly of apoB-containing lipoproteins, and that defects in MTP are the proximal cause of abetalipoproteinemia.

## PATHOLOGICAL CONSEQUENCES OF A DEFECT IN MTP

### Clinical Description of Abetalipoproteinemia

Abetalipoproteinemic subjects have fat malabsorption that results in abdominal discomfort, diarrhea, and steatorrhea following a meal with a normal fat content. Intestinal enterocytes are fat laden. Subjects also have malabsorption of fat-soluble vitamins—vitamins E, A, K, and, to a lesser extent, D. Plasma cholesterol and triglyceride levels are very low (approximately 40 and 10mg/dl, respectively), and plasma triglyceride levels do not increase following a meal. Apolipoprotein B,

the major protein component of VLDL and chylomicrons, is virtually absent. Although HDL levels may be only moderately decreased, they have an abnormal composition, including an elevated total cholesterol, free cholesterol/esterified cholesterol ratio, and sphingomyelin content.

Half or more of the erythrocytes in abetalipoproteinemic subjects have irregular cytoplasmic projections. These abnormal erythrocytes, referred to as acanthocytes, are thought to arise from an altered membrane composition. A moderate anemia may occur due to hemolysis and a shortening of the red blood cell resident time in the circulation. Coagulation abnormalities secondary to a deficiency of vitamin K may occur. Liver biopsies show evidence of steatosis. Elevated transaminase levels and fibrosis have been reported. There are also rare reports of serious liver pathology, including cirrhosis.

The neurological abnormalities are usually the most severe consequence of the syndrome and include spinal–cerebellar degeneration and peripheral neuropathies, which can lead to a loss of reflexes, altered sensation, muscle weakness, and ataxia that can progress to a point at which the affected individual is unable to walk. Degenerative pigmentary retinopathy, retinitis pigmentosa, leads to decreased night and color vision. If left untreated, this will progress to blindness.

### Treatment

The gastrointestinal side effects of the disease can be controlled by avoiding high-fat meals. This is particularly important early in life, when severe gastrointestinal side effects and malabsorption can result in a failure to thrive. Therapy for abetalipoproteinemic subjects includes fat-soluble vitamin supplements, including massive levels of vitamin E (up to 20,000mg/day) and more usual replacement doses of vitamins A and K. Vitamin D does not usually need to be supplemented. The neurological and ophthalmological manifestations of abetalipoproteinemia are similar to those found in animals with vitamin E deficiency. Vitamin E, which plays an important role in preventing lipid peroxidation, requires chylomicrons and VLDL for efficient absorption and transport throughout the body. Although plasma vitamin E levels remain far below normal following vitamin E therapy, tissue levels may increase to near normal levels. Vitamin E and A therapy may slow the progression or even stabilize the neurological and retinal consequences of the disease.

## See Also the Following Articles

Anderson's Disease (Chylomicron Retention Disease) • Dysbetalipoproteinemia and Type III Hyperlipidemia • Familial Low Cholesterol Syndromes, Hypobetalipoproteinemia • Hypercholesterolemias, Familial Defective ApoB (FDB) and LDL Receptor Defects • Low HDL/High HDL Syndromes

## Further Reading

- Berriot-Varoqueaux, N., Aggerbeck, L. P., Samson-Bouma, M.-E., and Wetterau, J. R. (2000). The role of the microsomal triglyceride transfer protein in abetalipoproteinemia. *Annu. Rev. Nutr.* **20**, 663–697.
- Gordon, D. A., and Jamil, H. (2000). Progress towards understanding the role of microsomal triglyceride transfer protein in apolipoprotein-B lipoprotein assembly. *Biochim. Biophys. Acta* **1486**, 72–83.
- Kane, J. P., and Havel, R. J. (1995). Disorders of the biogenesis and secretion of lipoproteins containing the B apolipoproteins. In “The Metabolic and Molecular Bases of Inherited Disease” (C. R. Scriver, A. L. Beaudet, W. S. Sly, and D. Valle eds.), 7th ed., Vol. 2 pp. 1853–1885. McGraw-Hill, New York.
- Rader, D. J., and Brewer, H. B., Jr. (1993). Abetalipoproteinemia: New insights into lipoprotein assembly and vitamin E metabolism from a rare genetic disease. *J. Am. Med. Assoc.* **270**, 865–869.
- Wetterau, J. R., Lin, M. C. M., and Jamil, H. (1997). Microsomal triglyceride transfer protein. *Biochim. Biophys. Acta* **1345**, 136–150.



# Acetylation

Paolo Portaleone  
University of Turin, Turin, Italy

## Glossary

**arylalkylamine** Ammonia ( $\text{NH}_3$ ) derivative with aromatic (aryl) and aliphatic (alkyl) radicals.

**choline** Alcoholic molecule derived from phosphatidylcholine (or lecithine) supplied by food.

**enzyme** A protein or proteinaceous substance produced by a living cell that catalyzes a specific reaction.

**histone** Eight proteic sequences (octamer) constituting the core of the nucleosome (chromatin unit) around which the DNA double helix is wrapped.

**peptide hormone** Proteic endocrine messenger with low molecular weight, from 2 to several 10s of amino acid sequence linked by amino acid bond.

Acetylation is a biochemical reaction, catalyzed by specific enzymes (acetyltransferases), that consists in the transfer of an acetyl radical ( $\text{CH}_3\text{-CO-}$ ) from a donor (e.g., acetyl coenzyme A) to an acceptor molecule.

## INTRODUCTION

Acetylation is a general metabolic reaction common to both the plant and animal kingdoms. Nevertheless, acetyltransferase enzyme systems are involved not only in the biotransformation of xenobiotics but also in activating–deactivating processes of endogenous active agents such as proteins (e.g., histones, peptide hormones), alcohols, amino acids, and amines (e.g., choline, 5-hydroxy-tryptamine).

In this regulatory function, it is evident that acetylation holds a crucial role in basic processes of cellular life, from DNA transcription, replication, and repair to the control of selective messenger information through neurohormones and neurotransmitters. For the relevance of the acetyltransferase regulatory role, its involvement in several pathologies, from functional or degenerative diseases (e.g., headache, depression,

Parkinson's disease, Alzheimer's disease) to neoplastic growth, has been suggested.

## HISTONE ACETYLATION

The nucleosome, the fundamental unit of chromatin in eukaryotic cells, holds a core particle consisting of a proteic histone octamer, two copies each of H2A, H2B, H3, and H4, around which 146 bp of DNA is wrapped. The whole chromatin components are responsible for DNA dynamic behavior, but histone modifications have the main influence on DNA transcription, repair, and replication. Histones are modified by means of the addition of several chemical radicals such as phosphate, methyl, acetyl, ribosyl, and ubiquitin groups.

Acetylation of core histones is associated with transcriptional activation. In contrast to cotranslational N-terminal  $\alpha$ -acetylation of many proteins, histone acetylation occurs posttranslationally and reversibly on the  $\epsilon\text{-NH}_3^+$  groups of highly conserved lysine residues of the N-terminal tails of core histones.

The enzymes catalyzing reversible histone acetylation are histone acetyltransferases (HATs) and deacetylases (HDACs). HATs catalyze the transfer of the acetyl moiety from acetyl coenzyme A (Ac-CoA) to the  $\epsilon\text{-NH}_3^+$  of lysine residues, whereas in the opposing deacetylation reaction, HDACs remove the acetyl groups and thereby reestablish the positive charge in the histones. Thus, acetylation neutralizes the positive charge increasing the hydrophobicity of the histones, leading to “opening up” the chromatin complex through a reduced affinity of acetylated N-terminal domains of histones to DNA. Therefore, the modification in nucleosomal structure is considered to play a causative role in activating transcription. Another possibility, or rather an additional effect, has been suggested: histone acetylation could represent a signal “flag” recognized by other molecules linked to transcriptional activation.

Two different classes of HATs have been described: type A and type B. Type A HATs are localized in the

nuclei and acetylate nucleosomal histones leading to transcriptional activation, whereas type B HATs can be found in cytoplasmic fractions and acetylate newly synthesized histones before chromatin assembly during DNA replication.

Several known transcriptional regulators in mammals have been found to possess intrinsic type A HAT activity, and among them the best understood family is the Gcn5-related N-acetyltransferase (GNAT) family. No type B HAT has been identified and characterized in mammals.

Not only nuclear histones are substrates for the acetyltransferase, but also numerous nonhistone proteins involved in transcription regulation, such as p53, E2F1, EKLF, TFIIE $\beta$ , TFIIF, TCF, GATA1, HMGI(Y), and ACTR, or even non-nuclear proteins, such as  $\alpha$ -tubulin, are substrates for the acetyltransferase. There are three possible consequences of the acetylation processes, depending on where acetylation takes place within the protein: increased or decreased DNA binding, protein-protein interaction regulation, and protein stability.

If there has been an explosion of studies on HAT activity during the past decade or so, only a few lines of evidence indicate the regulation of the enzymatic activity of acetylases. A bromodomain (a specific HAT protein structure) is present in many transcriptional activators with HAT activity, and it seems to be a requisite for targeting the enzyme to the substrate. The regulation of HAT activity is carried out by proliferation and differentiation signals by means of phosphorylation or hormonal signaling.

Because acetylation can regulate such wide and different cellular functions, both nuclear and cytoplasmic (including the circadian clock in DNA transcription), it is obvious that its dysfunctioning could be at the origin of the different pathologies.

## PEPTIDE HORMONE ACETYLATION

N-terminal acetylation is a nearly selective posttranslational processing event among peptide hormones, and among the end products of the pro-opiomelanocortin (POMC) biosynthetic pathway, only  $\alpha$ -melanocyte-stimulating hormone ( $\alpha$ -MSH) and  $\beta$ -endorphin undergo this posttranslational modification. The relevance of the POMC-derived peptide N-acetylating mechanism, under a phylogenetic point of view, is supported by its persistence as an "ancestral" mechanism throughout vertebrate evolution. In mammals,  $\alpha$ -MSH and  $\beta$ -endorphin, as final products of a set of cleavage reactions of POMC, were found not only in

secretory granules of pituitary neurointermediate lobe but also in anterior pituitary lobe and in neurons, mainly of the hypothalamic arcuate nucleus.

The N-acetylation reaction, which requires acetyl-CoA as a coenzyme, occurs on the serine-NH<sub>2</sub> terminal for  $\alpha$ -MSH and on the tyrosine-NH<sub>2</sub> terminal for  $\beta$ -endorphin. On the basis of the various amino-terminal targets of N-acetyltransferase (NAT) and its different regional distributions in the pituitary and brain, two distinct enzymes, an  $\alpha$ -MSH-acetyltransferase (MAT) and a  $\beta$ -endorphin-acetyltransferase (EAT), have been proposed. Actually, two forms of NAT have been found. One enzyme, specifically localized in secretory granules of the neurointermediate lobe (NIL) of the pituitary gland with an optimal pH of 6.0 to 6.6, is inhibited by several solubilizing detergents and possesses similar characteristics in the acetylation process of both the serine of  $\alpha$ -MSH and the tyrosine of  $\beta$ -endorphin. Therefore, this single NAT capable of acetylating the opioid and melanotropic peptides has been termed opiomelanotropin acetyltransferase (OMAT). A second enzyme, with an optimal pH of 7.4, is inhibited by Mg<sup>2+</sup>, shows different anatomical and subcellular (cytosol) distribution, and has a more general acetyltransferase activity (GAT).

The POMC-derived peptide NAT is coexpressed with the POMC gene and undergoes the same regulatory control of POMC synthesis by inhibitors and activators (e.g., glucocorticoids, sex steroids, and dopamine as inhibitors; adrenalectomy, castration, and dopamine receptor antagonists as activators).

The N-acetylation of des-acetyl- $\alpha$ -MSH and  $\beta$ -endorphin substantially alters the physiological responses produced by both peptides. The acetylated form of  $\alpha$ -MSH, in fact, is about 10 to 100 times more effective than its des-acetylated form in increasing arousal, memory, and attention in the Y-maze visual discrimination task and in eliciting excessive grooming. Moreover, the *in vitro* lipolytic activity on rabbit adipose tissue slices and the *in vitro* melanotropic activity in frog skin are markedly reduced after removing the acetyl group from monoacetyl- $\alpha$ -MSH. Conversely, the acetylation of  $\beta$ -endorphin completely eliminates the opiate analgesic activity of the peptide and markedly reduces its affinity in binding to opiate receptors. Furthermore, a substantial decrease of NAT activity has been observed during the lifetime, fitting in with the decreased concentrations of  $\alpha$ -MSH found in rat aged brain.

The physiological role of POMC-derived peptide NAT, with its ambivalent effects, activating des-acetyl- $\alpha$ -MSH, and deactivating  $\beta$ -endorphin, is very difficult to interpret, and the lacking characterization of the

enzyme, very unstable and ubiquitous, makes the task arduous. Considering that the POMC-derived peptides (adrenocorticotropin hormone [ACTH],  $\alpha$ -MSH, and  $\beta$ -endorphin) are the main effectors in the “organized stress response” coordinating the biological and behavioral adaptive effects, we suggest that the deactivating–activating N-acetylation regulates the chronology of the adaptive response sequence with a rapid inactivation of ACTH and  $\beta$ -endorphin effects and with a potentiation of the  $\alpha$ -MSH long-lasting adaptive activity. The progressively lower synthesis and activating–deactivating NAT activity of POMC-derived peptides during aging could be related to the reduced adaptive capabilities of aged individuals.

## CHOLINE ACETYLATION

Acetylcholine (ACh) is the neurotransmitter of the parasympathetic nervous system and is synthesized from choline and Ac-CoA in a single-step reaction catalyzed by the enzyme choline acetyltransferase (ChAT) that is expressed selectively in cholinergic neurons, where it serves as a phenotypic marker. The substrate choline, derived from phospholipids and the ACh hydrolysis, is taken up by the high-affinity sodium ( $\text{Na}^+$ )-dependent choline transporter. The newly synthesized neurotransmitter is accumulated in synaptic vesicles by means of a specific vesicular acetylcholine transporter (VACHT), a 12-transmembrane domain protein that uses the electrochemical gradient generated by a proton ATPase to exchange two protons by one ACh molecule.

ChAT is encoded by a single gene and is coexpressed with VACHT. The gene of VACHT is embedded in the first intron of the ChAT gene. This unique organization was named “cholinergic gene locus,” suggesting reciprocal posttranslational regulation between the two proteins. In humans, four of the six identified transcripts translate to the same 69-kDa protein. The fifth and sixth transcripts yield 82- and 74-kDa forms of ChAT. Until now, the mechanisms regulating production of these different transcripts and their physiological roles have not been elucidated.

ChAT exists in two forms in cholinergic nerve terminals: a soluble form (80–90% of the total enzyme activity) and a membrane-bound form (10–20%). Moreover, the 82-kDa ChAT has been localized to the nucleus, whereas the 69-kDa enzyme is largely cytosolic. Because it is unclear whether ACh synthesis occurs in the nucleus, different functional roles for 82-kDa ChAT must be considered. The 69-kDa ChAT is the form more represented and responsible for the majority of ACh biosynthesis.

The fact that ChAT is not saturated at the substrate concentrations in *in vitro* kinetic studies means that the enzyme would be in kinetic excess. Based on these *in vitro* data, it is accepted that the neuronal ChAT levels might not be rate limiting in ACh synthesis; therefore, the availability of the substrates would regulate ACh production. On the other hand, some data support the regulatory role of ChAT to maintain the homeostatic levels of ACh under some conditions of neuronal activity and demand for ACh synthesis.

Short- and long-term regulation of ChAT activity has been described. Thomas Dobransky and R. Jane Rylett reviewed the role of phosphorylation in ChAT short-term regulation, showing that ChAT is phosphorylated by several protein kinases at various sites of ChAT primary structure in response to different functional states of neurons. Taken together, these lines of evidence indicate that phosphorylation of ChAT is physiologically significant and could serve as a regulatory mechanism. Neurotrophic factors (neurotrophins) such as nerve growth factor (NGF) are responsible for long-term ChAT regulation, enhancing the enzyme expression and/or its activity. Clarifying the molecular mechanisms of ChAT regulation and its dysfunctions may be helpful in explaining the possible cellular mechanisms responsible for the loss of cognitive attributes associated with cholinergic deficit in Alzheimer’s disease.

## ARYLALKYLAMINE N-ACETYLATION

The acetylation of serotonin (5-hydroxy-tryptamine [5-HT]) in the pineal gland is another acetylation reaction essential to regulating the biological circadian rhythms in vertebrates by means of the production of the pineal hormone melatonin.

Melatonin is synthesized from serotonin, which is acetylated into N-acetylserotonin (NAS) by the arylalkylamine N-acetyltransferase (AA-NAT) and then is methylated by the hydroxyindole-O-methyltransferase enzyme (HIOMT). The rate of O-methylation is largely a function of substrate availability, whereas the N-acetylation step is regulated by the amount and activation of AA-NAT protein. Melatonin synthesis occurs in darkness under the control of norepinephrine (NE) that is released from sympathetic fibers originating in the superior cervical ganglia and that is regulated, through a polysynaptic neuronal pathway, by a circadian oscillator located in the suprachiasmatic nuclei (SCN). NE activates pinealocytes acting on  $\alpha_1$ - and  $\beta_1$ -adrenergic receptors and subsequent increases in the intracellular concentration of calcium ions ( $[\text{Ca}^{2+}]_i$ ) and cyclic AMP (cAMP), respectively, leading

a strong AA-NAT activation at both the transcriptional and posttranscriptional levels. cAMP-increased levels induce the phosphorylation protein kinase A-dependent of the stimulatory transcription factor cAMP response element-binding protein (CREB). Phosphorylated CREB binds to a cAMP response element (CRE) element in the AA-NAT gene, resulting in increased transcription and accumulation of AA-NAT mRNA and then in increased expression and activity of the AA-NAT enzyme. The activation of  $\alpha_1$ -adrenergic receptors induces increased concentrations of  $\text{Ca}^{2+}$  and diacylglycerol (DAG), leading to activation of protein kinase C (PKC) potentiating  $\beta_1$ -adrenergic receptor stimulation of adenylate cyclase (AC) through a postreceptor mechanism.

Furthermore, there is posttranslational regulation of AA-NAT protein levels by means of cAMP-dependent inhibition of proteosomal proteolysis. This mechanism involves phosphorylation-dependent binding of AA-NAT to 14-3-3 proteins, shielding the enzyme from proteolysis. In rodents, cAMP stimulation causes AA-NAT mRNA to increase more than 150-fold at night, whereas in ungulates and primates, the night/day ratio is approximately 1.5. This difference may be explained by the fact that in ungulates and primates, AA-NAT protein levels are regulated primarily at the posttranslational level by controlled proteosomal proteolysis.

AA-NAT is a member of a large superfamily of proteins referred to as the GNAT family and acts through catalysis of the transacetylation of serotonin to N-acetylserotonin with Ac-CoA as a donor. AA-NAT is a globular protein with a molecular weight of 23 kDa and consisting of eight stranded  $\beta$ -sheets containing Ac-CoA-binding sites. It is described in two different conformational states influencing its functional efficiency.

In addition to NE-induced activation of the melatonin-generating system, several other neuroactive substances have been shown to influence melatonin synthesis. These include vasoactive intestinal peptide (VIP), pituitary adenylate cyclase-activating polypeptide (PACAP), and ACh acting on AA-NAT

mRNA expression and/or on second messenger-induced AA-NAT stimulation.

It has been hypothesized that alterations in AA-NAT synthesis and/or activity can represent the pathophysiological basis not only of circadian rhythm disturbances but also of several pathologies such as migraine, depression, and insomnia.

### See Also the Following Articles

ACTH,  $\alpha$ -MSH, and POMC, Evolution of • Alzheimer's Disease and Hormones

### Further Reading

- Cangemi, L., Adage, T., Morabito, A., and Portaleone, P. (1995). N-acetyltransferase mechanism for  $\alpha$ -melanocyte stimulating hormone regulation in rat ageing. *Neurosci. Lett.* **201**, 65–68.
- Davie, R. J., and Spencer, V. A. (1999). Control of histone modifications. *J. Cell. Biochem. Suppl.* **32/33**, 141–148.
- Dobransky, T., and Rylett, R. J. (2003). Functional regulation of choline acetyltransferase by phosphorylation. *Neurochem. Res.* **28**, 537–542.
- Dores, R. M., Stevenson, T. C., and Price, M. L. (1993). A view of the N-acetylation of  $\alpha$ -melanocyte-stimulating hormone and  $\beta$ -endorphin from phylogenetic perspective. *Ann. NY Acad. Sci.* **680**, 161–174.
- Ganguly, S., Coon, S. L., and Klein, D. C. (2002). Control of melatonin synthesis in the mammalian pineal gland: The critical role of serotonin acetylation. *Cell Tissue Res.* **309**, 127–137.
- Hasan, S., and Hottiger, M. O. (2002). Histone acetyl transferases: A role in DNA repair and DNA replication. *J. Mol. Med.* **80**, 463–474.
- Kuo, M.-H., and Allis, C. D. (1998). Roles of histone acetyltransferases and deacetylases in gene regulation. *BioEssays* **20**, 615–626.
- O'Donohue, T. L., Handelsmann, G. E., Chaconas, T., Miller, R. L., and Jacobowitz, D. M. (1981). Evidence that N-acetylation regulates the behavioral activity of  $\alpha$ -MSH in the rat and human central nervous system. *Peptides* **2**, 333–344.
- Prado, M. A. M., Reis, R. A. M., Prado, V. F., de Mello, M. C., Gomez, M. V., and de Mello, F. G. (2002). Regulation of acetylcholine synthesis and storage. *Neurochem. Intl.* **41**, 291–299.
- Schomerus, C., Laedtke, E., Olcese, J., Weller, J. L., Klein, D. C., and Korf, H.-W. (2002). Signal transduction and regulation of melatonin synthesis in bovine pinealocytes: Impact of adrenergic, peptidergic, and cholinergic stimuli. *Cell Tissue Res.* **309**, 417–428.



Achondroplasia

see Chondrodysplasias



# Acromegaly, Clinical Features of

Ariel L. Barkan

University of Michigan, Ann Arbor, Michigan, United States

## Glossary

**arthropathy** Abnormality in the condition of a joint of the body; enlargement, swelling, degeneration, or other such disturbance in a joint.

**growth hormone (GH)** A polypeptide hormone that is secreted mainly by the pituitary gland, although it is also produced by other cell types, such as lymphoid cells. Its actions are related mainly to growth (soft tissues, long bones, etc.) and to metabolism. It belongs to a family of hormones that includes prolactin and placental lactogens as well as other placental factors. Acromegaly results from GH hypersecretion.

**insulin-like growth factor (IGF)** One of a class of hormones structurally related to insulin, but exhibiting proliferative and differentiative, rather than metabolic, effects. IGF-I overproduction is associated with acromegaly.

**multiple endocrine neoplasia (MEN) syndromes** A group of genetically distinct familial diseases in which two or more endocrine glands develop excess normal tissue (hyperplasia) and/or adenoma (tumor).

**sleep apnea** A sleep disorder in which the subject has intermittent periods of a failure to automatically control respiration; these involuntary pauses in breathing may occur repeatedly during a given period of sleep.

From the purely clinical perspective, acromegaly is the most spectacular endocrine disease. Humanity has always been fascinated with giants, and from the biblical Goliath to James Bond's foe, Jaws, they populate the lore of virtually every era and culture.

## INTRODUCTION

Acromegaly is due to growth hormone (GH) hypersecretion and the resultant secondary overproduction of insulin-like growth factor-1 (IGF-1). In virtually all cases (>99%), the source of excessive GH is a benign pituitary tumor of purely somatotroph or mixed cellular origin. Rarely, somatotroph tumors arise in an ectopic pituitary, a remnant of the primitive

Rathke's pouch, and are found in the posterior pharynx, sphenoid bone, or sphenoid sinus or even within the sella, separate from the normal pituitary gland. Ectopic production of GH-releasing hormone (GHRH) leads to pituitary somatotroph hyperplasia with subsequent adenoma formation. Carcinoids and islet cell tumors are the most frequent sources of ectopic GHRH. Hypothalamic/pituitary gangliocytomas or choristomas, or in one case the pituitary adenoma itself, were shown to be a source of excessive GHRH production. Ectopic GH secretion was documented in only two cases: one by a malignant islet cell tumor and another by a non-Hodgkin's lymphoma. Acromegaly may be a part of well-defined MEN syndromes such as MEN-1 (parathyroid, pituitary, pancreas), McCune-Albright, and Carney syndromes. In some instances, the clinical syndrome of acromegaly may be overshadowed by other manifestations of a malignant or polyglandular disease.

Clinical manifestations of acromegaly correlate better with the prevailing levels of IGF-1 than with GH, and the duration of GH/IGF-1 excess may play a major role. A clinical and biochemical syndrome of acromegaly may be transiently expressed during normal puberty or pregnancy. This is due to physiological overproduction of GH by the normal pituitary gland during sexual maturation or by the placental synthesis of the GH variant.

In most cases, acromegaly is an insidious disease, and its early clinical manifestations usually go unnoticed by the patient, the patient's family, and/or the patient's family physician. Retrospective questionnaires and the inspection of old photographs usually set the clinical onset of disease at 5 to 10 years prior to the diagnosis.

Clinical presentation of acromegaly consists of the mass effects of the tumor itself, the manifestations of the abnormal growth affecting virtually all organs and tissues, and the metabolic derangements effected by GH itself. Despite the seemingly straightforward pathophysiological mechanisms, the clinical picture of acromegaly is often protean, and the correct

and timely diagnosis requires significant clinical acumen.

### MASS EFFECTS OF THE TUMOR

Most pituitary somatotropinomas (60–80%) are large at the time of diagnosis (macroadenoma, >10 mm in the largest diameter) and are often invasive. Visual field defects were present at diagnosis in 90% of patients in the past, but this figure decreased to ~10 to 20% due to earlier recognition of the disease. Ophthalmoplegia (III, IV, VI, and V<sub>1</sub>, V<sub>2</sub> nerves) is rare and, if present, suggests recent rapid expansion of the tumor by a hemorrhage. Headache is present in 50 to 60% of patients and may be severe in about half of them. Headache may be present even in patients with relatively small tumors, where mass effect is unlikely to provide an explanation. Hypopituitarism (ACTH and TSH deficiency) in acromegaly occurs less frequently than in patients with nonfunctioning tumors of similar size. However, hypogonadism is frequent (~50%), likely as a consequence of coexisting hyperprolactinemia and inherent lactogenic effect of GH itself. Symptomatic pituitary hemorrhage occurs in less than 5% of patients, but asymptomatic events may occur in 30 to 40% of the cases.

### ABNORMAL GROWTH

Onset of pathological GH hypersecretion before puberty results in an augmented statural growth. Concomitant hypogonadism prevents epiphyseal closures, and gigantism ensues. Postpubertal onset of disease leads to disproportionate growth and dysmorphic features, that is, true acromegaly (“large extremities”). Eventually, even pituitary giants develop an acromegalic appearance.

#### Face

A combination of bone and soft tissue overgrowth leads to a typical “acromegalic” face: large nose, thick lips, exaggerated nasolabial and frontal skin furrows, mandibular overgrowth and prognathism, teeth separation, and frontal bossing. These features are seen in 98 to 100% of patients (see Fig. 1).

#### Extremities

Hands and feet become very “fleshy.” An increase in finger circumference and a widening of the hands and feet develop in 98 to 100% of patients. Patients



**Figure 1** Appearance of a patient with florid acromegaly. Courtesy of Professor Stefan S. Fajans.

routinely recall repeat resizing of rings and changes in shoe size (mostly widening).

#### Skin and Appendages

Skin is characteristically thickened because of excessive deposition of the glycosoaminglycans, hyaluronic acid, chondroitin sulfate, and dermatan sulfate in the papillary and upper reticular dermis. These compounds are very hydrophylic, causing the appearance of a nonpitting edema. Skin thickening at the vertex

causes a peculiar appearance of cutis verticis gyrata (skin folds at the top of the head). Hair growth is increased, and women complain of hirsutism. As opposed to the androgen-related hirsutism, this is pronounced even on the forearms and forelegs. Many women with acromegaly have exceedingly thick scalp hair growth. Hair loss after successful therapy is often a cause of concern but is essentially a physiological return to normalcy.

The functional capacity of sweat and sebaceous glands is increased, resulting in excessive perspiration, often with offensive odor, and in oily skin. Skin tags are frequently present, particularly on the neck. Whether their presence and number can be a marker for colonic neoplasia is uncertain.

### Neuromuscular

The muscle mass is increased, but this is primarily due to increased intracellular water, so that muscle strength is either normal or low. In fact, many patients have clinically obvious proximal myopathy. Muscle biopsy often shows hypertrophy of type I and/or atrophy of type II fibers.

Compression neuropathies are common (30–50%); the median nerve is most often affected. It was believed that carpal tunnel syndrome was due to external compression by the components of the wrist compartment, but MRI data showed that the nerve itself is swollen. Occasionally, distal symmetric polyneuropathy may be present.

### Oral

Prognathism and widening of the interdental spaces have already been mentioned. Importantly, the size of the tongue is usually enlarged and contributes to the obstruction of the pharynx, with the resultant sleep apnea and impaired mastication. The abnormal oral anatomy often results in speech disturbances. The diameter of the trachea is increased, and the vocal cords are thickened. Together with grossly enlarged sinus cavities, this results in a low and hollow voice pitch. Salivary glands are typically enlarged, and their size is a convenient measure of the GH effect on parenchymatous organs.

### Articular

GH receptors are present on all major cell types comprising the skeletal system: fibroblasts, chondrocytes, and osteoblasts. These cells readily produce IGF-1 and are targets for both endocrine and autocrine IGF-1

effects. Thus, arthropathy is a frequent (60–80%) symptom of acromegaly, affecting both axial and appendicular skeleton. The degree and severity of arthropathy best correlate with the duration of disease. Joint pain and low back pain may be experienced soon after the clinical onset of acromegaly but are often fully reversible with successful therapy. However, clinical duration of acromegaly in excess of 10 years is often associated with clinical and radiographical joint deformities that are only minimally affected by the GH-lowering therapy.

#### **Appendicular Arthropathy**

The knee is the joint most frequently affected, followed by the shoulder, hip, ankle, elbow, and small hand joints. Radiographic changes are usually seen even in clinically unaffected joints. Initially, GH excess causes cartilage hypertrophy and laxity of the ligaments. The combination of altered geometry of the joint and its instability leads to repeat trauma to the cartilage. The ensuing cartilage fissures are filled by regenerative fibrocartilage, with the subsequent calcification, formation of osteophytes, and exposure of the subchondral bone. Eventually, the articular cartilage becomes thinned and the joint space narrows. The end-stage acromegalic arthropathy looks essentially like degenerative osteoarthritis. Clinical and radiological reversibility of arthropathy can be seen after successful therapy in the joints exhibiting cartilage thickening, but late stages may require joint prosthesis for relief of pain and functional mobility.

#### **Axial Arthropathy**

Lumbar involvement is most common, followed by thoracic and cervical arthropathy. Overall, approximately 50% of patients complain of back pain and limitation of movements. Thickened intervertebral discs and lax paraspinal ligaments contribute to abnormal joint mobility. End-stage arthropathy is characterized by the narrowing of the intervertebral space. Ossification of the anterior aspect of the vertebral bodies with exuberant osteophyte formation often bridges the disc space, mimicking diffuse idiopathic skeletal hyperostosis (DISH). Vertebral deformities often lead to kyphosis that may become almost grotesque in some patients.

#### **Bone Metabolism**

Biochemical markers of bone remodeling are increased in patients with acromegaly, but histomorphometric data are conflicting; cortical bone shows predominance of bone formation over resorption, whereas trabecular bone has the opposite pattern. Because most patients

with acromegaly have concomitant hypogonadism, these data should be interpreted with caution.

Bone density studies are equally controversial, but overall it appears that cortical bone mineral density may be normal or even increased, whereas trabecular bone mineral density is normal or decreased. The “normalcy” of the latter is questionable because of the interference by osteophytes. In any case, patients with acromegaly do not have altered fracture rates, and the assessment of bone mineral density in acromegaly may be of academic interest only.

### Cardiovascular

Increased mortality of untreated or poorly treated acromegaly is almost completely attributable to cardiovascular disease. Hypertension is seen in 20% of patients and does not appear to be reversible by GH/IGF-1-lowering therapy. Concentric left ventricular hypertrophy is seen in approximately two-thirds of patients, and its occurrence is best related to the duration of disease. Hypertrophy frequently occurs even in normotensive patients. Early stages of acromegaly, typical of young patients with short disease duration, are characterized by tachycardia and increased systolic output (hyperkinetic syndrome). Progressively, cardiac hypertrophy and diastolic dysfunction ensue, and end-stage disease is characterized by impaired systolic function and heart failure. The development of hemodynamic abnormalities is augmented by valvular involvement in approximately 20% of patients. Normalization of GH and IGF-1 can reduce the degree of left ventricular hypertrophy within 2 to 4 weeks.

Rhythm abnormalities, occurring in 40% of patients, are more ominous. Ectopic beats, paroxysmal atrial fibrillation, paroxysmal supraventricular tachycardia, sick sinus syndrome, ventricular tachycardia, and bundle branch blocks all are seen with increased frequencies in patients with acromegaly and are exacerbated by physical exercise. Disturbingly, biochemical control of acromegaly might not improve conduction abnormalities. Late ventricular potentials (low-amplitude, high-frequency waves in the terminal phase of QRS complexes) are strong predictors of future arrhythmic events and are seen in 50% of patients with active acromegaly.

### Sleep Apnea

Sleep apnea occurs in 75% of patients. In most, it is of an obstructive nature due to soft tissue hypertrophy of the pharynx. Interestingly, in approximately one-third

of patients, there is a central component of sleep apnea with decreased ventilatory drive. The pathogenesis of central sleep apnea in acromegaly is unknown, but both subtypes may be ameliorated by GH-lowering therapy. Sleep apnea is manifested as snoring, daytime sleepiness, fatigue, and headache. It is a strong predictor of future cardiovascular events, hypertension, or stroke. Many patients are unaware of their snoring; family members provide markedly more reliable information.

### Renal

Kidney size and glomerular filtration rate are characteristically increased. Hypercalciuria may lead to kidney stone formation (in approximately 10% of patients).

### Gastrointestinal

Liver and spleen sizes are normal. Hepatomegaly in patients with acromegaly always should be assumed to result from another disease process and should be vigorously investigated. Similarly, the incidence of cholelithiasis is not increased relative to that in the general population. Patients with acromegaly often suffer from constipation due to long and tortuous colon. The incidence of colonic polyps appears to be increased.

### METABOLIC

Impaired glucose tolerance is present in 40% of patients, and frank diabetes (DM-2) is present in 30% of patients. These often improve or vanish altogether after successful GH-lowering therapy. Other metabolic abnormalities include hypertriglyceridemia, hypercalciuria, and hyperphosphatemia. Hypercalcemia is not a feature of acromegaly per se, and its presence should suggest another pathological process.

### NEUROPSYCHIATRIC

Similar to any chronic disease associated with physical discomfort and lifelong therapy, acromegaly is associated with decreased quality of life. However, there are no specific neuropsychiatric features attributable to the disease.

### CANCER

Because patients with acromegaly have increased incidence of colonic polyps, some investigators assumed that the incidence of colonic carcinoma will also be increased. This was supported by the data from several small studies. Large-scale data, however, failed to

document increased incidence of colon cancer in acromegaly. Similarly, the incidence of breast and prostate cancers was not increased. This, of course, does not apply to patients with genetic forms of acromegaly, such as MEN-1, in which there is a predictable component of coexisting neoplastic diseases.

### See Also the Following Articles

Acromegaly, Diagnosis of • Acromegaly, Therapy for • Growth Hormone (GH) • Pituitary Tumors, Clonality • Pituitary Tumors, Molecular Pathogenesis

### Further Reading

- Barkan, A. (1997). Acromegalic arthropathy and sleep apnea. *J. Endocrinol.* **155**(Suppl. 1), S41–S44.
- Colao, A. (2001). Are patients with acromegaly at high risk for dysrhythmias? *Clin. Endocrinol.* **55**, 305–306.
- Colao, A., Marzullo, P., Di Somma, C., and Lombardi, G. (2001). Growth hormone and the heart. *Clin. Endocrinol.* **54**, 137–154.
- Ezzat, S., Forster, M. J., Berchtold, P., Redelmeier, D. A., Boerlin, V., and Harris, A. G. (1994). Acromegaly: Clinical and biochemical features in 500 patients. *Medicine* **73**, 233–240.
- Melmed, S. (2001). Acromegaly and cancer: Not a problem? *J. Clin. Endocrinol. Metab.* **86**, 2929–2934.



# Acromegaly, Diagnosis of

Klaus von Werder

Humboldt University, Berlin, Germany

## Glossary

**acromegaly** Clinical consequence of growth hormone hypersecretion.

**growth hormone (GH)** Hormone of the anterior pituitary lobe that is essential for normal growth in children. GH is also an important metabolic hormone in adults.

**insulin-like growth factor-I (IGF-I)** Peptide hormone produced in the liver and other tissues following activation of the GH-receptor.

**E**ndocrinological investigation using hormone measurements and function tests to document active acromegaly is indicated whenever symptoms and clinical stigmata suggestive for the disease are present. Thus, the diagnosis of acromegaly is made by measuring elevated, often glucose nonsuppressible, growth hormone levels and elevated insulin-like growth factor-I levels.

## MEASUREMENT OF GROWTH HORMONE LEVELS

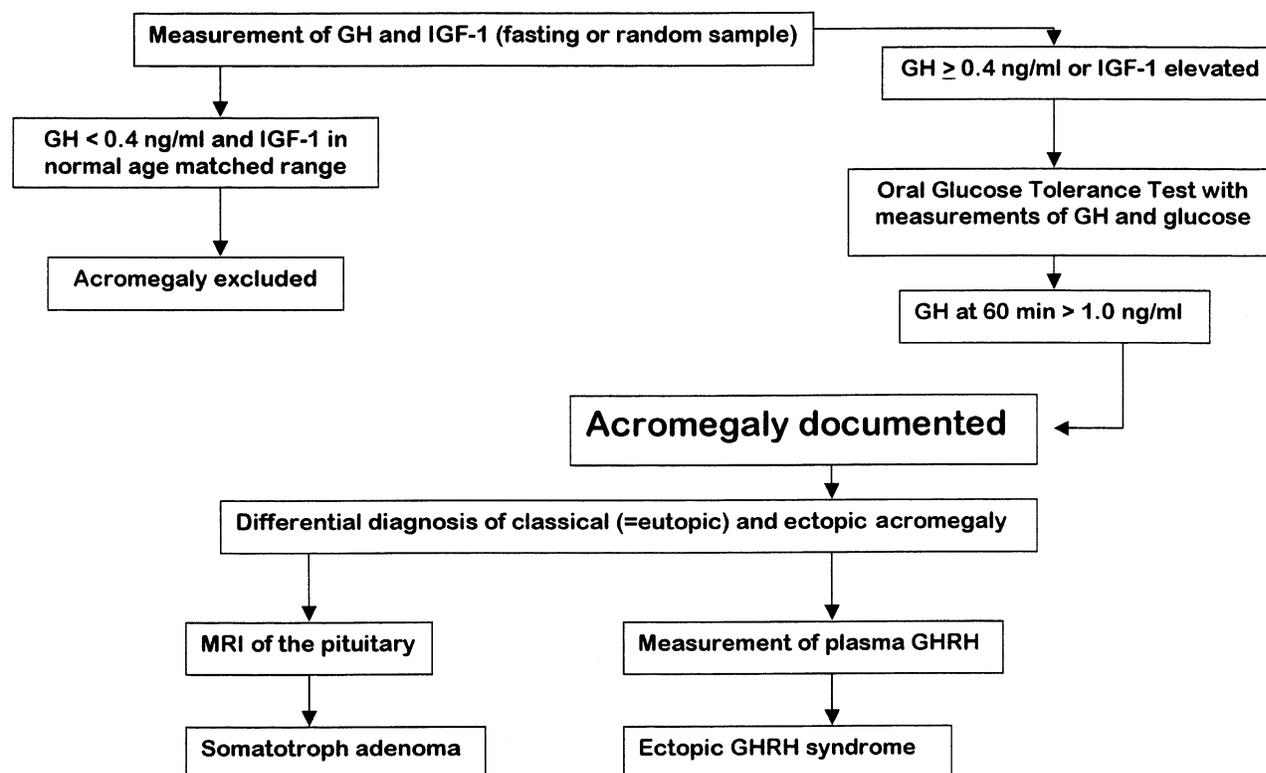
Growth hormone (GH) levels are measured by immunoassay or immunoradiometric assay. Most centers use polyclonal antibodies in their assay systems which measure higher GH levels compared to assay systems that use oligo- or monoclonal antibodies. Cutoff values that differentiate between normal and abnormally elevated GH levels are based on polyclonal assay systems according to the consensus conference on diagnostic aspects of acromegaly held in Cortina d'Ampezzo in 1999. Using these systems, fasting morning or random GH levels are usually clearly elevated in patients with acromegaly of at least more than 0.4 ng/ml (Fig. 1). Using the more sensitive and better standardized assay system with oligo- or monoclonal antibodies, these cutoff values will have to be redefined in the future. Since GH is episodically secreted in normal subjects, GH levels considerably

higher than the cutoff levels are not indicative for acromegaly unless the insulin-like growth factor-1 (IGF-1) level is also elevated. Because of the spontaneous fluctuation of GH levels in nonacromegalic subjects, many investigators recommend an oral glucose tolerance test as a suppression test for GH secretion.

Acromegaly is documented in patients in whom GH levels are higher than 1.0 ng/ml and who have elevated IGF-1 levels simultaneously. Blood sugar measurements during the oral glucose tolerance test demonstrate the different degrees of carbohydrate intolerance that are observed in active acromegalics. Measurement of insulin levels demonstrating insulin antagonism in active acromegalics is not required for making the diagnosis. Furthermore, the cumbersome measurement of GH 24-h profiles is not necessary, nor is it helpful to document the inappropriate stimulation of GH after thyrotropin releasing hormone (TRH) or gonadotropin releasing hormone (GnRH) injection that occurs in 70 and 40% of all active acromegalics, respectively. Measurement of urinary GH is also not useful for documentation of a GH hypersecretory state.

## MEASUREMENT OF IGF-I LEVELS

Serum IGF-1 levels are invariably elevated in active acromegaly, showing a certain degree of correlation with GH levels. Since IGF-1 levels are also elevated during puberty and pregnancy, they are specific for active acromegaly in the nonpregnant adult. Since IGFBP-3 is concomitantly produced and secreted with IGF-1, IGFBP-3 levels are also elevated in acromegalic patients. Measurement of the binding protein, however, does not add to the accuracy of the diagnosis. This pertains also to the acid labile subunit (ALS), which is also elevated as part of the ternary complex consisting of IGF-1, IGFBP-3, and ALS. In addition, measurement of free IGF-1, which is only possible in some laboratories, is not required for making the diagnosis of active acromegaly.



**Figure 1** Algorithm for the diagnostic workup of a suspected acromegalic patient. According to the consensus statement (Giustina *et al.*, 2000.)

## MEASUREMENT OF PROLACTIN AND OTHER ANTERIOR PITUITARY HORMONES

Since some GH-producing tumors also produce and secrete prolactin (PRL), basal PRL should be measured in all acromegalics. PRL levels higher than 200  $\mu\text{g/liter}$  are usually indicative of a somatomammotrophic adenoma. The identification of a somatomammotrophic tumor may be important with regard to medical therapy since the PRL-cosecreting adenomas respond particularly well to dopamine agonists.

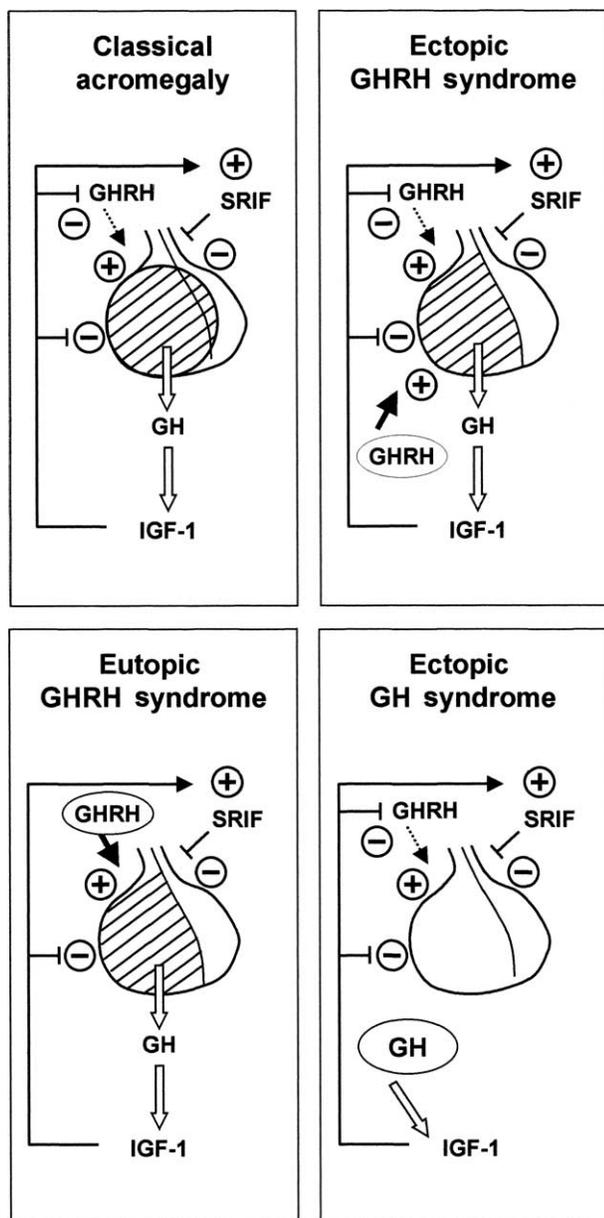
Slightly elevated PRL levels can result from impingement of the somatotroph tumor on the pituitary stalk leading to unrestrained PRL release from the nontumorous pituitary. Tests to evaluate the other pituitary functions—the gonadal, adrenal, and thyroid axis—are also indicated.

## DIFFERENTIAL DIAGNOSIS OF ACROMEGALY

In more than 99% of all patients with acromegaly, a monoclonal GH-producing tumor is the cause of the

disease (Fig. 2). The imaging method of choice is magnetic resonance imaging (MRI), which in most cases allows visualization of a macroadenoma and, less frequently, a microadenoma. This situation has been referred to by Losa and von Werder as classical acromegaly. In contrast to classical acromegaly, there are rare cases who have extra pituitary lesions causing GH hypersecretion, leading to the clinical picture of acromegaly. The ectopic growth hormone-releasing hormone (GHRH) syndrome is most common (Fig. 2). In this syndrome, GHRH is secreted from benign or malignant, often neuroendocrine, tumors, leading to somatotroph hyperplasia with consecutive GH hypersecretion. GH and IGF-1 levels are not different from those of classical acromegaly. On MRI, no pituitary adenoma can be demonstrated, although sometimes these patients have suprasellarly extending lesions due to somatotroph hyperplasia.

The diagnosis of ectopic GHRH syndrome cannot be made using function tests, although many of these patients do not respond to exogenous GHRH administration but show hypersecretion of GH after TRH. The diagnosis of ectopic GHRH secretion is made by the measurement of elevated peripheral GHRH levels (Fig. 1), which are in the nanogram range, in contrast



**Figure 2** Different causes of acromegaly. In more than 99% of all cases, a monoclonal pituitary tumor can be demonstrated (classical acromegaly). The ectopic GHRH syndrome is the second most common cause of GH hypersecretion. Eutopic GHRH secretion from hypothalamic lesions cannot be differentiated from classical acromegaly before operative therapy. The extremely rare ectopic GH syndrome has been demonstrated unequivocally in only two cases. SRIF, somatotrophic hormone release inhibiting factor.

to classical acromegaly, in which GHRH levels do not exceed 100 pg/ml. Often, the ectopic GHRH source can be localized by somatostatin receptor scintigraphy since these tumors seem to express somatostatin receptor types 2 and 5.

In contrast, patients with hypothalamic GHRH-secreting tumors, so-called eutopic GHRH syndrome, do not have elevated GHRH levels in the periphery (Fig. 2). This diagnosis is usually made after transphenoidal operative therapy. In these cases, no tumors which are separated from the anterior pituitary lobe, but in which somatotroph hyperplasia is found, are sometimes intermingled with neuroendocrine tissue expressing GHRH. Occasionally, GHRH-producing cells are found in the pituitary specimen interspersed within the somatotroph cells.

The ectopic GH syndrome is extremely rare. Only two cases, one with pancreatic cancer and the other with non-Hodgkin lymphoma (both shown to produce and secrete GH), have been reported. The diagnosis can be suspected when GH levels are elevated, there is no evidence of a pituitary lesion (either adenoma or somatotroph hyperplasia), and GHRH levels are normal.

When there is a family history of acromegaly, one has to consider familial acromegaly or multiple endocrine neoplasia type 1. In the latter case, one needs to exclude primary hyperparathyroidism and islet cell tumors.

Occasionally, acromegaly occurs as part of the McCune-Albright syndrome, with very severe bone deformations caused by acromegaly accompanied by osteofibrotic bone disease, which is the hallmark of this rare syndrome.

### See Also the Following Articles

Acromegaly, Clinical Features of • Acromegaly, Therapy for • Growth Hormone (GH) • Insulin-like Growth Factors • McCune-Albright Syndrome • Prolactin (PRL)

### Further Reading

- Giustina, A., Casanueva, F. F., *et al.* (2000). Consensus: Criteria for cure of acromegaly: A consensus statement. *J. Clin. Endocrinol. Metab.* **85**, 526–529.
- Harris, A. G. (1996). "Acromegaly and Its Management." Lippincott-Raven, Philadelphia.
- Lamberts, S. W. J. (1998). Acromegaly. In "Clinical Endocrinology" (A. Grossman, ed.), 2nd ed., pp. 170–183. Blackwell, Oxford, UK.
- Landolt, A. M., Vance, M. L., and Reilly, P. S. (eds.) (1996). "Pituitary Adenomas." Churchill Livingstone, New York.
- Losa, M., and Werder, K. von (1997). Pathophysiology and clinical aspects of the ectopic GH-releasing hormone syndrome. Review. *Clin. Endocrinol.* **47**, 123–135.
- Melmed, S. (ed.) (2002). "The Pituitary." Blackwell, Oxford, MA.
- Wass, J. (ed.) (2001). "Handbook of Acromegaly." Bioscientifica, Bristol, UK.



# Acromegaly, Therapy for

Anat Ben-Shlomo and Shlomo Melmed

University of California, Los Angeles School of Medicine, Cedars-Sinai Research Institute, Los Angeles, California, United States

## Glossary

**arthralgia** Pain in a joint.

**bradycardia** A slower than normal contraction of the heart muscle causing a pulse of less than 60 beats per minute.

**dopamine** A hormone-like substance that is an important neurotransmitter in both the central and peripheral nervous systems; occurs as an intermediate in epinephrine and norepinephrine biosynthesis.

**gallstone** An amalgamous concretion, commonly of cholesterol crystals, bilirubin, and protein, that is formed in the gallbladder or bile duct.

**A** cromegaly, a somatic growth and proportion disorder, is a rare and insidious disease caused by growth hormone (GH)-secreting pituitary tumors or (rarely) extrapituitary disorders. Elevated levels of GH and insulin-like growth factor-1 (IGF-1) are the hallmarks of this syndrome. Clinical manifestations include skeletal and soft tissue growth and deformations as well as cardiac, respiratory, neuromuscular, endocrine, and metabolic complications. Early diagnosis and aggressive control of the disease significantly attenuate or even abolish the increased morbidity and mortality from the disease. Although transsphenoidal surgery is considered the primary treatment of choice, it is apparent that complete macroadenoma resection, especially if invasive, is difficult even for skilled surgeons, hence their low rate of biochemical control.

## INTRODUCTION

When using strict criteria for disease control to define a cure for acromegaly (e.g., random serum growth hormone [GH] levels < 2.5 ng/ml, GH levels after oral glucose tolerance test [OGTT] < 1 ng/ml, gender- and age-matched insulin-like growth factor-1 [IGF-1] levels in the normal range), approximately 70% of patients were in remission during the short period after surgical intervention alone and 40% were in remission up to 16 years postsurgery. Radiotherapy,

either conventional external deep X-ray or heavy-particle (proton-beam) irradiation, is reserved mainly as adjuvant therapy for patients who have failed surgical or medical treatment. Therefore, pharmacological treatment has assumed more importance in managing patients with acromegaly. This article briefly discusses the pharmacotherapy available for acromegaly, including somatostatin analogues, GH receptor antagonists, and dopamine agonists.

## SOMATOSTATIN AND ITS SYNTHETIC ANALOGUES

Somatostatin, a peptide that suppresses GH secretion, was discovered in the hypothalamus some 30 years ago and was shown to reduce GH serum levels. The two major somatostatins that are enzymatically cleaved from the large preprosomatostatin precursor molecule are somatostatin-18 and somatostatin-28. These are produced in multiple tissues and induce their variable actions through a family of five major receptors, SSTR1 to SSTR5. Because GH cell pituitary adenomas express SSTRs (mainly SSTR5 and SSTR2 subtypes) that regulate GH secretion, the pharmacological use of somatostatin binding to these receptors has been employed for treating these tumors. Endogenous somatostatin-14 has a high affinity to all SSTRs, produced locally and rapidly degraded to prevent systemic effects. Exogenous administration of the peptide is not applicable for prolonged treatment of acromegaly due to an extremely short half-life and multiple side effects. The synthetic compounds octreotide (SMS201-995) and lanreotide (BIM23014) offer a greater metabolic stability and subtype specificity for the treatment of acromegaly (e.g., high affinity to SSTR2 and SSTR5 and moderate affinity to SSTR3 vs somatostatin-14).

Octreotide is a short-acting molecule, with a half-life of 2 h, that binds with high affinity to SSTR2 and, to a lesser extent, to SSTR5. This analogue was the first to be used clinically and has been proved safe and effective for medical therapy of acromegaly.

Reduction of GH levels is achieved with octreotide doses ranging from 50 to 500  $\mu\text{g}$  given subcutaneously every 8 h (in most patients, 100  $\mu\text{g}$  trice daily). The drug will achieve its maximal suppressive effect on GH levels 2 to 6 h after injection. With continuous treatment, the extent by which GH levels rise between injections is reduced. After 6 months of octreotide treatment, integrated mean GH levels decreased to less than 5 ng/ml in 53% of 115 patients with acromegaly, and IGF-1 levels were normalized in 68% of the patients. After 30 months of treatment, GH levels were reduced to less than 5 ng/ml in 65% and to less than 2 ng/ml in 40% of 97 patients. In a summary of 11 studies, serum IGF-1 levels were normalized in 53% (range 42–80%) of 417 patients with acromegaly treated with 100 to 1500  $\mu\text{g}$  octreotide daily for a period of 3 to 57 months. Improved clinical symptoms, including headache, perspiration, fatigue, arthralgia, and cystic acne, were observed in up to 78% of 115 patients treated with 750  $\mu\text{g}$  octreotide for 6 months and in up to 95% of 103 patients in another study. Adverse drug effects include diarrhea (60%), abdominal discomfort (45%), loose stool (32%), nausea, headache, dizziness, flatulence, and constipation. These symptoms usually subside after 1 to 3 weeks of treatment. Clinically nonsignificant bradycardia develops in approximately 25% of treated patients. Asymptomatic cholesterol gallstones form in approximately 25% of treated patients, usually during the first 2 years of therapy.

Octreotide LAR is a depot intramuscular preparation (D,L-lactide-coglycolide-glucose) that provides sustained slow release of the drug peaking at 28 days.

Injections of 20 to 30 mg octreotide LAR at 28-day intervals suppresses GH levels and generally achieves steady-state drug levels after two or three injections. As demonstrated in Table I, GH levels are reduced to less than 2.5 ng/ml in 70 to 90% of patients, and IGF-1 levels are normalized in approximately 65% or even as many as 88% of patients treated with 20 to 40 mg octreotide LAR monthly injections for 12 to 30 months. Octreotide responders respond well to the LAR form. Moreover, reduction of GH levels to less than 5 ng/ml was observed in more patients treated with octreotide LAR (94%) than in those having received subcutaneous octreotide (82%). IGF-1 levels were normalized in 65% of patients treated with octreotide LAR, as compared with 50% of those treated with subcutaneous octreotide. Marked improvement of carpal tunnel syndrome, paresthesias, perspiration, arthralgia, headache, and fatigue was reported, as was approximately 20% tumor shrinkage. Adverse effects are mostly mild, last for 1 to 2 days, and include diarrhea (45%), abdominal pain (32%), and flatulence (35%). These effects are transient, and their frequency decreases with treatment extension to 11, 3, and 8% of patients, respectively. Gallbladder abnormalities that were apparent in 26% of patients treated for 30 months included asymptomatic cholelithiasis, sediment, sludge, and biliary or gallbladder dilatation. More recently, a case of probable partial tachyphylaxis to both depot preparations of somatostatin was reported along with demonstration of antisomatostatin analogue antibodies.

Slow-release lanreotide is a cyclic octapeptide somatostatin analogue administered intramuscularly

**Table I** Summary of Prospective Open-Label Studies on the Effects of Long-Term Treatment (3–89 months) of Patients with Acromegaly with Octreotide LAR, Lanreotide SR, and Pegvisomant on Serum GH and/or IGF-1 Levels

	Number of patients	Treatment outcome		
		GH < 5 ng/ml	GH < 2.5 ng/ml	Normalized IGF-1
Lanreotide SR	298 <sup>a</sup>	225 (76)	197 (53)	297 (55)
	369 <sup>b</sup>			
	538 <sup>c</sup>			
Octreotide LAR	360 <sup>a</sup>	337 (94)	260 (59)	308 (69)
	439 <sup>b</sup>			
	449 <sup>c</sup>			
Pegvisomant	208			193 (93)

*Note.* Percentages are in parentheses.

<sup>a</sup>Total number of patients who achieved GH levels of < 5 ng/ml.

<sup>b</sup>Total number of patients who achieved GH levels of < 2.5 ng/ml.

<sup>c</sup>Total number of patients who achieved normal IGF-1 levels.

at a dose of 30 mg every 10 or 14 days. To summarize 10 open-labeled prospective studies, 71% of patients treated with lanreotide for at least 1 year exhibited GH levels less than 5 ng/ml (43% < 2.5 ng/ml) and 58% exhibited normal IGF-1 levels. Of 23 newly diagnosed, nonoperated or irradiated patients treated with lanreotide SR, 5 (22%) had significant tumor shrinkage (> 20%) after 6 months of therapy. Clinical improvement of perspiration, headache, and tissue swelling was reported in 4 of 13, 4 of 6, and 5 of 5 patients, respectively, after 1 month of treatment. Arthralgia improved in 3 of 5 patients after 9 months of therapy. Patients treated with long-term subcutaneous octreotide and then subsequently with lanreotide reported similar improvement with both therapies, including improved headache (81% of 22 patients), paresthesia (73%), and soft tissue swelling (61%). Lanreotide Autogel, a new delivery formulation of lanreotide administered by deep subcutaneous injection at a dose of 60 to 120 mg every 28 days, was introduced recently. Studied in 107 patients with acromegaly for 3 months, lanreotide Autogel proved to be at least as efficacious and well tolerated as 30 mg lanreotide. Improvement of clinical symptoms, including headache (30%), night sweats (20%), and joint pain (40%), was reported after 3 months of treatment with both preparations. Similar clinical improvement was reported with lanreotide SR during long-term treatment. Side effects include mainly diarrhea (38%), abdominal pain (22%), gallbladder lithiasis or sludge (27%), and sludge (11%).

## GROWTH HORMONE RECEPTOR ANTAGONIST

Pegvisomant (B2036-PEG) is a genetically engineered analogue of human GH. This mutated GH not only prevents the homodimerization of the two GH receptors, thereby preventing signaling, but it also has a higher affinity to growth hormone-binding protein (GHBP), thereby prolonging its circulating half-life. This GH receptor antagonist is not yet approved for clinical use in the United States. Subcutaneous daily injection of up to 40 mg normalized serum IGF-1 levels in 89% of 80 patients treated for 12 weeks and in 97% of 152 patients treated for 12 to 18 months. The biochemical improvement was accompanied by a clinical alleviation of soft tissue swelling, excessive perspiration, and fatigue. Serum GH levels were increased nearly twofold in the treated patients, and serum anti-GH antibodies were detected in 8% of the patients after 12 weeks of therapy and in 17%

after longer treatment. Adverse effects included headache (26%); infection (mainly upper respiratory tract infection) (33%); injection site reaction (11%); pain in scalp, neck, shoulders, arms, and/or legs (23%); diarrhea (13%); asthenia (13%); arthralgia (12%); sinusitis (10%); and hypercholesterolemia (14%). In addition, one patient exhibited a significant increase in transaminases levels. No tumor shrinkage was observed; moreover, two patients had an increase in tumor size under treatment (1.6- and 1.8-fold increases). Despite being the most effective of all drugs at reducing IGF-1, and despite the promising results when used as adjunct therapy, there is insufficient data supporting the use of pegvisomant as primary pharmacotherapy for acromegaly and safety for long-term treatment.

## DOPAMINE AGONISTS

Dopamine agonists decrease GH levels in some patients with acromegaly in contrast to those in normal individuals. Bromocriptine, long-acting bromocriptine LAR, pergolide, and cabergoline are ergot derivative dopamine agonists, whereas quinagolide is a nonergot derivative. Only bromocriptine and quinagolide are approved for clinical use in the United States. In contrast to their dramatic effect on prolactinomas, their effect on GH-secreting adenomas is limited and largely less effective than somatostatin analogues. GH levels decreased in 20% and IGF-1 levels decreased in 5 to 10% of patients treated with high doses of bromocriptine (up to 60 mg daily) and in 34 and 43% of those receiving cabergoline (up to 7 mg weekly) and quinagolide (up to 0.6 mg daily), respectively. Because of high drug doses required to achieve hormone reductions, the incidence of adverse effects is high and often unacceptable to patients. Side effects include gastrointestinal discomfort, including transient nausea and vomiting, dizziness (due to postural hypotension), headache, nasal congestion, and mood disorders. Addition of dopamine agonists, especially longer acting ones, may improve therapeutic efficacy of other modes of therapy for the treatment of acromegaly.

## CONCLUSION

The main goal of acromegaly therapy is tight control of GH and IGF-1 levels, hence reducing GH serum levels to less than 1 ng/ml after OGTT, normalizing age- and gender-matched IGF-1 levels, and reducing tumor mass as much as possible. Treatment of the disease's complications is also of great importance.

The preferred primary treatment is transsphenoidal surgery by an experienced surgeon. Patients who were not cured by surgery, or who cannot or do not wish to be operated on, may be controlled by pharmacotherapy. First choice for medical treatment is the long-acting somatostatin analogues octreotide LAR and lanreotide SR. Other somatostatin analogues, such as SOM203, are at different stages of development and are not yet in clinical use. Dopamine agonists can be added, especially if the patient has coexisting hyperprolactinemia. Until side effects and efficiency of GH receptor antagonist are better established, this treatment is reserved for patients who cannot be cured by surgery and/or somatostatin and dopamine agonists.

### See Also the Following Articles

Acromegaly, Clinical Features of • Acromegaly, Diagnosis of • Growth Hormone (GH) • Insulin-like Growth Factors • Pituitary Tumors, Clonality • Pituitary Tumors, Molecular Pathogenesis

### Further Reading

- Abs, R., Verhelst, J., Maiter, D., van Acker, K., Nobels, F., Coolens, J. L., Mahler, C., and Beckers, A. (1998). Carbegoline in the treatment of acromegaly: A study in 64 patients. *J. Clin. Endocrinol. Metab.* **83**, 374–378.
- Ambrosio, M. R., Franceschetti, P., Bondanelli, M., Doga, M., Maffei, P., Baldelli, R., Tamburrano, G., Siculo, N., Giustina, A., and degli Uberti, E. C. (2002). Efficacy and safety of the new 60-mg formulation of the long-acting somatostatin analog lanreotide in the treatment of acromegaly. *Metabolism* **51**, 387–393.
- Attanasio, R., Barausse, M., and Cozzi, R. (2001). GH/IGF-1 normalization and tumor shrinkage during long-term treatment of acromegaly by lanreotide. *J. Endocrinol. Invest.* **24**, 209–216.
- Ayuk, J., Stewart, S. E., Stewart, P. M., and Sheppard, M. C. (2002). Long-term safety and efficacy of depot long-acting somatostatin analogs for the treatment of acromegaly. *J. Clin. Endocrinol. Metab.* **87**, 4142–4146.
- Baldelli, R., Colao, A., Razzore, P., Jaffrain-Rea, M. L., Marzullo, P., Ciccarelli, E., Ferretti, E., Ferone, D., Gaia, D., Camanni, F., Lombardi, G., and Tamburrano, G. (2000). Two-year follow-up of acromegalic patients treated with slow release lanreotide (30 mg). *J. Clin. Endocrinol. Metab.* **85**, 4099–4103.
- Caron, P., Beckers, A., Cullen, D. R., Goth, M. I., Gutt, B., Laurberg, P., Pico, A. M., Valimaki, M., and Zgliczynski, W. (2002). Efficacy of the new long-acting formulation of lanreotide (lanreotide Autogel) in the management of acromegaly. *J. Clin. Endocrinol. Metab.* **87**, 99–104.
- Colao, A., Ferone, D., Marzullo, P., et al. (2001). Long-term effects of depot long-acting somatostatin analog octreotide on hormone levels and tumor mass in acromegaly. *J. Clin. Endocrinol. Metab.* **86**, 2779–2786.
- Herman-Bonert, V. S., Zib, K., Scarlett, J. A., and Melmed, S. (2000). Growth hormone receptor antagonist therapy in acromegalic patients resistant to somatostatin analogs. *J. Clin. Endocrinol. Metab.* **85**, 2958–2961.
- Ho, K. K. (2001). Place of pegvisomant in acromegaly. *Lancet* **358**, 1743–1744.
- Kendall-Taylor, P., Miller, M., Gebbie, J., Turner, S., and Al-Maskari, M. (2000). Long-acting octreotide LAR compared with lanreotide SR in the treatment of acromegaly. *Pituitary* **3**, 61–65.
- Melmed, S. (1998). Tight control of growth hormone: An attainable outcome for acromegaly treatment. *J. Clin. Endocrinol. Metab.* **83**, 3409–3410.
- Melmed, S., Casanueva, F. F., Cavagnini, F., Chanson, P., Frohman, L., Grossman, A., Ho, K., Kleinberg, D., Lamberts, S., Laws, E., Lombardi, G., Vance, M. L., von Werder, K., Wass, J., and Giustina, A. (2002). Consensus: Guidelines for acromegaly treatment. *J. Clin. Endocrinol. Metab.* **87**, 4054–4058.
- Pradhananga, S., Wilkinson, I., and Ross, R. J. M. (2002). Receptor antagonists: Pegvisomant—Structure and function. *J. Mol. Endocrinol.* **29**, 11–14.
- Shimon, I., Taylor, J. E., Dong, J. Z., Bitonte, R. A., Kim, S., Morgan, B., Coy, D. H., Culler, M. D., and Melmed, S. (1997). Somatostatin receptor subtypes specificity in human fetal pituitary cultures: Differential role of SSTR2 and SSTR5 for growth hormone, thyroid-stimulating hormone, and prolactin regulation. *J. Clin. Invest.* **99**, 789–798.
- Shimon, I., Yan, X., Taylor, J. E., Weiss, M. H., Culler, M. D., and Melmed, S. (1997). Somatostatin receptor (SSTR) subtype-selective analogues differentially suppress *in vitro* growth hormone and prolactin in human pituitary adenomas: Novel potential therapy for functional pituitary tumors. *J. Clin. Invest.* **100**, 2386–2392.
- Trainer, P. J., Drake, W. M., Katznelson, M. B. L., Freda, P. U., Herman-Bonert, V., van-der Lely, A. J., Dimaraki, E. V., Stewart, P. M., Friend, K. E., Vance, M. L., Besser, G. M., and Scarlett, J. A. (2000). Treatment of acromegaly with growth hormone-receptor antagonist pegvisomant. *N. Engl. J. Med.* **342**, 1171–1177.



# ACTH (Adrenocorticotropin Hormone)

George Mastorakos

University of Athens, Athens, Greece

Evangelia Zapanti

Alexandra Hospital, Athens, Greece

## Glossary

**adrenocorticotropin hormone (ACTH)** A 39-amino-acid peptide hormone (MW 45,000) that is part of the proopiomelanocortin precursor molecule. It controls the function of the adrenal cortex.

**proopiomelanocortin (POMC)** A precursor molecule (MW 28,500) that, in the anterior lobe of the pituitary, is processed to adrenocorticotropin hormone and  $\beta$ -lipotropin and further processed to  $\beta$ -endorphin, an endogenous opioid peptide of 31 amino acids.

**A**drenocorticotropin is the hypophyseal hormone that controls the function of the adrenal cortex.

## BIOCHEMISTRY

Adrenocorticotropin hormone (ACTH) is a 39-amino-acid peptide hormone (MW 45,000) that is part of the proopiomelanocortin (POMC) precursor molecule (MW 28,500). The POMC gene is located on chromosome 2. Proopiomelanocortin is expressed in the brain, skin, and immune system and in the anterior and intermediate lobes of the pituitary gland. In the anterior lobe, POMC is processed to ACTH and  $\beta$ -lipotropin ( $\beta$ -LPH), which is further processed to  $\beta$ -endorphin, an endogenous opioid peptide of 31 amino acids (Fig. 1). In the intermediate lobe, ACTH is processed to  $\alpha$ -melanocyte-stimulating hormone ( $\alpha$ -MSH) (ACTH 1-13) and corticotropin-like intermediate lobe peptide (CLIP) (ACTH 18-39). In species with developed intermediate lobes (rat and sheep), these fragments are secreted, whereas in humans they are normally found only during fetal life. The  $\text{NH}_2$ -terminal amino acids of ACTH are identical in all species thus far studied, but there are species differences in the COOH terminus of the molecule.

Interestingly, only the first 18 amino acids beginning with the  $\text{NH}_2$ -terminal amino acids of ACTH are required for biological activity. In the brain, POMC is expressed in the arcuate nucleus of the hypothalamus and the nucleus tractus solitarius of the caudal medulla. In the brain, POMC generates a range of bioactive peptides, including ACTH,  $\beta$ -endorphin, and  $\alpha$ -,  $\beta$ -, and  $\gamma$ -MSH.  $\beta$ -Lipotropin, a fragment with 91 amino acids, which contains  $\beta$ -MSH (41-58),  $\gamma$ -LPH (1-58), and  $\beta$ -endorphin, is secreted in equimolar quantities with ACTH. The endogenous opioid peptides  $\alpha$ -,  $\beta$ -, and  $\gamma$ -endorphin are derived from POMC and are composed of amino acids 61-76, 61-91, and 61-77 of  $\beta$ -LPH, respectively.

The biological effects of POMC-derived peptides are diverse and are largely mediated through melanocortin (MC) receptors (R), five of which have been described. MC1R, MC2R, and MC5R have established roles in the pigmentation in the skin, adrenal steroidogenesis, and thermoregulation, respectively. Proopiomelanocortin-derived peptides are also found in the placenta, pancreas, testes, and gastric antrum.

## PATHOLOGY AND EMBRYOLOGY

The synthesis and processing of POMC to ACTH and to the other POMC derivatives are executed within the corticotroph cells of the anterior pituitary lobe. This cell type was originally considered chromophobic and later was shown by light microscopic studies to be basophilic (Fig. 2). Corticotrophs represent 15-20% of adenohypophyseal cells. The embryological origin of corticotrophs is the intermediate lobe, but groups of cells migrate during development into regions of the anterior and posterior lobes. Ultrastructural studies have shown different morphologies in these two groups of cells, reflecting differences in the processing enzymes responsible for the production of the above-mentioned hormones. The cells in the anterior lobe appear irregularly shaped with

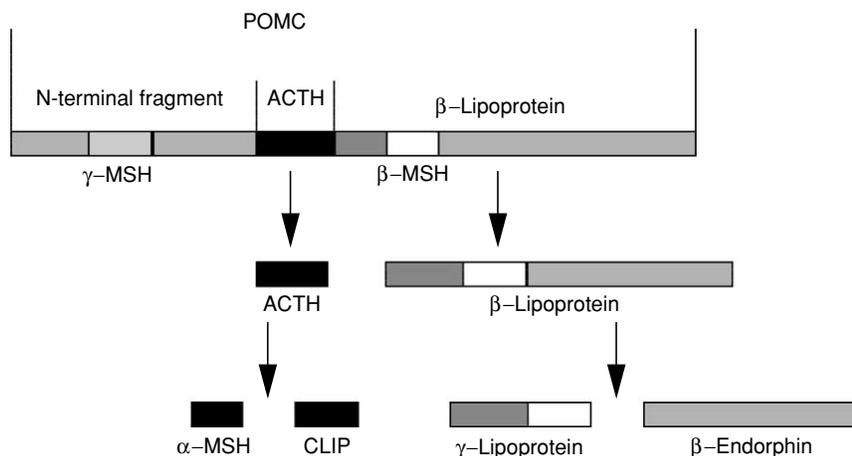


Figure 1 Enzymatic process of POMC in the pituitary.

sparse secretory granules and poor staining, whereas those of the intermediate and posterior lobes exhibit dense granulation. In states of glucocorticoid excess (endogenous or exogenous), corticotrophs undergo degranulation and microtubular hyalinization, known as Crooke's hyaline degeneration. In adrenal insufficiency, the corticotroph cells in the anterior lobe increase in number and the densely granulated cells of the intermediate and posterior lobes decrease in number.

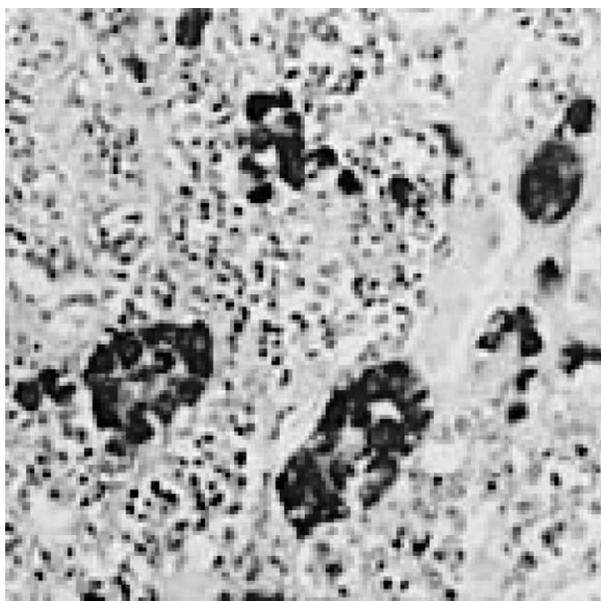


Figure 2 Corticotroph cells in the anterior pituitary, positively stained immunohistochemically with a basophilic stain.

## MEASUREMENT OF ACTH

The radioimmunoassay for the measurement of plasma ACTH is a highly sensitive and reliable assay for clinical use. Assays using other methodologies are also widely employed (i.e., enzyme-linked immunosorbent assay, chemiluminescence). The basal morning concentration of ACTH ranges from 9 to 80 pg/ml. The episodic secretion of ACTH causes fluctuations in plasma ACTH and cortisol levels. ACTH secretion has a distinct diurnal rhythm, with peak levels in the early morning and the lowest levels at approximately midnight. The interpretation of ACTH values requires a simultaneous cortisol determination. Provided that adrenocortical function is intact, plasma cortisol measurements, from a practical viewpoint, are a reliable index of ACTH secretion. There are other assays available for measuring ACTH (bioassays, radioreceptor assays, cytochemical assays), but measurement of ACTH by these techniques is performed only for research purposes, because of complexity and cost. The half-life of ACTH depends on the assay used for its measurement. Bioactive ACTH disappears from the circulation more rapidly (half-life of 3–9 min) than does immunoreactive ACTH (half-life of 7–12 min).

## PHYSIOLOGY

### Control of ACTH Secretion

The secretion of ACTH is controlled by an inherent diurnal rhythmicity; it is augmented by noxious stimuli that are neurally, hormonally, and biochemically

mediated, which is termed stress (open-loop component), and is inhibited by glucocorticoids (closed-loop, negative feedback). In normal circumstances, the daily secretion of ACTH and cortisol is episodic and variable. The highest burst of activity is observed in the early morning hours. Thereafter, the release of both ACTH and cortisol occurs only in 7–15 episodes per day and the levels of both hormones gradually decrease, reaching a nadir at approximately midnight.

### **Diurnal Rhythmicity**

The basis of the diurnal rhythm is poorly understood. There is evidence for the involvement of at least three factors in the regulation of the diurnal rhythmicity of ACTH: (1) intrinsic rhythmicity of the secretion of the corticotropin-releasing factor (CRH) as well as vasopressin (AVP); (2) light–dark exposure; and (3) feeding times. Diurnal variation frequently disappears during periods of stress and depression and is also changed by conditions that affect cortisol metabolism (liver disease, chronic renal failure, alcoholism).

### *Intrinsic Rhythmicity of Hypothalamic CRH and AVP*

Tests for the evaluation of the CRH secretion pattern have shown a diurnal intrinsic rhythmicity that persists even in hypophysectomized animals that are deprived of ACTH and glucocorticoid feedback. This hypothalamic rhythmicity appears to be neuronal but not hormonal. On the other hand, the presence of diurnal rhythmicity of ACTH in women during pregnancy when the increased circulating levels of placenta-derived CRH do not show a nycthemeral rhythm strongly suggests that hypothalamic AVP secretion plays a role in this diurnal regulation.

### *Light–Dark Cycles*

Secretory episodes of ACTH increase between the third and fifth hours of sleep and peak in the morning during the first several hours of wakefulness. This rhythm usually appears after the first year of life but may not be established until the age of 8 years. Reversal of the normal asleep–awake patterns, as occurs when an individual moves to a distant time zone, is followed by a corresponding change in the diurnal pattern of ACTH secretion over the course of 2 to 3 weeks.

### *Feeding Cycles*

Experiments in rats have shown that feeding schedule is more important than the light–dark cycle in determining the glucocorticoid diurnal secretory pattern. It appears that glucocorticoids tend to be released during fasting and decrease with feeding. Less is

known about the effects of feeding schedule on ACTH release in humans.

### **Open-Loop Control (Stress)**

The open-loop component of ACTH control may be initiated by noxious stimuli of various sorts, all of which represent types of physical or emotional stress, such as pain, fever, trauma, hypoglycemia, hypoxia, surgery, anxiety, and depression. All of these stimuli stimulate the secretion of ACTH via the release of CRH. Corticotropin-releasing hormone and AVP are probably the two major physiologic secretagogues of hypophyseal ACTH. Immunoreactive CRH is found in the human hypothalamus in the paraventricular, supraoptic, and infundibular nuclei and also in the human thalamus, cortex, cerebellum, and pons. However, most human CRH-secreting neurons are located in the anterior portion of the paraventricular nucleus and their nerve endings project to the external layer of the median eminence, where CRH is released into the portal hypophyseal circulation. The same neuronal bodies in the paraventricular nucleus also produce AVP. These AVP neurons are probably the most important in vasopressin control of ACTH release, but the major site of vasopressin neurons is the supraoptic nucleus. These vasopressin neurons are usually of the magnocellular type and most of them project to the neural lobe.

There is strong evidence that both  $\alpha$ - and  $\beta$ -adrenergic stimuli, cytokines, angiotensin II, and opiates are involved in the regulation of ACTH secretion. During infection, autoimmune processes, or trauma, a complex cascade of events ensues, characterized by fever, circulation of cytokines, and alterations in acute-phase proteins in plasma that are important to initiate, propagate, and terminate host defense mechanisms. In addition, it has been known for several decades that activation of the hypothalamic–pituitary–adrenal (HPA) axis occurs in parallel. It has become apparent that several mediators of inflammation play a major role in this phenomenon. Among all cytokines, three [tumor necrosis factor  $\alpha$  (TNF $\alpha$ ), interleukin-1 (IL-1), and interleukin-6 (IL-6)] are responsible for most of the stimulation of the HPA axis that is associated with the immune/inflammatory response. These three cytokines are produced at inflammatory sites and elsewhere in response to inflammation. Tumor necrosis factor  $\alpha$ , which has a tumoricidal activity and is responsible for cachexia, is the first to appear in the inflammatory cascade of the events and stimulates both IL-1 and IL-6; similarly, IL-1 stimulates both TNF $\alpha$  and IL-6. In contrast, IL-6, which participates in a major fashion in the

acute-phase reaction, inhibits the secretion of both of the other cytokines. All three inflammatory cytokines have been shown to activate the HPA axis, i.e., ACTH secretion *in vivo*, alone or in synergy with one another. This effect can be blocked significantly with CRH-neutralizing antibodies, glucocorticoids, and prostanoïd synthesis inhibitors. When administered to humans, both IL-1 and TNF $\alpha$  have significant toxicity, including fever, general malaise, and hypotension, at the doses needed to activate the HPA axis. In the past, it has been demonstrated that IL-6, with its ability to inhibit the two other inflammatory cytokines and its modest toxicity in experimental animals, was a potent stimulator of the HPA axis in humans, causing an impressively marked and prolonged elevation of plasma ACTH and cortisol when administered either subcutaneously or intravenously. The elevations of ACTH and cortisol attained after stimulation with IL-6 were well above those observed with maximal stimulatory doses of CRH, suggesting that parvocellular AVP and other ACTH secretagogues were also stimulated by this cytokine. In a dose-response study, maximal levels of ACTH were seen at doses at which no peripheral AVP levels were increased. At higher doses, however, IL-6 stimulated peripheral elevations of AVP, indicating that this cytokine might also be able to activate magnocellular AVP-secreting neurons. This suggested that IL-6 might be involved in the genesis of the syndrome of inappropriate secretion of antidiuretic hormone, which is observed in the course of infectious or inflammatory diseases or during trauma. It has been shown that IL-6, in patients with head trauma (an aseptic inflammatory state) and a syndrome of inappropriate secretion of antidiuretic hormone, is quantitatively correlated with AVP.

In addition to their hypothalamic effects, the inflammatory cytokines can apparently directly stimulate pituitary ACTH and adrenal cortisol secretion. This may be related to the chronicity of the elevation of the inflammatory cytokines or may be a dose-related phenomenon. It is noteworthy that IL-1 and IL-6 are themselves produced in the anterior pituitary and adrenal glands, where they may have autocrine/paracrine effects.

#### **Closed-Loop Feedback**

The negative feedback control of ACTH secretion is mediated by cortisol, which exerts inhibitory effects on both the central nervous system and the pituitary. Negative feedback occurs via three mechanisms: (1) fast feedback, which is sensitive to changes in the levels of circulating cortisol, (2) intermediate

feedback, and (3) slow feedback, which is sensitive to the absolute cortisol level. Increased concentrations of glucocorticoids accelerate the progression from fast to slow feedback. Both fast feedback and intermediate feedback appear to be mediated by inhibition of the release of the existing CRH and ACTH rather than by inhibition of their synthesis. Slow feedback is characterized by decreased synthesis of ACTH, complete suppression of POMC gene transcription, and a lack of responsiveness of the corticotroph to the administration of CRH. The last result is mediated by the direct inhibitory effect of cortisol on the pituitary and there is evidence suggesting that this inhibitory effect of cortisol at the level of the pituitary constitutes the most important cortisol negative feedback in the physiological regulation of ACTH secretion. Glucocorticoids decrease the hypothalamic content of CRH and AVP and also decrease AVP mRNA content, but they have only a minimal effect on CRH mRNA. Electrophysiological studies have shown that there are hypothalamic as well as extrahypothalamic sites of cortisol feedback, which serve to suppress the release of CRH.

#### **Action of ACTH**

The adrenal cortex is the principal target organ for ACTH. ACTH stimulates the synthesis and release of steroids by binding to high-affinity plasma membrane receptors of adrenocortical cells. The ACTH-receptor interaction then activates adenylyl cyclase and therefore stimulates the production of intracellular cyclic AMP (cAMP). The cAMP formed activates a number of intracellular phosphoprotein kinases that mediate both acute and chronic effects on steroidogenesis.

#### **Acute and Chronic Actions of ACTH**

ACTH stimulates the synthesis and release of cortisol within 2 to 3 min by increasing free cholesterol formation as a consequence of increased cholesterol esterase activity and decreased cholesteryl ester synthetase activity. ACTH rapidly promotes the transport of cholesterol across the mitochondrial membranes, facilitates the binding of cholesterol to the cytochrome P450<sub>scc</sub>, and facilitates the release of newly synthesized pregnenolone from the mitochondria. ACTH also stimulates the release of adrenal mineralocorticoids and androgens, as well as the release of various intermediate products. Chronic actions of ACTH are exerted on both adrenal architecture and steroidogenesis. ACTH chronically stimulates low-density lipoprotein (LDL) uptake and metabolism and the synthesis of the LDL receptor and of other factors, so

it has tropic effects on all known early steps in steroidogenesis. Chronic effects of ACTH on steroidogenesis occur mainly by promoting the transcription of the genes that encode steroidogenic enzymes and other factors. ACTH increases the transcription of the genes for P450scc, P450c17, P450c21, and P450c11 and stimulates the accumulation of human P450scc mRNA and human P450scc activity. The exact mechanisms of ACTH stimulation of the side-chain cleavage enzyme P450scc remain to be elucidated. ACTH promotes both adrenal cellular hypertrophy and hyperplasia. ACTH at physiologic concentrations can promote the synthesis of insulin-like growth factor-II (IGF-II) and also the synthesis of basic fibroblast growth factor and epidermal growth factor, which may act with IGF-II to stimulate adrenal growth.

### See Also the Following Articles

ACTH,  $\alpha$ -MSH, and POMC, Evolution of • Adrenal Cortex, Anatomy • Adrenal Cortex, Physiology • Adrenal Suppression • Circadian Rhythms: Hormonal Facets • Corticotropin-Releasing Hormone, Family of • Glucocorticoids, Overview • Stress and Endocrine Physiology

### Further Reading

- Chrousos, G. P. (1995). The hypothalamic–pituitary–adrenal axis and the immune/inflammatory reaction. *N. Engl. J. Med.* **332**, 1351–1362.
- Desir, D., Van Cauter, E., Fang, V., *et al.* (1981). Effects of “jet lag” on hormonal patterns. I. Procedures, variations in total plasma proteins and disruption of adrenocorticotropin–cortisol periodicity. *J. Clin. Endocrinol. Metab.* **52**, 628–641.
- DiBlasio, A. M., Voutilainen, R., Jaffe, R. B., and Miller, W. L. (1987). Hormonal regulation of mRNAs for P450scc (20,22 desmolase) and (17 $\alpha$ -hydroxylase/17,20 lyase) in cultured fetal adrenal cells. *J. Clin. Endocrinol. Metab.* **65**, 170–175.
- Favrod-Coune, C. A., Gaillard, R. C., Langevin, H., Jaquier, M. C., Doci, W., and Muller, A. F. (1986). Anatomical localization of corticotropin-releasing activity in the human brain. *Life Sci.* **39**, 2475–2481.
- Gionis, D., Ilias, I., Moustaki, M., *et al.* (2003). Hypothalamic–pituitary–adrenal axis and interleukin-6 activity in children with head trauma and syndrome of inappropriate secretion of anti-diuretic hormone. *J. Pediatr. Endocr. Metab.* **16**, 49–54.
- Grino, M., Dakine, N., Paulmyer-Lacroix, O., and Oliver, C. (2001). Ontogeny of the hypothalamic–pituitary–adrenal axis. In “Adrenal Disorders” (A. N. Margioris and G. P. Chrousos, eds.), pp. 1–10. Humana Press, Totowa, NJ.
- Jefcoate, C. R., McNamara, B. C., and DeBartolomeis, M. S. (1986). Control of steroid synthesis in adrenal fasciculata cells. *Endocr. Res.* **12**, 315–350.
- Krieger, D. T. (1979). Rhythms in CRF, ACTH and corticosteroids. In “Endocrine Rhythms” (D. T. Krieger, ed.), pp. 123–142. Raven Press, New York.
- Malee, M. P., and Mellon, S. H. (1991). Zone-specific regulation of two distinct messenger RNAs for P450c11 (11/18-hydroxylase) in the adrenals of pregnant and non-pregnant rats. *Proc. Natl. Acad. Sci. USA* **88**, 4731–4735.
- Mastorakos, G., Chrousos, G. P., and Weber, J. S. (1993). Recombinant interleukin-6 activates the hypothalamic–pituitary–adrenal axis in humans. *J. Clin. Endocrinol. Metab.* **77**, 1690–1694.
- Mesiano, S., Mellon, S. H., Gospodarowicz, D., DiBlasio, A. M., and Jaffe, R. B. (1991). Basic fibroblast growth factor expression is regulated by corticotropin in the human fetal adrenal: A model for adrenal growth regulation. *Proc. Natl. Acad. Sci. USA* **88**, 5428–5432.
- Miller, W. L., and Tyrell, J. B. (1995). The adrenal cortex. In “Endocrinology and Metabolism” (P. Felig, J. D. Baxter, and L. A. Frohman, eds.), International Edition, pp. 555–680. McGraw Hill, New York.
- Smith, A. L., and Funder, J. W. (1988). Proopiomelanocortin processing in the pituitary, central nervous system, and peripheral tissues. *Endocr. Rev.* **9**, 159–179.
- Tanaka, K., Nicholson, W. E., and Orth, D. N. (1978). The nature of the immunoreactive lipotropins in human plasma and tissue extracts. *J. Clin. Invest.* **62**, 94–104.
- Whitfield, P. L., Seeburg, P. H., and Shine, J. (1982). The human proopiomelanocortin gene: Organization, sequence, and interspersed with repetitive DNA. *DNA* **1**, 133–143.



# ACTH, $\alpha$ -MSH, and POMC, Evolution of

Robert M. Dores and Phillip B. Danielson  
University of Denver, Denver, Colorado, United States

## Glossary

**adrenocorticotropin (ACTH)** An anterior pituitary polypeptide hormone that induces the production of cortisol by cells of the adrenal cortex.

**$\beta$ -lipotropin ( $\beta$ -LPH)** A biosynthetic intermediate derived from the proopiomelanocortin precursor that contains the sequence of  $\beta$ -melanocyte-stimulating hormone and  $\beta$ -endorphin.

**melanocortins** Adrenocorticotropin,  $\alpha$ -melanocyte-stimulating hormone (MSH),  $\beta$ -MSH,  $\gamma$ -MSH, and  $\delta$ -MSH.

**$\alpha$ -melanocyte-stimulating hormone ( $\alpha$ -MSH),  $\beta$ -MSH,  $\gamma$ -MSH,  $\delta$ -MSH** Polypeptide hormones produced in the intermediate pituitary that share the melanocortin core sequence HFRW.

**proopiomelanocortin (POMC)** The common precursor for melanocyte-stimulating hormone-related polypeptides and  $\beta$ -endorphin.

**P**roopiomelanocortin is the common precursor for the melanocortins and  $\beta$ -endorphin. This gene is expressed in corticotropic cells of the anterior pituitary and melanotropic cells of the intermediate pituitary. As a result of differential post-translational processing, these two regions of the pituitary will yield distinct sets of melanocortin-related and  $\beta$ -endorphin-related end products.

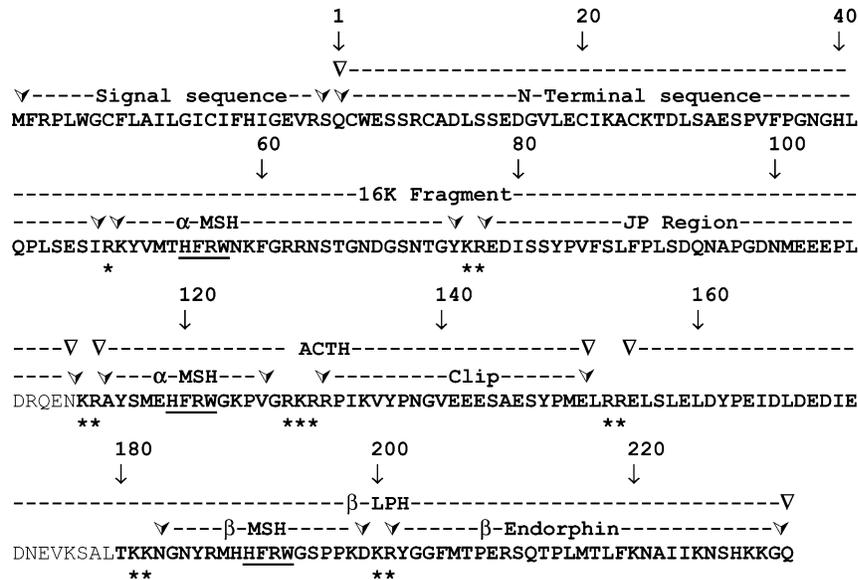
## INTRODUCTION

During the middle portion of the 20th century, prior to the implementation of immunocytochemistry and subsequent molecular biology procedures to analyze endocrine cells, several pituitary polypeptide hormones were biochemically characterized from extracts of the anterior and intermediate pituitary. These analyses revealed the surprising observation that adrenocorticotropin (ACTH) and  $\beta$ -lipotropin ( $\beta$ -LPH), polypeptides characterized from the anterior pituitary, shared the amino acid sequence motif HFRW with

$\alpha$ -melanocyte-stimulating hormone ( $\alpha$ -MSH) and  $\beta$ -melanocyte-stimulating hormone ( $\beta$ -MSH), polypeptides characterized from the intermediate pituitary. The fact that ACTH contained the complete sequence of  $\alpha$ -MSH and that  $\beta$ -LPH contained the complete sequence of  $\beta$ -MSH led to the hypothesis that all of these polypeptides had a common origin. Later immunocytochemical studies would show that ACTH-related immunoreactivity and  $\beta$ -LPH-related immunoreactivity were colocalized in the corticotropic cells of the anterior pituitary, whereas  $\alpha$ -MSH-related immunoreactivity and  $\beta$ -MSH-related immunoreactivity were colocalized in the melanotropic cells of the intermediate pituitary. From these observations, several research groups in the 1970s arrived at two conclusions: (1) ACTH,  $\beta$ -LPH,  $\alpha$ -MSH, and  $\beta$ -MSH must be derived from a common precursor and (2) the common precursor must undergo differential posttranslational processing to yield distinct sets of end products in the anterior pituitary and intermediate pituitary. The cloning and characterization of proopiomelanocortin (POMC) mRNA from the bovine pituitary by Nakanishi and colleagues in 1979 confirmed the common precursor hypothesis. In addition, that study revealed the presence of a third MSH sequence ( $\gamma$ -MSH) in the precursor. Since that time, the POMC gene has been analyzed in representative species from nearly every major group of vertebrate. This article focuses on the changes that have occurred in the POMC gene during the evolutionary radiation of the vertebrates.

## ORGANIZATION OF POMC IN TETRAPODS

The organization of the POMC precursor is remarkably conserved in tetrapods as diverse as amphibians and mammals. Tetrapod POMC can serve as a good model to illustrate the distribution of biologically active sequences and spacer regions within this precursor and to outline the major issues with respect to the origin and evolution of the POMC gene within



**Figure 1** The amino acid sequence of *Xenopus laevis* POMC. The deduced amino acid sequence for the pre-pro form of POMC (Accession No. X59370) is presented. Proposed endoproteolytic cleavage sites are marked by asterisks. MSH core sequences are underlined. The opioid core sequence for  $\beta$ -endorphin is Y<sup>202</sup>G<sup>203</sup>G<sup>204</sup>F<sup>205</sup>. NTS, N-terminal sequence; JP, joining peptide.

the phylum Chordata. Tetrapod POMC can be divided into three major regions: the 16K fragment region, the ACTH region, and the  $\beta$ -LPH region (Fig. 1). The 16K fragment is located in the N-terminal region of POMC immediately following the signal sequence. This region can be subdivided into the N-terminal sequence (NTS), the  $\gamma$ -MSH sequence, and the joining peptide sequence. The NTS (Q<sup>1</sup>-I<sup>47</sup>) contains four cysteine residues at positions 2, 8, 20, and 24. The number and location of these cysteine residues represent a feature common to all gnathostome POMC sequences and undoubtedly play a role in influencing the three-dimensional structure of the precursor. *In vitro* studies on developing mammalian pituitary cells indicate that the NTS can serve as a maturation factor to promote the development of prolactin-producing cells. Studies have not been performed *in vivo* to verify this activity. However, it is coincidental that in teleosts, vertebrates in which the anterior pituitary cell types segregate into zones, ACTH-producing cells are found in close proximity to prolactin-producing cells. The joining peptide region (E<sup>79</sup>-N<sup>111</sup>) is one of the most variable regions in POMC and is one of the spacer regions found in this precursor. No function has been found for this sequence. Located between the NTS and the joining peptide is the sequence of  $\gamma$ -MSH (K<sup>49</sup>-Y<sup>75</sup>).  $\gamma$ -MSH is one of three sequences within POMC that contains the MSH core sequence HFRW. Bioassay studies indicate that  $\gamma$ -MSH has weak

melanocyte-stimulating activity at nonphysiological concentrations. However, studies on rats indicate that this peptide may be involved in regulating blood pressure at sites outside of the central nervous system. The  $\gamma$ -MSH sequence is flanked by paired basic amino acids (R<sup>48</sup>K<sup>49</sup> and K<sup>76</sup>R<sup>77</sup>), and these sites undergo endoproteolytic cleavage in the intermediate pituitary but not in the anterior pituitary. In both mammals and anuran amphibians, the R<sup>48</sup>K<sup>49</sup> site is a functional monobasic cleavage site. Hence, for the *Xenopus laevis* POMC sequence presented in Fig. 1, the first residue in the  $\gamma$ -MSH sequence is K<sup>49</sup>. The reason for this particular cleavage event is not clear. Often, the presence of a proline residue C-terminal to a basic amino acid prevents the removal of the basic amino acid. However, for the R<sup>48</sup>K<sup>49</sup> sequence, the position C-terminal to K<sup>49</sup> is a tyrosine residue. In fact, as seen in Fig. 2, the tyrosine residue is conserved at this position in the  $\gamma$ -MSH sequences of vertebrates as diverse as sharks and amphibians. In contrast to the R<sup>48</sup>K<sup>49</sup> site, proteolytic cleavage at K<sup>76</sup>R<sup>77</sup> follows the conventional endoproteolytic mechanism and both amino acids are removed. Many gnathostome  $\gamma$ -MSH sequences have a potential internal cleavage site at R<sup>62</sup>R<sup>63</sup> that can be cleaved to yield an amidated form of  $\gamma$ -MSH that would correspond to K<sup>49</sup>-F<sup>60</sup>.

The ACTH region in the *X. laevis* POMC sequence extends from A<sup>113</sup> to L<sup>152</sup>. ACTH is a potent stimulator of cortisol production by adrenal cortical cells. This hormone plays a critical role in



found in all vertebrates at this position, apparently prevents cleavage at R<sup>131</sup>. The result of these proteolytic cleavage events yields ACTH(1–13) amide (A<sup>113</sup>–V<sup>126</sup>) and corticotropin-like intermediate lobe peptide (R<sup>131</sup>–L<sup>152</sup>) (CLIP) as products. ACTH(1–13) amide will undergo N-terminal acetylation to form  $\alpha$ -MSH, the second polypeptide sequence in amphibian POMC that has the MSH core sequence (H<sup>119</sup>F<sup>120</sup>R<sup>121</sup>W<sup>122</sup>).  $\alpha$ -MSH is a potent stimulator of physiological color change in several species of amphibians and reptiles. The function of CLIP is still an enigma.

The final major region of the POMC precursor is the  $\beta$ -LPH region (Fig. 1). Located in the C-terminal portion of POMC, this sequence has traditionally been divided into the  $\gamma$ -LPH sequence (E<sup>154</sup>–D<sup>199</sup>) and the  $\beta$ -endorphin sequence (Y<sup>202</sup>–Q<sup>232</sup>). The former sequence contains the  $\beta$ -MSH sequence (N<sup>183</sup>–D<sup>199</sup>), the third polypeptide with a MSH core sequence (H<sup>190</sup>F<sup>191</sup>R<sup>192</sup>W<sup>193</sup>) in tetrapod POMC.  $\beta$ -MSH has melanocyte-stimulating activity and appears to work in concert with  $\alpha$ -MSH to promote physiological color change in amphibians and reptiles. In all vertebrates in which MSH has been studied,  $\beta$ -MSH is an end product of the intermediate pituitary, but not of the anterior pituitary.

The other major product derived from  $\beta$ -LPH is  $\beta$ -endorphin. This polypeptide is an endogenous opiate-like chemical signal that functions as an inhibitory neurotransmitter when released from neurons located in the central nervous system. In the mammalian anterior pituitary,  $\beta$ -LPH is a major end product and  $\beta$ -endorphin is a minor end product. In the mammalian intermediate pituitary, just the opposite result is observed. In this tissue,  $\beta$ -LPH serves as a biosynthetic intermediate that is processed to yield N-terminally acetylated, C-terminally truncated forms of  $\beta$ -endorphin. The later peptides lack opiate receptor-binding activity.

The organization of tetrapod POMC is interesting for a number of reasons. This precursor contains three polypeptide sequences that have very distinct functions (ACTH,  $\alpha$ -MSH, and  $\beta$ -endorphin). The repeat of the MSH core sequence appears to be the result of a series of domain duplication events likely resulting from unequal crossover. This feature is not unique to POMC. Several neuropeptide precursors, such as proTRH or proenkephalin, contain repeats of a biologically important sequence. However, the presence of three MSH sequences in tetrapod POMC leads to the questions of when these sequences appeared and which of the polypeptides was the ancestral MSH sequence. To address these questions, a phylogenetic comparison of POMC sequences is required.

## PHYLOGENETIC ANALYSIS OF POMC SEQUENCES

The POMC gene has been analyzed in a jawless fish, two cartilaginous fishes, several ray-finned fishes, two lobe-finned fishes, several amphibians, and several mammals. Figure 2 provides a comparison of a few representative species. These species were selected because they represent five major taxonomic groups of vertebrates (i.e., jawless fish, cartilaginous fish, ray-finned fishes, lobe-finned fish, and tetrapods). The jawless fish were well established over 500 million years ago (mya) and are represented today by lamprey and hagfish. In the lamprey, *Petromyzon marinus*, two distinct POMC genes are expressed. The POM form of POMC, a gene expressed in the intermediate pituitary of this species, is presented in Fig. 2. The jawed vertebrates (gnathostomes) appear in the fossil record approximately 420 mya and two distinct groups were established by 400 million years ago: the cartilaginous fish, represented in Fig. 2 by the dogfish, and the bony fish. The bony fish radiated into two major groups: the ray-finned fish, represented by the gar and sockeye salmon, and the lobe-finned fish, represented by the Australian lungfish. The lungfish lineage can be traced back in the fossil record to 390 mya. Approximately 20 million years later, the first tetrapods were clearly present in the fossil record. *X. laevis* was included in Fig. 2 to provide a representative tetrapod POMC sequence. Two ray-finned fish POMC sequences are also included in Fig. 2 to illustrate some unique features associated with the radiation of the POMC gene in this group. The gar belongs to an older lineage of ray-finned fish that can be traced in the fossil record back to the Jurassic era. The sockeye salmon is a teleost, a relatively recent group of bony fish that first appears in the fossil record during the Cretaceous period.

A striking feature of the sequences presented in Fig. 2 is the presence of  $\alpha$ -MSH-like,  $\beta$ -MSH-like, and  $\beta$ -endorphin sequences in all taxa. There has been considerable divergence of the POMC sequence in vertebrates and without these conserved sites it would have been very difficult to align the agnathan POMC sequence to the gnathostome POMC sequences. Even with these conserved sites, it was still necessary to insert 11 gaps to facilitate the alignment of the sequences. The presence of  $\alpha$ -MSH-like,  $\beta$ -MSH-like sequences in the lamprey precursor indicates that the POMC gene must have been in the ancestral chordate. Indeed, this gene may have been in the ancestral eucoelomates. Based on this data set, it is not possible to ascertain when the  $\alpha$ -MSH/ $\beta$ -MSH duplication event occurred. One hypothesis would be

that the ACTH sequence appeared early in eucoelomate evolution and that  $\beta$ -MSH is the result of the duplication of the  $\alpha$ -MSH sequence. In the case of the lamprey, nucleotide insertions occurred in both the  $\alpha$ -MSH- and the  $\beta$ -MSH-coding regions of the lamprey POM gene. The gnathostome  $\alpha$ -MSH and  $\beta$ -MSH sequences are highly conserved in the taxa presented in Fig. 2. In addition, the  $\beta$ -endorphin region contains a remarkable number of conserved sites in all taxa. There must be selection pressure to retain certain features of the  $\beta$ -endorphin sequence that is not being exerted on other regions of the gene.

A sequence that is conspicuously absent from the lamprey POM sequence is  $\gamma$ -MSH. Based on this observation, it is reasonable to propose that the duplication event that yielded the  $\gamma$ -MSH sequence must have occurred in the ancestral gnathostomes. Although a  $\gamma$ -MSH-like sequence can be detected in dogfish, gar, and Australian lungfish POMC, this sequence is shorter than the tetrapod  $\gamma$ -MSH sequence. In addition, paired basic proteolytic cleavage sites are not found C-terminal to the  $\gamma$ -MSH sequence in the gar or lungfish sequences. In addition, the gar  $\gamma$ -MSH-like sequence does not have an intact MSH core sequence and the entire  $\gamma$ -MSH sequence is absent in teleost POMC. Apparently during the radiation of the ray-finned fish, selection pressures favored the gradual degeneration and eventual loss of the  $\gamma$ -MSH sequence.

Whereas the presence of  $\alpha$ -,  $\beta$ -, and  $\gamma$ -MSH is a feature common to many tetrapods, the dogfish POMC sequence has a fourth MSH sequence:  $\delta$ -MSH (D<sup>223</sup>-P<sup>237</sup>; Fig. 2). This sequence is located in a large insertion (V<sup>204</sup>-A<sup>242</sup>; Fig. 2) that appears to have occurred during the radiation of the cartilaginous fish and hence is limited to this group of gnathostomes.

The uniqueness of the  $\delta$ -MSH sequence is balanced by the high degree of primary sequence conservation observed for ACTH among the gnathostome taxa. For the gnathostome species presented in Fig. 2, 66% of the positions are identical in the ACTH sequence. In fact, 88% of the first 25 positions in gnathostome ACTH are

identical. For the  $\beta$ -endorphin region, 45% of the positions are identical in all taxa. This percentage rises to 67% when all gnathostome fish sequences are compared. Clearly, ACTH and  $\beta$ -endorphin must be physiologically important in all vertebrates.

## ORIGIN OF POMC

POMC is a member of the opioid/orphanin gene family that includes the proenkephalin gene, the prodynorphin gene, and the proorphanin gene. A unifying feature of this family is the presence of at least one Y(F)GGF sequence in each precursor. To date, members of this gene family have not been found in any prokaryotes and it appears that these genes are not found in pseudocoelomates such as *Caenorhabditis elegans*. Hence, the gene family appears to have evolved in the eucoelomates. Although POMC is clearly a member of this family because of the  $\beta$ -endorphin sequence, the presence of the melanocortin sequences (ACTH,  $\alpha$ -MSH,  $\beta$ -MSH,  $\gamma$ -MSH, and  $\delta$ -MSH) is a feature that is unique to the POMC gene. Since POMC is found in all vertebrate groups and perhaps is common to all chordates, the origin of the melanocortin sequences may provide the key to deciphering the genesis of this opioid/orphanin gene family.

## See Also the Following Articles

ACTH (Adrenocorticotrophic Hormone) • Pituitary Tumors, ACTH-Secreting

## Further Reading

- Cone, R. D. (ed.) (2000). "The Melanocortin Receptors." Humana Press, Totowa, NJ.
- Danielson, P. B., and Dores, R. M. (1999). Molecular evolution of the opioid/orphanin gene family. *Gen. Comp. Endocrinol.* **113**, 169–186.
- Hadley, M. E. (ed.) (1988). "The Melanotropic Peptides," Volumes 1, 2, and 3. CRC Press, Boca Raton, FL.
- Norman, A. W., and Litwack, G. (1997). "Hormones," 2nd ed. Academic Press, San Diego, CA.



Addison's Disease  
see Adrenal Insufficiency



# Adenylyl Cyclase

Ferenc A. Antoni

University of Edinburgh, Edinburgh, United Kingdom

## Glossary

### **cyclic nucleotide phosphodiesterases (PDEs)**

Enzymes that break down cyclic nucleotides into nucleotide monophosphates.

**guanosine-5'-triphosphate (GTP)** Required for the activity of G proteins to modulate effectors such as adenylyl cyclase; hydrolyzed by G proteins into guanosine-5'-diphosphate (GDP) before the GDP is released and new GTP is bound and effector activation is resumed ("GTP-ase switch").

**heterotrimeric G proteins** A group of proteins that consist of three subunits ( $\alpha$ ,  $\beta$ ,  $\gamma$ .) and couple cell surface receptors to their effector enzymes; on activation of cell surface receptors by their ligands, heterotrimeric G proteins dissociate into the  $\alpha$ -subunit and the  $\beta\gamma$ -complex, both of which regulate the activity of effector enzymes such as adenylyl cyclase in the cell membrane.

**protein kinases** Enzymes that tag proteins with phosphoryl residues derived from ATP; activity may be regulated by a variety of intracellular messengers, protein-protein interactions, or phosphorylation by other protein kinases.

**protein phosphatases** Reverse the action of protein kinase and remove phosphoryl groups from proteins.

**second messengers** Generated by extracellular stimuli arriving at the cell surface; transduce the extracellular signal toward the interior of the cell.

**A**denylyl cyclases (EC 4.6.1.1) are enzymes that generate the second messenger adenosine-3'/5'-monophosphate from ATP.

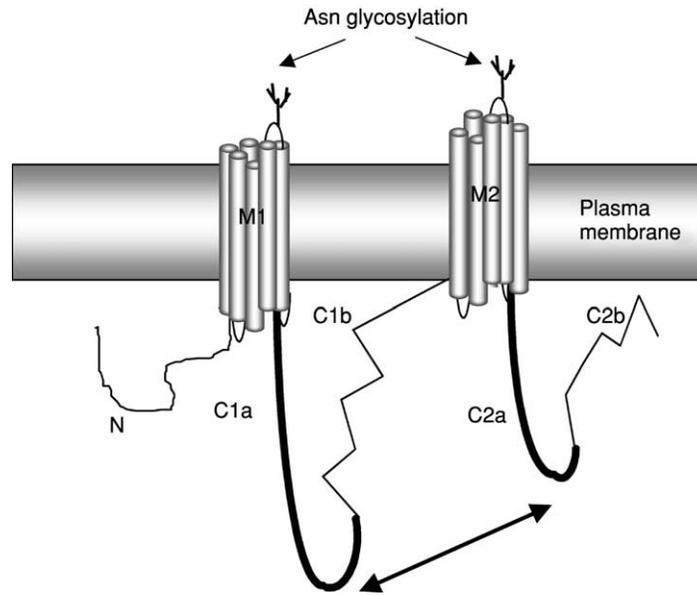
## INTRODUCTION

Adenosine-3'/5'-monophosphate (cAMP) was the first "second messenger" to be discovered in 1956 by Earl Sutherland. As a prototype and, for a long time, as the only identified intracellular signaling pathway, the

cAMP cascade was the subject of a large number of studies until the mid-1980s, when interest shifted toward novel signaling mechanisms such as polyphosphoinositides and receptor tyrosine kinases. The fact remains, however, that cAMP is a ubiquitous second messenger from gametogenesis to cognition. The application of molecular genetics to the field has highlighted the molecular diversity in the biosynthetic and catabolic arms (cyclic nucleotide phosphodiesterase, PDE) of cAMP metabolism. Ten adenylyl cyclase (AC) genes have been identified in mammals. Diversification of AC genes is already apparent in *Drosophila melanogaster* and *Caenorhabditis elegans*, indicating a distinct adaptational advantage of this feature. Taking into account the currently known molecular variants of AC and PDE, the turnover of cAMP in a mammalian cell may occur in up to 500 different ways. The biological significance of this diversity is only slowly becoming apparent and is currently under investigation in numerous biological systems.

## ADENYLYL CYCLASE GENE AND PROTEIN STRUCTURE

Two types of ACs have been identified in mammals: a membrane-bound species representing 9 of the 10 known ACs and a soluble enzyme. All of the membrane-bound ACs have the same predicted structure (Fig. 1), resembling that of ion channel/transporter proteins. The N-terminal end is cytoplasmic and continues in the M1 intramembrane segment consisting of six membrane-spanning helices. This is followed by a large cytoplasmic loop (C1), which can be subdivided into C1a and C1b on the basis of the conservation of the C1a segment in other ACs as well as a corresponding region (C2a) in the second cytoplasmic domain. The C1b portion of the cytoplasmic loop is a nonconserved isotype-specific segment. The second half of the molecule is analogous to the first half. The six transmembrane helices of the M2



**Figure 1** Schematic representation of mammalian ACs. N: N-terminal cytoplasmic loop; M1 and M2: intramembrane domains; C1a and C2a: homologous catalytic domains (thick lines); C1b and C2b: nonhomologous regulatory domains (thin lines). The C1a and C2a domains must come into physical contact to form the catalytic core. This is facilitated by  $G\alpha$  that binds to the C2a domain. Most ACs are inhibited by  $G\alpha$  that binds to a site on the C1a domain.

intramembrane segment are followed by a cytoplasmic tail designated C2, which is usefully divisible into C2a and C2b because C2a is relatively well conserved in all known ACs and also shows significant homology to C1a. Both intramembrane domains are predicted to be glycosylated on the extracellular surface.

The C1 and C2 segments of ACs also show highly significant homology to guanylyl cyclases. In fact, a point mutation in the C1a region is sufficient to change an AC into a guanylyl cyclase enzyme. Note that for historical reasons, the sequence motif corresponding to the conserved catalytic core in ACs is called guanylyl cyclase motif in protein structure databases. Mini-protein heterodimers consisting of the C1a and C2b domains of ACs are catalytically active and retain numerous regulatory properties of the holoenzyme. Furthermore, high-resolution X-ray crystallographic maps of the secondary structure of C1a and C2a heterodimers have been reported. The results concur that the catalytic core of the enzyme, including the substrate binding site, requires both domains. The heterodimer C1a:C2a is a symmetrical array; guanosine-5'-triphosphate (GTP)- $G\alpha$  is bound by the C1a domain, whereas forskolin, an alkaloid known to stimulate ACs, binds to C1a and C2a in the vicinity of the  $G\alpha$  binding site in C1a (Fig. 2). The C1-C2 complex also contains two metal-binding sites, which under physiological circumstances are occupied by  $Mg^{2+}$ . One of these sites attracts  $Mg^{2+}$

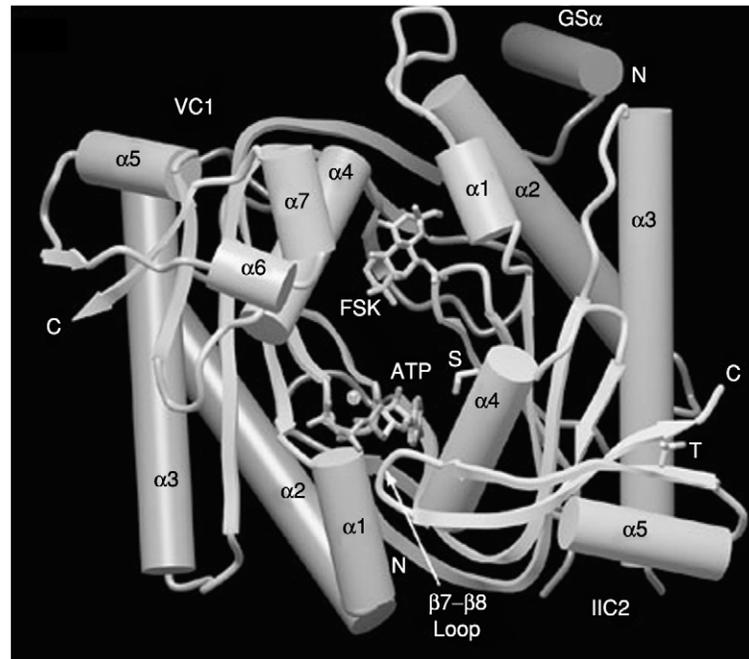
in complex with ATP, whereas the other site binds  $Mg^{2+}$  and has a preference for  $Mn^{2+}$ , explaining early observations on the stimulatory action of  $Mn^{2+}$  on ACs.

The 10th mammalian AC is a soluble enzyme that is structurally more closely related to cyanobacterial ACs than to the membrane-spanning family of mammalian ACs. The crystal structure of this enzyme was not yet published at the time of this writing.

## REGULATION OF ADENYLYL CYCLASE

### General Considerations

There are currently 10 mammalian AC genes, each of which encodes a different protein. These are differentially controlled by heterotrimeric G proteins, intracellular  $Ca^{2+}$ , and protein phosphorylation (Table I). In addition, ACX is uniquely controlled by bicarbonate ions. Hence, several signaling pathways converge on ACs, which act as molecular signal integrators. In a further aspect, AC isotypes have distinct tissue distributions, indicating nonredundant physiological roles. At the single-cell level, expression of multiple AC isotypes is the rule rather than the exception. Thus, polarized distribution of ACs within cellular microdomains or compartments is yet another element of functional diversity in the cAMP signaling cascade.



**Figure 2** The structure of the AC catalytic core. The C1a and C2a domains of dog ACV (VC1) and rat ACII (IIC2) in complex with GTP-Gs $\alpha$  are shown. Note the pseudo-symmetrical arrangement of  $\alpha$ -helices (represented by cylinders) numbered 1 to 7 and  $\beta$  strands (represented by arrows) of the C1a and C2a domains. ATP and forskolin (FSK) are drawn as stick models; Gs $\alpha$  is represented as a cylinder. The S and T denote serine and threonine residues that are putative phosphorylation sites in ACIII and ACII, respectively. Note that residues from both polypeptide chains are required to form the substrate-binding site. Reproduced from Tesmer, J. J., and Sprang, S. R. (1998). The structure, catalytic mechanism, and regulation of adenylyl cyclase. *Curr. Opin. Struct. Biol.* 8, 714, Elsevier.

## G Proteins

The activity of all transmembrane ACs is stimulated by Gs $\alpha$  combined with GTP. The structure-activity analysis of the AC catalytic core by various methods has identified the site of interaction with GTP-Gs $\alpha$  (Fig. 2) in the cytosolic C2a domain. The binding of GTP-Gs $\alpha$  to C2a is thought to facilitate the association of C1 and C2. The activity of some but not all ACs is inhibited by Gi $\alpha$  (Table I), thereby opposing the stimulation by Gs $\alpha$ .

In addition to G protein alpha subunits, G $\beta\gamma$  subunits have a direct influence on the activity of some ACs. Above all, in the presence of Gs $\alpha$ , activity of ACII and ACIV is enhanced further by G $\beta\gamma$ . Thus, ACII and ACIV are coincidence detectors for receptors coupled to Gs and Gi/o, respectively. It is important to note that G $\beta\gamma$  derived from Gs or Gq is unlikely to exert this type of effect given that the EC<sub>50</sub> of G $\beta\gamma$  for this action is approximately 100 nmol/L, whereas the stimulatory effect of Gs $\alpha$  is already maximal at approximately 10 nmol/L and the tissue levels of this G protein are low compared with those of Gi and Go. The potential for G $\beta\gamma$  to stimulate ACs is remarkable because it shows that the direction as well

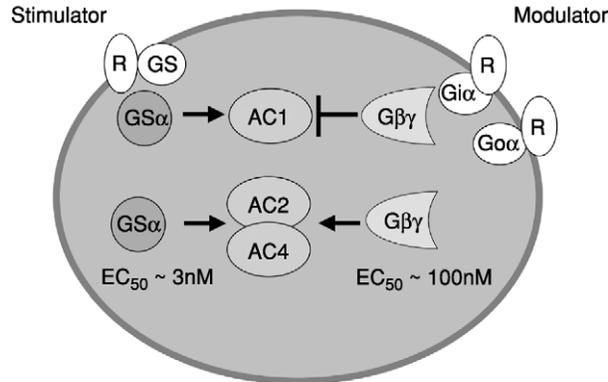
as the magnitude of the effects of hormones on cAMP synthesis can be dependent on the context of the stimulus applied. In contrast to ACII and ACIV, G $\beta\gamma$  subunits reportedly inhibit the activation of ACI by Gs $\alpha$ . A summary of the actions of G proteins is shown in Fig. 3.

## Calcium

ACI was the first AC protein to be purified and the first AC cDNA to be cloned. It is an enzyme stimulated by Ca<sup>2+</sup>-calmodulin and is highly abundant in the brain. The calmodulin modulatory site is in the C1b domain of ACI. The Ca<sup>2+</sup>-calmodulin complex and GTP-Gs $\alpha$  produce synergistic stimulation of ACI. ACVIII is also activated by Ca<sup>2+</sup>-calmodulin, but apparently through a different, IQ-type, calmodulin-binding domain, and no synergy with Gs $\alpha$  has been reported. Thus, these enzymes are capable of directly responding to a rise of intracellular-free Ca<sup>2+</sup> from a variety of sources, especially plasma membrane calcium channels. Another mode of regulation by Ca<sup>2+</sup> is an apparently direct inhibition of activity. ACVI is the best characterized example of

Table I Mammalian AC Genes

Cyclase	G proteins	Calcium-Calmodulin	Phosphorylation	Facets of tissue distribution	Human chromosomal localization	Other
ACI	G $\alpha$ stimulates, G $\beta\gamma$ inhibits	Calmodulin-dependent stimulation, synergy with G $\alpha$	Inhibition by CaMKIV	Mainly in the brain	7p13-p12	Participates in synaptic plasticity
ACII	G $\alpha$ stimulates, G $\beta\gamma$ synergizes with G $\alpha$		Stimulation by protein kinase C in synergy with G $\alpha$	Brain, lung, uterus pituitary	5p15.3	
ACIII	G $\alpha$ stimulates		Inhibition by CaMKII	Brain, kidney, testis, vasculature, pancreas beta cells	2p24-p22	Important for olfaction, involved in insulin release
ACIV	G $\alpha$ stimulates		Inhibition by protein kinase C	Lungs	14q11.2	Very little known about this enzyme
ACV	G $\alpha$ stimulates, G $\alpha$ inhibits	Direct inhibition by Ca <sup>2+</sup>	Stimulation by protein kinase C	Brain, heart	3q13.2-q21	Effect of protein kinase C not dramatic in vivo, has splice variant, required for EGF stimulation of cAMP
ACVI	G $\alpha$ stimulates, G $\alpha$ inhibits	Direct inhibition by Ca <sup>2+</sup>	Inhibition by protein kinase A and C, enhancement by tyrosine kinase	Widespread, including pituitary, adrenal cortex	12q12-q13	
ACVII	G $\alpha$ stimulates		Not known	Brain, retina, pituitary, hematopoietic system	16q12-q13	Has splice variant
ACVIII	G $\alpha$ stimulates	Calmodulin-dependent stimulation		Brain specific	8q24.2	Has splice variants
ACIX	G $\alpha$ stimulates, G $\alpha$ inhibits		Inhibited by dephosphorylation pathway involving calcineurin	Widespread, including major endocrine glands, prostate, uterus	16p13.3	Participates in adrenal corticosteroid feedback
ACX (soluble)	N/A	Not known	Not known	Testis, also present in other tissues	1q24	Modulated by bicarbonate, highly expressed in spermatogonia

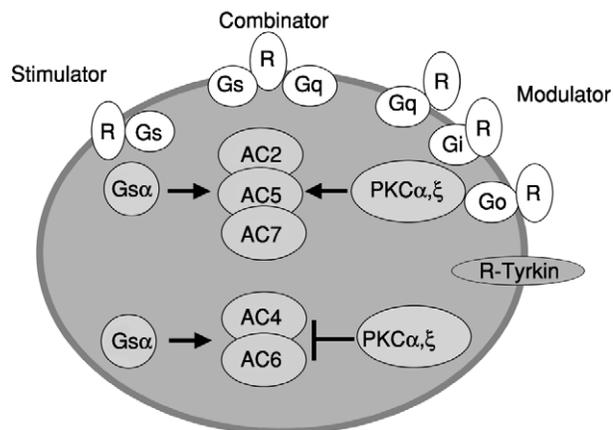


**Figure 3** Summary of the control of AC isotypes by G proteins. Arrows indicate stimulation; T bars indicate inhibition; R indicates G-protein-coupled receptor.

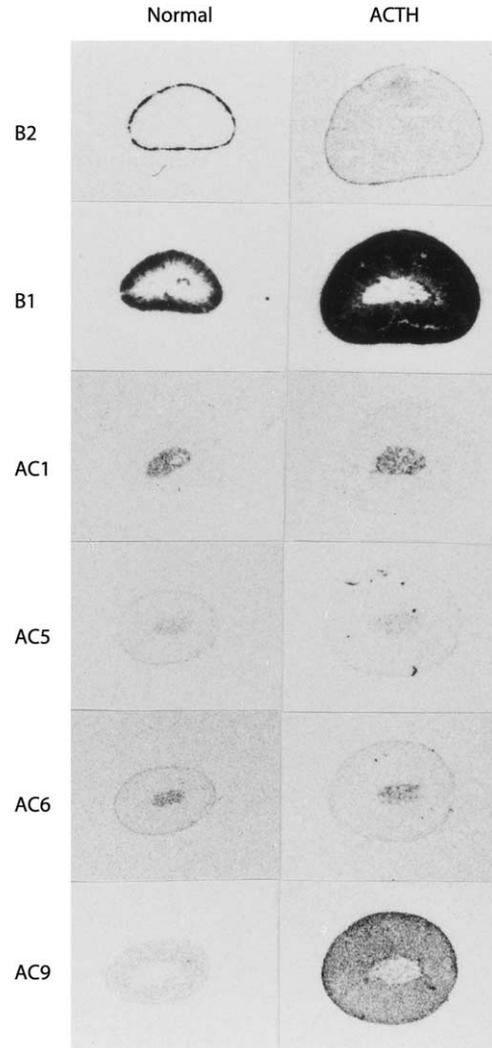
this control. The inhibition is apparent between sub-micromolar (>100 nmol/L) to low-micromolar concentrations of  $Ca^{2+}$ . The structural requirements for direct inhibition of ACs by  $Ca^{2+}$  are not known.

### Protein Phosphorylation

Typically, ACs contain a multitude of potential consensus phosphorylation sites for a variety of protein kinases. Because these are large molecules that cannot be assayed reliably in single-cell assays, the characterization of control by phosphorylation has been very slow to appear. The most striking effects have been reported for ACII and ACVII. The activity of both enzymes is markedly stimulated by protein kinase C. Whether protein kinase C phosphorylation alone is sufficient to activate the enzymes in normal cells is unclear, but synergistic stimulation by GTP-Gsα and



**Figure 4** Summary of the modulation of AC activity by protein kinase C. The α and ξ variants of protein kinase C have been implicated in the control of ACs.



**Figure 5** Distribution of mRNAs for cytochrome CYP11B2, which is responsible for the synthesis of aldosterone in the zona glomerulosa, cytochrome CYP11B1, which in turn is responsible for the synthesis of corticosterone in the zona fasciculata-reticularis and various ACs in the rat adrenal gland detected by *in situ* hybridization histochemistry. Animals were treated with vehicle or ACTH for 10 days before processing of the adrenal gland to *in situ* analysis. Reproduced from *Endocrinology* (1997), Vol. 138, p. 4596, with permission of The Endocrine Society, Bethesda, MD.

activators of protein kinase C is well established. In the case of ACII, the site of phosphorylation has been localized to the C2b domain. Inhibition of ACVI by protein kinase C has also been reported; here there are multiple sites of phosphorylation, one of which is in the N-terminal segment of the protein. The effects of protein kinase C activation on ACs are summarized in Fig. 4. ACVI is also inhibited by protein kinase A (cAMP-dependent protein kinase), and mutagenesis studies indicated that the site of phosphorylation is in the C1b segment. Thus, ACVI is inhibited by

cAMP-dependent negative feedback. In a similar vein, ACI is inhibited by calmodulin-dependent protein kinase IV. Calmodulin-dependent protein kinase II is inhibitory to ACIII, which appears to be important in the olfactory system of rodents where very rapid restoration of cellular responsiveness is vital. Yet another mode of  $\text{Ca}^{2+}$ -mediated inhibition is apparent in the case of ACIX, which is inhibited by a pathway involving the  $\text{Ca}^{2+}$ -calmodulin-dependent protein kinase calcineurin (protein phosphatase 2B). The nature of the presumed stimulatory phosphorylation is not known.

## ADENYLYL CYCLASES IN ENDOCRINE SYSTEMS

It is quite clear that without ACs, hormonal control would simply not function. Several hypothalamic-releasing hormones exert their actions through cAMP, and the effects of the majority of pituitary-tropic hormones that regulate the peripheral endocrine glands are mediated, at least in part, by cAMP. Interestingly, no obvious endocrine abnormalities have emerged from the three gene deletion studies published as of this writing that targeted ACI, ACIII, and ACVIII, respectively. This is plausibly due to the fact that these ACs are found mainly in the central nervous system or that, as in the cases of ACI and ACVIII, a double gene deletion was necessary to demonstrate a clear behavioral phenotype.

Data on the role of AC isotypes in endocrine systems are very limited indeed. The rat adrenal gland is the best explored tissue (Fig. 5) from this respect. There is striking compartmentalization of the ACs in the medulla and cortex. Furthermore, within the cortex, there are differences between the

aldosterone-synthesizing zona glomerulosa, which expresses mainly ACVI and the zona fasciculata, where the predominant AC appears to be ACIX. These characteristics are retained and accentuated after treatment of the animals with adrenocorticotrophic hormone. It was also shown that sodium restriction causes a marked increase of ACVI mRNA expression in the zona glomerulosa. Other studies of AC expression in endocrine tissues have not been consistent; however, there is little doubt that similar segregation of functionally distinct ACs is not restricted to the adrenal gland.

## See Also the Following Articles

G Protein-Coupled Receptors • G Proteins and Effectors • Lipid Second Messengers and Receptors • Receptor Tyrosine Kinase

## Further Reading

- Antoni, F. A. (2000). Molecular diversity of cyclic AMP signaling. *Front. Neuroendocrinol.* **21**, 103–132.
- Cooper, D. M. F., Karpen, J. W., Fagan, J. W., and Mons, N. E. (1998).  $\text{Ca}^{2+}$ -sensitive adenylyl cyclases. In "Advances in Second Messenger and Phosphoprotein Research" (D. M. F. Cooper, ed.), Vol. 32, pp. 23–52. Lippincott-Raven, Philadelphia, PA.
- Patel, T. B., Du, Z. Y., Pierre, S. C., Cartin, L., and Scholich, K. (2001). Molecular biological approaches to unravel adenylyl cyclase signaling and function. *Gene* **269**, 13–25.
- Tang, W.-J., Yan, S., and Drum, C. L. (1998). Class III adenylyl cyclases: Regulation and underlying mechanisms. In "Advances in Second Messenger and Phosphoprotein Research" (D. M. F. Cooper, ed.), Vol. 32. Lippincott-Raven, Philadelphia, PA.
- Taussig, R., and Gilman, A. G. (1995). Mammalian membrane-bound adenylyl cyclases. *J. Biol. Chem.* **270**, 1–4.
- Tesmer, J. J., and Sprang, S. R. (1998). The structure, catalytic mechanism, and regulation of adenylyl cyclase. *Curr. Opin. Struct. Biol.* **8**, 713–719.



# Adipocytokines

Tohru Funahashi and Ichihiro Shimomura

Osaka University Graduate School of Medicine, Osaka, Japan

Yuji Matsuzawa

Osaka University Graduate School of Medicine, Sumitomo Hospital, Osaka, Japan

## Glossary

**adipovascular axis** Adipose tissue secretes various vasoactive substances. Dysregulation of these adipocyte-derived vasoactive molecules in obesity directly contributes to the development of vascular diseases.

**lipodystrophy** A disease in which there is disturbed differentiation of adipocytes. Human generalized lipodystrophy and bioengineered mice lacking fully differentiated adipocytes exhibit phenotypes resembling that of obesity, such as insulin resistance, dyslipidemia, and fatty liver.

**metabolic syndrome** The condition characterized by upper body obesity, insulin resistance, dyslipidemia, and hypertension. The metabolic syndrome is a common cause of atherosclerotic vascular diseases.

**visceral fat** Adipose (fat) tissue located in the omentum and mesenterium. Accumulation of visceral fat is a common cause of the metabolic syndrome.

Obesity is a common risk factor for type 2 diabetes and cardiovascular diseases, and it is a major health problem in industrialized countries; however, the molecular basis for the link between obesity and obesity-related diseases has been unclear. Traditionally, adipose tissue has been regarded as an organ passively storing excess energy. However, research on adipocyte biology has found that adipocytes produce and secrete a variety of biologically active molecules, including growth factors, cytokines, and complement factors, in the immune system.

## INTRODUCTION

The adipocyte-derived factors (Fig. 1) affect the function of adipocyte in an autocrine and a paracrine fashion and affect whole body homeostasis through the bloodstream. These findings indicate that adipose tissue is an endocrine organ and the adipocyte-derived bioactive substances are adipocytokines. Because these adipocyte-derived substances include the molecules belonging to

the strict cytokines, they are also called adipokines. Every cell type secretes various cytokines and bioactive substances to the surrounding milieu to maintain normal functions. The importance of adipocytokines is highlighted by the fact that adipose tissue is the largest organ in the body. Although the amount of adipocytokines produced by a single adipocyte is small, the total amount in the body greatly affects whole body functions. Another notable feature is that adipose tissue is supplied by the bloodstream, and adipocytokines released from adipocytes easily flow into the systemic circulation. Findings obtained from bioengineered mice with undifferentiated adipose tissue revealed the importance of adipocytokines in maintaining metabolic homeostasis. Adipocytokines play a significant role in the self-defense system against metabolic overload and probably also in the immune system.

There is substantial evidence that dysregulated production of adipocytokines is involved in the development of obesity-related diseases. Tumor necrosis factor- $\alpha$  (TNF- $\alpha$ ), leptin, and resistin affect insulin sensitivity in the whole body. Overproduction of TNF- $\alpha$  has been suggested to contribute at least in part to the development of insulin resistance, which is a major clinical feature of obesity. Furthermore, adipose tissue also produces various vasoactive substances, such as plasminogen activator inhibitor

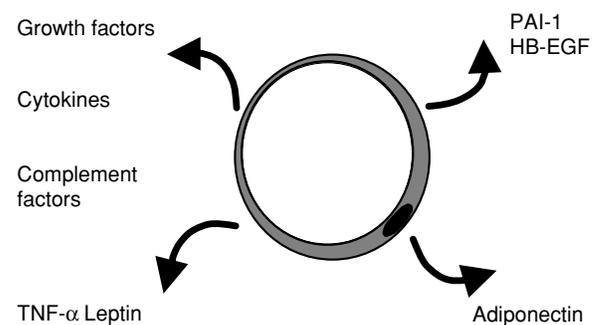


Figure 1 Adipocytokines.

(PAI-1), an inhibitor of the fibrinolytic system, and heparin-binding epidermal growth factor-like growth factor (HB-EGF), a potent growth factor for vascular smooth muscle cells. These factors are also overproduced in obesity. These findings suggest the existence of an adipovascular axis in which adipocyte-derived factors affect vascular functions independent of the coronary risk factors frequently associated with obesity.

Although the total amount of adipose tissue is increased in obesity, not all adipocytokines are overproduced. For example, adiponectin is factor D of the complement system. The expression of adiponectin in adipose tissue is severely impaired in obesity and results in the reduction of plasma concentration. The physiological meaning of the decreased plasma adiponectin in obesity is unclear. Next, adiponectin belongs to the soluble defense collagen superfamily. Although the protein possesses antiatherogenic and antidiabetic properties, its plasma concentration is decreased in obesity. The proteome analysis proved that there is impaired secretion of various molecules from the adipocytes of obese subjects. Not only hypersecretion of offense adipocytokines but also hyposecretion of defense adipocytokines affects the development of obesity-related disorders.

## LEPTIN

Leptin is a 16-kDa protein secreted primarily from adipocytes. Rodents defective in leptin synthesis (*ob/ob* mice) or leptin receptor function (*db/db* mice, Zucker *fa/fa* rats, and Koletsky rats) are obese and develop hyperinsulinemia and insulin resistance. Leptin suppresses food intake and increases energy expenditure by enhancing thermogenesis and metabolic rate. These functions seem to be mediated mainly by the central nervous system because intracerebroventricular injection of leptin produced significant effects with much smaller amounts than those required by systemic injection.

Leptin receptors (OB-R) are single membrane-spanning receptors with homology to members of the cytokine receptor superfamily. Seven different leptin receptors, produced by alternative splicing, have been identified. The receptors containing transmembrane domains can be divided into two groups. One group has a short amino acid residue intracellular domain (OB-Ra, OB-Rc, OB-Rd, and OB-Rf). The other group has a long intracellular domain (OB-Rb). OB-Rb is mainly expressed in hypothalamus, whereas the short forms are expressed in a variety of tissues.

The long form of the receptor has two Janus kinase sites in the intracellular domain and the short form has one. Only the long form can activate the signal transducers and activators of the transcription family (STAT). C57B1/Ks *db/db* mice, which lack the long form of the receptor and have intact short forms, exhibited almost identical phenotype to *ob/ob* mice, which lack leptin.

There is accumulating evidence indicating that leptin mimics some of the insulin actions in liver, adipose tissue, and muscle. In diabetic rats, leptin increased glucose uptake in muscle and brown adipose tissue and normalized hyperglycemia. In liver and hepatocytes, leptin lowered hepatic glucose production by decreasing glycogenolysis and increasing glycogen synthesis.

The physiological significance of leptin was emphasized by the successful administration of leptin to treat the metabolic disorders of lipodystrophy in mice and humans. Generalized lipodystrophy is a disorder characterized by a paucity of adipose (fat) tissue accompanied by a severe resistance to insulin, leading to hyperinsulinemia, hyperglycemia, and an enlarged fatty liver. In the two genetically engineered mouse models of lipodystrophy, the plasma concentration of leptin was very low due to the loss of mature adipose tissues. Leptin supplementation overcame insulin resistance, diabetes, hyperlipidemia, and fatty liver in both models. These effects were not observed with chronic food restriction. The results support the theory that leptin exerts favorable effects on glucose and lipid metabolism independently of its effect on food intake. Based on these observations, human lipodystrophy patients were recently treated with leptin, resulting in an improvement in metabolic disorders and fatty liver.

## TNF- $\alpha$

It has been shown that the adipose TNF- $\alpha$  mRNA and plasma TNF- $\alpha$  protein are increased in most animal models and human subjects with obesity and insulin resistance. Neutralizing the blood TNF- $\alpha$  in obese rats with a soluble TNF- $\alpha$  receptor-IgG fusion protein markedly improved insulin resistance. These results indicated that the higher production of TNF- $\alpha$  in accumulated adipose tissue was causative for obesity-associated insulin resistance.

TNF- $\alpha$  treatment reduced insulin-stimulated autophosphorylation of the insulin receptor and IRS-1 phosphorylation in various tissue cultured cells, including adipocytes, fibroblasts, and hepatoma cells. These disturbances of insulin signaling were also

observed in the muscle and adipose tissues of the obese and insulin-resistant *fa/fa* rats. It has been reported that TNF- $\alpha$  induces serine phosphorylation of IRS-1 in cultured adipocytes and hepatocytes. The serine phosphorylated form of IRS-1 was also increased in the muscle and adipose tissues of *fa/fa* rats. The modified IRS-1 inhibited the autokinase activity of the insulin receptor, resulting in the deterioration of insulin signaling.

Hotamisligil and colleagues fed wild-type and TNF- $\alpha$  knockout mice with a high-fat diet. Both types of mice became similarly obese, but obese TNF- $\alpha$ (-/-) mice maintained high insulin sensitivity. These results demonstrated that TNF- $\alpha$  deficiency blocked the development of insulin resistance associated with diet-induced obesity. In the *ob/ob* mice with targeted mutations in both p55 and p75 TNF- $\alpha$  receptors, the signaling and function of TNF- $\alpha$  were completely abolished. Without the changes in body weight, the *ob/ob* mice with null mutations of TNF- $\alpha$  receptors showed higher insulin sensitivity than the *ob/ob* mice with normal TNF- $\alpha$  receptors. These results further support the theory that the enhanced TNF- $\alpha$  protein and TNF- $\alpha$  signaling are involved in obesity-associated insulin resistance.

## RESISTIN

Resistin is a 12.5-kDa, cysteine-rich protein identified by screening for the genes that were induced during the differentiation of the adipocytes but were down-regulated in mature adipocytes exposed to glitazone, an insulin-sensitizing drug. Mouse resistin contains 114 amino acids and circulates as a homodimer of two peptides. Administration of resistin to mice impaired glucose tolerance and insulin action. Plasma concentrations of resistin were higher in genetic and diet-induced obese mice with insulin resistance. Administration of the neutralizing antibody against resistin increased insulin sensitivity in obese mice. These results suggest that resistin is a fat-derived factor causing insulin resistance in obesity and that the insulin-sensitizing effect of glitazone can be attributed to its inhibition of resistin expression.

One study, however, showed contradictory results. It was reported that resistin expression was decreased in obese mice and increased in response to glitazone.

The human homologue of resistin is located on chromosome 19p13.3, a region not previously implicated in the susceptibility to obesity, insulin resistance, or diabetes. Its significance in humans remains to be clarified.

## PAI-1

PAI-1 is an inhibitor of plasminogen activators and fibrinolytic activity. Although PAI-1 is synthesized in endothelial cells and liver, adipose tissue is also a main source of plasma PAI-1. Expression of PAI-1 is augmented in adipose tissue, especially intraabdominal visceral fat in obesity. The amount of visceral fat is positively correlated with the plasma level of PAI-1 and is one of major determinants of plasma PAI-1. It is well-known that the plasma concentration of PAI-1 is elevated in subjects with type 2 diabetes and hypertriglyceridemia, although the precise mechanism is unclear. Type 2 diabetes and hypertriglyceridemia are often associated with visceral fat accumulation. PAI-1 produced by accumulated visceral fat may explain the high plasma PAI-1 in these conditions. The reason why adipocytes secrete PAI-1 has not been clarified. Adipocytes dramatically change cell size in response to nutritional conditions. Plasmin works to destroy the basement membrane to facilitate cell expansion. PAI-1 may control the activity of plasminogen activators to prevent the overproduction of plasmin.

## ADIPONECTIN

Adiponectin is an adipocyte-derived factor identified through the extensive search of adipose tissue transcripts in the human genome project. The expression of adiponectin mRNA is exclusive in adipose tissue. The protein is composed of two structurally distinct domains—the C-terminal collagen-like fibrous domain and the complement C1q-like globular domain. Adiponectin is abundant in the circulating plasma in a multimeric form. Interestingly, plasma concentrations of adiponectin are decreased in obese subjects despite its restricted expression in adipocytes. Plasma adiponectin levels are also lower in patients with coronary artery disease and type 2 diabetes than those in body mass index-matched subjects. Physiologically, adiponectin inhibits the differentiation of adipocytes in a paracrine manner via the COX-2 pathway. When the endothelial barrier is injured, adiponectin accumulates in the subendothelial space of the vascular walls. The protein has anti-atherogenic properties, such as suppression of monocyte attachment to vascular endothelial cells via the reduced expression of adhesion molecules, suppression of foam cell formation and TNF- $\alpha$  secretion of macrophages, and suppression of the growth factor-induced proliferation of vascular smooth muscle cells. Administration of adiponectin also improves fatty

oxidation and insulin resistance in dietary-induced and genetically obese animals. Adiponectin is cleaved between fibrous and globular domains. The globular form of adiponectin has more potent activity for insulin sensitization than the whole protein. Thus, adiponectin has a dual function on insulin sensitivity and vascular functions. Hypoadiponectinemia in obesity may be a key factor in the metabolic syndrome, which is often accompanied by insulin resistance and cardiovascular diseases. Interestingly, patients with genetic hypoadiponectinemia caused by a missense mutation in the adiponectin gene also exhibit the clinical phenotype of the metabolic syndrome. Studies of mice genetically lacking adiponectin confirmed the significance of adiponectin in the metabolic syndrome. Two cohort studies of a specialized population report that hypoadiponectinemia is a risk for cardiac death or the developments of type 2 diabetes. The genetic variation in the adiponectin gene is also associated with an increased risk of type 2 diabetes in the Japanese population. Further

studies are necessary to verify the significance of adiponectin in the metabolic syndrome and obesity-related diseases.

### See Also the Following Articles

Diabetes, Type 2 • Leptin • Obesity Regulation • Tumor Necrosis Factor (TNF)

### Further Reading

- Friedman, J. M., and Halaas, J. L. (1998). Leptin and the regulation of body weight in mammals. *Nature* **395**, 763–770.
- Funahashi, T., Nakamura, T., Shimomura, I., Maeda, K., Kuriyama, H., Takahashi, M., Arita, Y., Kihara, S., and Matsuzawa, Y. (1999). Role of adipocytokines on the pathogenesis of atherosclerosis in visceral obesity. *Intern. Med.* **38**, 202–206.
- Matsuzawa, Y. (1997). Pathophysiology and molecular mechanisms of visceral fat syndrome: The Japanese experience. *Diabetes Metab. Rev.* **13**, 3–13.
- Spiegelman, B. M., Choy, L., Hotamisligil, G. S., Graves, R. A., and Tontonoz, P. (1993). Regulation of adipocyte gene expression in differentiation and syndromes of obesity/diabetes. *J. Biol. Chem.* **268**, 6823–6826.



# Adrenal Androgens

Lucia Ghizzoni and Alessandra Vottero  
University of Parma, Parma, Italy

## Glossary

**adrenal gland** Triangular, endocrine, hormone-secreting gland located on top of each kidney.

**androgens** Sex hormones that stimulate the development of secondary masculine characteristics (i.e., deep voice, facial hair).

**receptor** Molecule on the cell surface or inside the cell that recognizes a specific antigen or hormone.

The adrenal androgens (AAs), normally secreted by the fetal adrenal zone or the zona reticularis, are steroid hormones with weak androgenic activity. Although AAs do not appear to play a major role in the fully androgenized adult man, they seem to play a role in the adult woman and in both sexes before puberty. Girls, women, and prepubertal boys may be negatively affected by AA hypersecretion, in contrast to adult men. This article discusses AA biosynthesis, regulation, physiology, and biological action.

## ADRENAL ANATOMY AND ANDROGEN BIOSYNTHESIS

### Adrenal Gland Anatomy

The adrenal glands, consisting of the cortex and medulla, have a roughly pyramidal shape and lie above the upper poles of the kidneys. The two zones receive their blood supply from branches of the phrenic arteries, aorta, and renal arteries. Arterial blood enters from the outer cortex, flows through fenestrated capillaries between the cords of cells, and drains into venules in the medulla. On the right, the adrenal vein directly enters the inferior vena cava; on the left, it usually drains into the left renal vein. The adrenal cortex is divided into three histologic and functional zones: the outer, aldosterone-secreting zona glomerulosa; the intermediate, cortisol-secreting zona fasciculata; and the inner, androgen-secreting zona reticularis. Whereas the zona glomerulosa is primarily

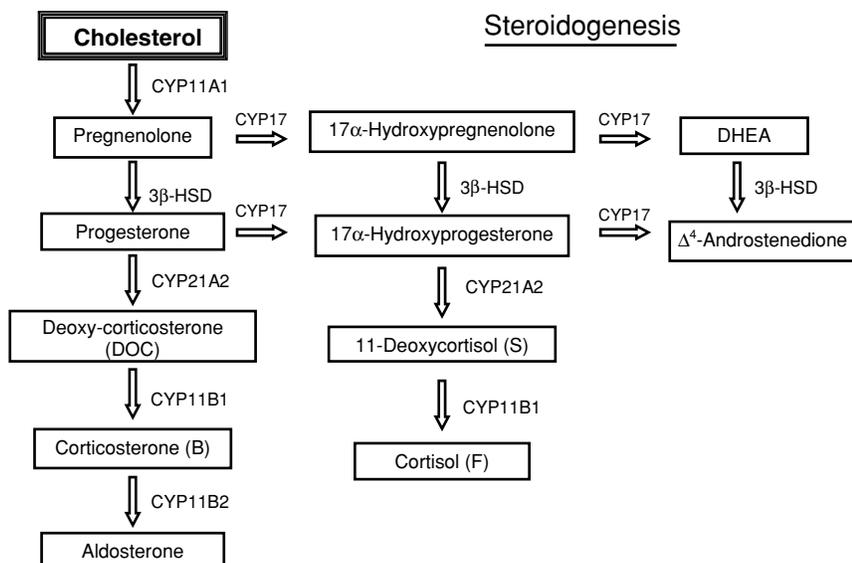
regulated by angiotensin II, both the zona fasciculata and the zona reticularis are regulated by adrenocorticotrophic hormone (ACTH).

### Fetal Adrenal and Development

The fetal adrenal cortex arises from mesodermal cells migrating from the celomic epithelium very early in the embryonic period. It consists of both the adult adrenal zona glomerulosa and zona fasciculata and an inner large adrenal zone, which virtually disappears within weeks after birth. The active secretion of steroids occurs by week 6 from the provisional zone, which represents the functional cortex in the fetal period. Remaining cell foci from the fetal adrenal zone presumably give rise to the adrenal zona reticularis, starting at the age of 4 to 5 years in both sexes. This zone continues to grow until young adulthood (20 to 25 years), remains at a plateau for 5 to 10 years, and regresses gradually after the age of 35 years. Aging results in a reduction in the size of the zona reticularis and a relative increase in the outer cortical zones with no significant difference in the total width of the cortex.

### Biosynthesis

Like all human steroid hormones, AAs are derived from cholesterol, which can be synthesized within the adrenal from acetyl coenzyme A, but mainly (80%) is derived from circulating plasma lipoproteins (low-density lipoproteins). The major androgens secreted by the adrenals are dehydroepiandrosterone (DHEA), DHEA sulfate (DHEAS), and androstenedione ( $\Delta^4$ -A) (Fig. 1). Production of testosterone (T) by these glands is minimal. DHEA and DHEAS are mainly products of the zona reticularis;  $\Delta^4$ -A and T are secreted by both the zona reticularis and the zona fasciculata. The enzymes responsible for the synthesis are hydroxylases, dehydrogenases, isomerases, and desmolases (Fig. 1), most of which require NADPH or  $\text{NAD}^+$  as



**Figure 1** Adrenal androgen biosynthetic pathway. CYP11A1, cholesterol side-chain cleavage enzyme; desmolase; CYP17, 17 $\alpha$ -hydroxylase/17,20-lyase; 3 $\beta$ -HSD, 3 $\beta$ -hydroxysteroid dehydrogenase; CYP21A2, 21-hydroxylase; CYP11B1, 11 $\beta$ -hydroxylase; CYP11B2, aldosterone synthase, corticosterone 18-methylcorticosterone oxidase/lyase.

cofactors. The anatomical alterations of the adrenal cortex, occurring mainly in the age groups of 20–30 years and 50–60 years, result in a marked decline in circulating adrenal C<sub>19</sub> steroids and their metabolites.

## BIOCHEMISTRY

The steroid hormones produced by the adrenal cortex are members of a large family of compounds derived from the cyclopentanoperhydrophenanthrene ring structure that comprises three cyclohexane rings and one cyclopentane ring. Three 19-carbon compounds are the principal androgens secreted by the adrenals: DHEA, DHEAS, and  $\Delta^4$ -A.

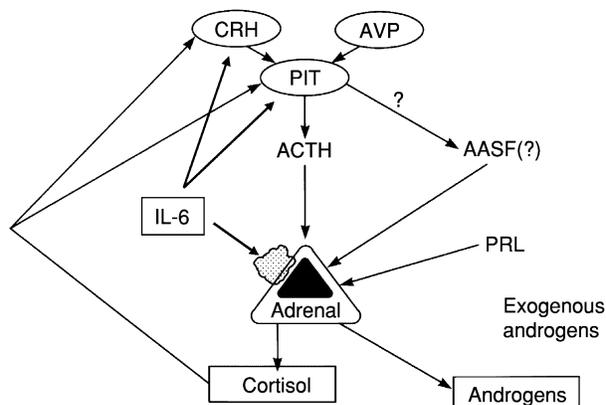
## ADRENAL ANDROGEN REGULATION AND PHYSIOLOGY

### Regulation

Adrenal androgens are secreted by the adrenal glands in response to ACTH. ACTH is a 39-amino-acid peptide derived from proopiomelanocortin, synthesized and secreted by the anterior pituitary under the regulation of corticotropin-releasing hormone (CRH) and arginine-vasopressin (AVP) (Fig. 2). Both CRH and AVP are produced by parvocellular neurons of the paraventricular nucleus of the hypothalamus and act in synergy with each other (Fig. 2). Under ACTH regulation, adrenal androgens such as DHEA,  $\Delta^4$ -A,

and T are secreted synchronously with cortisol both in secretory episodes and in a circadian pattern. The levels of plasma DHEAS do not exhibit a circadian rhythm because of its much longer circulating half-life.

Numerous other endocrine signals were proposed as regulators of adrenal androgen secretion. Glasgow *et al.* reported the presence of prolactin (PRL) receptors in the human adrenal gland and suggested a direct effect of PRL on adrenal steroidogenesis that may be of particular relevance in clinical disorders characterized by hyperprolactinemia. Interestingly, adults with



**Figure 2** Schematic representation of adrenal androgen regulation.

hyperprolactinemia have increased secretion of AAs by the zona reticularis, which is corrected by reduction of PRL secretion with bromocriptine. In women with PRL-secreting tumors, there is a correlation between PRL levels and DHEAS.

Together with ACTH and/or PRL, other factors including estrogen, epidermal growth factor, prostaglandins, angiotensin, growth hormone, gonadotropins,  $\beta$ -lipotropin, and  $\beta$ -endorphin may act as stimulators of androgen secretion. Interleukin-6 (IL-6) is also known to stimulate mineralocorticoid, glucocorticoid, and androgen production by acting through specific receptors expressed in the adrenals, mainly in the zona fasciculata and reticularis, but also to a lesser extent in the zona glomerulosa. The ability of IL-6 to stimulate mineralocorticoid, glucocorticoid, and androgen production suggests that IL-6 might play a role in coordinating the responses of all adrenocortical zones and the interaction of the adrenal function with the immune system.

Both ACTH and PRL stimulate AA secretion by the fetal adrenal zone. Placental CRH production, which rises exponentially during human pregnancy, may also play a key role in promoting DHEAS production by the fetal adrenals, leading to an increase in placental estrogen synthesis and contributing to the process of parturition in humans.

## Physiology

AAs are secreted in small amounts during infancy and early childhood and their secretion gradually increases with age, paralleling the growth of the zona reticularis. Adrenarche is the appearance of pubic hair (pubarche) resulting from a rise in adrenal androgen levels. The mechanisms by which the zona reticularis develops with age and by which adrenarche is regulated are not fully known. It was shown that children with premature pubarche have hormonal responses to a CRH stimulation test that are similar in magnitude to those of prepubertal children of comparable age, ruling out a prominent role for CRH in premature pubarche. Gell *et al.* suggested that as children mature, a decrease in  $3\beta$ -hydroxysteroid dehydrogenase activity in the adrenal reticularis occurs, resulting in the shift of pregnenolone through the  $17\alpha$ -hydroxylase/ $17,20$ -lyase pathway, leading to increased production of DHEA and DHEAS, as seen during adrenarche. Locally produced insulin-like growth factor type II (IGF-II) modulates fetal adrenocortical cell function by increasing responsiveness to ACTH via activation of the IGF type I receptor and increases the capacity of those cells for androgen synthesis by directly

augmenting the expression of P450c17. Thus, IGF-II may play a pivotal role in AA production, both physiologically *in utero* and at adrenarche, as well as under conditions of hyperandrogenemia.

Experiments by Miller and co-workers suggested that an increased serine phosphorylation of human P450c17 might play a role in the development of both human adrenarche and hyperandrogenism of polycystic ovary syndrome (PCOS), resulting in a substantial increase in  $17,20$ -lyase activity. P450c17 is the key enzyme that regulates androgen synthesis. It is the only enzyme known to have the ability to convert  $C_{21}$  precursors to the androgen prehormones, the  $17$ -ketosteroids. It is a single enzyme with two activities,  $17$ -hydroxylase and  $17,20$ -lyase, and serine phosphorylation appears to modulate its activity. In particular, it promotes  $17,20$ -lyase activity and at the same time inhibits the activity of the insulin receptor. It was postulated that a single abnormal serine kinase might hyperphosphorylate both P450c17 and the insulin receptor, accounting for the hyperandrogenism and the hyperinsulinism responsible for both the premature pubarche and later in life for PCOS. *In vitro* studies, however, failed to find evidence for hyperphosphorylation of insulin receptor- $\beta$  and P450c17 in PCOS.

## CIRCULATION

Circulating steroid hormones are largely bound to plasma proteins (binding globulins and albumin). Approximately 90% of DHEA, DHEAS, and  $\Delta^4$ -A is bound to albumin and 3% is bound to sex hormone-binding globulin. The binding globulins have high affinity and low capacity, whereas albumin has a low affinity and high capacity for steroids.

## ANDROGEN RECEPTOR

The inactive androgen precursors secreted by the adrenal, after conversion to T and  $5\alpha$ -dihydrotestosterone (DHT), exert their effects in most peripheral tissues by interacting with high-affinity receptor proteins. The androgen receptor (AR), encoded by the AR gene on the X chromosome, is a member of the steroid receptor superfamily. This gene contains a polymorphic CAG microsatellite repeat within exon 1, which codes for a variable length of polyglutamine chain at the amino terminus, the transactivation domain of the AR protein. Triplet-repeat DNA sequences can be sites of genetic instability and their expansion in a variety of genes has been associated with human genetic diseases, such as

fragile X syndrome and myotonic dystrophy. In the case of the AR gene, an inverse correlation between the number of CAG repeats and the risk for prostate cancer was described and its expansion was documented in Kennedy's disease (spinal and bulbar muscular atrophy), a disorder associated with primary hypogonadism. *In vitro* studies showed that progressive expansion of the repeat length in the AR is associated with a linear decrease in its transactivation function. These observations support the idea that there is an optimal number of repeats that varies in the population from 11 to 31 (average size:  $21 \pm 2$  repeats).

Methylation of deoxycytosine residues is another factor involved in the modulation of gene expression. Belmont *et al.* showed that the methylation of *HpaII* and *HhaI* sites near the polymorphic CAG repeats in the first exon of the human AR (HUMARA) locus correlated with X inactivation. Patients with idiopathic hirsutism were shown to have a normal number of CAG repeats but with a preponderance of the shortest and most active alleles. These patients also had a preferential methylation of the longer AR allele compared to normal subjects, leading to inactivation of the functionally weaker gene. This skewing could allow the shorter, more active AR allele to be preferentially expressed, explaining the peripheral hypersensitivity to androgens in hirsute patients.

Multiple coactivators that enhance the transcription of the AR gene have been identified, including AP-1, Smad3, nuclear factor  $\kappa$ B, sex-determining region Y, and the Ets family of transcription factors. The relative importance of these coactivators for any particular cell type remains unclear, since a putative coregulator that can alter transcriptional activity is typically included in transient transfection experiments. Although AR is normally thought to function as a homodimer, it was also shown to heterodimerize with other nuclear receptors including the estrogen receptor, glucocorticoid receptor, and testicular orphan receptor 4. One of the major mechanisms through which coregulators might function is by forming a bridge between the DNA-bound nuclear receptor and the basal transcriptional machinery (type I regulators). Coactivators may also facilitate ligand binding, promote receptor nuclear translocation, or mediate signal transduction (type II coregulators).

The role of corepressors in AR function is poorly defined. Three corepressors of androgen-bound AR have been identified thus far: cyclin D1, calreticulin, and HBO1 [histone acetyltransferase binding to ORC (origin recognition complex)]. However, relatively little is known about the mechanism(s) of their repressive effects.

## PERIPHERAL CONVERSION AND METABOLISM

DHEA, DHEAS, and  $\Delta^4$ -A are converted to the potent androgens T and DHT in peripheral tissues. Major conversions are those of  $\Delta^4$ -A to T and of T to DHT as carried out by the enzymes 17-hydroxysteroid dehydrogenase (17 $\beta$ -HSD) and 5 $\alpha$ -reductase, respectively. Major peripheral sites of androgen conversion are the hair follicles, the sebaceous glands, the prostate, and the external genitalia.

The active uptake of androgens and *in situ* estrogen synthesis occur in peripheral adipose tissue and are carried out by the enzymes 17 $\beta$ -HSD and aromatase, respectively. Peripheral conversion contributes significantly to circulating T levels in women, but not in men, in whom T is largely produced by the testis.

The AAs and their metabolites are inactivated or degraded in various tissues, including the liver and kidneys. Major biochemical routes for inactivation and excretion are conjugation of androgens to glucuronate or sulfate residues to produce hydrophilic glucuronides or sulfates that are excreted in the urine.

## BIOLOGICAL EFFECTS

In adult men, the conversion of adrenal  $\Delta^4$ -A to testosterone accounts for less than 5% of the production rate of the latter; thus, its role in the physiological androgenization of the male is negligible. Excessive AA secretion appears to have no major clinical consequences in the adult man, although this may be a matter of debate. AA hypersecretion in prepubertal boys, on the other hand, has clearly been associated with isosexual precocious puberty.

In adult women, adrenal  $\Delta^4$ -A and  $\Delta^4$ -A generated from the peripheral conversion of DHEA contribute substantially to total androgen production and its effect. In the follicular phase of the menstrual cycle, adrenal precursors account for two-thirds of testosterone production and one-half of dihydrotestosterone production. At midcycle, the ovarian contribution increases and the adrenal precursors account for 40% of testosterone production. In women, increased AA production may be manifested as cystic acne, hirsutism, male type baldness, menstrual irregularities, oligo-ovulation or anovulation, infertility, and/or frank virilization. Excessive adrenal androgen secretion in prepubertal or pubertal girls can cause heterosexual precocious puberty.

## See Also the Following Articles

Adrenal Cortex, Anatomy • Adrenal Cortex, Physiology • Adrenal Insufficiency • Adrenarche, Premature • Androgen Biosynthesis and Gene Defects • Androgen Biosynthesis Inhibitors and Androgen Receptor Antagonists • Androgen Insensitivity Syndrome • Androgens, Gender and Brain Differentiation

## Further Reading

- Auchus, R. J. (1998). The regulation of human P450c17 activity: Relationship to premature adrenarche and the polycystic ovary syndrome. *Trends Endocrinol. Metab.* **9**, 47–50.
- Ghizzoni, L., Virdis, R., Ziveri, M., Lamborghini, A., Alberini, A., Volta, C., and Bernasconi, S. (1989). Adrenal steroid, cortisol, adrenocorticotropin, and  $\beta$ -endorphin responses to human corticotropin-releasing hormone stimulation test in normal children and children with premature pubarche. *J. Clin. Endocrinol. Metab.* **69**, 875–880.
- Heinlein, C. A., and Chang, C. (2002). Androgen receptor (AR) coregulators: An overview. *Endocr. Rev.* **23**, 175–200.
- Longcope, C. (1986). Adrenal and gonadal androgen secretion in normal females. In “Clinics in Endocrinology and Metabolism” (R. Horton and R. A. Lobo, eds.), pp. 213–228. W. B. Saunders, Philadelphia, PA.
- Mastorakos, G., Weber, J. S., Magiakou, M. A., Gunn, H., and Chrousos, G. P. (1994). Hypothalamic–pituitary–adrenal axis activation and stimulation of systemic vasopressin secretion by recombinant interleukin-6 in humans: Potential implications for the syndrome of inappropriate vasopressin secretion. *J. Clin. Endocrinol. Metab.* **79**, 934–939.
- McKenna, T. J., Fearon, U., Clarke, D., and Cunningham, S. K. (1997). A critical review of the origin and control of adrenal androgens. *Bailliere Clin. Obstet. Gynecol.* **11**, 229–248.
- Nicholson, W. E., Liddle, R. A., Puett, D., and Liddle, G. W. (1978). Adrenocorticotropin hormone biotransformation, clearance, and catabolism. *Endocrinology* **103**, 1344–1351.
- Parker, L. N. (1991). Control of adrenal androgen secretion. *Endocrinol. Metab. Clin. North Am.* **20**, 401–421.
- Vale, W., Spiess, J., Rivier, C., and Rivier, J. (1981). Characterization of a 41-residue ovine hypothalamic peptide that stimulates secretion of corticotropin and  $\beta$ -endorphin. *Science* **213**, 1394–1397.
- Vottero, A., Stratakis, C. A., Ghizzoni, L., Longui, C. A., Karl, M., and Chrousos, G. P. (1999). Androgen receptor-mediated hypersensitivity to androgens in women with nonhyperandrogenic hirsutism: Skewing of X-chromosome inactivation. *J. Clin. Endocrinol. Metab.* **84**, 1091–1095.



# Adrenal Cortex Development, Regulation of

Michel Grino

Université de la Méditerranée, Marseille, France

## Glossary

**adrenocorticotrophic hormone (ACTH)** A 39-amino acid peptide secreted by the anterior pituitary that acts primarily on the adrenal cortex, stimulating its growth and its secretion of corticosteroids.

**corticotropin-releasing hormone (CRH)** A 41-amino acid peptide secreted by the hypothalamus and the placenta; hypothalamic CRH acts primarily on the anterior pituitary to stimulate adrenocorticotrophic hormone (ACTH) synthesis and secretion, whereas placental CRH acts on the fetal adrenals to regulate corticosteroid secretion and on the reproductive tract to regulate the processes of parturition.

**estrogens** Female sex hormones that are responsible for the development of the female secondary sex characteristics, the regulation of the menstrual cycle, the production of an environment suitable for the implantation of the early embryo, the maintenance of pregnancy, and the development of fetal and maternal tissues.

**growth factors** Secreted proteins that exert diverse effects on cell growth, metabolism, and differentiation.

Steroids synthesized by the adrenal cortex, in particular glucocorticoids, play an important role in the maintenance of pregnancy and the maturation of maternal and fetal tissues during fetal life and in the adaptation to extrauterine life after birth. An adequate exposure to glucocorticoids relies on coordinate regulation of adrenal growth, maturation, and steroid biosynthesis. Anterior pituitary adrenocorticotrophic hormone (ACTH) is believed to be the major trophic regulator of adrenal development and glucocorticoid synthesis and secretion. However, because ACTH is not a mitogen per se, it is now generally accepted that some of the trophic actions of ACTH on the fetal adrenal cortex are mediated indirectly via tissue growth factors. In addition, other hormonal factors originating from the placenta, such as corticotropin-releasing hormone (CRH) and estrogens, are important for fetal adrenal maturation.

## HORMONAL REGULATION OF ADRENAL GROWTH AND STEROID BIOSYNTHESIS

### Adrenocorticotrophic Hormone

Adrenocorticotrophic hormone (ACTH) is the main stimulator of glucocorticoid synthesis and secretion in the adult. Exposure to robust ACTH secretion is required for normal adrenal growth and maturation. Indeed, disruption of the hypothalamo-pituitary-adrenal (HPA) function in the human fetus, experimental anencephaly in the fetal rhesus monkey or the fetal rat, or fetal hypophysectomy in sheep inhibits growth of the fetal adrenal cortex.

ACTH acts through a specific adrenal cortical cell surface G protein-coupled receptor that activates adenylate cyclase, leading to an increase in intracellular cyclic adenosine monophosphate (cAMP) that in turn activates protein kinase A and initiates the cascade of intracellular signaling events. In midgestation human fetal adrenals, ACTH receptor mRNA is localized in cells from all cortical zones; abundance is higher in the definitive zone than in the fetal zone. The effect of ACTH on adrenal growth and/or metabolism is zone specific. In anencephalic fetuses, the definitive zone appears to be normal. In fetal monkeys, blockade of endogenous fetal ACTH secretion by glucocorticoid treatment decreases expression of  $3\beta$ -hydroxysteroid dehydrogenase ( $3\beta$ -HSD) and eliminates the transitional zone but has no effect on the size of the definitive zone. Conversely, ACTH administration or stimulation of endogenous ACTH secretion by metyrapone treatment results in stimulation of  $3\beta$ -HSD expression and in an increase of the width of the transitional zone but not of the size of the definitive zone. The growth and the steroidogenic activity of the fetal zone are also controlled by ACTH. In fetuses with congenital adrenal hyperplasia, the fetal zone is hypertrophied and adrenal androgen concentrations are stimulated dramatically. In monkeys, inhibition of

endogenous ACTH decreases adrenal 17 $\alpha$ -hydroxylase/17-20 lyase (P450c17) expression, whereas metyrapone-induced stimulation of ACTH secretion results in stimulation of P450c17 gene transcription in the fetal zone.

The stimulatory effect of ACTH on adrenal growth and steroid biosynthetic capacities seems to depend both on the circulating levels of the hormone and on the sensitivity of the adrenal gland. In addition, these phenomena could be associated given that it is established that ACTH up-regulates the expression of its own receptor gene in both adult and fetal adrenocortical cells.

In human fetuses, the maximal rate of adrenal growth occurs during the time the plasma ACTH concentration is the lowest. In baboons, there is a temporal parallelism between ACTH receptor mRNA expression in the fetal zone and ACTH-stimulable dehydroepiandrosterone (DHEA) formation. It has been suggested that the increase in 3 $\beta$ -HSD expression and cortisol production observed during late gestation results from the ACTH receptor-mediated development and enhanced functional capacity of the transitional/definitive zone.

In sheep, between embryonic day 90 (E90) and E120, adrenal growth and adrenal cell division occurs at the lowest rate compared with any other time during gestation, and there is a decreased expression of steroidogenic enzymes. Between E130 and postnatal day 2 (P2), there is a phase of rapid adrenal growth and functional maturation. Circulating ACTH increases steadily between E110 and E145, whereas plasma cortisol concentrations are low between E110 and E130 and increase abruptly between E130 and E145. Interestingly, the lowest level of adrenal expression of the ACTH receptor occurs at E90, whereas there is increased ACTH receptor mRNA, ACTH binding, and ACTH-induced adenylate cyclase activity after E125. The ACTH sensitivity of fetal adrenal cells incubated *in vitro* parallels the evolution of the density of ACTH binding sites. Ovine fetuses that have undergone hypothalamo-pituitary disconnection at E115 show decreased ACTH and cortisol secretion at E126 but normal ACTH and decreased cortisol secretion at E145. Hypophysectomy in ovine fetuses induces a decrease in adrenal expression of the side-chain cleavage enzyme (P450ssc), 3 $\beta$ -HSD, and P450c17 mRNAs that is reversed by ACTH<sub>1-24</sub> infusion. Other pro-opiomelanocortin (POMC)-derived peptides can participate in adrenal maturation because, in fetal sheep, infusion of N-POMC 1-77 stimulates adrenal growth and 21-hydroxylase (P450c21) mRNA expression.

In fetal rats, plasma corticosterone levels rise progressively from E16 to E19. The pattern of plasma ACTH levels parallels that of corticosterone, indicating that the observed adrenal hyperactivity during late gestation is driven by increased ACTH secretion from the corticotropes. In newborn rats, circulating ACTH levels are low during the first 10 days after birth, the so-called stress hyporesponsive period, and the density of adrenal ACTH-binding sites and basal and ACTH-stimulated adenylate cyclase activity are decreased in the adrenals of P7 rats. Chronic administration of ACTH or POMC-derived peptides (e.g., Lys- $\gamma$ <sub>3</sub>-melanocyte-stimulating hormone) during this period has a trophic effect on the adrenal and potentiates the subsequent corticosterone response to stress or to ACTH injection.

## Angiotensin II

In adults, angiotensin II, acting through membrane G protein-coupled receptors, is one of the most important factors involved in the regulation of aldosterone biosynthesis and secretion. Two subtypes of angiotensin II receptors have been described: AT<sub>1</sub> and AT<sub>2</sub>.

AT<sub>1</sub> and AT<sub>2</sub> receptors are present on human gestational week 16 (W16) to W18 fetal adrenocortical cells. AT<sub>1</sub> receptors, which mediate most of the known actions of angiotensin II, are located in the definitive zone, whereas AT<sub>2</sub> receptors are present throughout the gland, with a predominant labeling in the fetal zone. It has been proposed that AT<sub>2</sub> receptors are involved in the apoptotic process observed in the human fetal adrenal gland and could participate, after birth, in the involution of the fetal zone.

In sheep, isolated adrenal cells obtained from E40 to E90 fetuses secrete aldosterone and respond to relatively high doses of angiotensin II. At E100 to E130, adrenal glands become unresponsive to angiotensin II. *In vivo*, during late gestation, fetuses are less responsive to infused angiotensin II, in terms of aldosterone secretion, than are adult sheep. In addition, *in vitro*, angiotensin II inhibits ACTH-induced cortisol secretion and P450c17 expression. This phenomenon may be mediated via AT<sub>1</sub> receptors. The mRNA coding for the AT<sub>1</sub> receptor is first detected in the unzoned gland as early as E40 and is present at high levels in the zona glomerulosa and, to a lesser extent, in the zona fasciculata at E60 to E105. During late gestation (E120–E135), the AT<sub>1</sub> receptor hybridization signal decreases before showing a further increase to reach adult values by P2. The fall in AT<sub>1</sub> receptor expression during late gestation could allow ACTH to override angiotensin

II-induced inhibition of cortisol secretion and P450c17 expression. The mRNA for the AT<sub>2</sub> receptor is present in the same location from E40 to E130 and declines to extremely low levels after E140.

In the rat adrenal zona glomerulosa, angiotensin II receptor content decreases from very high levels at birth to adult levels by P20. This decrease is due to a decline in AT<sub>2</sub> receptors. In P7 rats, aldosterone secretion is more sensitive to the stimulatory effect of ACTH than of angiotensin II.

### Placental Factors

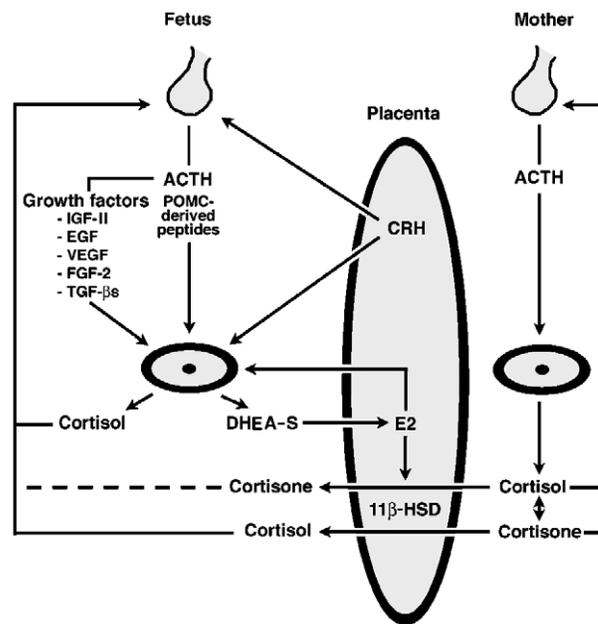
Functional interactions exist between the placenta and the fetal adrenals. The placenta synthesizes corticotropin-releasing hormone (CRH), the major hypothalamic neuropeptide that stimulates anterior pituitary POMC synthesis and ACTH secretion. Estrogens, which originate mainly from the placental conversion of fetal DHEA sulfate (DHEA-S), have a direct effect on fetal adrenal activity and regulate placental metabolism of maternal glucocorticoids (Fig. 1).

#### Corticotropin-Releasing Hormone

In primates, CRH and POMC synthesized in the placenta are released in both the maternal and fetal circulations. Plasma CRH increases exponentially during gestation, peaking at labor. The role of placental CRH in the regulation of fetal adrenal function is not clear. A circulating binding protein is present in the human fetal circulation and is capable of inactivating a large amount of CRH. CRH may modulate fetal adrenal function indirectly by stimulating placental or fetal anterior pituitary ACTH and POMC-derived peptide release, or it may do so directly by regulating the fetal adrenal cortex. Indeed, human midgestation fetal adrenals express CRH receptor mRNA. CRH stimulates DHEA-S and cortisol production, stimulates P450<sub>ssc</sub> and P450<sub>c17</sub> expression, and increases ACTH responsiveness in cultured human fetal adrenocortical cells.

#### Estrogens

Estrogens play an important role in regulating, either directly or indirectly, cortisol and DHEA synthesis and secretion. The primate fetal adrenal begins to synthesize cortisol *de novo* from cholesterol between mid- and late gestation. As a consequence, at midgestation, most if not all of the glucocorticoids in the fetal circulation originate from the mother. Maternal glucocorticoids are metabolized in the placenta, which



**Figure 1** Hormonal regulation of fetal adrenal growth and steroid biosynthesis. It is proposed that fetal anterior pituitary adrenocorticotrophic hormone (ACTH), either directly or mediated by growth factors, and pro-opiomelanocortin (POMC)-derived peptides have a stimulatory effect on fetal adrenal growth and steroid biosynthesis. Placental estrogens, originating from local conversion of the fetal adrenal androgen precursor dehydroepiandrosterone (DHEA), can regulate fetal adrenal metabolism directly or can act through a modulation of placental 11 $\beta$ -hydroxysteroid dehydrogenase (11 $\beta$ -HSD) activity. Between early and midgestation, 11 $\beta$ -HSD shows preferential reduction of cortisone to cortisol, leading to decreased pituitary ACTH secretion and subsequent reduced adrenal maturation and cortisol biosynthesis. With advancing gestation, the increase in estrogen production enhances placental 11 $\beta$ -HSD oxidation of cortisol to cortisone, leading to decreased cortisol feedback at the fetal pituitary level and acceleration of adrenal maturation and stimulation of cortisol and DHEA secretion, thereby creating a positive feedback loop. Placental corticotropin-releasing hormone (CRH) may modulate fetal adrenal function indirectly by stimulating fetal anterior pituitary ACTH and POMC-derived peptide release or directly by regulating fetal adrenal cortex. IGF-II, insulin-like growth factor-II; EGF, epidermal growth factor; VEGF, vascular endothelial growth factor; FGF-2, fibroblast growth factor-2; TGF- $\beta$ s, members of the transforming growth factor- $\beta$  family; DHEA-S, dihydroepiandrosterone sulfate; E2, estradiol.

synthesizes the two isoforms of 11 $\beta$ -hydroxysteroid dehydrogenase (11 $\beta$ -HSD). 11 $\beta$ -HSD-1, which is expressed in placental intermediate trophoblast cells and in the vascular endothelium, has both oxidase (active cortisol-to-inactive cortisone) and reductase (cortisone-to-cortisol) activities. 11 $\beta$ -HSD-2, which is present in placental syncytiotrophoblast cells, exhibits only oxidase activity. Because 11 $\beta$ -HSD activity

Table I Main Characteristics of the Major Growth Factors Involved in Fetal Adrenal Growth and Steroidogenesis

Growth factor	Site of synthesis	Receptors location	Role in growth	Role in steroidogenesis	Regulated by
IGF-II	C, DZ, TZ, FZ	DZ, TZ, FZ	Stimulates FZ proliferation	Stimulates P450ssc and P450c17 activity	ACTH
EGF	Mediated by TGF- $\alpha$	DZ, FZ	Stimulates FZ and DZ proliferation; synergistic with IGF-II	Stimulates 3 $\beta$ -HSD activity in DZ and TZ; stimulates cortisol and aldosterone secretion	ND
VEGF	Mainly FZ	ND	Angiogenic factor	ND	ACTH forskolin
FGF-2	Cortex	ND	Stimulates FZ and DZ proliferation; synergistic with IGF-II	ND	ACTH cAMP
TGF- $\beta$ 1	ND	Neocortex	Inhibits basal and EGF-stimulated growth of FZ and TZ	Inhibits P450ssc and P450c17 activity; inhibits cortisol and DHEA secretion	ND
Activin	DZ, TZ	DZ, FZ	Inhibits basal and EGF-stimulated FZ proliferation	Inhibits cortisol secretion	ACTH

*Note.* IGF-II, insulin-like growth factor-II; EGF, epidermal growth factor; VEGF, vascular endothelial growth factor; FGF-2, fibroblast growth factor-2; TGF- $\beta$ 1, transforming growth factor- $\beta$ 1; C, capsule; DZ, definitive zone; TZ, transitional zone; FZ, fetal zone; P450ssc, side-chain cleavage enzyme; P450c17, enzyme complex having 17 $\alpha$ -hydrolase and 17,20-lyase activities; 3 $\beta$ -HSD, 3 $\beta$ -hydroxysteroid dehydrogenase; DHEA, dehydroepiandrosterone; ACTH, adrenocorticotropic hormone; cAMP, cyclic adenosine monophosphate; ND, not determined.

shows preferential reduction of cortisone to cortisol between early and midgestation, increased circulating fetal cortisol leads to decreased pituitary ACTH synthesis and secretion and to subsequent reduced adrenal maturation and cortisol biosynthesis capacity. With advancing gestation, the increase in estrogen production enhances placental 11 $\beta$ -HSD oxidation of cortisol to cortisone, leading to decreased cortisol feedback at the fetal pituitary level. The subsequent increase in anterior pituitary ACTH synthesis and secretion induces acceleration of adrenal maturation and stimulation of cortisol secretion by the definitive zone and of DHEA production by the fetal zone, creating a positive feedback loop. In addition to their effects on placental glucocorticoid metabolism, estrogens can act directly at the adrenal level. Indeed, the primate fetal adrenal contains estrogen receptors  $\beta$ . Also, *in vivo*, estrogen treatment increases the responsiveness of the fetal adrenal gland to ACTH, presumably through stimulation of protein kinase A activity.

In sheep, during late gestation, estrogens also have a stimulatory effect on fetal ACTH and cortisol secretion. However, unlike in primates, the stimulatory effect of estrogens on the HPA axis takes place at the central nervous system level, that is, on the hypothalamic ACTH secretagogues arginine vasopressin and CRH.

## ROLE OF GROWTH FACTORS

As mentioned previously, ACTH is the primary regulator of adrenal growth and steroid biosynthesis. However, because ACTH is not a mitogen per se, it is now generally accepted that some of the trophic actions of ACTH on the fetal adrenal cortex are mediated indirectly via tissue growth factors such as insulin-like growth factors (IGF-I and IGF-II), epidermal growth factor (EGF), vascular endothelial growth factor (VEGF), fibroblast growth factor-2 (FGF-2), members of the transforming growth factor- $\beta$  (TGF- $\beta$ ) family, and adrenomedullin (Table I).

### Insulin-like Growth Factors

IGF-I and IGF-II are mitogenic peptides, structurally related to proinsulin, that affect growth and function in a wide variety of cell types and can act as autocrine, paracrine, or endocrine factors. The biological actions of IGF-I and IGF-II are modulated by insulin-like growth factor-binding proteins (IGFBPs). IGF-I mediates many of the somatotrophic actions of growth hormone, whereas IGF-II is important in the regulation of fetal development. The effect of IGFs on adrenal cortical cells is most likely mediated through

IGF type 1 receptors because this subtype has been identified in both adult human and fetal nonhuman primate adrenal glands and because IGF type 2 receptor does not bind IGF-II.

Circulating concentrations of IGF-II are high during fetal life in several species (e.g., primates, sheep, rodents) and decrease after delivery. IGF-I and IGF-II mRNAs are present in adrenals from mid-gestation human fetuses. IGF-I mRNA is detected only in the capsule and not in the cortical zones, whereas IGF-II mRNA is detected in the definitive and fetal zones as well as in the capsule. The IGF-II mRNA is present in high abundance in human fetal adrenals and is barely detectable in adult adrenals, whereas the mRNA encoding IGF-I is expressed at low levels in fetus adrenals and at high levels in adult adrenal glands. In adrenals obtained from nonhuman primates, IGF-II mRNA is abundant from early (E60) to late (E165) gestation and is localized in the definitive, transitional, and fetal zones. The mRNAs coding for IGFBP-2 and IGFBP-6, which have been shown to bind preferentially IGF-II, are expressed in adrenals obtained from monkey fetuses in the definitive, transitional, and fetal zones. In the developing ovine adrenal, IGF-II mRNA is highest in E60 fetuses, decreases slightly between E60 and E100, remains relatively constant until term, and decreases significantly after birth, with IGF-I mRNA being expressed at very low levels. At all gestational ages, IGF-II mRNA and protein are localized in the capsule and mesenchymal cells surrounding the gland and in the steroidogenic cells of the zona glomerulosa and zona fasciculata.

IGFs have profound effects on adrenal growth. Transgenic mice that are IGF-I, IGF-II, or IGF type 1 receptor-deficient have intrauterine growth retardation. However, IGF-II-deficient mice have near-normal postnatal growth due to IGF-I. In ovine fetuses, chronic infusion of either IGF-I or IGF-II between E120 and E130 results in an increase in adrenal growth. *In vitro*, IGF-I and/or IGF-II stimulate proliferation of fetal adrenocortical cells obtained from sheep or human fetuses. As mentioned previously, IGF-II may mediate the trophic actions of ACTH on the fetal adrenal gland. Treatment of rhesus monkey fetuses with metyrapone, which likely increases pituitary ACTH secretion, induces in the adrenals hypertrophy of all cortical zones together with an increase in the concentrations of IGF-II and IGF type 1 receptor mRNAs. *In vitro*, in human fetal adrenal cells in culture, ACTH increases IGF-II mRNA levels. In addition to its effects on adrenal growth,

IGF-II regulates adrenal steroidogenesis. It has been demonstrated, using primary cultures of adrenal cortical cells obtained from human mid-gestation fetuses, that IGF-I and IGF-II stimulate basal and ACTH-, forskolin-, or cAMP-induced cortisol and DHEA production. Under the same experimental conditions, IGF-II increases adrenocortical responsiveness to ACTH without any effect on ACTH receptor mRNA, suggesting that IGF-II modulates ACTH sensitivity in fetuses by increasing ACTH signal transduction at some point distal to the ACTH receptor. In addition, IGF-II increases ACTH-stimulated abundance of P450<sub>ssc</sub> and P450<sub>c17</sub> mRNAs, thereby augmenting the potential for adrenal androgen synthesis.

### Epidermal Growth Factor

Human cord blood EGF levels have been shown to increase with progressive gestation, suggesting a functional role for EGF during the perinatal period. Knockout mice for the EGF receptor have intrauterine growth retardation. However, EGF and EGF precursor mRNA are expressed as low levels, and tissue EGF immunoreactivity appears late in rodent fetuses. Therefore, it has been proposed that TGF- $\alpha$ , a member of the EGF family, is the ligand for the fetal EGF receptor. TGF- $\beta$  has been identified in steroidogenic cells of adult adrenal cortex. EGF receptor concentrations increase steadily from E15 until birth in several mouse tissues. EGF receptors have been identified in both the definitive and fetal zones from mid-gestation human adrenal fetuses. EGF has been shown to stimulate the proliferation of human fetal adrenal cells *in vitro* and to act cooperatively with IGF-I and IGF-II. Late-gestation EGF-treated monkey fetuses show an increase in adrenal weight due to a hypertrophy of the definitive zone and stimulation of 3 $\beta$ -HSD immunoreactivity in the definitive and transitional zones. In EGF-infused fetal sheep, adrenal cortical hypertrophy is accompanied by increased cortisol and aldosterone secretion. EGF may act directly at the adrenal level or indirectly through stimulation of the hypothalamo-pituitary axis because EGF increases the secretion of CRH from the hypothalamus and of ACTH from the anterior pituitary. It has been demonstrated that, in fetal monkeys, EGF acts on the hypothalamic-pituitary axis to modulate adrenal cortical growth and functional maturation of the transitional zone, whereas EGF can act independently of the hypothalamic-pituitary axis to stimulate functional maturation of the definitive zone.

## Vascular Endothelial Growth Factor

VEGFs are a family of direct-acting endothelial cell mitogens and angiogenic factors that derive from a single gene by alternative splicing. The mRNAs coding for the four isoforms of VEGF and for the two VEGF receptors have been identified in the adult mouse adrenal gland. VEGF mRNA and protein have been detected in the adrenal cortex from midgestation human fetuses, predominantly in the fetal zone. The predominant staining for VEGF in the fetal zone correlates with the extensive vasculature of this zone. In primary cultures of human fetal adrenal cortical cells, ACTH and forskolin increase both VEGF mRNA levels and VEGF protein secretion.

## Fibroblast Growth Factor-2

FGF-2 (also called basic FGF) is a potent angiogenic molecule that belongs to the family of the FGF and interacts with four cell surface receptor subtypes. FGF-2 is synthesized by human W16 fetal adrenal cortex and is stimulated by ACTH and cAMP. FGF-2 mRNA increases steadily during gestation in adrenal glands obtained from nonhuman primate fetuses. *In vitro*, FGF-2 stimulates the proliferation of human fetal adrenal cells from the definitive or fetal zone, with its effect being additive to those of IGF-I and IGF-II.

## Transforming Growth Factor- $\beta$

TGF- $\beta$ s are a family of potent multifunctional cytokines that modulate a wide variety of cellular activities. Members of the superfamily include TGF $\beta$ -1, activin, and inhibin.

### TGF- $\beta$ 1

TGF- $\beta$ 1 immunoreactivity has been detected in the adrenal cortex from adult and neonatal mice. However, expression of TGF- $\beta$ 1 in human adrenals during development remains to be determined. TGF- $\beta$ 1-binding sites have been identified in fetal human cortical cells. Several reports indicate that TGF- $\beta$ 1 inhibits basal and EGF-stimulated growth of human fetal cortical cells, in both the definitive and fetal zones, possibly through a stimulation of apoptosis. This inhibitory effect is significantly blunted by ACTH. TGF- $\beta$ 1 is also a potent inhibitor of adrenal steroidogenesis. Incubation of primary cultures of human fetal cortical cells with TGF- $\beta$ 1 results in a decrease of both basal and ACTH-induced DHEA secretion and P450c17 gene expression. In fetal sheep, adrenal TGF- $\beta$ 1 has an inhibitory effect on

basal P450ssc expression and on ACTH-induced cortisol secretion and P450c17 activity. The interplay between TGF- $\beta$ 1 and ACTH in regulating adrenal growth and steroid production does not involve mutual control of their respective receptors because TGF- $\beta$ 1 does not influence ACTH receptor gene expression and because ACTH increases TGF- $\beta$ 1 binding in the fetal adrenal cortex.

### Inhibins

Inhibins are dimers of an  $\alpha$ -subunit and either a  $\beta_A$ - or a  $\beta_B$ -subunit, whereas activins are homo- or heterodimers of either a  $\beta_A$ - or a  $\beta_B$ -subunit.  $\alpha$ -,  $\beta_A$ -, or  $\beta_B$ -subunit immunoreactivity and mRNAs, as well as activin type I/II receptor and inhibin receptor mRNAs, have been detected in the definitive and fetal zone of adrenals from midgestation human fetuses. *In vitro*, activin A inhibits basal and EGF-stimulated fetal zone proliferation, possibly through enhanced apoptosis, whereas inhibin A has no apparent mitogenic effect. The inhibitory effect of activin A on fetal adrenal cell proliferation is additive to that of TGF- $\beta$ . In addition, ACTH has a stimulatory effect on  $\alpha$  and  $\beta_A$  subunit mRNA levels and inhibin A/B secretion from primary cultures of human fetal adrenal cells.

## Adrenomedullin

Adrenomedullin is a multifunctional peptide, initially purified from an adrenal tumor of the medulla, that shows structural homology with calcitonin gene-related peptide (CGRP). In mice, adrenomedullin mRNA and protein are present in the adrenal primordia as early as E12. The observations that antagonism of adrenomedullin function during rat pregnancy causes fetal growth restriction, and that proadrenomedullin N-terminal 20 peptide enhances proliferation of adult rat zona glomerulosa cells by acting through CGRP1 receptors, suggest that proadrenomedullin-derived peptides may be important in regulating adrenal growth during development.

## See Also the Following Articles

ACTH (Adrenocorticotrophic Hormone) • ACTH,  $\alpha$ -MSH, and POMC, Evolution of • Adrenal Cortex, Anatomy • Adrenal Cortex, Development • Adrenal Insufficiency • Angiotensin, Evolution of • EGF and Related Growth Factors • Fibroblast Growth Factor (FGF) • Glucocorticoids, Overview • Insulin-like Growth Factors

## Further Reading

- Challis, J. R. G., Sloboda, D., Matthews, S. G., Holloway, A., Alfaidy, N., Patel, F. A., Whittle, W., Fraser, M., Moss, T. J. M., and Newnham, J. (2001). The fetal placental hypothalamic-pituitary-adrenal (HPA) axis, parturition, and postnatal health. *Mol. Cell. Endocrinol.* **185**, 135–144.
- Coulter, C. L., Ross, J. T., Owens, J. A., Bennett, H. P. J., and McMillen, I. C. (2002). Role of pituitary POMC-peptides and insulin-like growth factor II in the developmental biology of the adrenal gland. *Arch. Physiol. Biochem.* **110**, 99–105.
- Mesiano, S., and Jaffe, R. B. (1997a). Developmental and functional biology of the primate fetal adrenal cortex. *Endocrine Rev.* **18**, 378–403.
- Mesiano, S., and Jaffe, R. B. (1997b). Role of growth factors in the developmental regulation of the human fetal adrenal cortex. *Steroids* **62**, 62–72.
- Naaman Réperant, E., and Durand, P. (1997). The development of the ovine fetal adrenal gland and its regulation. *Reprod. Nutr. Dev.* **37**, 81–95.
- Pepe, G. J., and Albrecht, E. D. (1995). Actions of placental and fetal adrenal steroid hormones in primate pregnancy. *Endocrine Rev.* **16**, 608–648.



# Adrenal Cortex, Anatomy

Ilias Vrezas, Holger S. Willenberg, and Stefan R. Bornstein  
University of Düsseldorf, Düsseldorf, Germany

## Glossary

**adrenal cortex** The outer portion of the adrenal gland. It produces glucocorticoid and mineralocorticoid hormones and adrenal androgens.

**adrenal gland** A pair of small glands, each of which is located on top of one of the kidneys.

**cortisol** The major natural glucocorticoid in humans and the primary stress hormone.

The adrenal gland was first described by Bartholomeus Eustachius in 1563. However, it was Thomas Addison in 1855 who first recognized the importance of the adrenal glands, and in 1856 Charles Edward Brown-Séquard showed by bilateral adrenalectomy in experimental animals that the function of these glands was necessary for life. Using histochemical techniques developed in the mid-19th century, it was demonstrated that the adrenal medulla and the adrenal cortex have divergent cellular and functional properties.

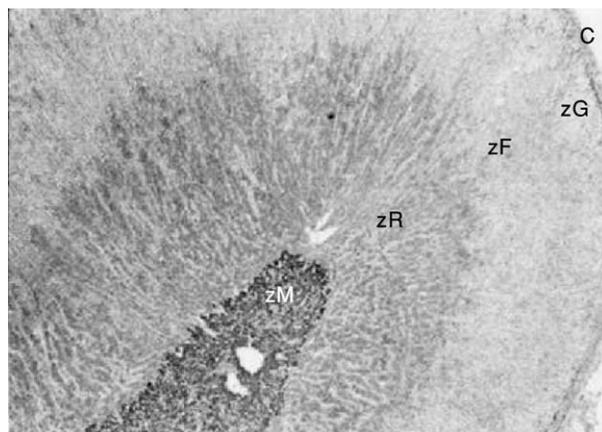
## HISTOLOGY

Whereas the fetal cortex mainly consists of the zona fetalis, the adult adrenal cortex consists of at least three anatomically distinct zones: the outer zona glomerulosa, which is the site of mineralocorticoid production (e.g., aldosterone); the central zona fasciculata, which is predominantly responsible for glucocorticoid production; and the inner zona reticularis, where adrenal androgens [predominantly dehydroepiandrosterone (DHEA), DHEA sulfate, and androstenedione] are located and some glucocorticoid synthesis (cortisol and corticosterone) occurs (Fig. 1). In rats, a fourth zone, the zona intermedia, can be discerned that is believed to contain adrenocorticoid stem cells and to be the region in which adrenocyte differentiation begins.

The adrenal cortex synthesizes exclusively steroid hormones, which are derived from various

modifications of the precursor cholesterol. Glucocorticoids are secreted during the course of stress regulation and act mainly on intermediary metabolism and the immune system. Mineralocorticoids exert their main action on salt and water homeostasis. Adrenal androgens show a testosterone-like effect but can also be precursors for aromatization to estrogens. The human adrenal cortex is able to synthesize more than 50 steroids, but not all of them are secreted into the blood circulation or are biologically active. In addition, adrenocortical cells can secrete active peptides, cytokines, and other hormones.

At the ultrastructural level, cells of the zona glomerulosa are rounded and smaller than the polyhedral cells of the zona fasciculata, which gradually extend into the zona reticularis. The latter zone consists of cells that appear identical to those of the zona fasciculata and also those of another type of smaller cells with a dark-staining nucleus. The cells of the adrenal cortex are arranged in a cord-like manner, extending from the adrenal capsule to the medulla and are surrounded by a capillary network. In adrenocortical cells, the mitochondria are particularly numerous, and the smooth endoplasmic reticulum is especially



**Figure 1** Cross-section of a human adrenal gland double immunostained for 17 $\alpha$ -hydroxylase and chromogranin A. zM, adrenal medulla; zR, zona reticularis; zF, zona fasciculata; zG, zona glomerulosa; C, adrenal capsule.



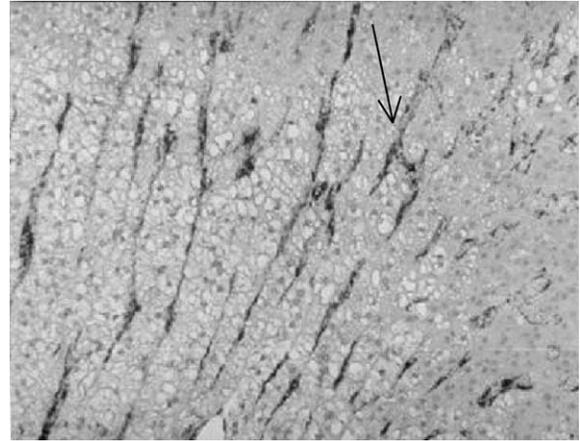
**Figure 2** Adrenal gland double immunostained for  $17\alpha$ -hydroxylase (zona reticularis and zona fasciculata, gray) and for tyrosine hydroxylase (adrenal medulla, black). Black islets of medullary cells are located in the gray adrenal cortex, and gray islets of adrenocortical cells are located in the black medulla.

abundant and forms a network of anastomosing tubules.

The classic view of a strict separation between the steroid-producing adrenal cortex and the catecholamine-producing medulla has been shown to be an oversimplification; displaced chromaffin cells have been found in all zones of the adult adrenal cortex and, similarly, cortical cells are found in the medulla (Fig. 2). The close anatomic colocalization of the cortical and medullary cells has been suggested to be a prerequisite for paracrine interactions.

## BLOOD SUPPLY

Relative to its small size, the adrenal gland is one of the most extensively vascularized organs in the body, with an estimated flow rate of 5 ml per minute. Each gland may be supplied by as many as 50 arterial branches, or arterioles, that arise directly from the aorta, the renal arteries, and the inferior phrenic arteries. The subcapsular arteriolar plexus receives this blood supply and distributes it via two types of vessels. Through the sinusoids, both the adrenal cortex and medulla are supplied, and through the medullary arteries there is a direct blood supply to the medulla. Blood converges at the corticomedullary junction, and through the



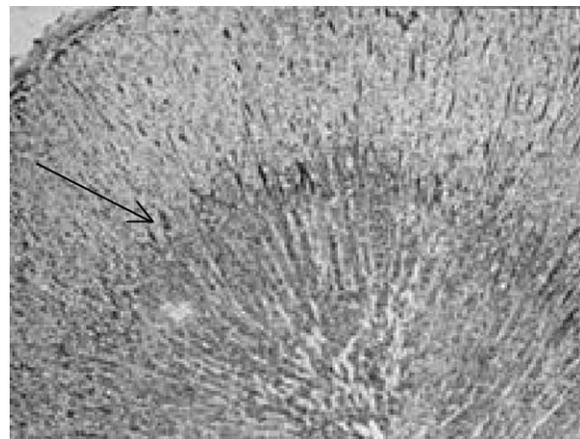
**Figure 3** Adrenocortical endothelial cells (arrow) immunostained for CD31.

central adrenal vein it is drained directly or indirectly (via the renal vein) into the inferior vena cava (Fig. 3).

## INNERVATION

The adrenal cortex receives extensive afferent innervation, with evidence of direct contact between nerve terminals and cortical cells. A possible efferent innervation has been reported, with the presence of baroreceptors and chemoreceptors in the adrenal cortex.

Adrenal innervation influences compensatory adrenal hypertrophy and has been implicated in the regulation of the diurnal variation of cortisol secretion. Moreover, splanchnic nerve activation has been demonstrated to regulate adrenal steroid release (Fig. 4).



**Figure 4** Nerve cells and fibers (arrow) silver stained.

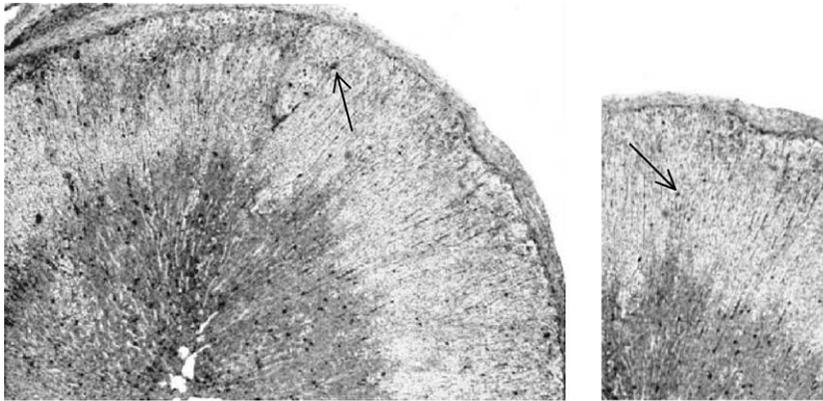


Figure 5 Lymphocytes (arrows) immunostained for CD45.

## IMMUNE CELLS

Macrophages are found within the adrenal cortex. Not only do they possess the ability to act as phagocytic cells but also they are able to produce and secrete a variety of different compounds, such as cytokines [interleukin-1 (IL-1), IL-6, and tumor necrosis factor- $\alpha$ ] and neuropeptides (VIP), that influence adrenocortical function. It has been demonstrated that lymphocytes infiltrate the adrenal cortex and have the ability to produce ACTH-like substances (Fig. 5). In addition, a cytokine-independent cell-cell-mediated regulation of adrenal androgen release has been demonstrated.

## See Also the Following Articles

ACTH (Adrenocorticotropic Hormone) • Adrenal Androgens • Adrenal Cortex, Development • Adrenal Cortex Development, Regulation of • Adrenal Cortex, Physiology • Adrenal Insufficiency • Adrenal Suppression

## Further Reading

Bornstein, S. R., Ehrhart-Bornstein, M., Scherbaum, W. A., Pfeiffer, E. F., and Holst, J. J. (1990). Effects of splanchnic nerve stimulation on the adrenal cortex may be mediated by chromaffin cells in a paracrine manner. *Endocrinology* **127**, 900–906.

Bornstein, S. R., González-Hernández, J. A., Ehrhart-Bornstein, M., Adler, G., and Scherbaum, W. A. (1994). Intimate contact

of chromaffin and cortical cells within the human adrenal gland forms the cellular basis for important intraadrenal interactions. *J. Clin. Endocrinol. Metab.* **78**, 225–232.

Dallman, M. F., Engeland, W. C., and McBride, M. H. (1977). The neural regulation of compensatory adrenal growth. *Ann. N. Y. Acad. Sci.* **297**, 373–392.

Dijkstra, I., Binnekade, R., and Tilders, F. J. H. (1996). Diurnal variation in resting levels of corticosterone is not mediated by variation in adrenal responsiveness to adrenocorticotropin but involves splanchnic nerve integrity. *Endocrinology* **137**, 540–547.

Dinarello, C. A. (1992). The biology of interleukin 1. In “Interleukins: Molecular Biology and Immunology” (T. Kishimoto, ed.), pp. 1–32. Karger, Basel.

Ehrhart-Bornstein, M., Hinson, J. P., Bornstein, S. R., Scherbaum, W. A., and Vinson, G. P. (1998). Intraadrenal interactions in the regulation of adrenocortical steroidogenesis. *Endocr. Rev.* **19**(2), 101–143.

González-Hernández, J. A., Bornstein, S. R., Ehrhart-Bornstein, M., Geschwend, J. E., Adler, G., and Scherbaum, W. A. (1994). Macrophages within the human adrenal gland. Morphological data for a possible local immune-neuroendocrine interaction. *Cell Tissue Res.* **278**, 201–205.

Ottaway, C. A. (1991). Vasoactive intestinal peptide and immune function. In “Psychoneuroimmunology” (R. Ader, D. L. Felten, and N. Cohen, eds.), pp. 225–262. Academic Press, San Diego.

Vinson, G. P., Hinson, J. P., and Tóth, I. E. (1994). The neuroendocrinology of the adrenal cortex. *J. Neuroendocrinol.* **6**, 235–246.

Wolkersdörfer, G. W., Lohmann, T., Marx, C., Schröder, S., Pfeiffer, R., Stahl, H.-D., Scherbaum, W. A., Chrousos, G. P., and Bornstein, S. R. (1999). Lymphocytes stimulate dehydroepiandrosterone production through direct cellular contact with adrenal zona reticularis cells: A novel mechanism of immune-endocrine interaction. *J. Clin. Endocrinol. Metab.* **84**, 4220–4227.



# Adrenal Cortex, Development

Michel Grino

Université de la Méditerranée, Marseille, France

## Glossary

**embryogenesis** Differentiation of the fertilized ovum during the period of most rapid development, i.e., after the long axis appears until all major structures are represented.

**nuclear receptor** Ligand-inducible transcription factor that specifically regulates the expression of target genes involved in metabolism, development, and reproduction.

**pregnancy or gestation** The condition of having a developing embryo or fetus in the body. Duration of pregnancy is 266 days in women, 184 days in baboons, 165 days in rhesus monkeys, 145–150 days in sheep, 21 days in rats, and 18.5 days in mice.

**steroid hormones** Lipophilic molecules, having a C<sub>17</sub> ring as the basis of their chemical structure, that freely cross the cell membrane and interact with nuclear receptors.

**transcription factor** Protein that directly affects the initiation of transcription of specific genes.

**M**aturation of the hypothalamic–pituitary–adrenal (HPA) axis, which is characterized by increased activity during late gestation, is essential for the development of the fetus and plays a critical role in preparing for its transition to extrauterine life. The fetal adrenal cortex synthesizes and secretes androgens and glucocorticoids. In primates, androgens are necessary for placental conversion to estradiol, a hormone that is crucial for the maintenance of pregnancy, the maturation of the fetal and maternal tissues, and immunosuppression, leading to implantation of the placenta and the fetus. Glucocorticoids are essential for the maturation of brain, lung, liver, gut, kidney, and the adrenal itself. In some species, a surge in fetal glucocorticoid secretion has been suggested as being integral to the cascade of events leading to the onset of parturition. However, premature or abnormal exposure of fetuses or newborns to high levels of glucocorticoids permanently programs the HPA axis, leading to an increased prevalence of metabolic and cardiovascular disease. This article details both morphological and functional aspects of adrenal cortex development in several types of

mammals, including primates, ruminants, and rodents. In particular, the role of nuclear receptors and transcription factors in the regulation of adrenocortical organogenesis and steroidogenesis is examined.

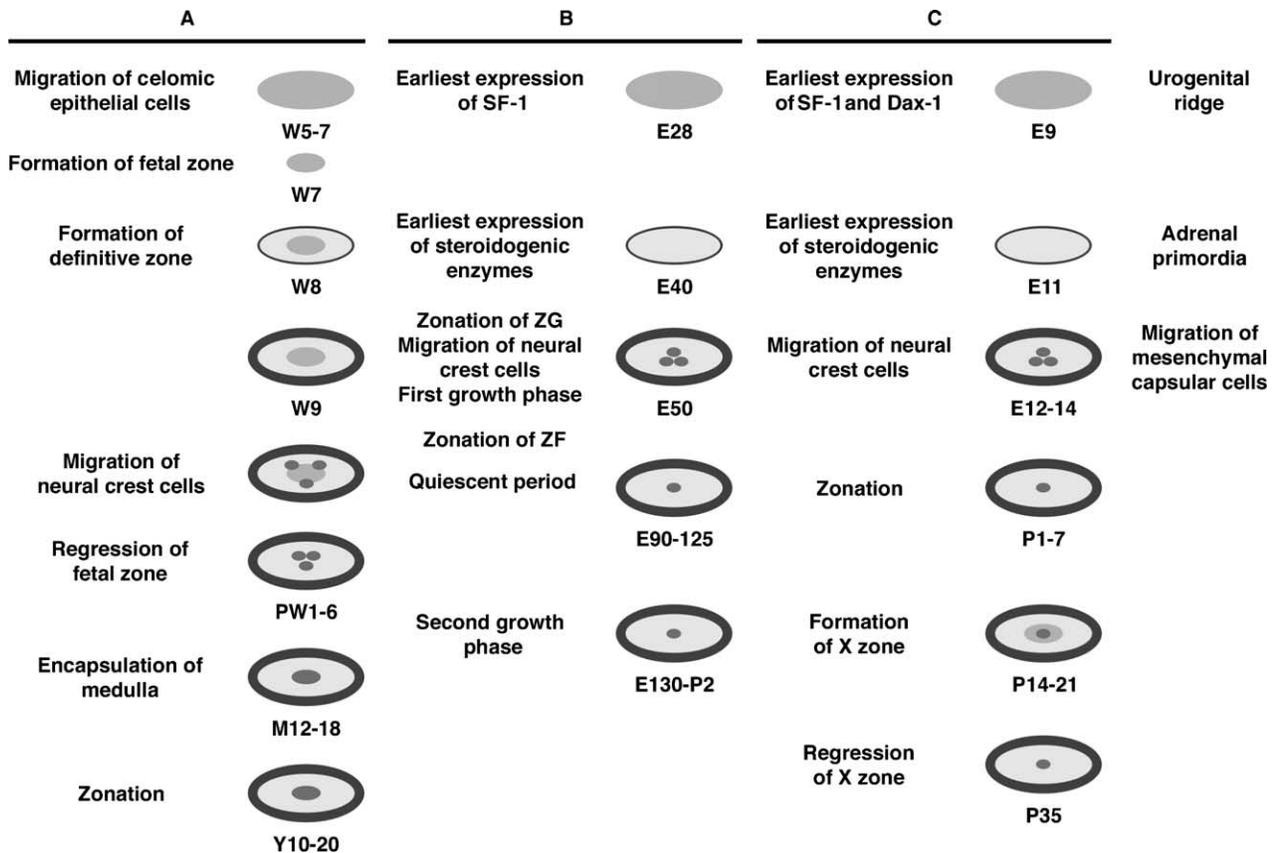
## EMBRYOGENESIS, DEVELOPMENT AND GROWTH

**Figure 1** depicts the major milestones of adrenal cortex development in primates, sheep, and rodents.

### In Primates

Human adrenal development begins at approximately the fourth week of gestation and continues into adult life. Adrenocortical cells derive from a single cell lineage that originates in the celomic epithelium in the notch between the primitive urogenital ridge and the dorsal mesentery. These cells are also the origin of gonadal and kidney structures. Five landmark phases have been described:

- Condensation of the celomic epithelium (3–4 weeks of gestation).
- Proliferation and migration of celomic epithelial cells (weeks 4–6) that stream medially and cranially, accumulating at the cranial end of the mesonephros, forming the adrenal blastema.
- Morphological differentiation of fetal adrenal cortical cells into two distinct zones (weeks 8–10): the fetal zone and the definitive zone. The fetal zone is an inner cluster of large, eosinophilic cells and represents the largest part (80–90%) of the adrenal cortex. The definitive zone is a thin outer band of small basophilic cells, densely packed, showing structural characteristics of proliferative cells that appear to function as a reservoir of progenitor cells that may populate the remainder of the gland. A third zone, located between the fetal and definitive zones, has been called the transitional zone. By week 30 of gestation, the definitive and transitional zones resemble the adult zona glomerulosa and the zona fasciculata,



**Figure 1** The adrenocortical development cascade in human (A), sheep (B), and mouse (C), showing progression from the urogenital ridge to final zonation. W, week of gestation; PW, postnatal week; M, month; Y, year; E, day of gestation; P, postnatal day; ZG, zona glomerulosa; ZF, zona fasciculata. Adapted from *Trends in Endocrinology and Metabolism* 13, Keegan, C. E., and Hammer, G. D. Recent insights into organogenesis of the adrenal cortex, pp. 200–208, copyright 2002, with permission from Elsevier.

respectively. A period of rapid growth begins at approximately week 10 and continues to term. The fetal zone grows by hypertrophy and limited proliferation, whereas growth in the definitive zone occurs mainly by hyperplasia. Several lines of morphological evidence indicate that the human fetal adrenal gland is a dynamic organ, in which proliferating cells located at the periphery migrate, differentiate, and finally undergo senescence in the central part of the gland.

- Decline and disappearance of the fetal zone (first 3 postnatal months); by this period, the primate adrenal cortex remodeling involves apoptosis of the fetal zone and expansion of the preexisting zona glomerulosa and zona fasciculata. Following the involution of the fetal zone, chromaffin elements, derived from the fetal ectoderm, begin to cluster around the central vein. The medulla acquires an adult-like pattern by 12–18 months.

- Establishment and stabilization of the adult zonal pattern (10–20 years of age); this leads to individualization of the three distinct cell layers: the outer zona glomerulosa, the central zona fasciculata, and the inner zona reticularis.

### In Sheep

In the ovine fetus, the adrenal gland can be identified as early as embryonic day (E) 28. Its development occurs during three phases, reflecting the interactions of the cellular kinetic phenomena of hyperplasia and hypertrophy:

- Establishment of functional zonation (E50–E90): this first growth phase is characterized by the separation of the medullary and cortical portions and the presence of a well-defined zona glomerulosa at

approximately E50 and zona fasciculata at approximately E90. During this period, the zona fasciculata grows mainly by hyperplasia.

- Quiescence and reactivation (E90–E125): during this period, adrenal growth and adrenal cell division occur at the lowest rate compared to any other time in gestation.

- Structural and functional maturation [E130 to postnatal day 2 (P2)]: this second growth phase is characterized by an increase in the rate of cell growth and cell division of the zona fasciculata, with the rate of increase in the volume of steroidogenic cells becoming greater than at any other stage. Cellular hypertrophy precedes cellular hyperplasia. The zona reticularis does not become apparent before 1 month of life.

### In Rodents

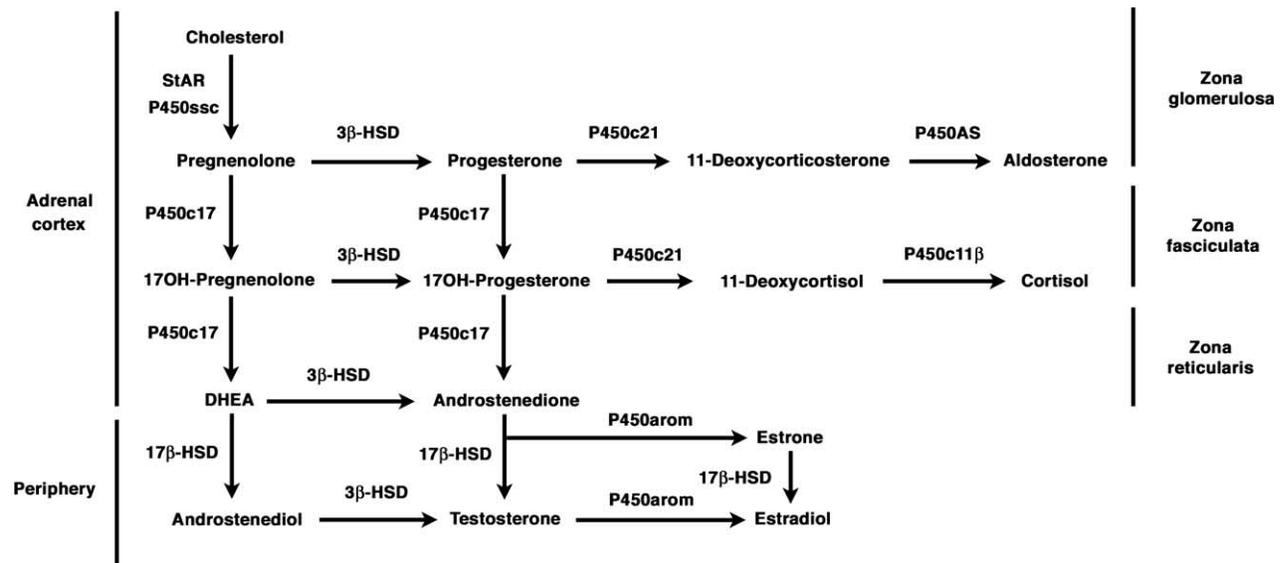
In mice, the adrenal gland starts to develop on E11 from the celomic epithelium, when the anlage of the adrenal cortex is formed. On E12, the cortical cells are found close to the adrenal medulla sympathoblasts. The capsule and the cortical capillaries have completed their development by E15, when the adrenal gland has become a morphologically distinct structure. The individualization of the different layers of the adrenal cortex is nearly completed by birth. The

mouse adrenal possesses a transient developmental zone between the cortical zones and the adrenal medulla: the X zone, which becomes histologically distinct at P10–P14 and enlarges until P21. The X zone subsequently degenerates. The function of the rodent X zone remains unclear.

### ONTOGENY OF STEROID BIOSYNTHESIS

The adrenals synthesize several classes of steroids: androgens, glucocorticoids, and mineralocorticoids. The major human adrenal steroidogenic pathways are summarized in Fig. 2. Schematically, the process of adrenal steroidogenesis has two major components. The first is quantitative; i.e., it regulates how much steroid can be made at a given moment. The second is qualitative; i.e. it regulates which particular steroid is made.

- First step: Conversion of cholesterol to pregnenolone. This step involves the delivery of the substrate cholesterol to the inner mitochondrial membrane, driven by steroid acute regulatory protein (StAR), and the conversion of cholesterol to pregnenolone, catalyzed by P450 steroid chain cleavage (P450ssc), adrenodoxin, and adrenodoxin reductase.



**Figure 2** Schematic view of human adult adrenal and peripheral steroidogenic pathways. StAR, steroid acute regulatory protein; P450ssc, side-chain cleavage enzyme; P450c17, enzyme complex having 17 $\alpha$ -hydroxylase and 17,20-lyase activities; 17 $\beta$ -HSD, 17 $\beta$ -hydroxysteroid dehydrogenase; 3 $\beta$ -HSD, 3 $\beta$ -hydroxysteroid dehydrogenase; P450c21, 21-hydroxylase; P450AS, aldosterone synthase (contains 11 $\beta$ -hydroxylase, 18-hydroxylase, and 18 $\beta$ -hydroxysteroid dehydrogenase activities); P450c11 $\beta$ , 11 $\beta$ -hydroxylase; P450arom, aromatase.

**Table I** Localization and Relative Expression of the Enzymes Involved in Fetal Adrenal Steroid Biosynthesis

Human	< Week 25			> Week 25		
	DZ	TZ	FZ	DZ	TZ	FZ
P450ssc	+	+	+	+	+	+
3 $\beta$ -HSD	-	-	-	++	++	-
P450c17	-	+++	++	-	+++	++
P450c21	++	++	+	++	++	+
P45011 $\beta$	-	+	-	++	+++	++
Sheep	E50		E90		E130	
	Whole cortex		ZG	ZF	ZG	ZF
P450ssc	+++		+	+	+++	+++
3 $\beta$ -HSD	+		+	++	++	++
P450c17	+++		-	++	-	+++
P450c21	+		++	++	+++	+++
P45011 $\beta$			++	++	++	++
Rodents	Whole Cortex					
	E16.5	E18.5	E21	P1	P10	P25
P450ssc	+	++	+++	+	+++	+++
3 $\beta$ -HSD	+	++	+++	+	++	+++
P450c17	+	-	-	-	-	-
P450c21	ND	+	ND	+	+	++
P45011 $\beta$	ND	+	++	+	++	+++

*Note.* E, day of gestation; P, postnatal day; DZ, definitive zone; TZ, transitional zone; FZ, fetal zone; P450ssc, side-chain cleavage enzyme; 3 $\beta$ -HSD, 3 $\beta$ -hydroxysteroid dehydrogenase; P450c17, enzyme complex having 17 $\alpha$ -hydroxylase and 17,20-lyase activities; P450c21, 21-hydroxylase; P450c11 $\beta$ , 11 $\beta$ -hydroxylase; ND, not determined.

- Second step: Transformation of pregnenolone to active hormones. The coordinate regulation of several enzymes will direct the transformation of pregnenolone toward a given class of steroid hormone—androgens, glucocorticoids, and mineralocorticoids. During adrenal cortex development, the synthesis of these various steroids is not temporally coordinated, is zone-dependent, and is species-specific. [Table I](#) summarizes the localization and relative expression of the enzymes involved in fetal adrenal steroid biosynthesis.

## Androgens

The conversion of cholesterol to pregnenolone is mature in the fetal zone during early pregnancy in humans. High levels of 17 $\alpha$ -hydroxylase/17,20-lyase (P450c17) are present in the fetal zone very early (as soon as 44 days postconception). As a consequence,

the fetal zone produces dihydroepiandrosterone (DHEA), the bulk of which appears to be secreted as a 3-sulfoconjugate DHEA sulfate (DHEAS) that is formed by the action of DHEA sulfotransferase and the cofactor 3'-phosphoadenosine 5'-phosphosulfonate. DHEAS is used by the placenta for conversion to estrogens. By term, the fetal zone produces approximately 200 mg DHEA per day. In fetal sheep and rodents, the adrenals do not synthesize DHEA.

## Glucocorticoids

### In Primates

Assays of steroids in cord blood have suggested that human fetus can synthesize cortisol as early as the 10th week of gestation. This hypothesis is supported by clinical and biological observations in patients with congenital adrenal hyperplasia (CAH) caused by deficiency of 21-hydroxylase (P450c21), 3 $\beta$ -hydroxysteroid dehydrogenase (3 $\beta$ -HSD), or 11 $\beta$ -hydroxylase (P450c11 $\beta$ ). Patients with this disorder have impaired glucocorticoid production with a compensatory increase in pituitary adrenocorticotrophic hormone (ACTH) secretion and subsequent stimulation of DHEA synthesis. Interestingly, female infants with CAH have ambiguous external genital development, indicating that, in unaffected female fetuses, cortisol is synthesized by the adrenal by the time of external genital differentiation, a process that is sensitive to androgens and begins at approximately week 10. These observations do not necessarily indicate that the week 10 fetal adrenal is able to synthesize cortisol *de novo* from cholesterol. Although StAR and P450ssc immunoreactivities are expressed in the cytoplasm of human fetal adrenocortical parenchymal cells of the transitional zone as soon as week 14, in early gestation the fetal adrenal can synthesize cortisol using placental progesterone as substrate, since high levels of progesterone are present in the fetal circulation. P450c21 and P450c11 $\beta$  immunoreactivities have been detected in the transitional zone of human fetal adrenals as soon as weeks 13–14 and P450c17 mRNA is present in the transitional zone of week 22 human fetal adrenal. This is consistent with the observation that progesterone infusion into human fetuses between weeks 16 and 18 results in cortisol synthesis. As a consequence, the key enzyme for the production of cortisol *de novo* from cholesterol during early gestation is 3 $\beta$ -HSD. It has been reported that 3 $\beta$ -HSD immunoreactivity and mRNA could not be detected in human adrenals before weeks 22–24, suggesting that the transitional zone of human fetus adrenals is

able to synthesize *de novo* cortisol beginning at week 24. Similarly, in nonhuman primates, 3 $\beta$ -HSD mRNA is undetectable during early gestation, starts to increase at midgestation, and is further stimulated at late gestation. 3 $\beta$ -HSD protein expression in the transitional zone of adrenals from fetal rhesus monkeys follows a comparable developmental pattern.

### **In Sheep**

Adrenal cortisol synthesis and secretion in ovine fetuses follow a triphasic pattern. P450ssc and P450c17 are expressed at a relatively high level in the whole adrenal cortex as early as E40–E60. In the zona fasciculata, the levels of P450ssc and P450c17 decrease between E90 and E120 before showing a further elevation between E130 and P2. At E90, P450c17 expression is confined to the zona fasciculata. 3 $\beta$ -HSD is present at uniformly moderate levels in the zona fasciculata at mid and late gestation. P450c21 mRNA shows a steady increase throughout gestation. Intense P450c11 $\beta$  immunoreactivity is consistently detected throughout the adrenal cortex as early as E90 and remains constant until birth. Cortisol is secreted from adrenal cells from E50 fetuses. Basal cortisol secretion decreases at E100 and increases subsequently, between E130 and E145.

### **In Rodents**

Adult mouse and rat adrenal cortex, which lack P450c17, produce mainly corticosterone. P450ssc and adrenodoxin are expressed at midgestation in the rodent adrenal cortex (E15–E16). 3 $\beta$ -HSD mRNA and protein have been detected in the fetal rat adrenal as early as E16; their labeling in the reticular and fascicular zones is at a higher level than in the glomerular zone at E18. P450c21 and P450c11 $\beta$  immunoreactivity and mRNA are present in the fetal rat adrenal at E18. P450c17 mRNA and activity are detectable in the adrenals of mouse fetuses at E12.5, increase in abundance from E12.5 to E14.5, and are then lost between E16.5 and E18.5, suggesting that the fetal mouse adrenal is able to synthesize cortisol during late gestation. Although the ontogeny of StAR expression in the rodent embryonic adrenal remains to be elucidated, it is known that fetal rat adrenals synthesize and secrete corticosterone as early as E13. The corticosterone contents of the fetal rat adrenal are high from E16 to E20 and plasma corticosterone concentrations rise progressively from E16 to E19. This phenomenon occurs following activation of pituitary ACTH and the hypothalamic neuropeptides that control ACTH synthesis and secretion.

During the first 10 days after birth, there is a marked decrease in adrenal corticosterone contents

and both basal and stress-induced circulating corticosterone levels. This period has been called the “stress hyporesponsive period” (SHRP). These low-circulating glucocorticoid levels are believed to be essential for normal brain and behavioral development. A decrease in the steroidogenic capacity of the newborn adrenal cortex may account, at least in part, for the SHRP. It has been demonstrated that StAR mRNA and protein are highly expressed in the adrenals at birth, decrease subsequently until P14, and increase thereafter. The level of expression of P450ssc is comparable to that of P1 in adults. At birth, adrenal 3 $\beta$ -HSD activity is low at P1, increases at P10, and remains stable until adulthood. P450c21 activity is low (approximately half of the adult values) on P1 and P10. P450c11 $\beta$  immunoreactivity has been detected in the adrenals of newborn rats and did not change during neonatal development. In P7 rats, P450c11 $\beta$  mRNA is present at high levels only in the zonae fasciculata and reticularis.

## **Mineralocorticoids**

### **In Primates**

Regarding the ontogeny of aldosterone secretion, assays of steroids in cord blood have suggested that human fetus can synthesize aldosterone as early as weeks 16–20. However, *in vitro* experiments have demonstrated that at midgestation human fetal adrenal tissues do not produce detectable levels of aldosterone, under basal or stimulated conditions. The primary steps in aldosterone synthesis, i.e., those driven by StAR, P450ssc, 3 $\beta$ -HSD, and p450c21, are mature in the definitive zone in human fetal adrenals at the end of the midgestation period. However, in human fetal adrenals obtained from second-trimester abortuses, aldosterone synthase (P450AS) immunoreactivity is absent in the definitive zone and P450AS mRNA is weakly detectable in the whole cortical zone. Similarly, P450AS immunoreactivity is absent in the definitive zone of fetal rhesus monkey adrenal until near term. Activation of the late gestation fetal rhesus monkey HPA axis (obtained after treatment with metyrapone, a compound that inhibits P450c11 $\beta$ ) was able to induce P450AS expression. Taken together, these data indicate that in the primate fetal adrenal gland, the definitive zone has the capacity to synthesize aldosterone, but not until term.

### **In Sheep**

In the fetal sheep, levels of aldosterone synthesis and secretion are very low until the final part of gestation.

Although P450AS is expressed as early as E90 in the fetal zona glomerulosa, there is weak 3 $\beta$ -HSD staining until term.

### In Rodents

In rats, P450AS immunoreactivity is detected in E16 adrenals, in small clusters of cells, dispersed throughout the gland. By E18–E19, the number of P450AS-labeled cells increases and these cells become localized in the outer cortex. P1 adrenals have a pattern of P450AS staining comparable to that of the adult gland. In P7, rats, P450AS mRNA is confined to the subcapsular zona glomerulosa. Aldosterone content in fetal adrenal homogenates increases between E17 and P1 and remains stable thereafter.

## ROLE OF NUCLEAR RECEPTORS AND TRANSCRIPTION FACTORS

The complex cascade of the development and differentiation of the adrenal cortex, from the formation of the urogenital ridge to the zonation of the fetal adrenal, involves several nuclear receptors and transcription factors (Table II).

## WT1 and WNT4

Two genes are crucial for the development of the urogenital ridge: WT1 (Wilm's tumor suppressor gene 1), which encodes a protein having many characteristics of a transcription factor, and WNT4 (wingless-related mouse mammary tumor virus integration site 4), a member of the WNT family of developmentally regulated signaling molecules. WT1 is developmentally one of the earliest known genes that specifies kidney, gonadal, and adrenal cell lineages. Mice fetuses knocked out for *Wt1* and rescued with a YAC (yeast artificial chromosome) construct spanning the *Wt1* locus have adrenal-like structures that are greatly reduced in size and express lower levels of P450<sub>ssc</sub> mRNA than wild-type mice. WNT4 is involved in the development of the kidney, pituitary gland, female reproductive system, and mammary gland. Wnt4 is expressed next to the anterior site of the mouse mesonephros on E11.5 and in the developing adrenal cortex from E12.5 onward. Adrenals of Wnt4 knockout mice are morphologically comparable to those of wild-type animals. However, Wnt4 knockout animals have reduced adrenal P450<sub>c21</sub> and P450AS mRNA concentrations and decreased aldosterone production.

Table II Transcription Factors and Nuclear Receptors Involved in Adrenal Differentiation and Steroidogenesis

Transcription factor or nuclear receptor	First detected on	Phenotype of rodent KO	Genes that are regulated in adrenal cells
WT1	Mouse: E9	Lethal: lacks adrenals, kidneys, and gonads	DAX-1
Wnt4	Mouse: E11.5; human: E33	Abnormal kidney and adrenal development	P450 <sub>c21</sub> P450AS
SF-1	Mouse: E11; human: E33	Lethal: lacks adrenals and gonads	StAR P450 <sub>ssc</sub> 3 $\beta$ -HSD P450 <sub>c17</sub> P450 <sub>c21</sub> P450AS P450 <sub>c11<math>\beta</math></sub> ACTH-R DAX-1
DAX-1	Mouse: E10.5	Lack of X zone regression	StAR P450 <sub>c17</sub> P450 <sub>c21</sub> ACTH-R

*Note.* WT-1, Wilm's tumor suppressor gene 1; Wnt4, wingless-related mouse mammary tumor virus integration site 4; SF-1, steroidogenic factor-1; DAX-1, dosage-sensitive sex reversal adrenal hypoplasia congenita critical region on the X chromosome, gene 1; E, day of gestation; StAR, steroid acute regulatory protein; P450<sub>ssc</sub>, side-chain cleavage enzyme; P450<sub>c17</sub>, enzyme complex having 17 $\alpha$ -hydrolase and 17,20-lyase activities; 17 $\beta$ -HSD, 17 $\beta$ -hydroxysteroid dehydrogenase; 3 $\beta$ -HSD, 3 $\beta$ -hydroxysteroid dehydrogenase; P450<sub>c21</sub>, 21-hydroxylase; P450AS, aldosterone synthase; P450<sub>c11 $\beta$</sub> , 11 $\beta$ -hydroxylase; ACTH-R, adrenocorticotropic hormone receptor.

## SF-1, DAX-1, and GATA Proteins

Several transcription factors, such as steroidogenic factor-1 (SF-1), dosage-sensitive sex-reversal adrenal hypoplasia congenita critical region on the X chromosome, gene 1 (DAX-1), and GATA proteins, play major roles in the development of the adrenal primordium, its functional zonation, and the regulation of fetal adrenal steroidogenic capacities.

### SF-1

SF-1, also called Ad4BP or NR5A1, belongs to the orphan receptor class of nuclear receptors. It has been proposed that SF-1 acts through recruitment of cofactors, including both coactivators (CREB-binding protein, WT1, and nuclear receptor coactivator 1) and corepressors (nuclear receptor corepressor, DEAD/H box polypeptide 20, COP9 constitutive photomorphogenic subunit 2, nuclear receptor-interacting protein 1, and DAX-1). In mouse, Sf-1 mRNA was detected in the adrenal primordium from E11. From E14–E14.5 onward, Sf-1 mRNA is restricted to the steroidogenic cells in the cortex. In humans, SF-1 was expressed as early as E33 in the presumptive adrenal primordium. Newborn Sf-1 knockout mice are devoid of adrenals and die from adrenocortical insufficiency shortly after birth. However, Sf-1 heterozygous mice (SF-1<sup>+/-</sup>) are viable but show adrenal disorganization (reduced adrenal size and hypoplastic zona fasciculata and adrenal medulla), altered adrenal gene expression (increased StAR mRNA and decreased ACTH receptor mRNA), and impaired basal and stress-induced glucocorticoid secretion. Studies of SF-1 gene mutations in humans demonstrated that SF-1 regulates adrenal development in a dose-dependent manner. A number of studies have demonstrated that SF-1 acts a global regulator of the proteins involved in adrenal steroidogenesis (StAR, P450<sub>ssc</sub>, 3β-HSD, P450c21, P450AS, P45011β, and P450c17). In addition, Sf-1 regulates adrenal ACTH sensitivity through a cell-specific modulation of both constitutive and cyclic AMP-induced expression of the ACTH receptor gene. In the late-gestation ovine fetus, Sf-1 mRNA is present in adrenal extracts; its expression is positively regulated by a pituitary-dependent factor, which is not ACTH.

### DAX-1

DAX-1 is an orphan nuclear receptor that regulates both adrenal development and functional zonation. Mutations in the DAX-1 gene are responsible for X-linked adrenal hypoplasia congenita, an inherited disorder in humans that is characterized by hypoplasia

of the fetal adrenal glands with absence of the definitive zone and persistence of the fetal zone. DAX-1 has been postulated to bind to hairpin loops of the StAR promoter and/or to function as an RNA-binding protein. In human fetuses, DAX-1 is expressed in the adrenal primordium as early as E33. In mouse, it has been shown that Dax-1 is colocalized with Sf-1 in the developing adrenal and that Sf-1 expression precedes or coincides with expression of Dax-1, suggesting that these two molecules could cooperate in regulating adrenal development. *In vitro*, Dax-1 inhibited Sf-1-induced stimulation of transcription of the adrenal promoter of StAR, P450c17, Dax-1 itself, possibly through interactions with Sf-1, and transcriptional corepressors such as NcoR and Alien, whereas Sf-1 stimulated expression of the Dax-1 promoter. *In vivo*, male mice with a single Dax-1-deleted allele (Dax-1<sup>-/-</sup>) have normal zona glomerulosa, but the zona fasciculata is less well developed and shows decreased staining for 3β-HSD and the X zone fails to regress. Dax-1<sup>-/-</sup> animals have increased stress-induced corticosterone secretion and enhanced adrenal responsiveness to exogenous ACTH stimulation, most probably following increased adrenal P450c21 and ACTH receptor expression. Dax-1<sup>-/-</sup> animals have normal expression of Sf-1 and expression of Dax-1 is maintained in Sf-1<sup>+/-</sup> mice. The absence of Dax-1 partially reverses adrenal growth defects in Sf-1<sup>+/-</sup> mice. However, the precise mechanisms governing the interplay between Sf-1 and Dax-1 in regulating adrenal development and steroidogenesis are not known.

### GATA Proteins

GATA proteins are transcription factors that bind to the consensus sequence (A/T)GATA(A/G) in the promoter and enhancer regions of their target genes. GATA-4 is present in human adrenocortical carcinoma and in a transgenic mouse model developing adrenocortical tumors. GATA-4 and GATA-6 are detectable from E14 and week 19 onward in the mouse and human adrenal cortex, respectively. After birth, GATA-4 expression decreases, whereas GATA-6 continues to be expressed. Studies in chimeric mouse embryos demonstrated that GATA-4 is not essential for early adrenocortical differentiation. The exact role of GATA-6 in adrenal development remains to be determined.

## See Also the Following Articles

ACTH (Adrenocorticotrophic Hormone) • Adrenal Androgens  
• Adrenal Cortex, Anatomy • Adrenal Cortex Development,

Regulation of • Adrenal Cortex, Physiology • Adrenal Insufficiency • Adrenal Suppression • Thyroid Gland Development, Molecular Biology

### Further Reading

- Challis, J. R. G., Sloboda, D., Matthews, S. G., Holloway, A., Alfaidy, N., Patel, F. A., Whittle, W., Fraser, M., Moss, T. J. M., and Newnham, J. (2001). The fetal placental hypothalamic-pituitary-adrenal (HPA) axis, parturition and post natal health. *Mol. Cell. Endocrinol.* **185**, 135–144.
- Grino, M., Dakine, N., Paulmyer-Lacroix, O., and Oliver, C. (2001). Ontogeny of the hypothalamo-pituitary-adrenal axis. In "Contemporary Endocrinology: Adrenal Disorders" (G. P. Chrousos and A. N. Margioris, eds.), pp. 1–9. Humana Press, Totowa, NJ.
- Keegan, C. E., and Hammer, G. D. (2002). Recent insights into organogenesis of the adrenal cortex. *Trends Endocrinol. Metab.* **13**, 200–208.
- Matthews, S. G. (2002). Early programming of the hypothalamo-pituitary-adrenal axis. *Trends Endocrinol. Metab.* **13**, 373–380.
- Mesiano, S., and Jaffe, R. B. (1997). Developmental and functional biology of the primate fetal adrenal cortex. *Endocr. Rev.* **18**, 378–403.
- Naaman Réperant, E., and Durand, P. (1997). The development of the ovine fetal adrenal gland and its regulation. *Reprod. Nutr. Dev.* **37**, 81–95.
- Parker, K. L., and Schimmer, B. P. (1997). Steroidogenic factor 1: A key determinant of endocrine development and function. *Endocr. Rev.* **18**, 361–377.



# Adrenal Cortex, Physiology

Christian A. Koch

University of Leipzig, Leipzig, Germany

## Glossary

**adrenocorticotropic hormone or corticotropin (ACTH)** Synthesized as part of a 241-amino acid precursor, proopiomelanocortin. Usually, ACTH derives from the pituitary gland to stimulate the adrenal glands through the melanocortin-2 receptor (rarely, ACTH is produced ectopically, i.e., in ectopic Cushing's syndrome).

**corticotropin-releasing hormone (CRH)** Hypothalamic 41-amino acid peptide that usually stimulates the pituitary to release ACTH.

**hypothalamic–pituitary–adrenal (HPA) axis** Hypothalamic-releasing factors including CRH are influenced by central nervous system afferents. For instance, stress can trigger CRH release. The major secretagogue of CRH is ACTH, which subsequently stimulates the adrenals to produce and release cortisol. Cortisol, on the other hand, can inhibit further release of CRH and ACTH in a negative feedback loop.

**steroids** Hormones produced by the adrenal gland. Members of a large family of compounds that are derived from the cyclopentanperhydrophenanthrene ring structure that consists of three cyclohexane rings and one cyclopentane ring. The nomenclature denotes rings by a letter and the individual carbon atoms by a number. Gonane is the unsaturated 17-carbon ring structure. Estranes are steroids with 18 carbons (C18 steroids) by adding a methyl group at C13. Androstane is a C19 steroid with two methyl groups. Pregnane is a C21 steroid with methyl and ethyl groups. The ending-ene denotes a double bond of the parent compound.

The *Tabulae Anatomicae* mention Bartholomeo Eustachius, who apparently first described the anatomy of the adrenal glands in 1563. A central physiologic role for the adrenals was presented in 1849 by Thomas Addison, and today Addison's disease refers to adrenal insufficiency, a life-threatening condition. To better understand pathophysiologic states of the adrenal glands, one must recall their normal anatomy and physiology. The adrenal glands are divided into cortex and medulla. The adrenal cortex is composed of three zones—the glomerulosa, fasciculata, and reticularis. The largest and

most important zone is the fasciculata, where glucocorticoids including cortisol are produced. Extracellular volume status is influenced by aldosterone, the hormone of the zona glomerulosa. Androgens such as dehydroepiandrosterone are produced by the zona reticularis, which starts to grow at approximately age 4, shortly before adrenarche occurs. Corticotropin, usually coming from the pituitary, is the principal stimulus for cortisol secretion and is released under stress and other stimuli through hypothalamic corticotropin-releasing hormone (CRH) secretion. Corticotropin can also stimulate the zona glomerulosa (to a minor extent) and reticularis to secrete aldosterone and adrenal androgens. A feedback loop exists between the hypothalamus (H), pituitary (P), and adrenal gland (A)—the HPA axis. High peripheral cortisol levels inhibit further release of CRH and corticotropin. At the extreme (e.g., in conditions of supraphysiologic exogenous glucocorticoid administration), the secretion of hypothalamic CRH and pituitary corticotropin becomes severely suppressed, with subsequent atrophy of the adrenal glands due to the lacking stimulus corticotropin. The adrenal medulla forms postnatally and exerts effects on the adrenal cortex and vice versa. Adrenomedullary chromaffin cells are intermingled with the adrenal cortex, facilitating an interaction between the two.

## FETAL ADRENAL GLAND

### Fetal Adrenal Steroidogenesis

The fetal adrenal cortex plays a critical role in regulating intrauterine homeostasis and the maturation of fetal organ systems that are necessary for extrauterine life. The important mediators for these functions are steroid hormones from the fetal adrenal. Throughout gestation and postnatally, the fetal adrenal gland undergoes morphological and functional changes during its transformation to the adult adrenal gland. The progenitor cells of the adrenal cortex stem from a cell lineage that also leads to steroid-secreting cells of the gonads. By week 8 of gestation, the adrenal cortex-forming progenitor cells build an inner cluster (the fetal zone) consisting of large eosinophilic steroid-secreting cells that express high levels of steroid 17 $\alpha$ -hydroxylase (CYP17) and an outer zone (the definitive zone) consisting of cells that do not express

CYP17. Between the fetal and definitive zone lies the transitional zone, which is composed of cells similar to those of the zona fasciculata of the adult adrenal gland. At gestational week 8, chromaffin cells enter the rudimentary adrenal gland and remain there as discrete islands until day 8 postnatally, before they form a rudimentary adrenal medulla. During gestation, the fetal zone represents 85% of the cortical volume and produces large amounts of dehydroepiandrosterone (DHEA) sulfate, a hormone that serves as a precursor for other androgens synthesized in the periphery. Before birth, DHEAS is used by the placenta to synthesize estrogen. Placental estrogen, on the other hand, supports the fetal adrenal to synthesize cortisol. Several endocrine, paracrine, and autocrine factors influence the steroidogenesis of the fetal adrenal cortex. Within the first postnatal year, the fetal zone degenerates. During the third trimester, aldosterone production starts by the zona glomerulosa, derived from the outer zone of the fetal adrenal cortex (without CYP17). A transitional area between the inner and outer zones gives rise to the zonae fasciculata and reticularis, which secrete glucocorticoids and DHEA sulfate under stimulation and control by corticotropin. Certain transcription factors (e.g., steroidogenic factor 1, DAX1, and WT1) are essential in adrenal development. Adrenal dys- or agenesis with subsequent adrenal insufficiency and death can result from nonfunction of these factors, for instance, by mutations in the steroidogenic factor 1.

### Growth Factors

During the first trimester of pregnancy, human chorionic gonadotropin (HCG) regulates the growth of the fetal adrenal. Adrenocorticotrophic hormone (ACTH) is critical for growth, steroidogenesis, and differentiation of the fetal adrenal gland and becomes the main force for further growth after the fifth month of gestation. ACTH deficiency leads to increased apoptosis and subsequent atrophy of the adrenal glands, whereas ACTH excess (e.g., in Cushing's syndrome or congenital adrenal hyperplasia) can cause hyperplastic adrenal glands. In addition to HCG, ACTH, and its receptor ACTHR, local growth factors are important for steroidogenesis, growth, and development of the adrenal gland.

Insulin-like growth factors 1 and 2 (IGF-1 and IGF-2), their receptors, and binding proteins are all expressed in the fetal adrenal. IGF-1 not only amplifies the effect of ACTH on the adrenal but also enhances steroid production of the adrenal gland by increasing

the activities of  $17\alpha$ , 21-, and  $11\beta$ -hydroxylase. Similarly, IGF-2 promotes ACTH action. In addition, IGF-2 helps the fetal adrenal to synthesize cortisol and androgen by regulating the enzymes p450scc, p450c17, and  $3\beta$ -hydroxysteroid dehydrogenase.

Basic fibroblast growth factor is a mitogenic protein and is more effective (stimulating proliferation) on adrenal cells of the definitive zone than those of the fetal zone. Epidermal growth factor (EGF) and transforming growth factor- $\alpha$  (TGF- $\alpha$ ) are growth factors with sequence homology that both activate the EGF receptor. This may lead to hypertrophy of the fetal adrenal. Although EGF stimulates hypothalamic corticotropin-releasing hormone (CRH) release, it does not cause subsequent ACTH secretion from the pituitary. The TGF- $\beta$  family of growth factors, including activin, inhibin, and TGF- $\beta_1$ , are paracrine/autocrine regulators of growth and steroidogenesis in the fetal adrenal cortex. Activin increases ACTH-stimulated cortisol production but not DHEAS production in fetal zone cells. In the adult adrenal cortex, activin has no effect on growth or steroidogenesis. In fact, activin may lead to apoptosis and involution of the fetal adrenal cortex postnatally. TGF- $\beta_1$  appears to decrease fetal and definitive zone cell proliferation and steroidogenesis.

### Nuclear Receptors

Steroidogenic factor 1 (SF-1), DAX-1, and the estrogen receptor (ER) belong to the nuclear receptor superfamily. Members of this family are transcription factors that are important for regulating expression of genes involved in cellular growth control and differentiation. SF-1 is classified as an orphan receptor because its ligand is unknown. The human cDNA sequence of SF-1 is highly homologous (>95%) to murine and bovine sequences. Human adrenal cortex, ovaries, testes, and spleen show high SF-1 mRNA expression. In human placenta, SF-1 is not or only minimally expressed. SF-1 plays an essential role in the organogenesis of the fetal adrenal gland and also in regulating genes that code for steroidogenic enzymes. SF-1 stimulates the promoter activities of genes encoding steroidogenic acute regulatory (StAR) protein, the scavenger receptor-type class BI (SR-BI), and the ACTH receptor. StAR protein is critical in the translocation process of cholesterol from the outer to the inner mitochondrial membrane. In contrast to the adult adrenal gland, the fetal adrenal uses low-density lipoprotein (LDL) rather than high-density lipoprotein cholesterol as the main source for steroid

biosynthesis. It appears that SR-BI binds to LDL with high affinity. SF-1 influences the constitutive activity of the human ACTH receptor gene promoter and regulates steroid hydroxylase enzymes. In addition, SF-1 regulates the genes coding for the  $\beta$ -subunit of luteinizing hormone, the  $\alpha$ -subunit of the glycoprotein hormones, gonadotropin-releasing hormone receptor, prolactin receptor, oxytocin, Mullerian inhibiting substance, and aromatase. Furthermore, SF-1 interacts with other proteins and cofactors.

Gene deletions at Xp21/22, where the DAX-1 gene is located, may lead to X-linked adrenal hypoplasia congenita, glycerol kinase deficiency, hypogonadotropic hypogonadism, and/or Duchenne's muscular atrophy. DAX-1 (dosage-sensitive sex reversal, adrenal hypoplasia congenita, X-linked) is highly expressed in the fetal and adult adrenal gland, hypothalamus, pituitary, testes, and ovaries. Together with SF-1, it may coregulate steroidogenesis as well as adrenal and gonadal organogenesis. DAX-1 can block steroidogenesis by inhibiting the activity of StAR and the expression of p450scc and  $3\beta$ -hydroxysteroid dehydrogenase.

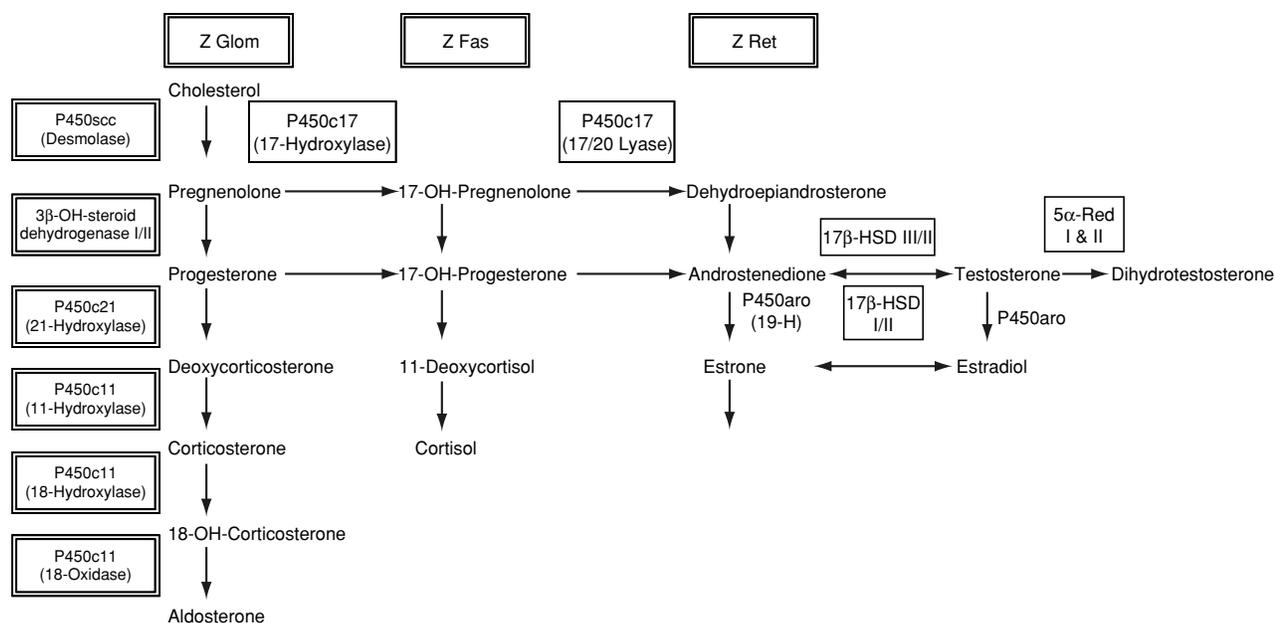
Estrogens are important in cell differentiation, growth, and function of various tissues. Estrogen receptors are members of the steroid receptor superfamily

and mediate the action of estrogens. ER $\beta$  is highly expressed in the fetal adrenal gland, in contrast to ER $\alpha$ . As mentioned previously, estrogen is critical in fetal steroidogenesis and organogenesis.

## ADULT ADRENAL CORTEX

### Steroid Biosynthesis and Regulation of Cortisol Production

The normal adult human adrenal gland weighs approximately 5 g. Ninety percent of this weight is due to the adrenal cortex, which is composed of three zones (from outside to inside): the zona glomerulosa, the zona fasciculata, and the zona reticularis. The adrenal medulla forms postnatally and is composed of chromaffin cells, some of which may still be intermingled and spread within the adrenal cortex. The adult adrenal cortex produces glucocorticoids, mineralocorticoids, and adrenal androgens (Fig. 1 and Table I). The blood flow in the adrenal gland is centripetal (from outside to inside), which exposes the inner zones and the adrenal medulla to increasing concentrations of adrenal steroids. High cortisol levels in the medulla are needed to induce enzymes for epinephrine biosynthesis. Seventy-five percent



**Figure 1** Adrenal steroidogenesis. Z Glom, zona glomerulosa; Z Fas, zona fasciculata; Z Ret, zona reticularis; 19-H = 19-hydroxylase; HSD, hydroxysteroid dehydrogenase; P450aro, aromatase; 5 $\alpha$ -Red, 5 $\alpha$ -reductase. The three adrenal cortex zones Z Glom, Z Fas, and Z Ret are listed above the hormones that are produced in the respective zones. The steroidogenic enzymes on the left, starting with P450scc (desmolase), are listed in order of vertical and horizontal reading—that is, desmolase converts cholesterol to pregnenolone,  $3\beta$ -OH-steroid dehydrogenase I/II convert pregnenolone to progesterone, 17-OH-pregnenolone converts progesterone to 17-OH-progesterone, P450c11 converts deoxycorticosterone to 18-OH-corticosterone and 11-deoxycortisol to cortisol, etc.

Table I Key Human Steroidogenic Enzymes and Cofactor Proteins<sup>a</sup>

Protein	Gene	Chromosomal locus	Substrate	Activities
StAR	<i>STAR</i>	8p11.2	Cholesterol flux within mitochondria	Sterol delivery to P450 <sub>scc</sub>
P450-OR	<i>CPR</i>	7p15-q35	Microsomal P450s	Electron transfer
Cytochrome b <sub>5</sub>	<i>CYB5</i>	18q23	Aids P450c17	Augments 17,20-lyase activity
Adrenodoxin reductase	<i>ADR</i>	17q24–25	Mitochondrial P450s	Electron transfer
Adrenodoxin	<i>ADX</i>	11q22	Mitochondrial P450s	Electron transfer
11 β-HSDI	<i>HSD11B1</i>	1	Cortisol, cortisone, corticosterone, 11-De B	11 β-Ketosteroid dehydrogenase
11β-HSD II	<i>HSD11B2</i>	16p22	Cortisol, corticosteron	11β-Hydroxysteroid dehydrogenase
5α-Reductase Type I	<i>SRD5A1</i>	5p15	Testosterone, C21-steroids	5α-Reduction
5α-Reductase Type II	<i>SRD5A2</i>	2p23	Testosterone, C21-steroids	5α-Reduction
17β-HSDI	<i>HSD17B1</i>	17q21	Estrone	17β-Ketosteroid reductase
17β-HSD II	<i>HSD17B2</i>	16q24	Testosterone, estradiol, DHT	17β-Hydroxysteroid dehydrogenase
17β-HSD III	<i>HSD17B3</i>	9q22	Androstenedione	17β-Ketosteroid reductase
17β-HSD V	<i>HSD17B5</i>	10p14–15	Androstenedione, DHT, 3α-asdiol, asone Asdione, 3α-androstan	17β-Ketosteroid reductase 3α-Hydroxysteroid Dehydrogenase
3β-HSD I	<i>HSDB1</i>	1p13	Preg, 17OH-Preg, DHEA, adiol DHEA, asdiol	3β-Dehydrogenase 5/4-Isomerase
3β-HSD II	<i>HSDB2</i>	1p13	Preg, 17OH-Preg, DHEA, asdiol	3β-Dehydrogenase 5/4-Isomerase
P450 <sub>scc</sub>	<i>CYP11A1</i>	15q23–24	Cholesterol Hydroxysterols	22R-Hydroxylase 20, 22-Lyase 20R-Hydroxylase
P450c17	<i>CYP17</i>	10q24.3	Preg, 17OH-Preg Prog, DHEA	17α-Hydroxylase 17,20 Lyase
P450c21	<i>CYP21B</i>	6p21.1	Prog, 17OH-Prog	21-Hydroxylase
P450c11β	<i>CYP11B1</i>	8q21–22	11-Deoxycortisol 11-DOC	11-Hydroxylase
P450c11AS	<i>CYPB2</i>	8q21–22	Corticosterone 11-DOC	11-Hydroxylase 18-Hydroxylase
P450aro	<i>CYP19</i>	15q21.1	Androstenedione Testosterone	19-Hydroxylase 19-Oxidase, aromatization

<sup>a</sup>Modified from Auchus and Miller (2001). Androstan, androstanediol; Preg, pregnenolone; Prog, progesterone; DOC, deoxycorticosterone; HSD, hydroxysteroid dehydrogenase; DHT, dihydrotestosterone; 11-De B, 11-dehydrocorticosterone B; 17OH-Prog, 17α-hydroxyprogesterone; 3α-Asdiol, 3α-androstanediol; Asdione, androstenedione.

of the adrenal gland weight is due to the zona fasciculata, the largest zone and the one that synthesizes glucocorticoids. The zona fasciculata also produces DHEA and DHEAS, whereas the zona reticularis also produces cortisol. In contrast to the zona fasciculata, the zona reticularis is small and not very involved in adrenal androgen production until adrenarche (approximately 6 years of age). The precursor for glucocorticoid production is cholesterol, which in a first step is converted to pregnenolone in the adrenal cortex. Steroids derive from the cyclopentanoperhydrophenanthrene four-ring hydrocarbon nucleus, a relatively inert structure. Depending on the presence

of several enzymes in the respective adrenal cortex zone, several steroid hormones can then be synthesized. Cytochromes P450 are categorized into two classes: type 1 enzymes that reside in the mitochondria and type 2 enzymes located at the smooth endoplasmic reticulum (Table II).

The secretion and synthesis of cortisol are regulated by the hypothalamic–pituitary–adrenal (HPA) axis. Certain stimuli including stress lead to the release of CRH in the hypothalamus. CRH then stimulates ACTH release from the pituitary (Fig. 2). ACTH binds to ACTH receptors located on adrenocortical cells and stimulates them to release cortisol

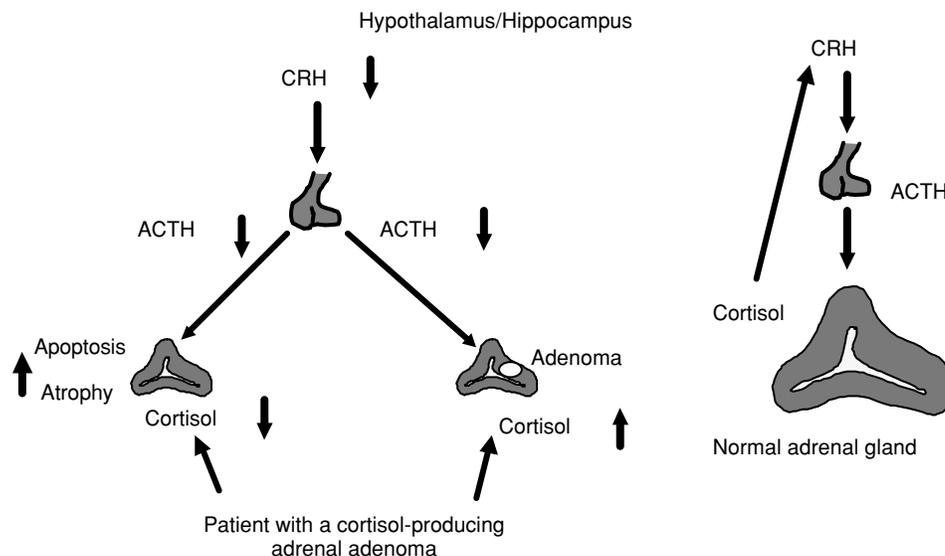


Figure 2 The HPA axis. Modified from Koch and Chrousos (2001).

through cAMP. Cortisol can then increase energy-providing compounds, including glucose, free fatty acids, and free amino acids. As mentioned previously, ACTH is also growth promoting on the adrenal cortex; that is, continuous stimulation by ACTH may cause adrenal hypertrophy, whereas a lack of ACTH may lead to adrenal atrophy. The HPA axis is very sensitive to exogenous glucocorticoid administration, which can easily lead to ACTH suppression through a negative feedback loop on CRH. In normal individuals who are not working in (night) shifts, there is a diurnal variation of cortisol production, with serum cortisol being highest in the morning and lowest at midnight. In patients with Cushing's syndrome (hypercortisolism), these normal physiologic circuits are disturbed (i.e., there is no suppression of

the HPA axis and cortisol production by administration of exogenous glucocorticoids such as dexamethasone). Prolonged (7–48 h) increases in ACTH lead to an increased synthesis of all the steroidogenic enzymes, especially P450<sub>scc</sub>, as well as an increased uptake of cholesterol from the circulation. Chronic lack of ACTH (e.g., through exogenous glucocorticoid administration) leads to adrenal atrophy. Therefore, the exogenous glucocorticoid has to be tapered to allow the pituitary and adrenal gland to recover in order to synthesize normal levels of cortisol on its own. Depending on the level of suppression, this may take weeks or months.

### Biosynthesis and Regulation of Aldosterone Production

Aldosterone, the major human mineralocorticoid, is produced in the zona glomerulosa of the adrenal cortex. Its secretion is stimulated mainly by angiotensin II (and III) through the renin–angiotensin–aldosterone system and, to a lesser extent, by ACTH. Chronic infusion of ACTH stimulates aldosterone secretion for only 24 h. Less potent stimulators of aldosterone secretion are endothelin and serotonin. Also, increases in potassium concentrations stimulate aldosterone production. An increase in serum potassium of 0.1 mmol/liter can elevate plasma aldosterone by 35%. On the other hand, a decrease in serum potassium of 0.3 mmol/liter can reduce plasma aldosterone by 46%. Aldosterone decreases the absorption

Table II Locations of Steroidogenic Proteins

Endoplasmic reticulum	Cytoplasm	Mitochondria
11 $\beta$ -HSD I and II	3 $\alpha$ -HSD	3 $\beta$ -HSD II
5 $\alpha$ -Reductase I and II	17 $\beta$ -HSD V	StAR
17 $\beta$ -HSD I–III	17 $\beta$ -HSD I	Adrenodoxin reductase
3 $\beta$ -HSD II	3 $\beta$ -HSD II	P450 <sub>c11AS</sub>
Cytochrome b <sub>5</sub>	StAR	P450 <sub>c11<math>\beta</math></sub>
P450 oxidoreductase	Adrenodoxin	P450 <sub>scc</sub>
P450 <sub>aro</sub>		
P450 <sub>c21</sub>		
P450 <sub>c17</sub>		

Modified from Auchus and Miller (2001).

of potassium and promotes sodium reabsorption and fluid retention, thereby increasing the extracellular fluid volume. However, after a few days of extracellular fluid expansion by increased aldosterone levels, the individual will be protected from continuous expansion through a so-called “escape” mechanism that denotes attaining a new sodium balance and the formation of a new steady state.

Target tissues of aldosterone, including kidney (distal tubules and cortical collecting ducts), colon, and salivary glands, have mineralocorticoid receptors that bind aldosterone. In the distal nephron, cortisol is a potent agonist at the mineralocorticoid receptor. Among inhibitors of aldosterone secretion are atrial natriuretic peptide (ANP) and dopamine. ANP strongly inhibits stimulated (e.g., low sodium intake) aldosterone secretion, with much less effect on basal (e.g., normal or high sodium intake) activity. Chronic sodium restriction leads to increased activity of aldosterone synthase and a higher content of this enzyme in the zona glomerulosa. The first steps of aldosterone biosynthesis are identical to those of cortisol biosynthesis (Fig. 1). The synthesis of cortisol, however, depends on  $17\alpha$ -hydroxylation of pregnenolone by  $17\alpha$ -hydroxylase (P450c17), which is exclusively expressed in the zona fasciculata. On the other hand, aldosterone synthase is normally expressed only in the zona glomerulosa.

### Regulation of Adrenal Androgen Production

At approximately 4 years of age, in both sexes the zona reticularis forms and continues to grow until the mid-20s. After age 40, this zone gradually regresses. Corticotropin and prolactin stimulate adrenal androgen secretion in the fetal adrenal zone. Postnatally, the zona reticularis responds to ACTH, as exemplified in congenital adrenal hyperplasia in which ACTH and androgen hypersecretion can occur. During infancy, only small amounts of androgens are secreted, and it is unknown how adrenarche, the time point at which a slight amount of pubic hair develops, is regulated. Seventy percent of circulating testosterone in women with normal menstrual cycles derives from the conversion of adrenal DHEA. The principal androgens secreted by the adrenals are DHEA, DHEAS, androstenedione, and (minimally) testosterone. DHEAS per se has only weak androgenic effects. Peripheral conversion of the aforementioned precursors leads to more potent androgens, such as testosterone and dihydrotestosterone. Major conversion sites include

the hair follicles, sebaceous glands, external genitalia, and prostate. Peripheral adipose tissue can convert androgens into estrogens by the highly active enzymes aromatase and  $17$ -ketosteroid reductase. Glucocorticoids stimulate aromatase. Inactivation or degradation of androgens and their metabolites occur at different sites, including the liver and kidneys. Exogenous adrenal androgen administration can suppress gonadotropin secretion. Excess endogenous androgen production can be caused by several conditions, including congenital adrenal hyperplasia and adrenal tumors.

### Impact of the Sympathoadrenal System on the Regulation of Adrenocortical Function

Adrenocortical steroid hormones influence the differentiation and hormone production of adrenal chromaffin cells. On the other hand, the sympathoadrenomedullary system modulates diurnal variations of steroidogenesis in the adrenal cortex. The adrenal cortex is innervated by neurons originating in cell bodies within the adrenal medulla and by nerves that have cell bodies outside the adrenal, reaching the cortex via blood vessels. Adrenal chromaffin cells contain many neuropeptides that regulate adrenocortical steroid production in many species. Adrenomedullary cells are found throughout the adrenal cortex, which facilitates the paracrine action of their products. Another avenue for adrenomedullary secretory products reaching the adrenal cortex is the lymphatics.

### CONCLUSION

The adrenal cortex fulfills important functions before and after birth. Prenatally, the fetal adrenal cortex is large and mainly consists of the fetal zone, which produces high amounts of DHEAS, a hormone that serves as a precursor for other androgens. DHEAS is used by the placenta to synthesize estriol and to regulate intrauterine homeostasis as well as maturation of fetal organ systems that are necessary for life after birth. Postnatally, the adrenal cortex becomes a three-zoned structure, with the largest zone being the zona fasciculata. From outside to inside, the three zones are the zona glomerulosa, the zona fasciculata, and the zona reticularis. The zona glomerulosa produces aldosterone. Production of this hormone is stimulated by angiotensin II and, to a lesser extent, ACTH, potassium, endothelin, and serotonin. Aldosterone secretion is inhibited by dopamine and atrial natriuretic peptide. The zona fasciculata produces mainly cortisol, the stress hormone.

Secretion of this hormone is regulated by the HPA axis. Chronic stress can overstimulate the HPA axis, leading to depression. The zona reticularis produces adrenal androgens and can be stimulated by ACTH. All three zones gradually regress during the life span. Interaction between the adrenal cortex and the adrenal medulla exists and is facilitated by the intermingling of adrenomedullary chromaffin cell islets with the adrenal cortex. The adrenal medulla matures within the first 18 months postnatally and produces the stress hormones epinephrine and norepinephrine as well as other catecholamines.

### See Also the Following Articles

ACTH (Adrenocorticotrophic Hormone) • Adrenal Androgens • Adrenal Cortex, Anatomy • Adrenal Cortex, Development • Adrenal Cortex Development, Regulation of • Adrenal Insufficiency • Adrenal Suppression

### Further Reading

Auchus, R. J., and Miller, W. L. (2001). The principles, pathways, and enzymes of human steroidogenesis. *In* "Endocrinology" (L. J. DeGroot and J. L. Jameson, eds.), 4th ed. Saunders, Philadelphia.

Crowder, R. E. (1957). The development of the adrenal gland in man, with special reference to origin and ultimate location of cell types and evidence in favor of the cell migration theory. *Contemp. Embryol.* **251**, 195–209.

Erhart-Bornstein, M., Hinson, J. P., Bornstein, S. R., Scherbaum, W. A., and Vinson, G. P. (1998). Intraadrenal interactions in the regulation of adrenocortical steroidogenesis. *Endocr. Rev.* **19**, 101–143.

Koch, C. A., and Chrousos, G. P. (2001). Is the diminuto/dwarf1 gene involved in physiologic adrenocortical size regulation and tumor formation? *J. Clin. Endocrinol. Metab.* **86**(11), 5127–5129.

Kvetnansky, R., Pacak, K., Fukuhara, K., Viskupic, E., Hiremagalur, B., Nankova, B., Goldstein, D. S., Sabban, E. L., and Kopin, I. J. (1995). Sympathoadrenal system in stress: Interaction with the hypothalamic–pituitary–adrenocortical system. *Ann. N. Y. Acad. Sci.* **771**, 131–158.

Margioris, A. N. and Chrousos, G. P. (eds.) (2001). "Adrenal Disorders." Humana Press, Totowa, NJ.

Mesiano, S., and Jaffe, R. B. (1997). Development and function of the primate fetal adrenal cortex. *Endocr. Rev.* **18**, 378–404.

Mortensen, R. M., and Williams, R. H. (2001). Aldosterone action. *In* "Endocrinology" (L. J. DeGroot and J. L. Jameson, eds.), 4th ed. Saunders, Philadelphia.

Orth, D. N., and Kovacs, W. J. (1998). The adrenal cortex. *In* "Williams Textbook of Endocrinology" (J. D. Wilson, D. W. Foster, H. M. Kronenberg, and P. R. Larsen, eds.), 9th ed. Saunders, Philadelphia.

Parker, K. L., and Schimmer, B. P. (2001). Genetics of the development and function of the adrenal cortex. *Rev. Endocr. Metab. Dis.* **2**, 245–252.



## Adrenal Hypoplasia

see Congenital Adrenal Hypoplasia



# Adrenal Insufficiency

Evangelia Charmandari and George P. Chrousos

NIH/National Institute of Child Health and Human Development, Bethesda, Maryland, United States

## Glossary

**Cushing's syndrome** A condition caused by increased adrenocortical secretion of cortisol resulting from adrenocortical hyperplasia or tumor and characterized by obesity, acne, amenorrhea, hypertension, abdominal pain, and weakness.

**HIV infection** Human immunodeficiency virus, the causative agent of AIDS.

**thrombosis** The formation, development, or presence of a thrombus, or blood clot, in a blood vessel or the heart.

Adrenal insufficiency is a disorder characterized by impaired adrenocortical function and decreased production of mineralocorticoids, glucocorticoids, and/or adrenal androgens.

## INTRODUCTION

Adrenal insufficiency can be caused by diseases affecting the adrenal cortex (primary), the pituitary gland and the secretion of adrenocorticotrophic hormone (ACTH) (secondary), or the hypothalamus and the secretion of corticotrophic-releasing hormone (CRH) (tertiary). This article provides a brief overview of the etiology, clinical manifestations, diagnosis, and treatment of adrenal insufficiency.

## CAUSES OF ADRENAL INSUFFICIENCY

### Primary Adrenal Insufficiency

#### **Autoimmune Adrenitis**

This condition is the result of an autoimmune process that destroys the adrenal cortex. Both humoral and cell-mediated immune mechanisms directed at the adrenal cortex are involved. Antibodies that react with several steroidogenic enzymes as well as all three zones of the adrenal cortex are detected in 60–75%

of patients with autoimmune primary adrenal insufficiency but only rarely in patients with other causes of adrenal insufficiency or normal subjects. Approximately 50% of patients with autoimmune adrenal insufficiency have one or more other autoimmune endocrine disorders, whereas patients with the more common autoimmune endocrine disorders, such as type 1 diabetes mellitus, chronic autoimmune thyroiditis, or Graves' disease, rarely develop adrenal insufficiency. The combination of autoimmune adrenal insufficiency with other autoimmune endocrine disorders is referred to as the polyglandular autoimmune syndromes types I and II (Table I).

#### **Infectious Adrenitis**

Many infectious agents may affect the adrenal gland and result in adrenal insufficiency, including tuberculosis, disseminated fungal infections, and HIV infection.

#### **Hemorrhagic Infarction**

Bilateral adrenal infarction caused by hemorrhage or adrenal vein thrombosis may lead to adrenal insufficiency. Adrenal hemorrhage has been mostly associated with meningococemia (Waterhouse–Friderichsen syndrome) and *Pseudomonas aeruginosa* infection.

#### **Drugs**

Drugs may cause adrenal insufficiency by inhibiting cortisol biosynthesis, particularly in individuals with limited pituitary and/or adrenal reserve, or by accelerating the metabolism of cortisol and most synthetic glucocorticoids following induction of hepatic mixed function oxygenase enzymes.

### Secondary and Tertiary Adrenal Insufficiency

Secondary adrenal insufficiency may be caused by any disease process that affects the anterior pituitary and

**Table I Etiology of Adrenal Insufficiency**

<b>Primary adrenal insufficiency</b>
Autoimmune (polyglandular failure)
Tuberculosis
Sarcoidosis, amyloidosis, hemochromatosis
Hemorrhage (meningococemia, anticoagulants, and trauma)
Fungal infections
Metastatic neoplasia/infiltration
Congenital adrenal hyperplasia
Congenital adrenal hypoplasia
Congenital unresponsiveness to ACTH (glucocorticoid deficiency and ACTH resistance)
Adrenoleukodystrophy/adrenomyeloneuropathy
Acquired immunodeficiency syndrome
Bilateral adrenalectomy
Steroid synthesis inhibitors (e.g., metyrapone, ketoconazole, and aminoglutethimide)
Adrenolytic agents (o,p'DDD, suramin)
Glucocorticoid antagonists (RU486)
<b>Secondary and tertiary adrenal insufficiency</b>
Following discontinuation of exogenous glucocorticoids or ACTH
Following the cure of Cushing's syndrome
Pituitary and hypothalamic lesions
Tumors
Inflammation
Infections
Autoimmune lesions
Granulomatous infiltration
Trauma
Congenital aplasia, hypoplasia, dysplasia, ectopy
Pituitary–hypothalamic surgery
Pituitary–hypothalamic radiation
Pituitary–hypothalamic hemorrhage (apoplexy)
Acquired isolated ACTH deficiency
Familial corticosteroid-binding globulin deficiency

interferes with ACTH secretion. The ACTH deficiency may be isolated or occur in association with other pituitary hormone deficits (Table I). On the other hand, tertiary adrenal insufficiency can be caused by any process that involves the hypothalamus and interferes with CRH secretion. The most common causes of tertiary adrenal insufficiency are abrupt cessation of high-dose glucocorticoid therapy and treatment of Cushing's syndrome.

## CLINICAL MANIFESTATIONS OF ADRENAL INSUFFICIENCY

The clinical manifestations of adrenal insufficiency depend on the extent of loss of adrenal function and

whether mineralocorticoid production is preserved. The onset of adrenal insufficiency is often gradual and may go undetected until an illness or other stress precipitates an adrenal crisis.

## Adrenal Crisis

Adrenal crisis or acute adrenal insufficiency may complicate the course of chronic primary adrenal insufficiency and may be precipitated by a serious infection, acute stress, bilateral adrenal infarction, or hemorrhage. It is rare in patients with secondary or tertiary adrenal insufficiency. The main clinical manifestation of adrenal crisis is shock, but patients may also have nonspecific symptoms, such as anorexia, nausea, vomiting, abdominal pain, weakness, fatigue, lethargy, confusion, or coma. Hypoglycemia is rare in acute adrenal insufficiency but more common in secondary adrenal insufficiency. Hypoglycemia is a common manifestation in children and thin women with the disorder. Hyperpigmentation due to chronic ACTH hypersecretion and weight loss are indicative of long-standing adrenal insufficiency, and additional symptoms and signs relating to the primary cause of adrenal insufficiency may also be present.

The major factor precipitating an adrenal crisis is mineralocorticoid deficiency and the main clinical problem is hypotension. Adrenal crisis can occur in patients receiving appropriate doses of glucocorticoid if their mineralocorticoid requirements are not met, whereas patients with secondary adrenal insufficiency and normal aldosterone secretion rarely present in adrenal crisis.

## Chronic Primary Adrenal Insufficiency

Patients with chronic primary adrenal insufficiency may have symptoms and signs of glucocorticoid, mineralocorticoid, and androgen deficiency. In contrast, patients with secondary or tertiary adrenal insufficiency usually have normal mineralocorticoid function. The onset of chronic adrenal insufficiency is often insidious and the diagnosis may be difficult in the early stages of the disease.

The most common clinical manifestations of chronic primary adrenal insufficiency include general malaise, fatigue, weakness, anorexia, weight loss, nausea, vomiting, abdominal pain, diarrhea that may alternate with constipation, hypotension, electrolyte abnormalities (hyponatremia, hyperkalemia, or metabolic acidosis), hyperpigmentation, autoimmune manifestations (vitiligo), decreased axillary and pubic hair, and loss of libido and amenorrhea in women.

## Secondary or Tertiary Adrenal Insufficiency

The clinical features of secondary or tertiary adrenal insufficiency are similar to those of primary adrenal insufficiency. However, hyperpigmentation is not present because ACTH secretion is not increased. Also, because the production of mineralocorticoids by the zona glomerulosa is mostly preserved, dehydration and hyperkalemia are not present, and hypotension is less prominent. Hyponatremia and increased intravascular volume may be the result of an “inappropriate” increase in vasopressin secretion. Hypoglycemia is more common in secondary adrenal insufficiency, possibly due to concomitant growth hormone insufficiency, and in isolated ACTH deficiency. Clinical manifestations of a pituitary or hypothalamic tumor, such as symptoms and signs of deficiency of other anterior pituitary hormones, headache, or visual field defects, may also be present.

## DIAGNOSIS

The clinical diagnosis of adrenal insufficiency can be confirmed by demonstrating inappropriately low cortisol secretion, determining whether the cortisol deficiency is secondary or primary and, hence, dependent or independent of ACTH deficiency, and detecting the cause of the disorder.

### Cortisol Secretion

The diagnosis of adrenal insufficiency depends on the demonstration of inappropriately low cortisol secretion. Serum cortisol concentrations are normally highest in the early morning hours (4:00–8:00 AM) and increase further with stress. Serum cortisol concentrations of less than 3  $\mu\text{g}/\text{dl}$  (80 nmol/liter) at 8:00 AM are strongly suggestive of adrenal insufficiency, whereas values less than 10  $\mu\text{g}/\text{dl}$  (275 nmol/liter) make the diagnosis likely. Basal urinary cortisol and 17-hydroxycorticosteroid excretion is low in patients with severe adrenal insufficiency but may be low-normal in patients with partial adrenal insufficiency. Generally, baseline urinary measurements are not recommended for the diagnosis of adrenal insufficiency.

### ACTH Secretion

Inappropriately low serum cortisol concentrations in association with increased plasma ACTH concentrations determined simultaneously are suggestive of

primary adrenal insufficiency. On the other hand, inappropriately low baseline morning cortisol and ACTH concentrations indicate secondary or tertiary disease. Given that plasma ACTH measurements depend on proper preparation of the sample and may not be readily available, confirmation of the diagnosis requires stimulation of the adrenal glands with exogenous ACTH.

### Short ACTH Stimulation Tests

A short ACTH stimulation test should be performed in all patients suspected of having adrenal insufficiency. It involves the intravenous administration of synthetic ACTH(1–24) (cosyntropin), which has the full biologic potency of native ACTH(1–39), and subsequent measurement of serum cortisol concentrations at regular intervals for up to 1 h (usually at 0, 30, and 60 min).

#### **High-Dose ACTH Stimulation Test**

This test consists of determining serum cortisol responses immediately before and 30 and 60 min after intravenous administration of 250  $\mu\text{g}$  of cosyntropin. This dose of cosyntropin results in pharmacologic plasma ACTH concentrations for the 60-min duration of the test, which may be too high to detect cases of chronic partial and mild pituitary ACTH deficiency. Therefore, this test may miss mild cases of adrenal insufficiency. Also, in early acute secondary or tertiary adrenal insufficiency, as in Sheehan syndrome, the test is not reliable because it takes several days for the adrenal cortex to atrophy. The advantage of the high-dose test is that the cosyntropin can be injected intravenously or intramuscularly since pharmacologic plasma ACTH concentrations can be achieved by either route.

An increase in serum cortisol concentration after 30 or 60 min to a peak of 18–20  $\mu\text{g}/\text{dl}$  (500–550 nmol/liter) or more is considered a normal response to the high-dose ACTH stimulation test and excludes the diagnosis of primary adrenal insufficiency and almost all cases of secondary adrenal insufficiency. However, if secondary adrenal insufficiency is of recent onset, the adrenal glands will have not yet atrophied and will still be capable of responding to ACTH stimulation normally. In these cases, a low-dose ACTH test or insulin-induced hypoglycemia may be required to confirm the diagnosis.

#### **Low-Dose ACTH Stimulation Test**

This test theoretically provides a more sensitive index of adrenocortical responsiveness because it

results in physiologic plasma ACTH concentrations. It is performed by measuring serum cortisol concentrations immediately before and 30 min after intravenous injection of cosyntropin in a dose of 1.0  $\mu\text{g}$  (160 mIU) per 1.73  $\text{m}^2$ . This dose stimulates maximal adrenocortical secretion up to 30 min postinjection and, in normal subjects, results in a peak plasma ACTH concentration approximately twice that of insulin-induced hypoglycemia. A value of 18  $\mu\text{g}/\text{dl}$  (500 nmol/liter) or more at any time during the test is indicative of normal adrenal function. The advantage of this test is that it can detect partial adrenal insufficiency that may be missed by the standard high-dose test. The low-dose test is also preferred for patients with secondary or tertiary adrenal insufficiency.

### Prolonged ACTH Stimulation Tests

Prolonged ACTH stimulation tests are rarely performed because the history and physical examination, the CRH test, and/or determination of cortisol and ACTH concentrations in association with the low-dose ACTH test may provide all necessary information. Prolonged stimulation with exogenous ACTH is used to differentiate between primary and secondary or tertiary adrenal insufficiency. In secondary or tertiary adrenal insufficiency, the adrenal glands display cortisol secretory capacity following prolonged stimulation with ACTH, whereas in primary adrenal insufficiency, the adrenal glands are partially or completely destroyed and do not respond to ACTH.

#### ***Eight-Hour ACTH Stimulation Test***

This test consists of administering 250  $\mu\text{g}$  (40 IU) of cosyntropin intravenously as an infusion for 8 h and determining serum cortisol and 24-h urinary cortisol and 17-hydroxycorticoid (17-OHCS) concentrations before and after the infusion. In normal subjects, the 24-h urinary 17-OHCS excretion increases three- to fivefold above the baseline. Serum cortisol concentrations reach 20  $\mu\text{g}/\text{dl}$  (550 nmol/liter) at 30–60 min and exceed 25  $\mu\text{g}/\text{dl}$  (690 nmol/liter) 6–8 h after initiation of the infusion.

#### ***Two-Day ACTH Stimulation Test***

The 2-day ACTH stimulation test is similar to the 8-h infusion test, except that 250  $\mu\text{g}$  of ACTH(1–24) is infused for more than 24 h on 2 (or 3) consecutive days. This test may be helpful in distinguishing primary from secondary/tertiary adrenal insufficiency. In primary adrenal insufficiency there is no or a minimal response of plasma or urinary cortisol and urinary

17-OHCS. Increases in these values during the 2 or 3 days of the test are indicative of a secondary/tertiary cause of adrenal insufficiency.

### CRH Stimulation Test

This test is used to differentiate between secondary and tertiary adrenal insufficiency. In both conditions, cortisol levels are low at baseline and remain low after CRH. In patients with secondary adrenal insufficiency, there is little or no ACTH response, whereas in patients with tertiary disease there is an exaggerated and prolonged response of ACTH to CRH stimulation, which is not followed by an appropriate cortisol response.

Additional methods to identify the cause of adrenal insufficiency vary depending on whether the disease is primary, secondary, or tertiary.

## TREATMENT

Adrenal insufficiency is a potentially life-threatening condition. Treatment should be initiated as soon as the diagnosis is confirmed or sooner if the patient presents in adrenal crisis.

### Adrenal Crisis

Adrenal crisis is a life-threatening emergency that requires immediate treatment. If the diagnosis is suspected but not known, blood samples should be obtained for measurement of cortisol concentrations.

#### ***Initial Treatment***

The aim of initial management in adrenal crisis is to treat hypotension (i.e., to correct hypovolemia) and to reverse the electrolyte abnormalities and cortisol deficiency. Large volumes of normal saline solution should be given intravenously. The glucocorticoid deficiency should be treated by immediate intravenous administration of dexamethasone sodium phosphate or hydrocortisone sodium succinate. Dexamethasone may be preferred because it has a long duration of action and does not interfere with the measurements of serum or urinary steroids during subsequent ACTH stimulation tests. After the initial treatment is provided, the cause of the adrenal crisis should be sought and treated.

#### ***Subsequent Treatment***

Once the patient's condition is stable and the diagnosis has been confirmed, parenteral glucocorticoid

therapy should be tapered over 3 or 4 days and converted to an oral maintenance dose. Patients with primary adrenal insufficiency require life-long glucocorticoid and mineralocorticoid replacement therapy.

### Chronic Adrenal Insufficiency

One of the important aspects of the management of chronic primary adrenal insufficiency is patient and family education. Patients should understand the reason for life-long replacement therapy and the need to increase the dose of glucocorticoid during minor or major stress and to inject hydrocortisone, methylprednisolone, or dexamethasone in emergencies.

#### **Emergency Precautions**

Patients should wear a medical alert (Medic Alert) bracelet or necklace and carry the Emergency Medical Information Card, which should provide information on the diagnosis, medications and daily doses, and the physician involved in the patient's management. Patients should also have supplies of dexamethasone sodium phosphate and should be educated about how and when to administer them.

#### **Glucocorticoid Replacement Therapy**

Patients with adrenal insufficiency should be treated with hydrocortisone, the natural glucocorticoid. The hydrocortisone daily dose is 25–30 mg (10–15 mg/m<sup>2</sup> body surface area) and can be given in two or three divided doses. A longer acting synthetic glucocorticoid, such as prednisolone, prednisone, or dexamethasone, may be employed but should be avoided because their longer duration of action may produce manifestations of chronic glucocorticoid excess, such as loss of lean body mass and bone density and gain of visceral fat. The usual oral replacement dosages are 5–7.5 mg of prednisolone or prednisone or 0.25–0.75 mg of dexamethasone once daily.

#### **Glucocorticoid Replacement during Minor Illness or Surgery**

During minor illness or surgical procedures, the dosage of glucocorticoid can be increased up to three times the usual maintenance dosage for 3 days. Depending on the nature and severity of the illness, additional treatment may be required.

#### **Glucocorticoid Replacement during Major Illness or Surgery**

During major illness or surgery, high doses of glucocorticoid up to 10 times the daily production rate are required to avoid an adrenal crisis. A continuous infusion of 10 mg of hydrocortisone per hour or the equivalent amount of dexamethasone or prednisolone eliminates the possibility of glucocorticoid deficiency. This dose can be halved on postoperative day 2, and the maintenance dose can be resumed on postoperative day 3.

#### **Mineralocorticoid Replacement Therapy**

Mineralocorticoid replacement therapy is required to prevent sodium loss, intravascular volume depletion, and hyperkalemia. It is given in the form of fludrocortisone (9 $\alpha$ -fluorohydrocortisone) in a dose of 0.1 mg daily. The dose of fludrocortisone is titrated individually based on the findings of clinical examination (mainly body weight and arterial blood pressure) and the levels of plasma renin activity. Patients receiving prednisone or dexamethasone may require higher doses of fludrocortisone to lower their plasma renin activity to the upper normal range, whereas patients receiving hydrocortisone, which has some mineralocorticoid activity, may require lower doses. The mineralocorticoid dose may have to be increased during the summer, particularly if patients are exposed to temperatures higher than 29°C (85°F).

#### **Androgen Replacement**

In women, the adrenal cortex is the primary source of androgen in the form of dehydroepiandrosterone and dehydroepiandrosterone sulfate. Although the physiologic role of these androgens in women has not been fully elucidated, their replacement is being increasingly considered in the treatment of adrenal insufficiency.

### Chronic Secondary and Tertiary Adrenal Insufficiency

Glucocorticoid replacement in chronic secondary or tertiary adrenal insufficiency is similar to that in primary adrenal insufficiency. However, measurement of plasma ACTH concentration cannot be used to titrate the optimal glucocorticoid dose. Mineralocorticoid replacement is rarely required, whereas replacement of other anterior pituitary deficits may be necessary.

## See Also the Following Articles

ACTH (Adrenocorticotropic Hormone) • Adrenal Androgens • Adrenal Cortex, Anatomy • Adrenal Cortex, Development • Adrenal Cortex Development, Regulation of • Adrenal Cortex, Physiology • Adrenal Suppression

## Further Reading

- Abdu, T. A. M., Elhadd, T. A., Neary, E., and Clayton, R. N. (1999). Comparison of the low dose short synacthen test (1  $\mu$ g), the conventional dose short synacthen test (250  $\mu$ g), and the insulin tolerance test for assessment of the hypothalamic–pituitary–adrenal axis in patients with pituitary disease. *J. Clin. Endocrinol. Metab.* **84**, 838.
- Addison, T. (1855). “On the Constitutional and Local Effects of Disease of the Supra-renal Capsules.” Highley, London.
- Crowley, S., Hindmarsh, P. C., Honour, J. W., and Brook, C. G. (1993). Reproducibility of the cortisol response to stimulation with a low dose of ACTH(1–24): The effect of basal cortisol levels and comparison of low-dose with high-dose secretory dynamics. *J. Endocrinol.* **136**, 167.
- Gold, P. W., Kling, M. A., Khan, I., Calabrese, J. R., Kalogeras, K., Post, R. M., Avgerinos, P. C., Loriaux, D. L., and Chrousos, G. P. (1987). Corticotropin releasing hormone: Relevance to normal physiology and to the pathophysiology and differential diagnosis of hypercortisolism and adrenal insufficiency. *Adv. Biochem. Psychopharmacol.* **43**, 183–200.
- Longui, C. A., Vottero, A., Harris, A. G., and Chrousos, G. P. (1998). Plasma cortisol responses after intramuscular corticotropin 1–24 in healthy men. *Metabolism* **47**, 1419.
- Nye, E. J., Grice, J. E., Hockings, G. I., Strakosch, C. R., Crosbie, G. V., Walters, M. M., and Jackson, R. V. (1999). Comparison of adrenocorticotropin (ACTH) stimulation tests and insulin hypoglycemia in normal humans: Low dose, standard high dose, and 8-hour ACTH-(1–24) infusion tests. *J. Clin. Endocrinol. Metab.* **84**(10), 3648–3655.
- Oelkers, W. (1996). Adrenal insufficiency. *N. Engl. J. Med.* **335**(16), 1206–1212.
- Schulte, H. M., Chrousos, G. P., Avgerinos, P., Oldfield, E. H., Gold, P. W., Cutler, G. B., Jr., and Loriaux, D. L. (1984). The corticotropin-releasing hormone stimulation test: A possible aid in the evaluation of patients with adrenal insufficiency. *J. Clin. Endocrinol. Metab.* **58**, 1064.
- Stewart, P. M. (2003). The adrenal cortex. In “Williams Textbook of Endocrinology” (P. R. Larsen, H. M. Kronenberg, S. Melmed, and K. S. Polonsky, eds.), pp. 525–532. Saunders, Philadelphia.
- Thaler, L. M., and Blevins, L. S., Jr. (1998). The low dose (1- $\mu$ g) adrenocorticotropin stimulation test in the evaluation of patients with suspected central adrenal insufficiency. *J. Clin. Endocrinol. Metab.* **83**, 2726.



# Adrenal Suppression

Maria Alexandra Magiakou

*Athens University Medical School, Athens, Greece*

George P. Chrousos

*NIH/National Institute of Child Health and Human Development, Bethesda, Maryland, United States*

## Glossary

**ACTH (adrenocorticotropic hormone)** A hormone synthesized and secreted by the anterior pituitary that is primarily responsible for the regulation of cortisol and adrenal androgen secretion.

**adrenal crisis** A medical emergency in which the subject is affected by an extreme state of adrenal insufficiency, with symptoms resembling shock or coma; considered to be life-threatening and requiring immediate administration of fluids, electrolytes, glucose, and intravenous glucocorticoids. May occur in situations of extreme stress or after the abrupt cessation of a period of glucocorticoid therapy.

**Cushing's syndrome** A metabolic disorder relating to the chronic oversecretion of adrenal glucocorticoids (mainly cortisol); may be associated with the long-term administration of pharmacological amounts of synthetic glucocorticoids. A condition first described by U.S. neurosurgeon Harvey Cushing in 1912.

**HPA axis** An anatomical complex consisting of the hypothalamus, the pituitary gland, and the adrenal cortex; ultimately controls the secretion of glucocorticoids from the adrenal cortex and thus plays a key role in various neuroendocrine, behavioral, autonomic, and immune responses to alterations in homeostasis.

Glucocorticoids are produced by the cortices of the adrenal glands and secreted into the systemic circulation in a circadian fashion and in response to stressful stimuli. These steroid hormones play pivotal roles in the regulation of intermediary metabolism, maintenance of cardiovascular function, stimulation of behavior, and control of the immune inflammatory reaction.

## INTRODUCTION

The major endogenous glucocorticoid in humans is cortisol, the synthetic form of which has traditionally

been called hydrocortisone. Cortisone, the 2-keto form of cortisol, was first used therapeutically in the management of rheumatoid arthritis by Hench and coworkers in 1949. Since then, a large number of synthetic compounds with glucocorticoid activity have been developed, and glucocorticoids (administered systemically or in a compartmental fashion) have been used in the therapy of a broad spectrum of nonendocrine and endocrine diseases.

One of the adverse effects of long-term glucocorticoid therapy in supraphysiologic doses is suppression of the hypothalamic–pituitary–adrenal (HPA) axis, which can render the adrenal glands unable to generate sufficient cortisol if glucocorticoid treatment is abruptly stopped, and the patient may develop glucocorticoid deficiency manifestations. The true prevalence of clinically significant adrenal insufficiency is not known since physicians usually discontinue high glucocorticoids gradually to allow recovery of the HPA axis.

Some of the risk factors for HPA axis suppression are clearly defined, whereas others are less certain. Systemic glucocorticoid therapy is more likely to suppress the HPA axis than compartmentalized use of glucocorticoids, with the possible exception of intra-articular steroids. Systemic glucocorticoid potency is also known to correlate with risk for adrenal insufficiency.

Glucocorticoid treatment in endocrine and none-endocrine disorders, the side effects of these medications, their concomitant use and interactions with other drugs, adrenal suppression, and the glucocorticoid withdrawal syndrome are discussed in detail in this article.

## PHYSIOLOGY OF THE HPA AXIS

The adrenal cortex consists of three anatomic zones: the outer zona glomerulosa, the intermediate zona fasciculata, and the inner zona reticularis. The zona

glomerulosa is responsible for the production of aldosterone, the zona fasciculata for the production of cortisol, and the zona reticularis for the production of adrenal androgens. Corticotropin (ACTH), synthesized and secreted by the corticotropes of the anterior pituitary, is the primary regulator of cortisol and adrenal androgen secretion. Hypothalamic control of ACTH secretion is exerted primarily by corticotropin-releasing hormone (CRH), a 41-amino acid peptide produced by parvocellular neurons of the paraventricular nucleus and secreted into the hypophyseal portal system.

There are several regulatory negative feedback loops that function to constrain the activity of the HPA axis. Prominent negative feedback loops are those exerted by glucocorticoids on CRH and ACTH secretion. The adrenal cortisol secretion rate under basal conditions is 12–15 mg/m<sup>2</sup>/day. In normal individuals, the highest plasma cortisol levels occur between 6:00 and 8:00 AM and the lowest at approximately midnight. Cortisol secretion increases two- to fourfold under stress. Plasma cortisol concentrations are elevated during physical and/or emotional stress, including illness, trauma, surgery, and starvation.

## ADRENAL INSUFFICIENCY

Adrenal insufficiency results from inadequate adrenocortical function, which may be due to destruction of the adrenal cortex (primary adrenal insufficiency; Addison's disease), deficient pituitary ACTH secretion (secondary adrenal insufficiency), or deficient hypothalamic secretion of CRH or other ACTH secretagogues (tertiary adrenal insufficiency). Primary and secondary adrenal insufficiency related to natural causes are uncommon, whereas iatrogenic, tertiary adrenal insufficiency caused by suppression of HPA function by glucocorticoid administration is common.

## PATHOGENESIS OF GLUCOCORTICOID-INDUCED ADRENAL SUPPRESSION

Glucocorticoid treatment may not suppress the HPA axis at all, or it may cause central suppression or complete adrenal gland atrophy. Supraphysiologic glucocorticoid doses inhibit both CRH production in the hypothalamus and ACTH production in the pituitary gland. When this inhibition lasts longer than the duration of the glucocorticoid exposure, it is called adrenal suppression.

The most severely affected glucocorticoid-treated patients have complete HPA axis suppression, characterized by functional adrenal gland atrophy. Cortisol production seems to depend on intermittent but consistent exposure to circulating ACTH. In states of profound or prolonged ACTH deficiency, the adrenal glands atrophy and are unable to generate cortisol in response to exogenous ACTH. However, in reality, adrenal suppression is a central nervous system problem. The rate-limiting step in HPA axis recovery from suppression is located in the brain. Thus, chronic administration of ACTH does not accelerate recovery, even though it may return cortisol production to normal while administered.

## SYNTHETIC GLUCOCORTICOIDS

Since the introduction of glucocorticoids in the treatment of rheumatoid arthritis in 1949, intense efforts have been made by science and industry to maximize the beneficial effects and to minimize the side effects of glucocorticoids. Thus, many synthetic compounds with glucocorticoid activity have been manufactured and tested. The pharmacologic differences among these chemicals result from structure alterations of their basic steroid nucleus and its side groups. These changes may affect the bioavailability of these compounds (including their gastrointestinal or parenteral absorption, plasma half-life, and metabolism in the liver, fat, or target tissues) and their abilities to interact with the glucocorticoid receptor and to modulate the transcription of glucocorticoid-responsive genes. In addition, structural modifications diminish the natural cross-reactivity of glucocorticoids with the mineralocorticoid receptor, eliminating their undesirable salt-retaining activity. Other modifications increase glucocorticoids' water solubility for parenteral administration or decrease their water solubility to enhance topical potency.

Thus, prednisolone has the structure of cortisol with an additional double bond between C1 and C2, which increases its glucocorticoid and decreases its mineralocorticoid activity. Introduction of an  $\alpha$ -fluoro group at C9, on the other hand, enhances both glucocorticoid and mineralocorticoid activity, whereas addition of a hydroxyl or methyl group at C16 practically eliminates mineralocorticoid activity. Dexamethasone, a potent synthetic glucocorticoid that has a double bond at C1 and C2, a fluoro group at C9, and an  $\alpha$ -methyl group at C16, has 25–50 times the glucocorticoid potency of cortisol and a minimal mineralocorticoid effect. A double bond between C2

and C3 and methylation at the C2 or C16 positions significantly prolong the plasma half-life of a compound. A keto group at C11 is normally reduced by liver enzymes to an 11 $\beta$ -hydroxyl group, which is necessary for glucocorticoid activity. In contrast, a C11 hydroxyl group is oxidized to a keto group in the kidney, minimizing the access of the compound to the mineralocorticoid receptor, and thus its salt-retaining effect.

Most synthetic glucocorticoids (e.g., methylprednisolone and dexamethasone) are minimally bound to cortisol-binding globulin (transcortin) and circulate mostly bound to albumin or in the free form. The percentage of such glucocorticoids bound to plasma proteins is relatively constant, and because binding is concentration independent, the metabolic clearance rate of glucocorticoids remains constant regardless of dose. Table I shows the relative glucocorticoid and mineralocorticoid potencies of commonly used systemic glucocorticoids and their approximate plasma and biologic effect half-lives. Glucocorticoid activity has been mostly defined in rat bioassays and may not always pertain to human responses, especially the growth-suppressing properties of synthetic glucocorticoids, which have been markedly underestimated. Based on the biologic effect half-life of glucocorticoids, they are classified as short, intermediate, or long acting based on the duration of corticotropin suppression after a single dose of the compound.

## SYSTEMIC GLUCOCORTICOID ADMINISTRATION

### Therapeutic Indications

Glucocorticoids may be administered as replacement therapy in patients with primary or secondary adrenal insufficiency, as adrenal suppression therapy in congenital adrenal hyperplasia and glucocorticoid resistance, and as anti-inflammatory or immunosuppressant therapy in a broad range of mostly nonendocrine disorders affecting many different systems. Thus, glucocorticoids are used in endocrine, autoimmune, collagen, renal, gastrointestinal, respiratory, nervous, hematologic, and ophthalmic diseases and are used in the suppression of the host versus graft and graft versus host reaction in cases of organ transplantation. Neoplastic disorders of the lymphoid system, such as leukemia and lymphomas, are also treated with glucocorticoids, along with the appropriate chemotherapy.

Acute administration of pharmacologic doses of glucocorticoids is necessary in a small number of nonendocrine diseases, such as malignant hyperthermia, and in patients with craniospinal trauma or brain tumors or in those who are undergoing major neurosurgical operations to decrease the temperature and prevent destruction of neural tissue from the local edema and inflammatory reaction, respectively. In addition, glucocorticoids have been used in the

Table I Glucocorticoid Equivalencies<sup>a</sup>

	Equivalent dose (mg)	Glucocorticoid potency	Mineralocorticoid potency	Plasma half-life (min)	Biologic half-life (h)
Glucocorticoids					
Short acting					
Cortisol	20.0	1.0	2	90	8–12
Cortisone	25.0	0.8	2	80–118	8–12
Intermediate acting					
Prednisone	5.0	4.0	1	60	18–36
Prednisone	5.0	4.0	1	115–200	18–36
Triamcinolone	4.0	5.0	0	30	18–36
Methylprednisolone	4.0	5.0	0	180	18–36
Long acting					
Dexamethasone	0.5	25–50	0	200	36–54
Betamethasone	0.6	25–50	0	300	36–54
Mineralocorticoids					
Aldosterone	—	0.3	300	15–20	8–12
Fluorocortisone	2.0	15.0	150	200	18–36
Desoxycorticosterone acetate	—	0.0	20	70	—

<sup>a</sup>From Liapi and Chrousos (1992).

prevention of the respiratory distress syndrome in the premature neonate, when delivery is anticipated before week 34 of gestation. In this case, treatment of the pregnant woman with 12 mg of betamethasone, followed by 12 mg 18–24 h later, stimulates the production of pulmonary surfactant and the maturation of the fetal lungs.

Recently, glucocorticoid therapy has been re-introduced in the treatment of adult acute respiratory distress syndrome. The recommended methylprednisolone doses are moderate and treatment is given continuously during the course of the disease. Significantly improved morbidity and mortality have been observed. Similar glucocorticoid treatment is being studied in patients with systemic inflammation syndrome, septic shock, and multiple organ dysfunction with promising results.

### Side Effects

Side effects occur only with supraphysiologic doses of glucocorticoids and not with proper replacement, which is equivalent to 12–15 mg of hydrocortisone/m<sup>2</sup> body surface area/day. Major complications are unlikely for short-term treatment (<2 weeks) with high doses of glucocorticoids, although sleep disturbances and gastric irritation are common complaints, and depression, mania, or psychosis may be infrequently precipitated. On the other hand, many side effects are associated with long-term daily administration of pharmacologic amounts of glucocorticoids (Table II), including the development of varying degrees of Cushing's syndrome manifestations during therapy and secondary adrenal insufficiency (adrenal suppression) after discontinuation of treatment. Growth retardation is one of the major side effects of long-term daily glucocorticoid therapy in children.

The degree of cushingoid features and the severity and length of adrenal suppression depend on the type and dose of the specific compound used, the duration of treatment, the idiosyncrasy, and the stress status of the patient. Most complications of glucocorticoid treatment are totally or partially reversible after discontinuation of glucocorticoid administration, with the exception of posterior subcapsular cataracts and advanced bone necrosis.

High doses of glucocorticoids suppress the immune defenses of the organism. Thus, individuals on glucocorticoid therapy are particularly susceptible to viral diseases against which they have not been vaccinated or naturally immunized (e.g., varicella, which can be devastating). Also, such individuals are

**Table II Effects of Long-Term Glucocorticoid Therapy<sup>a</sup>**

<b>Endocrine and metabolic</b>	
	Suppression of HPA axis (adrenal suppression)
	Growth failure in children
	Carbohydrate intolerance
	Hyperinsulinemia
	Insulin resistance
	Abnormal glucose tolerance test
	Diabetes mellitus
	Cushingoid features
	Moon facies, facial plethora
	Generalized and truncal obesity
	Supraclavicular fat collection
	Posterior cervical fat deposition (buffalo hump)
	Glucocorticoid-induced acne
	Thin and fragile skin, violaceous striae
	Impotence, menstrual disorders
	Decreased thyroid-stimulating hormone and triiodothyronine
	Hypokalemia, metabolic alkalosis
<b>Gastrointestinal system</b>	
	Gastric irritation, peptic ulcer
	Acute pancreatitis (rare)
	Fatty infiltration of liver (hepatomegaly) (rare)
<b>Hemopoietic system</b>	
	Leukocytosis
	Neutrophilia
	Increased influx from bone marrow and decreased migration from blood vessels
	Monocytopenia
	Lymphopenia
	Migration from blood vessels to lymphoid tissue
	Eosinopenia
<b>Immune system</b>	
	Suppression of delayed hypersensitivity
	Inhibition of leukocyte and tissue macrophage migration
	Inhibition of cytokine secretion or action
	Suppression of the primary antigen response
<b>Musculoskeletal system</b>	
	Osteoporosis, spontaneous fractures
	Aseptic necrosis of femoral and humeral heads and other bones
	Myopathy
<b>Ophthalmic</b>	
	Posterior subcapsular cataracts (more common in children)
	Elevated intraocular pressure or glaucoma
<b>Neuropsychiatric disorders</b>	
	Sleep disturbances, insomnia
	Euphoria, depression, mania, psychosis
	Pseudotumor cerebri (benign increase of intracranial pressure)

<sup>a</sup>From Laue *et al.* (1989).

susceptible to contract or to sustain activation of dormant tuberculosis. Individuals treated with massive doses of glucocorticoids can also develop saprophytic, fungal, or protozoan infections such as those seen in patients with severe immunodeficiency.

Nonfluorinated glucocorticoids (cortisone, cortisol, prednisone, and prednisolone) cross the placenta poorly. Fluorinated steroids, on the other hand, cross the placenta readily and should be given cautiously to women during pregnancy. The ratio of maternal–fetal plasma concentration gradients is approximately 10:1 for cortisol and prednisolone and approximately 2.5:1 for betamethasone and dexamethasone. Newborns exposed to high doses of synthetic fluorinated corticosteroids *in utero* should be checked for signs of adrenal insufficiency, and a Cortrosyn stimulation test should be performed to assess the need for glucocorticoid replacement. Special precautions should be taken for premature infants because some glucocorticoid preparations containing benzyl alcohol have been associated with a fatal “gasping” syndrome.

To avoid complications, alternate-day administration of intermediate-acting glucocorticoids should be used, when possible, if long-term therapy is necessary. Frequently, such a regimen can control the activity of the disease under therapy without causing Cushing’s syndrome, growth retardation, or adrenal suppression. Termination of long-term daily therapy (>2 weeks) should be gradual to prevent development of acute adrenal insufficiency and to avoid reactivation of the disease under therapy. Switching to daily hydrocortisone replacement or to alternate-day administration of intermediate-acting glucocorticoids is an acceptable method for weaning patients from glucocorticoid therapy.

## COMPARTMENTAL GLUCOCORTICOID ADMINISTRATION

### Topical Glucocorticoids

Glucocorticoids are quite effective when applied topically and are nontoxic to the skin in the short term. The factors that determine local penetration are the structure of the compound employed, the vehicle, the basic additives, occlusion versus open use, normal skin versus diseased skin, and small areas versus large areas of application. Fluorinated steroids (e.g., dexamethasone, triamcinolone acetonide, betamethasone, and beclomethasone) penetrate the skin better than nonfluorinated steroids, such as hydrocortisone. However, fluorinated steroids also cause more local

complications and may be associated with systemic absorption and side effects.

The complications of chronic topical skin use of glucocorticoids are mostly local (e.g., epidermal atrophy and hypopigmentation, telangiectasia, or acne and folliculitis) or infrequently systemic, with the classic manifestations of Cushing’s syndrome, growth retardation in children, and adrenal suppression. The frequency of systemic effects by topical corticosteroids is increased in newborns and small children compared to adolescents and adults because glucocorticoids penetrate the skin of newborns and small children more easily and in larger amounts. Systemic effects may also be observed in patients with hepatic disease or idiosyncratically because of decreased drug metabolism. Although most types of dermatitis are generally responsive to topical glucocorticoids, there are rare cases in which intralesional injections might be considered (e.g., hypertrophic scars, acne cysts, or prurigo nodularis).

### Ophthalmic Glucocorticoids

Patients with autoimmune or idiopathic inflammation of the anterior segment of the eye (e.g., iritis and uveitis) may benefit from local administration of glucocorticoids. Also, patients with postsurgical or traumatic inflammation are given topical glucocorticoids to prevent local destruction from edema. Special care should be taken to avoid treating patients with herpes simplex conjunctivitis or keratitis during the infectious stage of the disease because major spread of the infection may be precipitated.

### Inhaled Glucocorticoids

Glucocorticoid inhalation therapy is widely used in patients with bronchial asthma and croup. The existing preparations at the recommended doses have a remarkable therapeutic effect without causing manifestations of Cushing’s syndrome, growth retardation, or clinically significant adrenal suppression. Systemic effects may be observed, however, as a result of increased intake of such preparations or altered steroid metabolism.

Inhaled glucocorticoids have also been used in ventilator-dependent preterm infants to reduce the severity of respiratory distress syndrome and to facilitate weaning from mechanical ventilation of infants with bronchopulmonary dysplasia. Their effect on the function of the immature HPA axis of these preterm neonates is unclear.

## Nasal Glucocorticoids

Aerosols containing glucocorticoids are available for the treatment of allergic rhinitis. Frequent and chronic use should be avoided to prevent local and systemic complications.

## Intra-articular Glucocorticoids

The intra-articular injection of glucocorticoids may be of value in carefully selected patients if strict aseptic techniques are used and repeated and frequent injections are avoided.

## MONITORING OF PATIENTS RECEIVING GLUCOCORTICOID TREATMENT

Patients receiving long-term treatment with glucocorticoids should adhere to a high-protein, calorie-restricted diet. The diet should also be rich in potassium and calcium and low in sodium. Adequate ambulation or exercise should be recommended to prevent muscular atrophy and osteopenia. Patients should concurrently take antacids or histamine antagonists to prevent gastric irritation or peptic ulcers. Young children should have their growth monitored every 3 months (until age 5), and older children should have their growth monitored every 6 months. For all patients, body weight, length or height, blood pressure, fasting and 2-hpostprandial blood glucose, serum electrolytes, and bone maturation and density should be measured. Because glucocorticoids decrease the organism's response to infection, care should be taken to determine whether latent infections, such as mycobacterial disease, are present before treatment begins.

## CONCOMITANT USE OF GLUCOCORTICOID WITH OTHER DRUGS

Special attention is required for the concomitant use of glucocorticoids with other drugs because of potential interactions and because some drugs may affect the metabolism of the steroids, which may lead to a decreased or increased glucocorticoid effect on their target tissues. Such interactions and effects are shown in [Tables III–V](#).

## PREDICTING GLUCOCORTICOID-INDUCED HPA AXIS SUPPRESSION

Several predictors of glucocorticoid-induced HPA axis suppression have been discussed. The following are the most important:

1. Kind of steroid used and glucocorticoid potency: The synthetic analogs of glucocorticoids are much better tolerated as anti-inflammatory agents because they cause significantly less sodium retention at supraphysiologic doses than hydrocortisone and cortisone acetate. Glucocorticoid potency ([Table I](#)) correlates positively with risk for adrenal insufficiency. Thus, hydrocortisone and cortisone acetate are the least potent and, therefore, least suppressive agents. Prednisone, prednisolone, methylprednisolone, and triamcinolone are moderately suppressive, and dexamethasone suppresses ACTH the longest.
2. Systemic vs compartmental therapy: Systemic glucocorticoid therapy is more likely to suppress the HPA axis than are intra-articular, inhalational, or topical glucocorticoids.
3. Alternate-day therapy: There is evidence that patients are at lower risk for adrenal insufficiency if they can take glucocorticoids on alternate days from the outset or if they can convert to alternate-day therapy before the HPA axis is suppressed.

Table III Interactions of Glucocorticoids with Other Drugs<sup>a</sup>

Drug	Side effect	Comment
Amphotericin B	Hypokalemia	Monitor potassium levels frequently
Digitalis glycosides	Digitalis toxicity	Monitor potassium levels frequently
	Hypokalemia	
Growth hormone	Ineffective	—
Potassium-depleting diuretics	Hypokalemia	Monitor potassium levels frequently
Vaccines from live attenuated viruses	Severe generalized infections	—

<sup>a</sup>From Liapi and Chrousos (1992).

Table IV Effects of Glucocorticoids on Blood Levels of Other Drugs<sup>a</sup>

Drug	Drug blood levels	Comments
Aspirin	Decreased	Increased metabolism or clearance; monitor salicylate levels
Coumarin anticoagulants	Decreased	Frequent control of prothrombin levels
Cyclophosphamide	Increased	Inhibition of hepatic metabolism; adjust the dosage of the drug
Cyclosporine	Increased	Inhibition of hepatic metabolism
Insulin	Decreased	Adjust the dosage of the drug
Isoniazid	Decreased	Increased metabolism and clearance
Oral hypoglycemic agents	Decreased	Adjust the dosage of the drug

<sup>a</sup>From Liapi and Chrousos (1992).

4. Once-a-day dosing in the morning or mimicking normal diurnal cortisol rhythms: Since evening doses of glucocorticoids tend to suppress the normal early morning surge of ACTH secretion, it is better, whenever possible, to cure patients with a single morning dose. Once-a-day dosing is usually feasible for prednisone, triamcinolone, and dexamethasone. The short-acting hydrocortisone and cortisone acetate are usually given twice a day, at waking and at approximately 5 P.M. To mimic normal diurnal cortisol rhythms, the morning dose is two-thirds and the afternoon dose is one-third of the total daily dose.

5. Duration and cumulative dose of glucocorticoid treatment: Although traditionally the duration of glucocorticoid therapy and the cumulative dose of glucocorticoid received have been considered as predictive of the likelihood of HPA axis suppression, several studies suggest that they only roughly predict HPA axis suppression. Adrenal insufficiency is extremely rare in patients treated for 1 week or less.

Perhaps the best predictor of HPA axis suppression is the patient's current glucocorticoid dosage. A strong correlation has been found between prednisone maintenance doses >5 mg/day and a subnormal ACTH stimulation test result.

### WEANING PATIENTS FROM GLUCOCORTICOID THERAPY

Termination of long-term daily glucocorticoid therapy (>2 weeks) should be gradual to prevent development of adrenal insufficiency and to avoid reactivation of the disease under therapy. The likelihood of the latter depends on the activity and natural history of the disorder. When there is a chance that the underlying illness may recur, the glucocorticoids should be withdrawn slowly over a period of weeks to months, with frequent reassessment of the patient's condition. Daily hydrocortisone replacement or double or triple replacement of intermediate-acting glucocorticoids

Table V Effect of Drugs on Plasma Glucocorticoid Concentrations<sup>a</sup>

Drug	Glucocorticoid blood levels	Comments
Antacids	Decreased	Possible physical adsorption to antacid
Carbamazepine	Decreased	Increased cytochrome P450 activity
Cholestyramine	Decreased	Decreased gastrointestinal absorption of glucocorticoids
Colestipol	Decreased	Decreased gastrointestinal absorption of glucocorticoids
Cyclosporine	Increased	Inhibition of hepatic metabolism
Ephedrine	Decreased	Probably increased metabolism
Erythromycin	Increased	Impaired elimination
Mitotane	Decreased, with elevated transcortin	Total plasma cortisol unreliable; adjust glucocorticoid levels
Oral contraceptives	Increased	Impaired elimination, increased protein binding
Phenobarbital	Decreased	Increased cytochrome P450 activity; adjust glucocorticoid dosage
Phenytoin	Decreased	Increased cytochrome P450 activity; adjust glucocorticoid dosage
Rifampin	Decreased	Increased cytochrome P450 activity(?); adjust glucocorticoid dosage
Troleandomycin	Increased	Partially resulting from impaired elimination

<sup>a</sup>From Liapi and Chrousos (1992).

given on alternate days are acceptable methods for weaning patients from glucocorticoid therapy.

## ACUTE ADRENAL CRISIS

Recovery of the HPA axis can take 12 months or longer. Abrupt cessation of glucocorticoid treatment or quick tapering can precipitate an acute adrenal insufficiency crisis. The main symptoms range from anorexia, fatigue, nausea, vomiting, dyspnea, fever, arthralgia, myalgia, and orthostatic hypotension to dizziness, fainting, and circulatory collapse. Hypoglycemia is occasionally observed in children and very thin adults. The diagnosis is a medical emergency, and treatment should consist of immediate administration of fluids, electrolytes, glucose, and parenteral glucocorticoids.

## GLUCOCORTICOID WITHDRAWAL SYNDROME

Glucocorticoid withdrawal can present as an acute adrenal crisis or with symptoms of chronic glucocorticoid deficiency. Thus, patients may suffer from anorexia, myalgia, nausea, emesis, lethargy, headache, fever, skin desquamation, arthralgias, weight loss, and postural hypotension. In addition, they may experience exacerbation of a previously present autoimmune disease (e.g., rheumatoid arthritis, atopic dermatitis, and asthma) or develop a new autoimmune disease (e.g., Hashimoto's thyroiditis and Graves' disease). Amatruda *et al.* first defined the steroid withdrawal syndrome as a symptom complex resembling true adrenal insufficiency, with nonspecific symptoms such as weakness, nausea, and arthralgias, occurring in patients who have completed a dosage reduction of glucocorticoid therapy and who respond normally to HPA axis testing.

The occurrence of the subjective component of the steroid withdrawal syndrome does not depend on the absence of cortisol from the circulation or an impairment of the HPA axis because these symptoms may occur while the patient is on proper glucocorticoid replacement or when the patient has a normal cortisol response to Cortrosyn. In this instance, the steroid withdrawal syndrome may be a result of difficulties in withdrawing from the high levels of glucocorticoids—a phenomenon that appears to be idiosyncratic.

However, when patients become ill after a dosage reduction, the physician should consider a differential diagnosis that includes true adrenal insufficiency, a

flare of the disease being treated, and steroid withdrawal syndrome. All three conditions resolve after patients are restarted on the glucocorticoid regimen that previously controlled their symptoms.

## BIOCHEMICAL DIAGNOSIS OF ADRENAL INSUFFICIENCY

As previously mentioned, glucocorticoid treatment may not suppress the HPA axis at all, or it may cause central suppression and adrenal gland atrophy of varying degrees. The insulin tolerance test and the metyrapone test have been employed in the diagnosis of adrenal suppression and are quite sensitive. However, the risks involved with both tests do not justify their use when a rapid ACTH stimulation test can distinguish clinically significant adrenal suppression.

To evaluate the adequacy of HPA axis recovery, the rapid Cortrosyn (or high-dose ACTH stimulation test) is most commonly used. An intravenous bolus of 250  $\mu\text{g}$  of corticotropin 1-24 is administered, and cortisol is measured after 30 or 60 min or both. A plasma cortisol concentration  $>18\text{--}20$   $\mu\text{g}/\text{dL}$  at these times indicates adequate recovery of the HPA axis. This test can also be done intramuscularly.

A modified Cortrosyn test has been recommended in lieu of the standard test. Only 1  $\mu\text{g}$  of corticotropin 1-24 is administered instead of 250  $\mu\text{g}$ . This test is fraught with technical errors as a result of multiple dilutions of the Cortrosyn preparations and adhesion of Cortrosyn to the tubing system. Its purported increased sensitivity may not necessarily signify a greater clinical prediction of adrenal suppression.

## See Also the Following Articles

ACTH (Adrenocorticotrophic Hormone) • Adrenal Cortex, Anatomy • Adrenal Cortex, Development • Adrenal Insufficiency • Corticotropin-Releasing Hormone (CRH) and Inflammation • Glucocorticoids, Overview

## Further Reading

- Ackerman, G. L., and Nolan, C. M. (1968). Adreno cortical responsiveness after alternate corticosteroid therapy. *N. Engl. J. Med.* **278**, 405–409.
- Axelrod, L. (1976). Glucocorticoid therapy. *Medicine* **55**, 39–65.
- Boumpas, D. T., Chrousos, G. P., Wilder, R. L., *et al.* (1993). NIH conference of the combined clinical staff. Glucocorticoid therapy of immune related diseases: Basic and clinical correlates. *Ann. Intern. Med.* **119**, 1198–1208.
- Christy, N. P. (1992). Pituitary adrenal function during corticosteroid therapy: Learning to live with uncertainty. *N. Engl. J. Med.* **326**, 266–267.

- Chrousos, G. P., and Harris, A. G. (1998). Hypothalamic pituitary adrenal axis suppression and inhaled corticosteroid therapy. Part I. General principles. *Neuroimmunomodulation* **5**, 287.
- Chrousos, G. P., and Harris, A. G. (1998). Hypothalamic pituitary adrenal axis suppression and inhaled corticosteroid therapy. Part II. Review of the literature. *Neuroimmunomodulation* **5**, 288–308.
- Gomez, M. T., Magiakou, M. A., Mastorakos, G., and Chrousos, G. P. (1993). The pituitary corticotroph is not the rate limiting step in the postoperative recovery of the hypothalamic pituitary adrenal axis in patients with Cushing's syndrome. *J. Clin. Endocrinol. Metab.* **77**, 173.
- Krasner, A. S. (1999). Glucocorticoid induced adrenal insufficiency. *J. Am. Med. Assoc.* **282**, 671.
- Laue, L., Kawai, S., Udelsman, R., et al. (1989). Glucocorticoid antagonists: Pharmacological attributes of the prototype antiglucocorticoid RU 486. In "Antiinflammatory Steroid Action: Basic and Clinical Aspects" (L. M. Lichtenstein, H. Claman, and A. Oronsky, eds.), pp. 303–329. Academic Press, New York.
- Liapi, C., and Chrousos, G. P. (1992). Glucocorticoids. In "Pediatric Pharmacology" (S. J. Jaffe, and J. V. Aranda, eds.), 2nd ed., pp. 466–475. Saunders, Philadelphia.
- Magiakou, M. A., and Chrousos, G. P. (1994). Corticosteroid therapy, nonendocrine disease and corticosteroid withdrawal. In "Current Therapy in Endocrinology and Metabolism" (C. W. Bardin, ed.), 5th ed., pp. 120–124. Mosby Yearbook, St. Louis.
- Magiakou, M. A., and Chrousos, G. P. (1996). Corticosteroid therapy and withdrawal. In "Endocrinology and Metabolic Diseases, Current Practice of Medicine," pp. 6.1–6.6. Palatino Helvetica, Philadelphia.
- Orth, D. N. (1994). Adrenal insufficiency. In "Current Therapy in Endocrinology and Metabolism" (C. W. Bardin, ed.), 5th ed., pp. 124–130. Mosby Yearbook, St. Louis.
- Rimsza, M. E. (1978). Complications of corticosteroid therapy. *Am. J. Dis. Child.* **132**, 806.
- Tyrell, J. B., and Baxter, J. D. (1987). Glucocorticoid therapy. In "Endocrinology and Metabolism" (P. Felig, J. B. Baxter, and A. E. Broadus, eds.), 2nd ed., pp. 788–817. McGraw-Hill, New York.



# Adrenal Tumors, Molecular Pathogenesis

Christian A. Koch

NIH/National Institute of Child Health and Human Development and University of Leipzig, Leipzig, Germany

George P. Chrousos

NIH/National Institute of Child Health and Human Development, Bethesda, Maryland, United States

## Glossary

**adrenocorticotrophic hormone or corticotropin (ACTH)** Stimulates the adrenal gland.

**first hit** According to Knudson's two-hit model of tumorigenesis, in tumor suppressor gene-related tumors, the first hit is the germ-line mutation of the defective tumor suppressor gene. The second hit consists most commonly of a deletion of the wild-type allele of the tumor suppressor gene in question and can cause biallelic inactivation of this gene with subsequent tumor formation in the affected cell and organ.

**hyperaldosteronism** State of aldosterone excess. The most common cause in primary hyperaldosteronism is an aldosterone-oversecreting adrenocortical mass that can lead to secondary hypertension in the affected patient.

**incidentaloma** Mass (e.g., in the adrenal gland) detected by chance during radiologic imaging workup.

**loss of heterozygosity (LOH)** This can occur through deletion of the wild-type allele of a gene that has a germ-line mutation in hereditary tumor syndromes or through deletion of a gene that has a somatic mutation in sporadic tumors. This will move the affected cell into the homozygous state for the respective gene.

Adrenal tumors are found in up to 9% of autopsy studies. Modern imaging modalities facilitate better detection of adrenal masses than in the past but also detect adrenal incidentalomas in patients who are not evaluated for adrenal tumors. Assessing such patients for the presence of subclinical disease caused by adrenal incidentalomas and the potential for malignancy is challenging. Although there are radiological (e.g., intratumor necrosis, irregular margins, and local metastases), histopathological (e.g., Weiss criteria), and genetic (e.g., overexpression of *IGF2* and loss of heterozygosity at 17p13) criteria to predict malignancy, no absolute specific marker and feature exist to reliably distinguish benign and malignant adrenal lesions, excluding pure adrenal ganglioneuroma, which by definition is

always benign. Although adrenal tumors have been observed in several familial syndromes, the majority of these lesions occur sporadically. If the gene defect in familial syndromes is known, timely identification of family members already affected by or at risk for the development of an adrenal tumor is possible by genetic screening tools including germ-line mutation analysis. Despite these molecular advances and tools, the pathogenesis of adrenal tumors remains widely unknown. The elucidation of adrenal tumorigenesis may be facilitated by studying adrenal tumors of patients with hereditary syndromes since in these tumors at least one hit, the inherited gene defect, is known and presumably represents the "first hit" in tumor evolution. In contrast, sporadic tumors have an unknown first hit, making it more difficult to determine the sequence of genetic events or hits. In this article, we discuss the molecular pathogenesis in hereditary and sporadic adrenal tumors.

## ADRENOMEDULLARY TUMORS

### Pheochromocytoma

#### **Von Hippel–Lindau Syndrome**

Von Hippel–Lindau (VHL) syndrome consists of a variety of masses, including renal carcinomas, hemangioblastomas, and pheochromocytomas. It affects approximately 1 in 36,000 individuals and is caused by mutations in the VHL tumor suppressor gene located at chromosome 3p25–26. Less than 26% of patients with a *VHL* germ-line mutation develop a pheochromocytoma, and up to one-third of these patients do not suffer from symptoms of catecholamine excess since they may have "silent" (not catecholamine-oversecreting) pheochromocytomas. Most VHL-associated pheochromocytomas have loss of function (LOH) at 3p25–26, leading to biallelic inactivation of the VHL gene, thereby following Knudson's two-hit model of tumorigenesis. A small subset of VHL pheochromocytomas have LOH at 1p or chromosome 11. VHL protein leads to degradation of certain proteins, most in the 26S proteasome complex (Table I).

Table 1 Gene Alterations in Pheochromocytoma<sup>a</sup>

Gene	Chromosomal location	Gene alteration	Prevalence
<b>Tumor suppressor genes</b>			
<i>VHL</i>	3p25.5	LOH of the wild-type allele in VHL tumors	33/46
		LOH in MEN2 tumors	10/18
		Somatic mutations in MEN2 tumors	3/21
		Somatic mutations in VHL tumors	0/36
		Somatic mutations in sporadic tumors	1/20
		LOH in sporadic tumors	35/102 benign, 3/10 malignant
<i>NF1</i>	17q11.2	LOH of the wild-type allele	3/7
		LOH in sporadic tumors	1/6
		Absent neurofibromin expression	6/6
		Reduced or absent neurofibromin expression	1/4 sporadic, 5/14 MEN-2, 1/2 VHL
<i>SDHD</i>	11q23	LOH in sporadic tumors	13/18
<i>p53</i>	17p13	Somatic mutations	1/18
		Somatic mutations	6/55
<i>MEN1</i>	11q13	LOH	10/46
		LOH of the wild-type allele	2/2
<i>RASSF1A</i>	3p21	Somatic mutations	0/23
<i>p16</i>	9p21	Hypermethylation	5/23
		LOH	0/26
<b>Oncogenes</b>			
<i>RET</i>	10q11.2	Duplication of the mutant <i>RET</i> allele in trisomy 10 or loss of the wild-type allele in MEN2 tumors	7/9
		LOH in MEN2 tumors	0/13
		Gain at 10q or 10 in sporadic tumors	2/10 malignant, 3/42 benign
		LOH in sporadic tumors	2/38 benign, 1/10 malignant
		Somatic mutations in MEN2 tumors	1/3
		Somatic mutations in VHL tumors	0/41
		Somatic mutations in sporadic tumors	17/149 benign, 1/29 malignant
<i>Ras</i>	11p15 ( <i>H-ras</i> )	Somatic mutations	0/8
	12p11 ( <i>K-ras</i> )	Somatic mutations	0/8
	20q13.2	Somatic mutations	0/10
<i>EGFR</i>	7p12	Overexpression	4/7

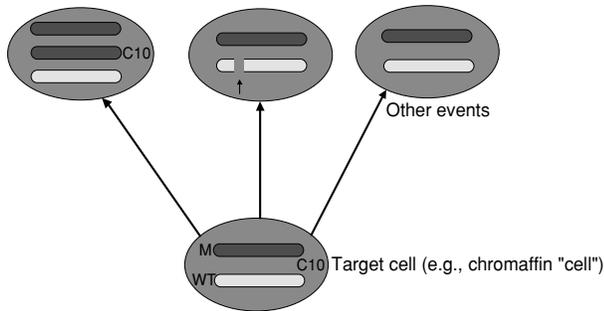
<sup>a</sup>Modified from Koch, Pacak, and Chrousos (2002).

### Multiple Endocrine Neoplasia Types 1 and 2

Multiple endocrine neoplasia type 1 (MEN1) is an autosomal dominant tumor syndrome caused by germ-line mutations in the gene *menin* located at chromosome 11q13. Characteristic endocrine tumors occur in the pituitary, parathyroid gland, and pancreas. Genotype–phenotype correlations in patients with germ-line mutations in the MEN1 gene do not exist. Adrenal nodules occur in patients with MEN1 approximately four times more often than in individuals without MEN1, suggesting a pathogenetic link. Pheochromocytomas in patients with MEN1 germ-line mutations have rarely been reported. LOH

at 11q13 in these tumors has been observed in pheochromocytomas.

Multiple endocrine neoplasia type 2 (MEN2) is caused by germ-line mutations in the *RET* protooncogene located at chromosome 10q11.2. It is an autosomal dominantly inherited cancer syndrome and affects approximately 1 in 40,000 individuals. MEN2 is characterized by the presence of medullary thyroid carcinoma, pheochromocytoma, and parathyroid hyperplasia. Approximately 50% of patients with germ-line mutations in *RET* develop a pheochromocytoma during their lifetime. There is a clear genotype–phenotype correlation in patients with MEN2.



**Figure 1** Model for tumorigenesis in MEN2-associated pheochromocytomas. In an individual with MEN2, each cell of the target organs (e.g., adrenal medulla) carries a *RET* germ-line mutation in the heterozygous state with one intact wild-type *RET* allele. Selected chromaffin cells undergo a second hit through duplication of the mutant *RET* allele in trisomy 10 or loss of the wild-type allele, giving these cells overrepresentation of mutant *RET* and a growth advantage with subsequent tumor formation. M, mutant *RET* allele; WT, wild-type *RET* allele; C10, chromosome 10. Modified from Koch, Pacak, and Chrousos (2002).

The same germ-line mutation in *RET* may lead to the development of a pheochromocytoma in one patient but not in another patient. Similarly, one patient with MEN2 may develop medullary thyroid carcinoma in the first month of life, whereas another patient with the same germ-line mutation in *RET* may develop medullary thyroid carcinoma at his or her 83rd birthday. It is therefore puzzling how germ-line mutations in *RET* lead to tumor formation in patients with MEN2. Recently, the model of a “second hit” in such patients was introduced. Selected cells in the target organs, such as the chromaffin cells and the C cells of the thyroid gland, may gain a growth advantage by overrepresentation of mutant *RET* by duplication of mutant *RET* in trisomy 10, loss of wild-type *RET*, or tandem amplification (Fig. 1). A subset of MEN2-related pheochromocytomas have somatic *VHL* gene alterations, possibly leading to an overrepresentation of *RET*. LOH at 1p has been reported in some MEN2-related pheochromocytomas, highlighting the developmental relationship between pheochromocytoma and neuroblastoma (neuroblastomas frequently have LOH at 1p). Also, reduced neurofibromatosis type 1 (*NF1*) expression occurs in a subset of MEN2 pheochromocytomas, suggesting a role for the *ras* pathway in tumor formation.

### Neurofibromatosis Type 1

Although this inherited tumor syndrome affects 1 in 4000 individuals, pheochromocytomas occur in less than 2% of patients with *NF1*. The *NF1* gene maps

to chromosome 17q11.2 and encodes neurofibromin, which is involved in controlling the *ras* signaling pathway. Mice heterozygous for one mutant *NF1* allele develop pheochromocytoma in 50% of cases. LOH at 17q11 has been observed in a subset of *NF1* pheochromocytomas. Because of the rarity of *NF1*-related pheochromocytomas, there is a lack of larger genetic studies and the molecular pathogenesis of *NF1* pheochromocytomas is largely unknown.

### *SDHX* Syndromes

Pheochromocytomas are also referred to as adrenal paragangliomas (“next to the ganglia”). Paragangliomas can be of sympathetic and parasympathetic origin. Sympathetic paragangliomas are mainly located in the retroperitoneum. Parasympathetic paragangliomas are often located in the neck from the skull base down to the aortic arch. Head and neck paragangliomas usually do not oversecrete catecholamines, whereas those below the neck commonly do. Carotid body paragangliomas develop from the carotid body, a structure that serves as an oxygen-sensing organ. Chronic hypoxia has been shown to be associated with enlargement of the carotid bodies, an observation that drew the attention of investigators to possible genetic defects of the oxygen-sensing and oxygen-signaling pathways. Subsequently, germ-line mutations in the succinate-ubiquinone oxidoreductase subunit D gene (*SDHD*), a gene belonging to the mitochondrial complex II that is involved in the Krebs cycle and in the aerobic electron transport chain, have been identified in extraadrenal pheochromocytomas. Researchers have also analyzed adrenal pheochromocytomas for mutations in *SDHD*. Subsequent investigations on genes encoding for the other three subunits of mitochondrial complex II led to the detection of germ-line mutations in *SDHB* and *SDHC* in patients with hereditary pheochromocytomas. To date, germ-line mutations in *SDHA*, the gene encoding for the flavoprotein, have not been found in adrenal pheochromocytomas or in “paragangliomas.” Of the five mitochondrial complexes (I–V), complex II is the only one with no subunits encoded by the mitochondrial genome. The *SDHD* gene is located at 11q23 and consists of four exons. Mutation analysis in pheochromocytomas has revealed missense and nonsense mutations. *SDHD* encodes for the small (cybS) subunit of cytochrome b in the succinate-ubiquinone oxidoreductase (complex II). Functional analysis revealed that inactivation of cybS by a nonsense mutation (R22X) abolishes the enzymatic activity of mitochondrial complex II and activates the hypoxia pathway (e.g., through increased expression of

VEGF). The *SDHB* gene is located at 1p35–36 and consists of eight exons. Mutation analysis showed inactivating mutations in hereditary pheochromocytomas and paragangliomas. *SDHB* encodes for the iron sulfur protein in the succinate–ubiquinone oxidoreductase (complex II). The *SDHC* gene is located at 1q21 and consists of six exons. Mutation analysis revealed a G-to-A transition in exon 1 in one family with paragangliomas. *SDHC* encodes for the large (cybL) subunit of cytochrome b in the succinate–ubiquinone oxidoreductase (complex II).

### Sporadic Pheochromocytomas

Allele losses at 1p, 3p, 3q, 17p, and 22q are commonly found in pheochromocytomas. Their role in tumorigenesis and tumor progression, however, remains unclear. Less than 10% of sporadic pheochromocytomas have somatic mutations in *RET*, *VHL*, *SDHD*, or *SDHB*. Among 91 sporadic pheochromocytomas investigated for *SDHD* mutations, only 1 carried a somatic mutation. Mutation analysis of *SDHB* in 24 sporadic pheochromocytomas revealed only 1 with a germ-line mutation. Mutations in *SDHC* have not been identified.

Notably, apparently sporadic pheochromocytomas may be part of a familial syndrome with germ-line mutations in *VHL*, *RET*, *SDHD*, or *SDHB*. In a recent study of 271 patients, Neumann *et al.* reported up to 24% of such cases. This may justify screening all patients presenting with a pheochromocytoma for germ-line mutations in the previously mentioned genes. Somatic mutations and genetic alterations in other genes, such as *p53*, *NF1*, *RASSF1A*, *c-erbB-2*, and *EGFR*, have been observed in sporadic pheochromocytomas, suggesting a role in the pathogenesis of these tumors. No markers reliably distinguish benign and malignant pheochromocytomas. The only criterion is clinical: the presence of metastases at locations at which there is usually no chromaffin tissue (e.g., liver, bone, and lung). Telomerase expression does not predict malignancy. Clonal analyses of adrenomedullary nodules in patients with a *RET* germ-line mutation suggest that these nodules are monoclonal.

### Ganglioneuromas and Ganglioneuroblastomas

Ganglioneuromas are neuroectodermal tumors related to neuroblastoma. Rarely, these tumors occur in the adrenal gland and are entirely benign except for mixed or so-called composite tumors, such as ganglioneuroblastomas. Most neuroblastomas show LOH at 1p36

and have a less favorable prognosis. Ganglioneuromas as mature tumors are not expected to reveal LOH at 1p36. A few tumors have low p53 content. The pathogenesis of these adrenal lesions is unclear.

## ADRENOCORTICAL TUMORS

### Li–Fraumeni Syndrome

This autosomal dominant familial cancer syndrome is associated with breast cancer, brain tumors, soft tissue sarcomas, leukemia, and adrenocortical carcinoma. Adrenal cancer occurs in approximately 1% of patients with the classic Li–Fraumeni syndrome. Germ-line mutations in *p53* at chromosome 17p13 are responsible for this syndrome and can lead to adrenocortical tumors as the sole manifestation. Frequently, adrenal tumors of patients with germ-line mutations in *p53* show LOH at 17p13 and, therefore, evidence of biallelic inactivation of this tumor suppressor gene (Table II).

### Beckwith–Wiedemann Syndrome

This syndrome shows variable expressivity and occurs sporadically. Affected patients are at increased risk for developing adrenal cancer. The gene locus is at 11p15.5, a chromosomal region that includes the *IGF2* and the *p57/KIP2* genes. p57 is a negative regulator of cell proliferation and inhibits G1 cyclin/cyclin-dependent kinase (CDK) complexes. Adrenocortical tumors in this syndrome show overexpression of *IGF2*, probably related to uniparental paternal isodisomy for the *IGF2* locus. Duplication of the paternal 11p15 allele containing the *IGF2* gene locus and/or loss of the maternal allele are frequently found in adrenal cancer. This is in contrast to adrenocortical lesions that are classified as benign and frequently do not demonstrate overexpression of *IGF2*.

### Carney Complex

First described in 1980, this autosomal dominant hereditary syndrome comprises primary pigmented nodular adrenocortical disease, growth hormone-secreting pituitary tumors, spotty skin pigmentations, atrial and peripheral myxomas, and psammomatous melanotic schwannomas. Adrenocortical disease occurs in approximately 26% of patients with this complex. At least two chromosomal loci have been identified: 2p16 and 17q22–24. A subset of patients with Carney complex have germ-line mutations in

Table II Genetic Alterations in Adrenocortical Tumors<sup>a</sup>

Gene	Chromosomal location	Gene alteration	Prevalence (tumors)
<b>Tumor suppressor genes</b>			
<i>p53</i>	17p13	LOH at the wild-type allele in Li-Fraumeni syndrome LOH in sporadic tumors Somatic mutations in sporadic tumors Gain at 17 in sporadic tumors	13/17 22/39 cancers 26/97 16/42 cancers 11/57 6/26 cancers 9/38
<i>MEN1</i>	11q13	LOH of the wild-type allele in MEN1 tumors LOH at 11q13 in sporadic tumors Somatic mutations in sporadic tumors	1/13 11/13 cancers 28/126 0/18 cancers 2/90
<i>p57/kip2</i> and <i>H19</i>	11p15	LOH in sporadic tumors Somatic mutations in sporadic tumors Low expression in sporadic tumors	21/38 cancers 35/155 0/61 16/17 cancers 8/19
<i>p21</i>	6p21	Overexpression in sporadic tumors	29/49 cancers
<i>p16</i>	9p21	LOH in sporadic tumors and low expression	1/7 3/7 cancers
<b>Oncogenes</b>			
<i>RET</i>	10q11.2	Somatic mutations in sporadic tumors <i>RET/PTC1</i> rearrangement	1/23 2/23
<i>RAS</i>	1p13.2 ( <i>N-ras</i> ) 12p11.2 ( <i>K-ras</i> ) 11p15.5 ( <i>H-ras</i> )	Somatic mutations in sporadic tumors Overexpression of <i>K-ras</i>	0/33 0/17 3/24 cancers 10/50 6/18
<b>Growth factors</b>			
<i>IGF2</i>	11p15.5	Overexpression in sporadic tumors with duplication of the paternal allele	51/57 cancers 31/161
<i>EGFR</i>	7p12	Overexpression	68/69 cancers 10/23
<i>EGF</i>	4q25	Expression	0/5 cancers 0/26
<i>TGF-<math>\alpha</math></i>	2p13	Overexpression	5/5 cancers
<b>Signal transduction molecules</b>			
<i>GNAS1</i>	20q13.2	Somatic mutations in sporadic tumors	8/55
<i>GNAI2</i>	3p21	Somatic mutations in sporadic tumors	3/47
<i>ACTH-R</i>	18p11.2	Somatic mutations Deletion	0/16 3/45
<i>ATRI</i>	3q21–25	Somatic mutations Moderate to high expression	0/17 0/1 cancer 21/76
<i>PRKAR1A</i>	17q22–24	LOH in tumors associated with Carney complex Increased activity by cAMP stimulation	2/2 3/3
Calcium-dependent protein kinase C	17q22–23 ( $\alpha$ ) 13p21 ( $\delta$ )	Normal activity	17/17

<sup>a</sup>Modified from Koch, Pacak, and Chrousos (2002).

*PRKARIA*, a gene that encodes protein kinase A regulatory subunit 1 $\alpha$ .

### Multiple Endocrine Neoplasia Type I

Approximately 35% of patients with MEN1 have adrenal nodules, most of which are adrenocortical. LOH at 11q13, the *menin* locus, has been identified in some adrenocortical tumors, but the precise pathogenesis of adrenal lesions in patients with MEN1 remains obscure.

### Familial Hyperaldosteronism

Hyperaldosteronism can occur in at least two familial syndromes: familial hyperaldosteronism type 1, caused by the formation of a hybrid gene with fusion of the corticotropin-regulated promoter of the 11 $\beta$ -hydroxylase gene and the angiotensin II-regulated aldosterone synthase gene at 8q24, and familial hyperaldosteronism type 2, which does not respond to dexamethasone administration by decreasing corticotropin (ACTH) secretion and whose responsible gene is linked to chromosomal subband 7p22 but is still unidentified. The molecular pathogenesis of adrenal nodule development in patients with hyperaldosteronism is unknown.

### Congenital Adrenal Hyperplasia

Almost all patients with this autosomal recessive disorder have germ-line mutations in the gene coding for 21-hydroxylase at 6p21.3. This genetic defect leads to impaired cortisol secretion with subsequent elevation of ACTH. Almost half of heterozygous carriers have macronodular adrenal disease, the rationale basis for investigators to search for mutations in the 21-hydroxylase gene in patients with apparently sporadic adrenal tumors. How these adrenal nodules develop remains unknown.

### McCune–Albright Syndrome

This syndrome is a sporadic postzygotic genetic disease that includes growth hormone-secreting pituitary adenomas and nodular adrenocortical disease. Somatic mutations in the  $\alpha$  chain of the stimulatory G protein *GNAS1* at 20q13.2 are responsible for this syndrome and result in stimulation of cAMP. Mutations in the cAMP inhibitory *GNAI2* gene occur in only a few adrenocortical tumors. The precise

mechanisms of adrenal nodule formation are unknown.

### Ectopic G Protein-Coupled Receptors

Aberrant expression of ectopic membrane hormone receptors, including gastric inhibitory polypeptide,  $\beta$ -adrenergic agonists, luteinizing hormone, vasopressin, and interleukin-1, has been found in some adrenal adenomas but not in adrenal cancer.

### Sporadic Tumors

Cell replication errors may lead to numerical changes of chromosomes, chromosomal translocations, amplification and/or loss of genes, somatic sequence alterations in specific genes including DNA repair genes, and other genomic changes. An important step in analyzing adrenal and other tumors is to determine whether they are mono- or polyclonal. The cellular origin of adrenocortical tumors is unknown, although clonal analyses of adrenal cortex lesions have helped better define the nature of these tumors. Usually, the development of cancer and tumors in general is regarded as a multistep process. One tumor-initiating mutation in a single cell may equip the cell with a selective growth advantage, enabling it to become a tumor. This tumor would then be called monoclonal since it derived from a single genetically aberrant cell. In contrast, a polyclonal tumor would develop from a group of aberrant cells arising in parallel. Only one of three studies on the clonal analysis of adrenocortical tumors noted the duration of follow-up for patients with benign adrenal adenomas/hyperplasias, even though this feature is important because differences in the rate of growth may distinguish benign from malignant adrenocortical tumors. It appears that adrenocortical carcinomas are monoclonal, whereas the majority of benign adrenocortical lesions are polyclonal. It can be speculated that tumorigenesis develops from polyclonal adrenocortical cell aggregates, some of which gain a selective growth advantage, giving rise to a monoclonal tumor. However, considering the issues of clonality interpretation, it is unclear whether a clonality assay helps differentiate benign and malignant adrenocortical lesions.

Comparative genomic hybridization (CGH) studies and allele typing using microsatellite markers to detect involvement of genes with tumor suppressor or oncogenic function may help in elucidating the pathogenesis of adrenocortical tumors. Studies on adrenocortical tumors must be considered carefully because

many investigators do not refer to a specific follow-up time; the natural history is important in classifying an adrenocortical tumor as benign or malignant. Sidhu *et al.* reported an equal distribution of chromosomal gains and losses in benign and malignant adrenocortical tumors, although the genetic events in both groups were quite different. Limitations of this study are the classification of benign vs malignant tumors, which was not based on the presence of metastases but rather on modified Weiss histologic criteria, and the variable follow-up period for the group of benign tumors, which was up to 41 months. These investigators analyzed 18 benign and 13 malignant adrenocortical tumors. Benign adrenal masses <5 cm frequently had gains in chromosome 4, whereas malignant (>5 cm) adrenal tumors had gains in chromosomes 5, 12, and 19. Losses in the benign group were limited to chromosome 3q, whereas they occurred at chromosomes 1p, 11, 17p, and 22 in the malignant group. This is in contrast to earlier CGH studies showing gains at chromosome 4 mainly or exclusively in malignant tumors. Sidhu *et al.* proposed activation of a protooncogene on chromosome 4 as an early event in adrenocortical tumorigenesis and that the presence of four or more CGH alterations in one tumor is suggestive of the malignant phenotype.

To elucidate the pathogenesis of adrenocortical tumors, the precise definition of adenoma vs carcinoma is of utmost importance, as is the identification of precursor lesions. Without this critical information, it is difficult to determine the sequence of events from tumor initiation to progression. The assumption that putative oncogenes at regions of chromosomal gain represent the first step in tumor development, and regions of chromosomal loss represent putative tumor suppressor genes allowing subsequent steps of tumor progression, may seem plausible to many investigators but it is speculative.

Candidate genes believed to be involved in tumor formation of sporadic adrenal tumors are those that are known to be associated with hereditary tumor syndromes, such as *p53*, *GNAS1*, *MEN1*, and *IGF2*, and genes and receptors involved in signal transduction systems. Mutations or other genetic alterations in the angiotensin II type 1 receptor and corticotropin receptor gene have rarely been found. Biallelic inactivation of the *MEN1* gene in sporadic adrenocortical tumors has not been identified, although either somatic mutation in *MEN1* or LOH at its locus 11q13 have been reported.

The roles of *IGF2*, *H19*, and *p57/KIP2*, all located at 11p15, in the pathogenesis of adrenocortical tumors were previously mentioned. Overexpression

of *IGF2* is frequently found in malignant sporadic adrenocortical tumors. Genetic alterations in *p57/KIP2* (CDK inhibitor 1C) were sought in adrenocortical tumors because adrenal cancers frequently showed allelic loss at 11p15. However, no somatic mutations were found in this gene in 75 sporadic adrenocortical tumors, but low *p57/KIP2* expression was demonstrated in 3 of 10 adrenal adenomas and 6 of 6 carcinomas. Reduced or absent function of this gene seems to lead to enhanced activity of G1 CDK complexes with possibly subsequent promotion of cell proliferation. Consequently, investigators examined adrenocortical tumors for abnormalities in other inhibitors of CDKs. One such protein, P16, whose gene is located at 9p21, is an inhibitor of the *CDK 2A* gene. One of 7 benign and 3 of 7 malignant adrenocortical tumors showed loss of one *p16* allele and absent p16 protein by immunohistochemistry. P21, a CDK inhibitor that can be induced by p53, was overexpressed in 70% (25/38) of adrenocortical cancer samples. Downregulation of *p21*, however, did not affect the prognosis of patients with adrenal cancer. This suggests a role for these CDK inhibitors in only a small subset of adrenocortical cancers.

In addition to IGF2, epidermal growth factor (EGF) is very important. Its receptor, EGFR, has been studied in a small number of adrenal tumors. By immunohistochemistry, EGFR was overexpressed in benign and malignant adrenocortical tumors, whereas EGF was not detected. Instead, transforming growth factor- $\alpha$  (TGF- $\alpha$ ) was overexpressed in adrenal cancer. TGF- $\alpha$  is a natural ligand for EGFR.

The *p53* gene has been studied in many tumors, including adrenocortical tumors, because of its frequent mutation in cancers and its known role in regulating the cell cycle. *p53* has been classified as a tumor suppressor gene and is mutated in the germ line of most patients with Li-Fraumeni syndrome. Since this syndrome is associated with adrenocortical tumors, it is conceivable that *p53* may also play a role in the pathogenesis of sporadic adrenal tumors. Although allelic loss at 17p13, the locus for the *p53* gene, and somatic *p53* mutations frequently occur in adrenal cancer, they are uncommon in benign adrenocortical tumors, suggesting that genetic alterations in *p53* are involved in tumor progression rather than initiation. Only one study from Taiwan reported *p53* mutations in adrenocortical adenomas, but the classification of benign vs malignant adrenal tumors, including the follow-up period of the affected Taiwanese patients, remains unclear. In contrast, *p53* mutations (predominantly in exons 5–8) were reported in up to 70% of adrenal cancers. However, it is unknown whether

biallelic inactivation of *p53* occurred in these tumors since investigators reported either allelic loss at 17p13 or somatic *p53* mutations and not whether both events (biallelic inactivation of *p53*) took place in the respective tumor. Nevertheless, LOH at 17p13 has been suggested as a molecular prognostic marker for sporadic adrenocortical tumors.

The *ras* gene family encodes G proteins, which are involved in signaling pathways through modification of intracellular cAMP concentrations. Overexpression of *ras* may lead to constitutive signal transduction and/or cell proliferation. *ras* mutations were first identified in 12.5% of adrenocortical tumors with an equal prevalence in benign and malignant tumors. A subsequent study reported somatic *K-ras* mutations and an overexpression of *K-ras* in 33% of benign adrenocortical tumors, whereas somatic *H-ras* mutations were not detected. However, previous studies did not find this prevalence rate of *ras* mutations.

The *RET* protooncogene at chromosome 10q11.2 encodes a tyrosine kinase receptor that is involved in the control of cell differentiation and proliferation. Germ-line mutations in *RET* are found in almost all patients with MEN2. Importantly, *RET* is only expressed in certain tissues, such as the neural crest-derived parafollicular C cells in the thyroid gland and extra- and intraadrenal chromaffin cells. Overrepresentation of mutant *RET* may initiate tumorigenesis of medullary thyroid carcinoma and pheochromocytoma. Although adrenocortical tumors are not part of MEN2, they may have genetic alterations involving *RET*. Analysis of 21 sporadic adrenocortical tumors revealed 1 aldosterone-producing tumor with a point mutation in *RET* and *RET/PTC1* rearrangements in 2 tumors—one cortisol-producing and one aldosterone-producing.

Telomerase is the ribonucleoprotein enzyme that elongates telomeres (i.e., chromosomal DNA ends). In most normal somatic cells, telomerase is repressed, whereas this enzyme is reactivated in transformed cells. Recently, a considerable amount of data have been obtained on telomerase activity in human cancers, including endocrine tumors. In general, the presence or absence of telomerase activity in adrenocortical tumors does not seem to be a prognostic marker, although some studies suggest that malignant adrenal tumors have increased telomerase activity.

## CONCLUSION

Most adrenal tumors are sporadic and associated with numerous somatic genetic alterations. The exact

order and time frame of these genetic events are difficult to determine. Therefore, one cannot make precise statements about the pathogenesis of tumor initiation and tumor progression. Identifying reliable genetic prognostic markers is also a challenge and strongly depends on the follow-up period of patients with adrenal tumors that are classified as benign or malignant. Focusing on adrenal tumors that occur in hereditary syndromes may facilitate the determination of their pathogenesis since in these adrenal masses at least one genetic hit, the inherited gene defect, is known. Subsequent genetic alterations may then be easier to place into context. In MEN2-associated pheochromocytomas, a second hit model leading to overrepresentation of mutant *RET* as a possible tumor-initiating event has been proposed. Somatic mutations in genes that are known to be involved in hereditary syndromes, such as *p53*, *VHL*, and *RET*, are rarely found in sporadic adrenal tumors. Future studies should focus not only on single genes and genetic alterations but also on the interaction of their encoded protein products.

## See Also the Following Articles

Beckwith-Wiedemann Syndrome (BWS) • McCune-Albright Syndrome • Multiple Endocrine Neoplasia (MEN) Type 2 • Neurofibromatosis • Pheochromocytoma • Von Hippel-Lindau Syndrome

## Further Reading

- Brodeur, G. M., Sawada, T., Tsuchida, Y., and Voute, P. A. (eds.) (2000). "Neuroblastoma." Elsevier, Amsterdam.
- Casey, M., Vaughan, C. J., He, J., Hatcher, C. J., Winter, J. M., Weremowicz, S., Montgomery, K., Kucherlapati, R., Morton, C. C., and Basson, C. T. (2000). Mutations in the protein kinase A R1alpha regulatory subunit cause familial cardiac myxomas and Carney complex. *J. Clin. Invest.* **106**, R31–R38.
- Celik, V., Unal, G., Ozgultekin, R., Goksel, S., Unal, H., and Cercel, A. (1996). Adrenal ganglioneuroma. *Br. J. Surg.* **83**, 263.
- Eng, C. (1999). *RET* proto-oncogene in the development of human cancer. *J. Clin. Oncol.* **17**, 380–393.
- Gicquel, C., Bertagna, X., Gaston, V., Coste, J., Louvel, A., Baudin, E., Bertherat, J., Chapuis, Y., Duclos, J. M., Schlumberger, M., Plouin, P. F., Luton, J. P., and Bouc, Y. L. (2001). Molecular markers and long-term recurrences in a large cohort of patients with sporadic adrenocortical tumors. *Cancer Res.* **61**, 6762–6767.
- Gimm, O., Koch C. A., Januszewicz, G., Opocher, G., and Neumann, H. P. H. (2003). The genetic basis of pheochromocytoma. *Frontiers Horm. Res.*, in press.
- Hedeland, H., Ostberg, G., and Hokfelt, B. (1968). On the prevalence of adrenocortical adenomas in an autopsy material in relation to hypertension and diabetes. *Acta Med. Scand.* **194**, 211–214.

- Huang, S. C., Koch, C. A., Vortmeyer, A. O., Pack, S. D., Lichtenauer, U. D., Mannan, P., Lubensky, I. A., Chrousos, G. P., Gagel, R. F., Pacak, K., and Zhuang, Z. (2000). Duplication of the mutant *RET* allele in trisomy 10 or loss of the wild-type allele in multiple endocrine neoplasia type 2-associated pheochromocytoma. *Cancer Res.* **60**, 6223–6226.
- Kirschner, L. S., Carney, J. A., Pack, S. D., Taymans, S. E., Giatzakis, G., Cho, Y. S., Cho-Chung, Y. S., and Stratakis, C. A. (2000). Mutations of the gene encoding the protein kinase A type I-alpha regulatory subunit in patients with the Carney complex. *Nature Genet.* **26**, 89–92.
- Kloos, R. T., Gross, M. D., Francis, I. R., Korobkin, M., and Shapiro, B. (1995). Incidentally discovered adrenal masses. *Endocr. Rev.* **16**, 460–484.
- Koch, C. A., Pacak, K., and Chrousos, G. P. (2002). The molecular pathogenesis of hereditary and sporadic adrenocortical and adrenomedullary tumors. *J. Clin. Endocrinol. Metab.* **87**, 5367–5384.
- Koch, C. A., Vortmeyer, A. O., Zhuang, Z., Brouwers, F. M., and Pacak, K. (2002). New insights into the genetics of familial chromaffin cell tumors. *Ann. N. Y. Acad. Sci.* **970**, 11–28.
- Koch, C. A., Huang, S. C., Azumi, N., Zhuang, Z., Chrousos, G. P., Vortmeyer, A. O., and Pacak, K. (2002). Somatic *VHL* gene deletion and mutation in MEN 2A-associated pheochromocytoma. *Oncogene* **21**, 479–482.
- Koch, C. A., Walther, M. M., and Linehan, W. M. (2002). Von Hippel-Lindau syndrome. In "Adrenal Physiology and Diseases" (G. P. Chrousos, ed.). www.endotext.org/adrenal.
- Koch, C. A., Brouwers, F. M., Rosenblatt, K., Burman, K. D., Davis, M. M., Vortmeyer, A. O., and Pacak, K. (2003). Adrenal ganglioneuroma in a patient presenting with severe hypertension and diarrhea. *Endocr. Rel. Cancer* **10**, 99–107.
- Lack, E. E. (1997). "Tumors of the Adrenal and Extraadrenal Paraganglia." Armed Forces Institute of Pathology, Washington, DC.
- Lacroix, A., Ndiaye, N., Tremblay, J., and Hamet, P. (2001). Ectopic and abnormal hormone receptors in adrenal Cushing's syndrome. *Endocr. Rev.* **22**, 75–110.
- Levy, A. (2001). Monoclonality of endocrine tumors: What does it mean? *Trends Endocrinol. Metab.* **12**, 301–307.
- Neumann, H. P., Bausch, B., McWhinney, S. R., Bender, B. U., Gimm, O., Franke, G., Schipper, J., Klisch, J., Althoefer, C., Zerres, K., Januszewicz, A., and Eng, C. (2002). Germ-line mutations in nonsyndromic pheochromocytoma. *N. Engl. J. Med.* **346**, 1459–1466.
- Schussheim, D. H., Skarulis, M. C., Agarwal, S. K., Simonds, W. F., Burns, A. L., Spiegel, A. M., and Marx, S. J. (2001). Multiple endocrine neoplasia type 1: New clinical and basic findings. *Trends Endocrinol. Metab.* **12**, 173–178.
- Sidhu, S., Marsh, D. J., Theodosopoulos, G., Philips, J., Bambach, C. P., Campbell, P., Magarey, C. J., Russell, C. F. J., Schulte, K. M., Roher, H. D., Delbridge, L., and Robinson, B. G. (2002). Comparative genomic hybridization analysis of adrenocortical tumors. *J. Clin. Endocrinol. Metab.* **87**, 3467–3474.
- Stratakis, C. A., Carney, J. A., Lin, J. P., Papanicolaou, D. A., Karl, M., Kastner, D. L., Pras, E., and Chrousos, G. P. (1996). Carney complex: A familial multiple neoplasia and lentiginosis syndrome: Analysis of 11 kindreds and linkage to the short arm of chromosome 2. *J. Clin. Invest.* **97**, 699–705.
- Stratakis, C. A., Jenkins, R. B., Pras, E., Mitsiadis, C. S., Raff, S. B., Stalboerger, P. G., Tsigos, C., Carney, J. A., and Chrousos, G. P. (1996). Cytogenetic and microsatellite alterations in tumors from patients with the syndrome of myxomas, spotty skin pigmentation, and endocrine overactivity (Carney complex). *J. Clin. Endocrinol. Metab.* **81**, 3607–3614.
- Tischler, A. S. (2000). Divergent differentiation in neuroendocrine tumors of the adrenal gland. *Semin. Diagn. Pathol.* **17**, 120–126.
- Van der Harst, E., de Krijger, R. R., Dinjens, W. N., Weeks, L. E., Bonjer, H. J., Bruining, H. A., Lamberts, S. W., and Koper, J. W. (1998). Germline mutations in the *VHL* gene in patients presenting with pheochromocytomas. *Int. J. Cancer* **77**, 337–340.
- Walther, M. M., Reiter, R., Keiser, H. R., Choyke, P. L., Venzon, D., Hurley, K., Gnarr, J. R., Reynolds, J. C., Glenn, G. M., Zbar, B., and Linehan, W. M. (1999). Clinical and genetic characterization of pheochromocytoma in von Hippel-Lindau families: Comparison with sporadic pheochromocytoma gives insight into natural history of pheochromocytoma. *J. Urol.* **62**, 659–664.
- Walther, M. M., Herring, J., Enquist, E., Keiser, H. R., and Linehan, W. M. (1999). Von Recklinghausen's disease and pheochromocytomas. *J. Urol.* **162**, 1582–1586.
- Weiss, L. M. (1984). Comparative histologic study of 43 metastasizing and nonmetastasizing adrenocortical tumors. *Am. J. Surg. Pathol.* **8**, 163–169.



# Adrenarche, Premature

Catherine Dacou-Voutetakis, Sarantis Livadas, Antony Voutetakis, and Maria Dracopoulou  
Athens University Medical School, Aghia Sophia Children's Hospital, Athens, Greece

## Glossary

**adrenarche** Initiation of function of the reticular zone of the adrenal gland.

**congenital adrenal hyperplasia** A disorder resulting from enzymatic defects in the pathway of cortisol synthesis.

**CYP21** A gene encoding the steroidogenic enzyme 21-hydroxylase.

**gonadarche** Initiation of function of the hypothalamic-pituitary-gonadal axis.

**premature pubarche** The appearance of pubic hair, with or without axillary hair growth or increased apocrine odor, in the absence of other secondary sexual characteristics, occurring prior to age 8 in girls and age 9 in boys.

**P**remature adrenarche refers to the premature activation of the zona reticularis of the adrenal gland, clinically manifested by pubic hair development.

## INTRODUCTION

Although more than 60 years have elapsed since the endocrinologist F. Albright coined the term adrenarche, the mystery of the mechanisms involved in its development has not yet been unraveled. In studying girls with gonadal dysgenesis, a natural experimental model of gonadism, Albright observed that pubic hair can develop under the influence of adrenal androgens and in the absence of gonadal hormones. He was thus the first to distinguish between gonadal and adrenal puberty (gonadarche and adrenarche, respectively). Subsequent studies of other clinical prototypes, including hypogonadotropic hypogonadism and isolated premature thelarche, confirmed his observation. Dissociation between adrenal androgens and cortisol secretion has also been documented by demonstrating that a rise in dehydroepiandrosterone

sulfate (DHEAS) during puberty is not accompanied by a rise in cortisol. Moreover, the rise in adrenocorticotropin hormone (ACTH) and cortisol in Cushing's syndrome occurs without an analogous increase in DHEAS.

## ADRENARCHE

### Definition

Adrenarche refers to the activation of the zona reticularis (ZR) of the adrenal gland, biochemically marked by a rise in dehydroepiandrosterone (DHEA) and DHEAS but not in other androgens. The clinical manifestation of adrenarche is termed pubarche and is characterized by the development of pubic hair with or without axillary hair growth or apocrine odor. The decline in the function of the ZR, later in life, is termed adrenopause.

### Phylogenetic Data on Adrenarche

Adrenarche and adrenopause appear to be phenomena that are unique to the highest order of primates and, therefore, represent a recent evolutionary development. The serum levels of DHEA, DHEAS and  $\Delta_4$ -adrostenedione ( $\Delta_4$ A) remain unchanged during sexual maturation in rats, hamsters, guinea pigs, sheep, pigs, goats, horses, and cows. Modest (twofold) changes in DHEA are observed in rabbits and dogs only after their sexual maturation. In primates, the phenomenon of adrenarche is observed only in the chimpanzee and not in other types (e.g., the rhesus monkey).

### Ontogenetic Data on Adrenarche

During the fifth gestational week, cells from the celomic epithelium migrate to the cephalic part of the mesonephros and form the distinct adrenal primordium. The central part of this primordium consists of large, eosinophilic cells that constitute the so-called

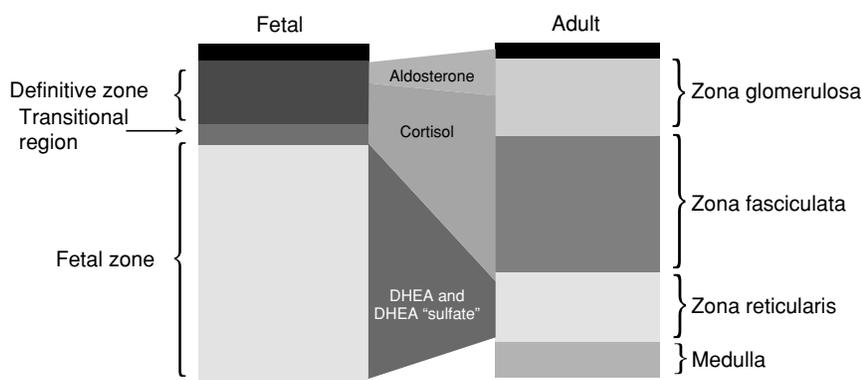
fetal zone, whereas the outer zone consists of densely packed cells that form the definitive zone. By the second trimester of pregnancy, the fetal zone of the human adrenal cortex represents approximately 80–90% of the fetal adrenal. In the fetal zone, CYP11A and CYP17 [but not 3 $\beta$ -hydroxysteroid dehydrogenase (3 $\beta$ -HSD)] and dehydroepiandrosterone sulfotransferase are expressed. The adult ZR, which also lacks 3 $\beta$ -HSD and expresses dehydroepiandrosterone sulfotransferase activity, is considered the equivalent of the fetal adrenal zone (Fig. 1). During the third trimester of pregnancy, the definitive zone of the fetal adrenal forms two, functionally separate units: the outer unit is the zona glomerulosa and the inner unit comprises the zona fasciculata and the ZR (Fig. 1). It must be underscored that the cortisol- and androgen-producing parts of the adrenal and the adrenal medulla are related ontogenetically, anatomically, and functionally, suggesting a possible link between adrenomedullary function and adrenarche.

### Postnatal Development of the Reticular Zone

During the first postnatal months in humans, the fetal zone regresses and almost disappears. Focal islands of the ZR are noted in the adrenals at the age of 3 years and a continuous ZR starts to develop at the age of 6, an age at which adrenal C<sub>19</sub> steroids begin to rise. Adrenarche is initiated at approximately age 6 in girls and approximately age 7 in boys and this is reflected in a gradual rise of the adrenal androgens DHEA and DHEAS. The driving force of these evolutionary events has not been determined.

The serum levels of DHEA and DHEAS start to rise approximately 2 years prior to the gonadal

activation (gonadarche), reaching their peak values at 20–25 years. Thereafter, the serum concentrations of the adrenal androgens gradually decline and in elderly people are 10–20% of those encountered in young adults (adrenopause). The DHEAS reduction in the elderly is coupled with a decrease in the width of the ZR without any impact on the size of the rest of the adrenal cortex. The values of cortisol and aldosterone do not show analogous changes. This observation strongly indicates that the events described are unique to the ZR. Some of the mechanisms involved in these alterations of the ZR are outlined herein. A histological study of adrenal samples from individuals between the ages of 4 months and 56 years showed a decrease in the enzymatic activity of 3 $\beta$ -HSD in the ZR, beginning at the age of 8–13 months with a further progressive decline up to the age of 25–26 years. Longitudinal clinical data obtained from subjects aged 2.9 to 12.3 years have demonstrated a progressive, age-related increase in the DHEAS values (approximately 22% per year) in parallel with an increased activity in 17,20-lyase and a decreased activity in 3 $\beta$ -HSD (an enzymatic profile encountered in both the fetal and the adult reticular zones). These changes are already evident at the pre-adrenarcheal stage of development. It can thus be deduced from this study that adrenarche is not the result of a sudden change in the activity of adrenal enzymes, at a particular period of time; it rather reflects a gradual maturational process that begins in early childhood. When controlled for chronological age, no association between weight, body mass index (BMI), and DHEAS (a marker of adrenarche) was evident. Nevertheless, changes in the nutritional status, measurable by changes in BMI, have been suggested as an important physiological regulator of adrenarche.



**Figure 1** The zones in the fetal and adult adrenal gland. Modified from *Topical Endocrinology* 21, p. 18, with permission of the publisher.

## Origin, Regulation, and Biological Significance of Adrenarche

### **Origin and Regulation**

The extra- or intra-adrenal factors involved in the development and function of the ZR have not yet been fully elucidated. It has long been postulated that a distinct pituitary factor [adrenal androgen-stimulating hormone (AASH)] conducts the development and function of the ZR, but this factor has not been isolated.

Studies of natural human models have provided valid data on the putative developmental and/or functional factors that influence the ZR. Among these models are anencephalic fetuses, patients with congenital adrenal hypoplasia (ACTH receptor defect), combined pituitary hormone deficiency, hypogonadotropic hypogonadism, or isolated premature thelarche, and subjects with precocious gonadarche on gonadotropin-releasing hormone analogue (GnRH $\alpha$ ) suppression. The study of adrenal androgens in these models suggests that DHEA and DHEAS synthesis and secretion are dependent on an intact corticotropin-releasing hormone (CRH)–ACTH axis, not only later in life but also in the latter part of pregnancy. It is most probable that the CRH–ACTH complex exerts a permissive effect and acts in synergy with a putative extra-adrenal factor to successfully orchestrate the development and function of the ZR.

An interesting model in which low DHEAS levels have been detected, despite the normal ACTH–adrenal axis, is the pituitary insufficiency associated with the Prop1 gene defect. The low level of DHEAS in these patients may indicate that the pituitary transcription factor Prop1 is necessary for the normal synthesis of the putative factor (AASH) that initiates adrenarche. Alternatively, the low DHEAS level in patients with the Prop1 gene defect could simply represent an early marker of incipient ACTH insufficiency, which is known to occur later in life in a number of these patients.

In addition to CRH–ACTH and the putative AASH, other extra-adrenal factors have been implicated in the development and function of the ZR: prolactin, estrogens, the epidermal growth factor, angiotensin, gonadotropins, proopiomelanocortin-related peptides, growth hormone (GH), insulin growth factor-I (IGF-I), insulin, and possibly adipose tissue factors. None of these factors, however, have been conclusively shown to regulate androgen secretion by the adrenal gland.

Age-related alterations in the expression of the adrenal enzymes have also been proposed as a mechanism

for the development and function of the ZR (intra-adrenal factors). These changes not only refer to the relative activity of the adrenal enzymes but also to their responsiveness to ACTH. Specifically, the increase in 17-hydroxylase and 17,20-lyase activity occurs along with a decrease in 3 $\beta$ -HSD activity, primarily evident in the developing ZR. It must be stressed, however, that these biochemical changes alone cannot fully explain the initiation of adrenarche.

### **Biological Significance of Adrenarche**

Since a human model of isolated absence of the ZR has not thus far been identified, the exact biological role of the ZR and the implications of its absence or insufficiency still remain enigmatic.

A small, transient increase in growth rate occurring at approximately age 7 (midchildhood growth spurt) has been attributed to the initiation of adrenarche. However, a cause and effect relationship between adrenarche and the midchildhood growth spurt has been disputed. It has also been shown that adrenarche is not a *sine qua non* for gonadarche since gonadal puberty proceeds normally in clinical entities in which adrenarche is absent.

The decline of DHEA coincides with signs of aging and has therefore been interpreted to indicate that aging is, at least in part, a DHEA deficiency syndrome. This observation has prompted studies on the effect of DHEAS replacement in the elderly and in young subjects with DHEAS deficiency of various etiologies, with equivocal results.

## PREMATURE ADRENARCHE–PUBARCHE

### **Definitions**

Premature adrenarche (PA) and premature pubarche (PP) are frequently used interchangeably. Nevertheless, they are not synonymous. PA refers to premature activation of the ZR of the adrenals and is marked by levels of DHEA and DHEAS that are high for the chronological age (CA) but appropriate for the stage of pubic hair development. PP is the term applied to characterize the clinical expression of PA, namely, the appearance of pubic hair, usually at the labiae, with or without axillary hair growth or increased apocrine odor, in the absence of other secondary sexual characteristics, prior to age 8 in girls and 9 in boys. This age cut-off point has been called into question but is still accepted. PP occurs more frequently in girls than in boys, with a male to female ratio of 1/5 to 1/10, for no apparent reason.

The term “exaggerated adrenarche” has been coined to describe a form of PP in which androgen levels, basal or post-ACTH stimulation, are above those expected for the stage of pubic hair development.

### Etiology of Premature Pubarche

Premature pubarche may be caused by: (1) premature activation of the ZR without any apparent pathological condition (idiopathic), (2) congenital adrenal hyperplasia (CAH), (3) virilizing adrenal or ovarian tumor, and (4) increased end-organ sensitivity to androgens.

In idiopathic PP, increased BMI or a sudden rise in BMI may constitute a trigger factor for its induction. Nevertheless, BMI and leptin levels only partially explain the increased DHEAS values. The GH-IGF-I axis and especially hyperinsulinism, as a consequence of insulin resistance, have been implicated in androgen production by the ZR and generally in the mechanism of PP initiation.

The most frequent pathological condition underlying PP, and the one usually creating diagnostic dilemmas, is defective adrenal steroidogenesis and, in particular, nonclassical CAH (NC CAH). This aspect of PP has been quite controversial. Based on the determination of basal and ACTH-stimulated adrenal androgen levels, defective steroidogenesis, indicative of the NC CAH (caused by 21-hydroxylase or 3 $\beta$ -HSD deficiency), ranges from 0 to 54% in the various published series. These huge differences are probably due to the variety of criteria used for recruitment and evaluation of the subjects participating and, most important, in the lack of confirmation by molecular analysis. In a study by Dacou-Voutetakis and Dracopoulou, 48 consecutive cases of PP were evaluated by molecular analysis of the CYP21 gene as well as by basal and ACTH-stimulated 17-OH progesterone (17OHP) values. A significantly increased incidence of mutations in the CYP21 gene was detected in PP children in comparison to the general population (heterozygous 37.5% and homozygous 8.3%). The 17OHP values on the ACTH test showed an overlap between carriers and noncarriers. The application of the receiver operating curve (ROC) curve in this study showed that the sum of the basal plus 60 min value of 17OHP was the best indicator of heterozygosity. For children over the cut-off point of 5 ng/ml (15 nmol/liter), there is a 76.5% certainty of heterozygosity for a CYP21 mutation. By using the nomogram proposed by New *et al.*, the heterozygote's values fell in the expected area, but this was also the case for the majority of normal values. A value of 17OHP 60 min post-ACTH stimulation equal to or greater than 10 ng/ml

(30 nmol/liter) was indicative of a homozygous mutation and occurred in 8.3% of the cases. The latter finding is accepted by most investigators. An increased incidence of CYP21 heterozygosity in either PP or functional hyperandrogenism in adolescents has also been reported by Witcell *et al.* and has been postulated by Knorr *et al.*, based on hormonal evaluation. Contrary to these findings, Potau *et al.* did not find higher incidence of CYP21 carriers in Spanish subjects with a history of PP. It is not possible to determine whether or not PP CYP21 heterozygote subjects have a higher probability of manifesting hyperandrogenism and the clinical syndrome related to this entity than PP girls who are not CYP21 carriers. Only a long-term follow-up study of such cases will provide a definitive answer to these important questions. Mutations in the 3 $\beta$ -HSD gene have also been detected in girls with PP, though infrequently. The new hormonal criteria for the diagnosis of 3 $\beta$ HSD deficiency in children with PP are as follows: (1) baseline 17OH pregnenolone (17P) and 17P/cortisol ratio >29 nmol/liter and >103, respectively, and (2) ACTH-stimulated 17P and 17P/cortisol ratio >294 nmol/liter and >363, respectively.

Androgen-producing tumors of the adrenals or ovaries are rarely a cause of PP, but they should be considered in the differential diagnosis.

Finally, in some children with PP, no androgen excess for either CA or pubertal stage is detected and the pubic hair growth is attributed to increased end-organ sensitivity to androgens.

### Long-Term Consequences of Premature Adrenarche–Pubarche

#### **Growth and Pubertal Development**

At presentation, children with PP show an acceleration of linear growth and skeletal maturation. In some children, the difference between their bone age (BA) and CA ( $\Delta$ BA–CA) can be up to 2 years, whereas in others the BA is comparable to CA. Long-term follow-up of children with PP has not shown impaired growth potentials: the final height (FH) attained is comparable to target height (TH). It seems that the linear growth pattern in children with PP is modified in that height velocity in the period preceding PP diagnosis is higher than that of controls, peak height velocity occurs at an earlier age, and growth during puberty is compromised. Despite reports indicating good FH prognosis, the physician must individualize the approach for each child with PP and closely watch patients with so-called “exaggerated adrenarche” and/or those who have BA advancement greater than 1

year, for the possibility of an unfavorable effect of PP on growth potentials, namely, a predicted height below the TH. The age of gonadarche and menarche in children with PP is not different than in the general population.

#### **Functional Ovarian Hyperandrogenism and Polycystic Ovary Syndrome**

In a percentage of girls with PP, functional ovarian hyperandrogenism (FOH) (excessive response of androgens to GnRH $\alpha$ ), along with an increased incidence of anovulation, hirsutism, and menstrual irregularities [characteristics of polycystic ovary syndrome (PCOS)], has been reported to occur already in late adolescence. It seems that girls with an exaggerated response of 17OHP to ACTH at the time of PP diagnosis are more likely to develop FOH and/or PCOS.

#### **Insulin Resistance, Hyperinsulinism, and the Metabolic Syndrome**

Acanthosis nigricans, decreased insulin sensitivity and consequent hyperinsulinism, and dyslipidemia have been found in a number of girls with PA. Pertinent data in boys are inconsistent. In certain studies, the hyperinsulinemia is already evident at the prepubertal stage, possibly conferring a higher risk for later development of diabetes mellitus type 2 (DM2). Contrary to the above findings, other investigators found no differences in glucose tolerance, insulin resistance indices, or lipid values between PP girls and control girls studied 5 years after menarche. It is quite possible that the stated differences reflect differences in the population (ethnic) groups studied and this must be seriously considered in the study and follow-up of PP girls. The contribution of intrauterine growth retardation to the occurrence of later adverse consequences, and especially FOH in girls with PP, is highly controversial and most authors have not found such an association.

#### **Does PA Represent a Pathologic Condition?**

The question of whether or not PP is a benign condition, simply reflecting premature activation of the ZR without long-term adverse consequences, cannot be convincingly answered yet. A number of girls with PP later present insulin resistance and hyperinsulinism, low sex hormone-binding globulin (SHBG), FOH, increased hirsutism, acne, or polycystic ovary-like syndrome with possible later occurrence of DM2 or the metabolic syndrome. Pertinent data in boys are not consistent.

No specific features at PP diagnosis have yet been identified to predict the child at risk of developing these pathological entities. Nevertheless, it seems that some of the late consequences are, to a large extent, restricted to certain population (ethnic) groups. In addition to ethnic group, other factors, such as obesity and a family history of DM2, might contribute to the occurrence of adverse consequences. Exaggerated adrenarche as well as low SHBG (an index of insulin resistance and free androgen levels) and glucose/insulin ratio  $<4.5$  at the time of PP diagnosis might confer a higher risk for the later development of FOH and/or the metabolic syndrome in children with PP.

#### **Evaluation, Management, and Follow-up of the Child with PP**

##### **Clinical Evaluation**

The clinical evaluation of the child with PP at presentation should include measurements of height, weight, waist circumference, BMI estimation, BA determination, and recording of relevant family history. Predicted height, TH, and the stage of pubertal development should also be assessed. The presence of acne, hirsutism, or virilization (clitoral enlargement) should be noted. If  $\Delta\text{BA}-\text{CA}$  is less than 1 year and the predicted height falls within the height range expected from parental height (TH), the possibility of isolated premature activation of the ZR is quite high (Fig. 2).

##### **Hormonal Determinations**

If basal values of 17OHP,  $\Delta_4$ -androstenedione, testosterone, and DHEAS taken at 8 AM fall into the range expected for the pubic hair developmental stage, the diagnosis of idiopathic PP is fairly well confirmed. If basal adrenal androgen levels are higher than the levels for the corresponding pubertal stage and are accompanied by advanced BA ( $\Delta\text{BA}-\text{CA} > 1$ ), the main possibilities are as follows: (1) exaggerated adrenarche, (2) defective adrenal steroidogenesis, or (3) virilizing tumor (adrenal or ovarian). In such cases, an intravenous Synacthen test (ACTH) is carried out (250  $\mu\text{g}$ , 150  $\mu\text{g}/\text{m}^2$ , or 10  $\mu\text{g}/\text{m}^2$ , the last dose being infrequently used). Blood samples, primarily for 17OHP determination, are obtained at 0 and 60 min. A 60 min value of 17OHP  $>10$  ng/ml strongly suggests a homozygous mutation of the CYP21 gene (NC form) and CYP21 gene analysis should be considered (Fig. 3). There is a 75% probability of existence of heterozygosity if the sum value of 0 and 60 min post-Synacthen is  $>5$  ng/ml (15 nmol/liter). Obviously, the latter information is of no immediate practical significance and may be of value only for genetic

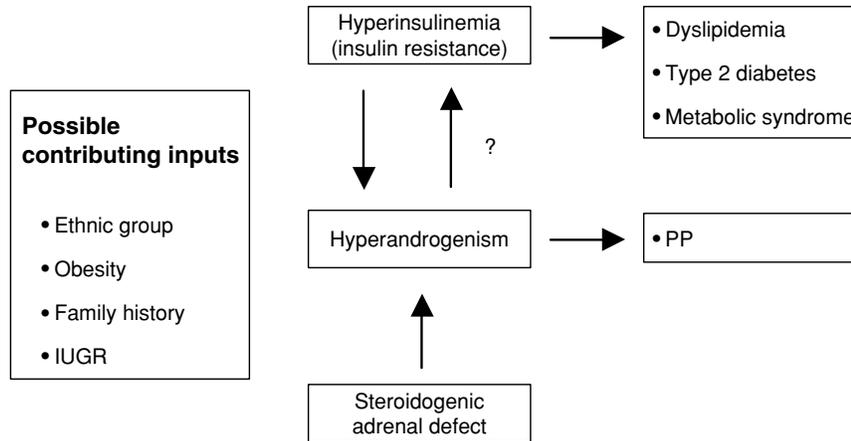


Figure 2 Possible pathogenetic mechanisms of premature pubarche (PP). IUGR, intrauterine growth retardation.

counseling. Determination of glucose/insulin value, SHBG, and ovarian sonography might constitute a good baseline record but should not be regarded as necessary diagnostic tools. It must be stated that an increased prevalence of sonographic findings of PCOS has been reported in girls with PP at the prepubertal stage, but the diagnostic and prognostic value of such a finding remains unclear. The dexamethasone suppression test or some form of imaging

study is rarely indicated for the remote possibility of a virilizing tumor. This latter possibility is much higher if pubarche occurs very early and there is an increased level of serum testosterone.

**Management and Follow-up**

The therapeutic approach for children with PP will be determined by the pathogenetic mechanism involved and the clinical findings. In most cases, however, drug

**Clinical and laboratory data**

**Diagnostic possibilities**

$\Delta BA-CA < 1$  year  
 Predicted height within target → High probability of idiopathic PP

**DHEAS,  $\Delta 4A$ , testosterone, 17OHP: Basal values**

• Within normal limits for CA → Increased end-organ sensitivity  
 • Within the range of pubertal stage → Idiopathic PP

• Above the values for pubertal stage and  $\Delta BA-CA > 1$

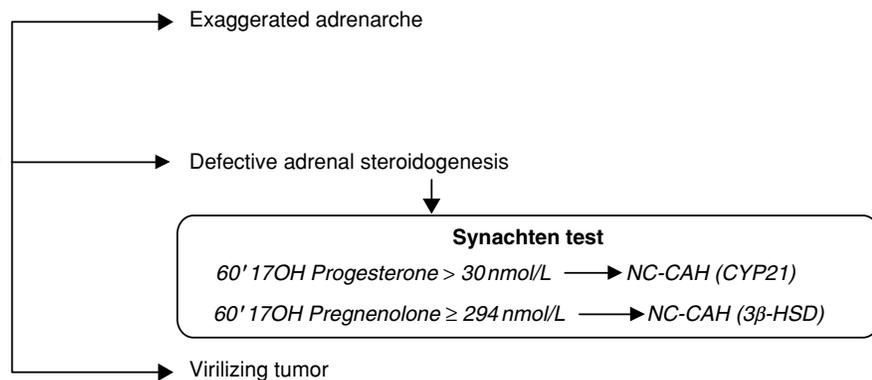


Figure 3 Premature pubarche (pp): diagnostic steps.

intervention is not required. A follow-up of children with PP must be assured, especially during puberty and the immediate postpubertal years. In general, both the child and the parents should be made aware of the good prognosis of most patients. Instructions concerning dietary habits and exercise should be given in cases with borderline or elevated BMI or a positive family history of DM2.

The physician should be aware of the possible long-term consequences of PP and should individualize the care and follow-up, keeping in mind that PP, in certain cases, might not represent an innocent temporal deviation of sexual maturation but, possibly, an early manifestation of a complex metabolic abnormality. Fasting glucose/insulin ratio <4.5, low SHBG values, and exaggerated adrenarche might constitute early markers conferring an increased risk for later development of FOH, PCOS, DM2, or the metabolic syndrome. However, it must be underscored that such recommendations are not strictly evidence based; rather, they constitute extrapolation from still controversial published data. Moreover, data related to long-term consequences seem to apply to certain population (ethnic) groups. The physician in charge, being aware of the trends concerning prognosis of PP, will determine the diagnostic studies necessary at presentation of the patient, as well as the frequency and duration of the follow-up visits.

### See Also the Following Articles

Adrenal Androgens • Congenital Adrenal Hyperplasia, Prenatal Diagnosis and Therapy • Congenital Adrenal Hypoplasia Syndromes • Delayed Puberty and Hypogonadism,

Female • Hypergonadotropic Hypogonadism • Precocious Puberty, Central (Female) • Steroid Metabolism and the Metabolic Syndrome

### Further Reading

- Dacou-Voutetakis, C., and Dracopoulou, M. (1999). High incidence of molecular defects of the CYP21 gene in patients with premature adrenarche. *J. Clin. Endocrinol. Metab.* **84**, 1570–1574.
- Ibanez, L., Dimartino-Nardi, J., and Saenger, P. (2000). Premature adrenarche—Normal variant or forerunner of adult disease? *Endocr. Rev.* **21**, 671–696.
- Likitmaskul, S., Cowell, C. T., and Silink, M. (1995). ‘Exaggerated adrenarche’ in children presenting with premature adrenarche. *Clin. Endocrinol.* **42**, 265–272.
- Lutfallah, C., Wang, W., and Pang, S. (2002). Newly proposed hormonal criteria via genotypic proof for type II 3 $\beta$ -hydroxysteroid dehydrogenase deficiency. *J. Clin. Endocrinol. Metab.* **87**, 2611–2622.
- Meas, T., Chevenne, D., and Levy-Marchal, C. (2002). Endocrine consequences of premature pubarche in post-pubertal Caucasian girls. *Clin. Endocrinol.* **57**, 101–106.
- Miller, W. L. (1999). The molecular basis of premature adrenarche: An hypothesis. *Acta Paediatr. Suppl.* **88**, 60–66.
- Palmert, M. R., Hayden, D. L., and Boepple, P. A. (2001). The longitudinal study of adrenal maturation during gonadal suppression: Evidence that adrenarche is a gradual process. *J. Clin. Endocrinol. Metab.* **86**, 4536–4542.
- Parker, C. R., Jr. (1999). Dehydroepiandrosterone and dehydroepiandrosterone sulfate production in the human adrenal during development and aging. *Steroids* **64**, 640–647.
- Voutetakis, A., Livadas, S., and Dacou-Voutetakis, C. (2001). Insufficient adrenarche in patients with combined pituitary hormone deficiency caused by a PROP-1 gene defect. *J. Pediatr. Endocrinol. Metab.* **14**, 1107–1111.
- Witchel, S. F., Lee, P. A., Suda-Hartman, M., and Hoffman, E. P. (1997). Hyperandrogenism and manifesting heterozygotes for 21-hydroxylase deficiency. *Biochem. Mol. Med.* **62**, 151–158.



# Adrenergic Mechanisms

José Marino-Neto, Marcelo Sabi, and Marta Aparecida Paschoalini  
*Federal University of Santa Catarina, Florianópolis, Brazil*

## Glossary

**energy balance** A dynamic matching of bodily demands and supplies of metabolic fuels (e.g., glucose [GLU], free fatty acids [FFAs]); this balance is accomplished by a variety of internally and externally generated signals processed by numerous, redundant, and widely distributed central and peripheral circuits controlling feeding behavior, cellular energy metabolism, and the stored and circulating GLU/FFA levels.

**glucoprivic state** A state of diminished intracellular metabolism of glucose [GLU] that may be produced by (1) a reduction in circulating GLU levels, (2) reduced access of circulating GLU to the intracellular metabolic machinery, or (3) agents that compete with GLU metabolic pathways (thereby inhibiting cellular GLU use) such as 2-deoxy-glucose (2-DG) and 5-thio-D-glucose (5-TG); systemic or central injections of these agents produce glucoregulatory responses (hyperglycemia and feeding) in mammals.

Adrenergic systems (ASs) are neural and endocrine circuits that use epinephrine (adrenaline [EPI]) and/or norepinephrine (noradrenaline [NE]) as neurotransmitters. The ASs include (1) EPI- and NE-producing neurons in the central nervous system (CNS-AS), (2) NEergic postganglionic sympathetic neurons, and (3) EPI- and NE-producing cells of the adrenal medulla. These systems play critical roles in mechanisms coordinating homeostatic responses to environmentally and systemically generated challenges.

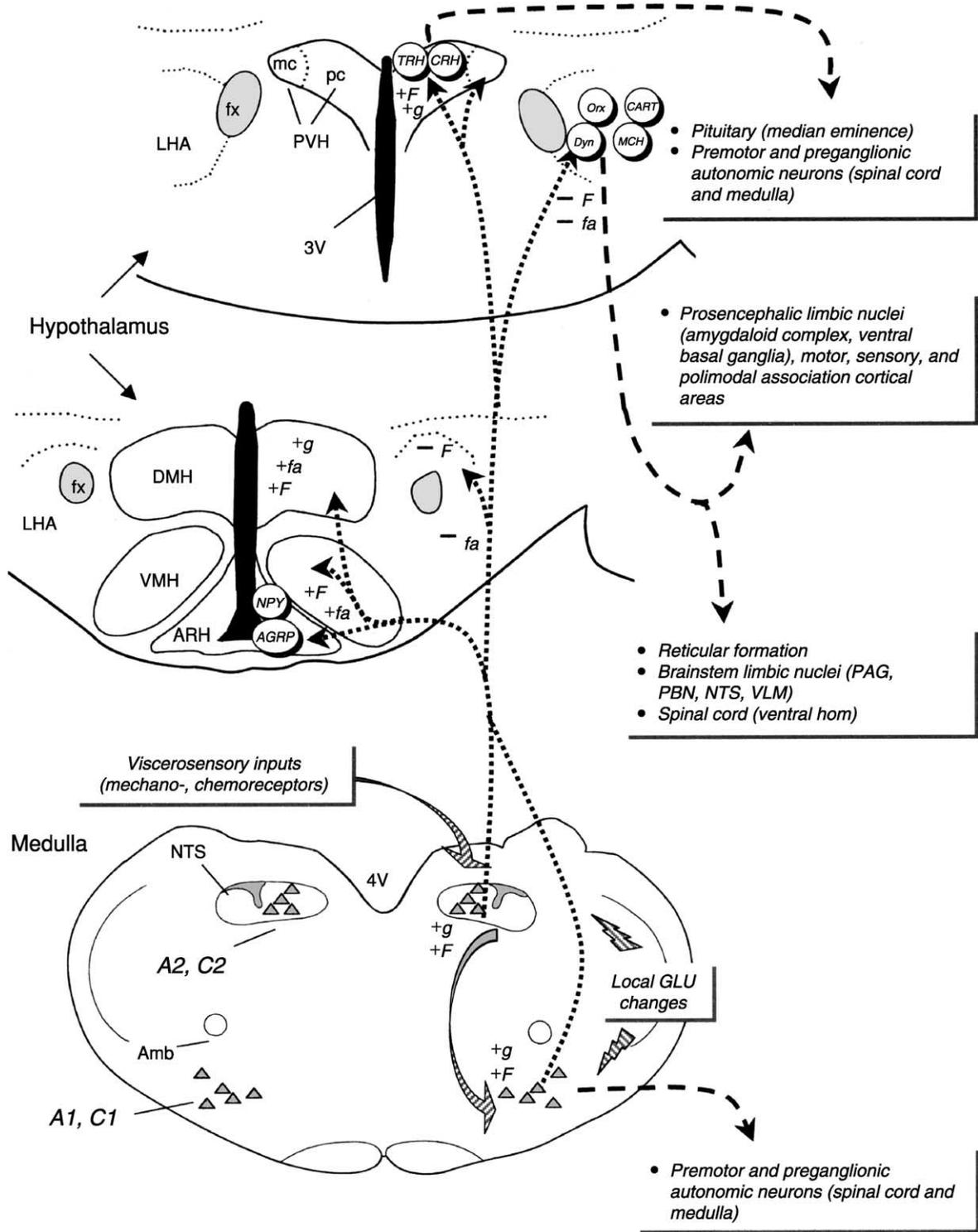
## INTRODUCTION

The catecholamines norepinephrine (noradrenaline [NE]) and epinephrine (adrenaline [EPI]) have been identified as neurotransmitters crucially involved in the control of cognitive, emotional, arousal, and memory processes. In addition, they have a role in reproductive, aggressive, and ingestive activities; in pain regulation; and in thermal, energetic, cardiovascular, and osmoregulatory functions. The

postganglionic sympathetic neurons and adrenal adrenergic systems (ASs) constitute the major effector mechanisms that elicit rapid visceral/metabolic adjustments in response to homeostatic challenges, originated internally or externally, that prepare the organism to engage in energetically demanding behaviors. Central nervous system (CNS)-ASs are crucially involved in the integration of viscerosensory information and in the coordination of neuroendocrine, autonomic, and behavioral responses to these challenges. This article concentrates on some of the neuroanatomical and neurophysiological information relevant to the role of CNS-ASs in mechanisms controlling energy balance.

## CENTRAL ADRENERGIC PATHWAYS: NEUROANATOMICAL OUTLINE

Simultaneous immunoreactivity for tyrosine hydroxylase (TH), the first and rate-limiting enzyme in the catecholamine synthetic pathway, and for phenylethanolamine-N-methyltransferase (PNMT), the enzyme that converts NE into EPI, has been used to identify EPI-synthesizing cells, whereas the absence of PNMT in cells that express TH and dopamine  $\beta$ -hydroxylase (DBH, which converts dopamine into NE) identifies NE-synthesizing neurons. During the past two decades or so, P. E. Sawchenko and colleagues have shown that important adrenergic influences on neuroendocrine/autonomic functions are originated in (1) brainstem NE-synthesizing cell groups referred to as A1, A2, and A5, located in the caudal ventrolateral medulla (cVLM), the nucleus of the solitary tract (NTS), and the ventrolateral pons, respectively, and (2) in EPI-synthesizing neurons that are found in the C1, C2, and C3 groups, located in the rostral VLM (rVLM), the NTS, and the rostral dorsomedial medulla, respectively. The A2/C2 neurons are innervated by primary viscerosensitive afferents carried by the vagus and the glossopharyngeal nerves that convey information from mechano- and chemoreceptors



**Figure 1** Schematic representation of the adrenergic mechanisms and circuitry involved in energy balance control systems. In the hypothalamus, a plus sign (+) indicates sites where NE and/or EPI infusions evoke increased feeding (F), hyperglycemia (g), or increased plasma FFA (fa) levels, whereas a minus sign (-) indicates sites where AS activation produces decreased feeding or FFA levels. In the medulla, the same symbols indicate that injections of the glucoprivic agent 5-thio-D-glucose into A1/C1 and A2/C2 evoke feeding and hyperglycemia. 3V and 4V, third and fourth ventricles, respectively; Amb, nucleus ambiguus; fx, fornix; pc and mc, parvo- and magnocellular cell groups of the PVH, respectively; PAG, periaqueductal gray matter; PBN, parabrachial nuclei. See text for further explanation and other abbreviations.

related to cardiovascular, pulmonary, and gastrointestinal systems (Fig. 1). NTS neurons receiving visceral afferents in turn project to A1 and C1 neurons. Descending projections from C1 cells to the spinal cord innervate sympathetic preganglionic neurons and so can directly activate catecholaminergically mediated responses to visceral changes.

Ascending projections from A1/C1 and A2/C2 cells travel through the ventral adrenergic bundle and terminate in several limbic nuclei and hypothalamic areas. AS terminals are found in the ventromedial (VMH), arcuate (ARH), dorsomedial (DMH), and lateral (LHA) hypothalamic nuclei, and a particularly dense innervation reaches the paraventricular nucleus (PVH). Neurons in these nuclei (particularly in the PVH and LHA) are in turn connected to (1) endocrine (pituitary) mechanisms (e.g., driving pituitary–adrenal and pituitary–thyroid activities), (2) mesencephalic and prosencephalic limbic nuclei (thereby influencing the expression of behavioral, learning/memory, and emotional responses), (3) premotor and preganglionic autonomic neurons (thereby controlling autonomic outflow), and (4) A1/C1/A2/C2 neurons (where they can modulate visceral reflexes mediated by these cells). These relationships are summarized in Fig. 1.

Other hypothalamic hypophysiotropic (somatostatin- and luteinizing hormone-releasing hormone [LHRH]-synthesizing) and neurosecretory (arginine vasopressin [AVP]- and oxytocin [OT]-containing) neurons are also innervated by EPI/NE terminals. These terminals provide the substrate for relevant adrenergic control of pituitary secretion of growth hormone/somatomedins (deeply affecting the balance between energy mobilization and storage in peripheral tissues in addition to the thyroid-stimulating hormone [TSH]- and adrenocorticotrophic hormone [ACTH]-related mechanisms mentioned previously) as well as of the plasmatic levels of luteinizing hormone/follicle-stimulating hormone (LH/FSH), AVP, and OT. Also important to point out is that although terminals from A6 (the locus ceruleus) NE-synthesizing neurons are ubiquitous in the CNS, these neurons have limited direct input to the hypothalamic circuitry. The locus ceruleus receives a major, essentially inhibitory innervation from EPI- and NE-producing medullary neurons, through which the latter ASs gain access to a “brain-wide web” of NE-mediated modulation of cognitive and arousal functions as well as to indirect influences on neuroendocrine mechanisms through prosencephalic limbic nuclei.

## CENTRAL ADRENERGIC SYSTEM MECHANISMS CONTROLLING FEEDING AND FUEL HOMEOSTASIS

### Brainstem EPI- and NE-Synthesizing Cells Manifest Glucoreceptive Mechanisms

Sue Ritter, Kadhija Yettefti, and their colleagues have independently gathered convincing evidence for glucoreceptive activity in medullary EPI/NE cells. Single-unit recording studies have identified neurons sensitive to peripheral or local changes of glucose (GLU) levels in A1/C1 and A2/C2 cell groups, whose responses are depressed by locally applied clonidine (an  $\alpha_2$ -adrenoceptor agonist). Both feeding and blood GLU increases were observed after local GLU deficits evoked by injections of 5-thio-D-glucose (5-TG), a glucoprivic agent, into the C1 and C2 areas. Systemic injections of 2-deoxy-glucose (2-DG), another glucoprivic agent, provoked feeding, hyperglycemia, and Fos expression predominantly in C1 and C2 (but also in A1 and A2) cells. Thus, these neurons are in a position to mediate metabolic and behavioral responses to central or systemic changes in GLU levels. A1/C1 and A2/C2 cells are also activated by other stressful situations such as hemorrhage, immune challenge, noise, restraint, and forced swim. In each case, activation of medullary ASs occurs either directly through visceral afferents or subsequent to higher level processing of the stressful stimulus and may recruit different, possibly stressor-specific A1/C1 and A2/C2 cell subpopulations. Thus, these circuits may be of importance under different conditions of stress and energy demand.

### NE and EPI Influences on Hypothalamic Circuits Regulating Fuel Homeostasis

#### **Feeding Behavior**

After the pioneering work by S. P. Grossman during the early 1960s, and thanks mainly to the systematic efforts of Sara Leibowitz, Paul Wellman, and their colleagues over the past 20 years or so, CNS-AS inputs to the hypothalamus were established as crucial to the control of feeding behavior and fuel metabolism. The PVH is by far the most sensitive (and the most studied) area with regard to feeding-related AS activity. Injections of NE or EPI, as well as injections of NE reuptake blockers (which increase local extracellular NE levels), into the PVH promptly elicit feeding. These effects are mediated by  $\alpha_2$ -adrenoceptors,

whereas activation of  $\alpha_1$  receptors suppresses food intake. NE release within the PVH reaches a maximum immediately before the active (dark) period and is positively correlated with the intense feeding that occurs at this time. The reactivity to exogenous NE and the number of  $\alpha_2$ -adrenoceptors in the PVH also increase at the beginning of the dark period. Furthermore, extracellular NE increases into the PVH were positively correlated with meal size and with the carbohydrate content of the meal. Conversely, decreases in NE release into the medial hypothalamus were observed after intraduodenal nutrient infusions. Thus, it is possible that endogenous NE acts in an  $\alpha_2$ -receptor-enriched PVH mainly to increase carbohydrate intake and restore bodily reserves of this macronutrient.

Ritter and her colleagues suggested that these medullary ASs may represent the main origin of NE/EPI inputs to the PVH that provoke feeding in response to glucoprivic situations. They showed that injections into the PVH of the toxin saporin conjugated to an antibody against DBH, DSAP (which is retrogradely transported to cell bodies and selectively destroys NE/EPI terminals and perikaria), dramatically reduced the number of TH-containing neurons in the A1, C1, A2, C2, and C3 groups and abolished 2-DG-induced feeding and Fos expression in the PVH. Neurons of the A1 group project mainly to hypothalamic magnocellular neurosecretory cell groups (which produce AVP and OT) either within or outside the PVH, whereas C1, C2, and A2 neurons project densely to parvocellular PVH neurons (whose terminals release corticotropin-releasing hormone [CRH] and thyrotropin-releasing hormone [TRH] into the median eminence and into premotor and preganglionic autonomic neurons). Thus, this circuit is in a position to regulate autonomic outflow as well as neuro- and adrenergic activities related to energy balance (as well as to other functional domains).

NE injections into the VMH acutely induce hyperphagia, and chronic infusions of NE into this nucleus provoke weight gain associated with a GLU-intolerant, diabetic-like state in normal rats. Furthermore, increases in NE levels and/or activity have been observed in this nucleus in a variety of animal models of insulin-resistant obesity. Extracellular NE increases into the VMH were observed after systemic 2-DG injections, suggesting a role for NE-mediated VMH circuits in behavioral and humoral responses to glucoprivic situations. Grossman's seminal study pointed to the DMH as a site where NE injections also increase food intake. Push-pull cannula studies revealed

substantial increases in the DMH content of EPI (and, to a lesser extent, of NE) during intense feeding episodes but not during other maintenance behaviors. NE/EPI neurons densely innervate the ARH, which is populated by neurons producing the orexigenic agents agouti gene-related protein (AGRP) and neuropeptide Y (NPY) and which is rich in leptin receptors. Systemic 2-DG injections induce activation of these AGRP and NPY neurons of the ARH, and 2-DG-induced increases in AGRP expression are inhibited by DSAP lesions in this nucleus, pointing to an important role for ASs in the control of feeding-related functions of the ARH.

Contrary to what is observed in medial hypothalamic nuclei, CNS-AS inputs to the LHA have hypophagic effects. NE and EPI injected into the perifornical nucleus (PF) of the LHA inhibit food intake, a response that is mediated by  $\beta_2$ -adrenoceptors. On the other hand, activation of  $\alpha$ -adrenoceptors in this area enhances feeding responses. The amphetamine-induced hypophagic effect is strongly associated with increases in extracellular levels of NE in the PF. Increases in NE release in the LHA appear to be associated with satiation signals brought about by a meal. NE release in the LHA increase after intraduodenal nutrient infusions, whereas food deprivation and systemic injections of 2-DG evoke a decrease in extracellular levels of NE. It is important to note that many LHA neurons have shown to express peptide systems involved in feeding behavior, including the melanin-concentrating hormone (MCH), the cocaine-amphetamine-regulated transcript (CART), dynorphin (Dyn), and orexins (Orx). These LHA neurons give rise to afferents to limbic, basal ganglia, and cortical structures and represent major output mechanisms by which AS-mediated inputs are translated into feeding/metabolic responses in addition to the pituitary- and autonomic-directed PVH outputs (Fig. 1).

### **Blood Glucose and Free Fatty Acids**

Data obtained by A. J. Scheurink and colleagues, as well as by Leibowitz's group, suggest an important role for CNS-AS-mediated circuits in the hypothalamic integration of blood levels of metabolic fuels to feeding responses. EPI or NE administration into the PVH (as well as into the DMH and preoptic area) has been shown to induce a potent hyperglycemic response in resting rats. Furthermore, injections of an  $\alpha$ -adrenoceptor blocker in the PVH and VMH suppress the exercise-induced increase in blood GLU levels. Physiological evidence for a catecholaminergic mechanism in the hypothalamic circuitry controlling

lipolysis in rats has also been described. Intracerebroventricular (ICV) administration of NE, as well as infusion of EPI or NE into the VMH or DMH, evokes an increase in plasma free fatty acid (FFA) concentration. In contrast, infusion of EPI or NE into the LHA decreased or left unchanged plasma FFAs, respectively. The increase in plasma FFAs evoked by swimming is potentiated by injections of an  $\alpha$ -adrenoceptor antagonist (phentolamine) into the VMH, whereas the infusion of a  $\beta$ -adrenoceptor antagonist (timolol) provoked a delay in this response. These data suggest the existence of an extensive hypothalamic catecholaminergic circuitry involved in an integrated control of fuel homeostasis.

However, AS-originated inputs to PVH might not be essential for 2-DG-induced hyperglycemia. Injections of DSAP into the PVH, which abolished 2-DG-induced feeding, failed to affect 2-DG-evoked hyperglycemia. DSAP injections into the spinal cord abolished 2-DG-induced hyperglycemia (but not feeding responses) and also deleted TH-containing cells in the A5, A6, A7, C1, and C3 neurons, indicating that glucoprivic-induced feeding and the glycemic response are mediated by different subpopulations of hindbrain catecholamine neurons. An important participation of ASs in VMH mechanisms of fuel control has been shown by A. Cincotta and colleagues. Chronic infusions of NE in the VMH increased plasma insulin, glucagon, leptin, triglyceride, abdominal fat, and lipogenic/lipolytic activities and also evoked insulin resistance and glucose intolerance. Furthermore, extracellular concentrations of NE in the VMH increased during insulin-induced hypoglycemia.

### Are There Functional Central Pathways That Use EPI as a Neurotransmitter?

The studies on central AS roles in neuroendocrine/autonomic mechanisms have focused on NE-mediated pathways rather than EPI-mediated pathways. Such a disregard for EPI central mechanisms may be related to the intimate neurochemical, neuroanatomical, and pharmacological correspondence between these catecholamines. EPI and NE systems (1) share most of their synthetic pathways, pre- and postsynaptic fates, and receptor mechanisms, (2) show considerable anatomical overlap in the brainstem and hypothalamus, and (3) have their central actions poorly differentiated by the available pharmacological tools. Furthermore, the presence of PNMT in a

neuron does not necessarily mean that it uses EPI as neurotransmitter. For example, A. F. Sved found very low EPI levels in the thoracic cord, which is densely innervated by TH-, DBH- and PNMT-containing C1 neurons. However, C. Routledge and C. A. Marsden showed that stimulation of the C1 region increases extracellular EPI (but not NE, as observed by intracerebral microdialysis with high-performance liquid chromatography [HPLC] detection) in the posterior hypothalamus of adrenalectomized animals. Leibowitz and colleagues showed that, at maximally effective doses, the hyperphagic and hyperglycemic effects of EPI injections into the PVH are twice those of NE, and EPI is even more potent than NPY in eliciting an increase in food intake. Ritter and colleagues demonstrated that 2-DG-evoked Fos expression affects predominantly EPI neurons. Although these data support the existence of functional pathways innervating the hypothalamus that use EPI as a neurotransmitter, the functional attributes of AS actions in central circuits actually attributable to EPI or NE systems are poorly understood.

### A Comparative Note

A role for CNS-ASs in feeding and fuel homeostasis is also evident in nonmammalian vertebrates with disparate feeding habits and metabolic demands. Research by our group, as well as by others, has indicated remarkable parallels between avian and mammalian AS-mediated energy balance mechanisms. The existence of medullary NE and EPI cell groups anatomically comparable to mammalian A1/C1 and A2/C2 with ascending projections to a hypothalamic paraventricular nucleus (PVN) corresponding to the mammalian PVH was observed in pigeons. Injections of NE or EPI into the PVN of pigeons evoke intense,  $\alpha$ -adrenoceptor-mediated feeding responses. Similar to what was observed in rats, hyperphagic responses after intra-PVN EPI injections in pigeons are significantly higher than those observed after injections of equimolar doses of NE. Also in pigeons, ICV or intra-PVN injections of EPI evoke a  $\beta$ -adrenoceptor-mediated increase in blood glucose (with no effect on FFA levels), whereas ICV injections of NE raise plasma FFA levels. These responses are mediated by both autonomic and pituitary mechanisms. These functional and anatomical attributes may represent a mechanism highly conserved during amniote (reptiles, birds, and mammals) evolution, possibly present in the last common ancestor of these vertebrate classes and probably still playing a fundamental role in the control of fuel homeostasis.

## See Also the Following Articles

Adrenergic Receptors • Antiadrenergic Agents • Insulin-Resistant States, Role of Free Fatty Acids (FFA) • Stress and Endocrine Physiology

## Further Reading

- Bernardis, L. L., and Bellinger, L. L. (1996). The lateral hypothalamic area revisited: Ingestive behavior. *Neurosci. Biobehav. Rev.* **20**, 189–287.
- Cincotta, A. H., Luo, S., Zhang, Y., Liang, Y., Bina, K. G., Jetton, T. L., and Scislowski, P. W. (2000). Chronic infusion of norepinephrine into the VMH of normal rats induces the obese glucose-intolerant state. *Am. J. Physiol.* **278**, R435–R444.
- Paez, X., Stanley, B. G., and Leibowitz, S. F. (1993). Microdialysis analysis of norepinephrine levels in the paraventricular nucleus in association with food intake at dark onset. *Brain Res.* **606**, 167–170.
- Ritter, S., Bugarith, K., and Dinh, T. T. (2001). Immunotoxic destruction of distinct catecholamine subgroups produces selective impairment of glucoregulatory responses and neuronal activation. *J. Comp. Neurol.* **432**, 197–216.
- Routledge, C., and Marsden, C. A. (1987). Adrenaline in the CNS: *In vivo* evidence for a functional pathway innervating the hypothalamus. *Neuropharmacology* **26**, 823–830.
- Sabi, M., Raimondi, C., Marques, S., Paschoalini, M. A., and Marino-Neto, J. (2002). Adrenergic control of glycemia in the paraventricular nucleus of the hypothalamus in pigeons. *NeuroReport* **13**, 871–875.
- Sawchenko, P. E. (1998). Toward a new neurobiology of energy balance, appetite, and obesity: The anatomists weigh in. *J. Comp. Neurol.* **402**, 435–441.
- Steffens, A. B., Scheurink, A. J. W., Luiten, P. G. M., and Bohus, B. (1998). Hypothalamic food intake regulating areas are involved in homeostasis of blood glucose and plasma FFA levels. *Physiol. Behav.* **44**, 581–589.
- Sved, A. F. (1990). Effect of monoamine oxidase inhibition on catecholamine levels: Evidence for synthesis but not storage of epinephrine in rat spinal cord. *Brain Res.* **512**, 253–258.
- Wellman, P. J. (2000). Norepinephrine and the control of food intake. *Nutrition* **16**, 837–842.



# Adrenergic Receptors

Linda F. Hayward

University of Florida College of Veterinary Medicine, Gainesville, Florida, United States

Patrick J. Mueller and Eileen M. Hasser

University of Missouri, Columbia, Missouri, United States

## Glossary

**$\alpha$ -adrenoceptor** Membrane-bound G protein-coupled receptor that mediates physiological and neurophysiological responses to epinephrine and norepinephrine; characterized by order of agonist potencies for stimulation: epinephrine  $\rightarrow$  norepinephrine  $\rightarrow$  isoproterenol.

**$\beta$ -adrenoceptor** Membrane-bound G protein-coupled receptor that mediates physiological and neurophysiological responses to epinephrine and norepinephrine; characterized by order of agonist potencies for stimulation: isoproterenol  $\rightarrow$  epinephrine  $\rightarrow$  norepinephrine.

**arrestin** Intracellular protein that, when bound to an adrenoceptor, promotes the receptor uncoupling from G proteins and receptor desensitization.

**autonomic nervous system** A component of the nervous system that is primarily involved in internal body functions and behavior.

Adrenergic receptors are membrane-bound proteins that mediate the peripheral and central actions of norepinephrine and epinephrine. By virtue of their location, either presynaptically or postsynaptically on neurons or effector organs such as the heart, vasculature, and adipose tissue, this class of receptors mediates a wide range of important homeostatic responses. Adrenergic receptors were originally divided into two major groups:  $\alpha$ - and  $\beta$ -adrenoceptors (ARs).  $\alpha$ -ARs demonstrate weak responses to the synthetic agonist isoproterenol but are very responsive to epinephrine and norepinephrine. In contrast,  $\beta$ -ARs respond potently to isoproterenol and are less sensitive to epinephrine and norepinephrine. During more recent years, both new pharmacological tools and the cloning of genes have revealed nine different AR subtypes: three  $\alpha$ 1-ARs ( $\alpha$ 1A,  $\alpha$ 1B, and  $\alpha$ 1D), three  $\alpha$ 2-ARs ( $\alpha$ 2A/D,  $\alpha$ 2B, and  $\alpha$ 2c), and three  $\beta$ -ARs ( $\beta$ 1,  $\beta$ 2, and  $\beta$ 3).

## $\beta$ -ADRENOCEPTORS

Peripherally,  $\beta$ 1-adrenoceptors (ARs) are localized to cardiac and adipose tissue, whereas  $\beta$ 2-ARs are found primarily in smooth muscle, liver, and human white cells but have also been localized on cardiac tissue. In the rat central nervous system (CNS), both  $\beta$ 1- and  $\beta$ 2-ARs are found with high densities in the cerebral cortex, caudate, and cerebellum. In amphibian and rat brains,  $\beta$ -ARs are located primarily on the postsynaptic membrane.  $\beta$ 3-ARs were not cloned until 1989, and receptor density appears to vary among species. Immunohistochemical and radioligand-binding studies suggest that  $\beta$ 3-AR mRNA is localized in both white and brown adipose cells as well as in skeletal muscle. Activation of this class of  $\beta$ -ARs is thought to mediate lipolysis and play a role in thermogenesis in these tissues. There is also evidence that  $\beta$ 3-ARs can be expressed in human atrial and ventricular cardiomyocytes under certain conditions.

## $\beta$ 1- and $\beta$ 2-Adrenoceptors

The postsynaptic mechanisms of  $\beta$ 1- and  $\beta$ 2-ARs have been studied the most extensively.  $\beta$ 1 receptors have approximately equal affinities for epinephrine and norepinephrine. Conversely,  $\beta$ 2 receptors have a higher affinity for epinephrine than for norepinephrine, suggesting that tissues with a predominance of  $\beta$ 2 receptors respond quickly to circulating epinephrine.  $\beta$ -AR stimulation leads to activation of adenylate cyclase and elevation of intracellular cyclic AMP (cAMP). This receptor effector system consists of three components: the  $\beta$ -AR, a guanine nucleotide regulatory protein that links the receptor to adenylate cyclase, and adenylate cyclase. When the G protein is in the stimulatory configuration, the binding of a neurotransmitter to the  $\beta$ -AR results in increased production of cAMP from adenosine triphosphate (ATP). In cardiac muscle, increased levels of cAMP

in turn activate protein kinase A (PKA), calcium channels are phosphorylated, and calcium entry increases through the sarcolemma, promoting a positive inotropic effect. Phospholamban is phosphorylated, promoting an increased rate of calcium reuptake into the sarcoplasmic reticulum and increasing the rate of myocardial relaxation (lusitropic effect). In addition, there is an increased rate of sinus node depolarization or a positive chronotropic effect. Cardiac  $\beta_2$ -ARs may also interact with an inhibitory G protein and may possibly modulate several G protein-independent pathways, including a possible link to a phosphatidylinositol 3' kinase pathway that has been linked with antiapoptotic mechanisms. In smooth muscle, the  $\beta$ -AR-stimulated increase in intracellular cAMP promotes relaxation through increased calcium reuptake into the sarcoplasmic reticulum and calcium extrusion to the extracellular space.  $\beta$ -receptor stimulation of smooth muscle also triggers activation of calcium-dependent potassium channels that hyperpolarize the membrane.

Prolonged agonist exposure to  $\beta$ -AR stimulation leads to receptor internalization and diminished response to subsequent stimulation. This process of down-regulation of the receptor primarily affects  $\beta_1$ -ARs and is mediated by increased activity of  $\beta_1$ -adrenergic kinase ( $\beta_1$ -ARK) during prolonged  $\beta_1$  agonist activity.  $\beta_1$ -ARK in turn phosphorylates the  $\beta_1$ -receptor. In the presence of  $\beta_1$ -arrestin, the receptor becomes uncoupled from the stimulatory G protein and is internalized. In failing human ventricles,  $\beta_1$ -ARK levels are elevated and the ratio of  $\beta_1$ : $\beta_2$  receptors decreases as  $\beta_1$ -receptors are internalized. In addition, there may be a functional uncoupling of remaining receptors. Although this mechanism of down-regulation is thought to be protective against the adverse effects of prolonged increased intracellular levels of cAMP and calcium, including apoptosis, the net outcome is a reduced response to  $\beta$ -receptor stimulation and systolic dysfunction. Transgenic mice with overexpression of human  $\beta_1$ - or  $\beta_2$ -ARs develop cardiomyopathic phenotype, including systolic dysfunction and chamber dilation, suggesting that overactivation of either receptor subtype can trigger internal pathways involved in receptor desensitization. The link between prolonged  $\beta$ -receptor stimulation and up-regulation of  $\beta_1$ -ARK also appears to be critical to the development of heart failure. Inhibition of  $\beta_1$ -ARK activity with a peptide inhibitor both enhances contractility in the normal heart and prevents the development of cardiomyopathy in a murine model of heart failure.

### $\beta_3$ -Adrenoceptors

$\beta_3$ -ARs have also been linked to cardiac function. In  $\beta_3$ -knockout mice, the inotropic response to isoproterenol is greater than that observed in wild-type controls. Alternatively, cardiac-specific overexpression of  $\beta_3$ -ARs results in a reduction in left ventricular function. Evidence suggests that  $\beta_3$ -ARs are up-regulated in human heart failure, whereas  $\beta_1$ -ARs are down-regulated, suggesting that  $\beta_3$ -ARs may constitute an opposing role to the excitatory effects of the other two  $\beta$ -ARs. The postsynaptic mechanisms of  $\beta_3$ -AR stimulation in cardiovascular tissue remains to be fully defined, yet there is evidence of coupling to both stimulatory and inhibitory G proteins as well as activation of nitric oxide synthase. It does not appear that  $\beta_3$ -ARs are targeted for intracellular phosphorylation or that they bind  $\beta$ -arrestin. As a result,  $\beta_3$ -ARs are relatively resistant to long-term down-regulation, although this resistance may be tissue specific.

$\beta$ -ARs are also extremely important in thermogenesis and lipolysis in adipocytes and possibly skeletal muscle. All three  $\beta$ -ARs appear to be present on brown and white adipocytes. Activation of  $\beta$ -ARs produces biphasic changes in intracellular cAMP (reflecting the coupling of the  $\beta_2$  and  $\beta_3$ -ARs to both stimulatory and inhibitory G proteins) and activates a mitochondrial uncoupling protein (UPC-1 in brown adipose tissue) that is important for the purpose of heat generation (thermogenesis) at the expense of coupled ATP production. In animals, stimulation of  $\beta_3$ -ARs has antiobesity and antidiabetic properties by decreasing fat content, increasing muscle glucose uptake, and decreasing hepatic glucose output. There is also a link between  $\beta_3$ -AR activation and decreased leptin expression and secretion from white adipose tissue. The mechanism underlying interaction between  $\beta_3$ -AR stimulation and leptin has yet to be fully determined but may involve  $\beta$ -AR activation of cAMP-dependent protein kinases. Alternatively, in animal models of obesity such as the ob/ob mouse and Zucker fatty (fa/fa) rat, where leptin release or receptor binding is dysfunctional, there is a decrease in both  $\beta_3$ - and  $\beta_1$ -AR expression. The mismatch between leptin levels and  $\beta$ -AR function may be related to sympathetically mediated  $\beta$ -AR down-regulation. However, in humans, the profound effect of  $\beta$ -AR down-regulation in obesity has yet to be linked specifically to  $\beta_3$ -AR deletion. The relative numbers of  $\beta_3$ -ARs in metabolically active tissues of rodents versus humans may explain these differences. Furthermore, there is evidence of significant

redundancy in the system given that deletion of all three  $\beta$ -AR subtypes is necessary to generate mice that are intolerant to cold exposure.

## $\alpha$ -ADRENOCEPTORS

$\alpha$ -ARs mediate most (but not all) of the excitatory effects of catecholamines, including contraction of vascular smooth muscle, contraction of uterine muscle, contraction of the urethra, and pupillary dilation. In general,  $\alpha$ 1-ARs are involved in mediating responses at effector organs. The  $\alpha$ 2-ARs are located primarily on the presynaptic terminal and modulate transmitter release, although they also are present postsynaptically and contribute to effector organ responses. Responses at many effector organs are mediated by more than one AR subtype. For example, several  $\alpha$ -AR subtypes are expressed in the vasculature and mediate different aspects of vascular function. Moreover, the relative roles of these subtypes vary among different vascular beds, at different levels of individual blood vessels, and are species dependent.

### $\alpha$ 1-Adrenoceptors

The  $\alpha$ 1-ARs are distributed in various organs and tissues, including the vasculature, heart, lung, kidney, liver, and brain. Within the brain, radioligand binding and autoradiographic techniques have indicated rich localization of  $\alpha$ 1-ARs in the rat cerebral cortex, hippocampus, certain thalamic nuclei (dorsal lateral geniculate), and dorsal raphe nucleus. Peripherally, this class of receptors is thought to be involved importantly in the control of vascular tone. The three different subtypes ( $\alpha$ 1A,  $\alpha$ 1B, and  $\alpha$ 1D) all have been implicated in vascular smooth muscle contraction, with varied roles depending on the species and the vascular bed. It appears that the  $\alpha$ 1A subtype may be involved more in the maintenance of basal tone, whereas the  $\alpha$ 1B subtype may participate more in the response to exogenous agonists. Different physiological conditions can influence the response to activation of these receptors, and expression of the different subtypes can be affected independently under different physiological and pathophysiological situations. It appears that expression of the  $\alpha$ 1B subtype may be especially subject to changes in level of expression. Thus, it is possible that this receptor may be involved in altered responses under different physiological and pathophysiological conditions.

The  $\alpha$ 1-ARs are G protein-coupled receptors, coupled primarily to Gq/11. Activation of this system

stimulates activity of phospholipase C, resulting in the formation of 1,4,5-trisphosphate and diacylglycerol. These molecules subsequently stimulate an increase in cytosolic  $\text{Ca}^{2+}$  and activation of protein kinase C. Although this is the primary coupling system for most  $\alpha$ 1-AR-mediated responses, evidence suggests that these receptors can also be coupled to other G proteins. Thus, it is likely that activation of  $\alpha$ 1-ARs is capable of influencing a number of different signaling pathways and of modulating many different types of responses.

An important aspect of  $\alpha$ 1-AR function, similar to  $\beta$ -ARs, is that they are highly subject to desensitization. Prolonged exposure to an agonist results in reduced responsiveness to subsequent stimulation of the receptor. The reduction in response to agonist stimulation appears to be due to binding of arrestin proteins to the receptor, which favors its uncoupling from the G protein. In addition, receptors are internalized within the cell, also contributing to decreased receptor activity. Phosphorylation of the receptor via activation of G protein receptor kinases is a primary event in the process of desensitization/internalization of these receptors.

### $\alpha$ 2-Adrenoceptors

$\alpha$ 2-ARs generally are considered to be located presynaptically, modulating neurotransmitter release. However, they also are present postsynaptically and on non-neuronal tissues. Like the  $\alpha$ 1-ARs, they are located both peripherally and throughout the central nervous system. The  $\alpha$ 2-ARs are important in a wide variety of responses, including inhibition of neurotransmitter release, control of vascular tone, regulation of renin release, inhibition of insulin secretion, inhibition of lipolysis, and platelet aggregation. Within the central nervous system,  $\alpha$ 2-ARs influence the control of arterial pressure, modulation of growth hormone release, sedation, analgesia, and a variety of effects on behavior and cognition. Prominent locations within the CNS are in cardiovascular regions such as the nucleus tractus solitarius and rostral ventrolateral medulla as well as the locus coeruleus. Central stimulation of  $\alpha$ 2-ARs can result in hyperpolarization and neuronal inhibition.

Peripherally, it is clear that the  $\alpha$ 2-ARs, like the  $\alpha$ 1-receptors, mediate vasoconstriction. The primary subtype involved in vasoconstrictor responses is the  $\alpha$ 2A/D receptor, although there are contributions from the other subtypes as well, in particular  $\alpha$ 2B receptors. Interestingly, the  $\alpha$ 2B-AR also appears to

be the dominant subtype that mediates the central hypotensive response to  $\alpha 2$ -AR agonists. Thus, the same receptor subtype appears to mediate vasoconstriction peripherally and hypotension centrally; however, the hypotensive response predominates. An interesting aspect of the vasoconstrictor effects of  $\alpha 2$ -ARs is that they appear to be the primary AR involved in the contraction of veins. Depending on the species and vascular bed, this venous constriction is mediated by different subtypes, primarily the  $\alpha 2A/D$ -ARs and the  $\alpha 2C$ -ARs.

The presynaptic effects of  $\alpha 2$ -ARs appear to be mediated predominantly by the  $\alpha 2A/D$  subtype, although some studies suggest that the  $\alpha 2C$  subtype also may be involved. These receptors act as autoreceptors, primarily mediating inhibition of transmitter release. They are present widely in the brain as well as in peripheral tissues, including the heart, kidney, and vasculature. These presynaptic receptors are thought to contribute, at least in part, to the hypotensive, sedative, and analgesic effects of  $\alpha 2$ -AR agonists.

The  $\alpha 2$ -ARs primarily are negatively coupled to adenylate cyclase. In addition, cellular effects of  $\alpha 2$ -ARs may be mediated through inhibition of voltage gated  $Ca^{2+}$  channels, activation of inwardly rectifying  $K^{+}$  channels, activation of phospholipase C, increased  $Ca^{2+}$  release from intracellular sources, and stimulation of mitogen-activated protein kinase (MAP kinase). Thus, activation of  $\alpha 2$ -ARs has the capacity to influence a number of signaling pathways, providing a modulatory effect on a variety of cellular functions.

## See Also the Following Articles

Adrenergic Mechanisms • Antiadrenergic Agents • Stress and Endocrine Physiology

## Further Reading

- Bachman, E., Dhillon, H., Zhang, C-Y., Cinti, S., Bianco, A. C., Kobilka, B. K., and Lowell, B. B. (2002). Beta-AR signaling required for diet-induced thermogenesis and obesity resistance. *Science* **297**, 843–845.
- Bylund, D., Regan, J. W., Faber, J. E., Hieble, J. P., Triggle, C. R., and Ruffolo, R. R., Jr. (1995). Vascular alpha-adrenoceptors: From the gene to the human. *Can. J. Physiol. Pharmacol.* **73**, 533–543.
- Calzada, B., and DeArtinano, A. A. (2001). Alpha-adrenoceptor subtypes. *Pharmacol. Res.* **44**, 195–205.
- Garcia-Sainz, J., Vazquez-Prado, J., and del Carmen Medina, L. (2000). Alpha-1 adrenoceptors: Function and phosphorylation. *Eur. J. Pharmacol.* **389**, 1–12.
- Gauthier, C., Langin, D., and Balligand, J-L. (2000). Beta3-adrenoceptors in the cardiovascular system. *Trends Pharmacol. Sci.* **21**, 426–431.
- Janssen, L. (2002). Ionic mechanisms and calcium regulation in airway smooth muscle contraction: Do the data contradict dogma? *Am. J. Physiol.* **282**, L1161–L1178.
- Port, J., and Bristow, M. R. (2001). Altered beta-adrenergic receptor gene regulation and signaling in chronic heart failure. *J. Mol. Cell. Cardiol.* **33**, 887–905.
- Rockman, H., Chien, K. R., Choi, D-J., Iaccarino, G., Hunter, J. J., Ross, J., Lefkowitz, R. J., and Koch, W. J. (1998). Expression of a beta-adrenergic receptor kinase 1 inhibitor prevents the development of myocardial failure in gene-targeted mice. *Proc. Natl. Acad. Sci. USA* **95**, 7000–7005.
- Small, K., and Leggett, S. B. (2001). Identification and functional characterization of alpha-2 adrenoceptor polymorphisms. *Trends Pharmacol. Sci.* **22**, 471–477.



# Adrenocorticotrop Hormone

see ACTH



# Aging and Longevity of Human Populations

Kenneth G. Manton

Duke University, Durham, North Carolina, United States

## Glossary

**antioxidant** A substance that chemically inhibits oxidation reactions.

**caloric restriction** Eating fewer calories than *ad libitum* while achieving adequate or optimal nutrition. This has extended both mean and maximum life span in various animal models.

**Gompertz function** A mathematical model developed by B. Gompertz in 1825 of the age-related increase in mortality during the adult life span.

**juvenile hormone** An insect hormone that is secreted by the corpora allata and plays a role in reproduction, gene expression, and metabolism.

**mitochondria** Any of various round or long cellular organelles of most eukaryotes that are found outside the nucleus, produce energy for the cell through oxidative processes, and are rich in fats, proteins, and enzymes.

**nuclear DNA** DNA found within the nucleus of the cell.

**oxidative theory of aging** The theory that declines in physiological function with age are due to damage accumulated from reactive oxygen species.

**superoxide dismutase (SOD)** A potent intracellular antioxidant.

**telomere** The natural end sequence of a eukaryotic chromosome.

Aging and longevity are fundamental characteristics of human populations. Aging typically means the loss with increased chronological age of the functional capability of different organs of the body due to the operation of a general process of biological senescence. Aging, or senescence, is usually viewed as distinct from major chronic diseases, such as cancer or cardiovascular diseases, although there may be shared exposure factors or mechanisms (e.g., oxidative processes causing mutations in DNA). Often, aging is defined at the cellular level, such as the replicative senescence of cells whose telomere reaches a critical length, or related to

mitochondrial dysfunction in energy production due to the accumulated damage of reactive oxygen species. Longevity, the number of years lived, is generally viewed as being determined by senescent processes reaching their end stage. However, it is clear that there are multiple determinants of longevity, and the correlation of life span with the progression of the different dimensions of senescence is more complex and plastic than had been previously thought and is not well understood.

## CELL-BASED SENESENCE

Biological models of human aging and longevity have often focused on the senescence and death of the individual cell—the fundamental building block of the human organism. For example, one model of senescent-limited longevity is based on a model of genetic restrictions on the number of cell replications that can occur under the so-called Hayflick limit. The physical basis for this phenomena is believed to be the telomere, the end sequence of the chromosome that tends to decrease in length with each replication of the cell. These cell-based models tend to underplay the significance of endocrine and growth factors.

One problem with this perspective is that an experimental analysis of the decline in the number of replications that a cell can perform (often estimated at 50–60) per year of age *in vitro* was estimated to be 0.2/year in humans aged 30–80, implying that this mechanism limited the human life span to 250–300 years. The highest reliable age reported at death is currently that of Madame J. Calment, who died at age 122. A second problem is that although the Hayflick limit may explain longevity bounds, it does not explain loss of function with age.

A recent analysis showed that in healthy human subjects, and controlling for the biopsy site, there was no significant negative correlation between the cell's replicative capacity in a given tissue and subjects' age *in vitro*. As a consequence, questions are raised about the importance and exact role of the loss of cell replicative capacity as a limiting factor in human longevity. Additionally, studies have found the correlation

of telomere length with replicative senescence to be complex because it varies across tissue types and because the telomere may have additional cellular functions, such as monitoring oxidative cell damage.

Further complicating analysis of the mechanisms controlling longevity is the fact that most tissues, including the brain, are now thought to have a small proportion of stem cells from which new cells can generate. It has been demonstrated that new neurons are generated in the human brain at all ages. Aerobic exercise may stimulate production of new cells in the hippocampus. One specific interesting stem cell is the SHED stem cell found in teeth. This stem cell has been shown to have the potential to form cells of a variety of tissue types. The intrinsic potential for regeneration of each tissue type and the tissue-specific hormonal and growth factor triggers that may cause new cell growth (or may restrict growth, e.g., the myostatin gene discovered in 1995) are not well understood with respect to their role in regulating senescence and determining human longevity.

### OXIDATIVE STRESS AND MITOCHONDRIAL INFLUENCES ON AGING

Another promising model of the cellular mechanisms limiting the human life span is the senescence of the human mitochondria. This is a specific form of the general theory of an oxidative basis for aging originally proposed by Harman. Mitochondrial DNA (mtDNA) has fewer error monitoring and correcting mechanisms than does nuclear DNA (nDNA). Furthermore, since the mitochondria is the primary center of energy production in cells, the organelle is thought to be especially vulnerable to the oxidative stress of reactive oxygen species (ROS), which are produced in the respiratory processes centered in the mitochondria.

Because the mutation rate of mtDNA is much greater (approximately 10-fold) than that of nDNA, in a variant of the oxidative theory of aging, “aging” of the mitochondria has been suggested to be the physiological process limiting human longevity to a maximum of 130 years. This is a more plausible limit than that derived from the Hayflick limit in *in vitro* studies focusing on the loss of replicative capacity of the cell and its correlation with the age of the cell donor.

Of interest is that the production of many mitochondrial proteins in humans has been taken over by nDNA, with 47 of 60 mitochondrial proteins produced in the nucleus, in which error monitoring and corrective mechanisms are more complete. An interesting corollary of

this model is that as aging degrades the efficiency of energy production, the physical activity of the organism is degraded, inducing a correlation of functional loss and longevity. Cell-specific failure can be communicated systematically by cytokines such as interleukin-8, which is increased by stress due to ROS, and interleukin-6, which is a proinflammatory cytokine.

### EXOGENOUS MODULATORS OF MITOCHONDRIAL FUNCTION

Problems arise with this model of longevity and senescence when one considers how “external” factors affect the process of mitochondrial senescence and functional degeneration. A common model for life-expectancy extension is caloric restriction. In this case, the lowering of caloric consumption is thought to decrease basal metabolism, reduce oxidative stress, and retard age-related degeneration. Since this involves energy production and metabolism, it is reasonable to consider the mitochondria as the focal point of these processes. Such an intervention has been shown to increase life expectancy in experimental models of rodents and other lower organisms. Additionally, several genetic mutations in *Caenorhabditis elegans* and *Drosophila* have shown that slowing metabolism, and reducing growth and reproduction, can significantly extend the life span. It is believed that longevity-extending mutations in these organisms either protect against ROS species (e.g., genes controlling SOD2 production) or alter insulin and insulin growth factor-1 pathways and signaling.

A significant counterexample to these experimental models is *Apis mellifera*, the common honeybee. Both bee workers and queens have identical genotypes. Their polyphenotypic differentiation occurs due to differences in larval nutrition whereby larva destined to become queens are fed royal jelly, which contains juvenile hormone that alters the developmental trajectory, increases body size and respiration/metabolic rate, increases life expectancy 15- to 30-fold over that of the worker bee, and increases the level of royal jelly produced in queens.

Juvenile hormone seems to affect basal metabolic rate by increasing the production of mitochondrial transcription factors and certain mitochondrial enzymes (Cox-1 and CytC) involved in the respiratory process. Thus, juvenile hormone in bees appears to play roles similar to those of thyroid hormones in humans, which can also affect the developmental process as well as the mitochondrial energy production function. Of interest is that the nuclear receptors

for thyroid hormone  $T_3$  also code for the type of mitochondrial translation proteins elevated in queen bees. One of the crucial findings of experimental studies of the queen bee was that the increases in respiration rate were due to enhanced mitochondrial function and not to increased numbers of mitochondrial organelles. The queen bee model raises the question of how thyroid hormones in humans might regulate senescence and affect longevity by controlling mitochondrial behavior and function (a process not extensively studied), especially regarding how nutritional and physical activity (energy expenditure) factors may affect mitochondrial function.

There is significant evidence that nutritional or dietary factors can alter mitochondrial function and, indeed, changes in mitochondrial functions with age in animal models. Studies of mitochondrial efficiency in experimental animal models suggest that oral ingestion of acetyl-L-carnitine (a substrate for fatty acid membrane transport in mitochondria) and  $\alpha$ -lipoic acid (a fat-soluble antioxidant) increased oxidative efficiency of the mitochondria in elderly experimental animals, increased their physical activity level, and improved memory. Thus, the age-related level of accumulated oxidative stress and degradation of mitochondrial function may be partly reversed by appropriate nutritional supplementation in these model systems. It is not known how such interventions perform in humans.

### TISSUE INTERDEPENDENCE AND HORMONAL CONTROL IN LONGEVITY

Any model that attempts to explain human longevity and senescence solely at a cellular or molecular level is likely to fail as an oversimplified representation of the physiological and molecular mechanisms involved. The fact is that the molecular functioning of a cell is important not only for its internal maintenance but also for the functioning of the organism. Thus, part of the question of the relation of senescence and longevity concerns the biological complexity of the organism due to the functional differentiation of tissue and organs and the hormonal synchronization of the wide variety of heterogeneous physiological functions that need to be performed.

One way to model such biological complexity is to explicitly introduce the effects of the endocrine system into a model of human aging and longevity. Although certain hormonal factors have been studied in experimental models of aging, they have not explicitly been

introduced in mathematical models of human aging and longevity in populations.

To illustrate, the two hazard functions most often used to describe the age trajectory of mortality in humans are the Gompertz and Weibull. One theoretical justification for these functions is that they may describe increases in the risk of death as a function of the thermodynamics of protein denaturation. As a consequence, the lowering of body temperature is thought to significantly increase human longevity; for example, a several-degree (centigrade) decline in body temperature is projected to increase longevity by approximately 20 years. Not specified in these models are the physiological mechanisms controlling core body temperature (possibly the thyroid gland), its effects on thermogenesis, and its regulation by feedback through other endocrine organs. Thyroid hormone ( $T_3$ ) effects energy production by regulating mitochondrial respiratory function and proliferation.

The hormone and signaling pathways studied most intensively for their effects on longevity and aging are insulin, insulin-like growth factor 1 (IGF-1), and growth hormone (GH) because of the experimental observation of increases in longevity due to caloric restriction. Among the experimental models most frequently used in such studies are *C. elegans*, *Drosophila melanogaster*, yeast, and rodents. IGF-1 and GH have been intensively studied because of the conceptual attractiveness of the oxidative theory of aging and because of the effects of caloric restriction on life span in experimental models (e.g., rodents), which are believed to be mediated, in part, by IGF-1, insulin, and/or GH. The logic of the relation is that caloric restriction slowed metabolism, the production of ROS, the accumulation of oxidative damage, and thus the aging rate. Genetic mutations that increased the production of antioxidants, such as catalase or SOD2, were also associated with significant increases in longevity in various experimental models.

It is unclear, however, whether such interventions work in the same way in humans because of their greater histological (especially hormonal) complexity. Genetic mutations that altered IGF-1 and GH production and reduced metabolic functions, although associated in simple experimental models with increased longevity, have been associated in humans with a number of adverse conditions, such as obesity and lipid dysfunction, resulting in a higher risk of death from circulatory disease and cancer.

An additional rationale for this line of research is reasoning based on evolutionary models that there is a trade-off of energy expended in growth and reproductive function and somatic maintenance. Reduced

somatic maintenance during periods of scarce food supplies causes metabolism in *C. elegans* to decrease and reduces the requirement for energy and may cause organisms to enter a quiescent, or dauer, state. This quiescent, low-energy consumption state is often associated with reduced oxidative stress and increased longevity but a lower rate of reproduction and less growth.

Hormonal control, however, is more complex in humans than in these experimental models. The IGF-1/insulin effect in the fly and the worm may favor localized hormonal actions. In worms and flies, there was a single receptor that favored local system actions of insulin and IGF peptides, permitting aging regulation to be dominated by a single tissue type (e.g., the neuroendocrine tissue of *C. elegans*). In tetrapods, in contrast, there are four candidate receptors—IGF-1, IGF-2, IR, and RR. This may have led to the developmental differentiation of insulin to control metabolic factors and IGF-1 to control mechanisms of growth in humans. As a consequence, in humans, caloric restriction may have become dissociated from factors controlling longevity, with factors linked to human longevity (e.g., DHEA, insulin, and body temperature) no longer simply associated with caloric restriction. Indeed, recent animal studies show that intermittent starvation, as opposed to caloric restriction, leads to preservation of growth while having even greater effects on insulin and insulin sensitivity and, consequently, on longevity. Another recent experiment in mice is also of interest. Mice were raised without a gene for the insulin receptor in adipose tissue. These mice were found to have greater longevity (+18%) than control mice. In addition, the gene knockout mice were lighter and much leaner than the control mice, even though they consumed more food. This suggests that they had a higher rate of respiration along with their greater average and maximal longevity. This study suggests the primary effect of caloric restriction in rodents is not in reducing oxidative processes but in reducing body fat. Indeed, increasing the efficiency of mitochondrial function and energy production appears to increase longevity. Thus, both insulin regulation and thyroid function appear to be factors that can strongly regulate human longevity and perhaps senescence, suggesting systematic multiorgan regulation of longevity rather than simple, universal cellular “clocks.”

Whereas insulin, IGF-1, and GH have been extensively studied, one endocrine system's effects on aging and longevity that has not been well studied is that of the thyroid. Thus, an interesting future focus of hormonal research on human longevity and aging is on

the effects of thyroid hormone,  $T_3$ , on energy production in the mitochondria. A recent review suggested it has at least three effects. The first, occurring within 30 min, is up-regulation of the energy production of the mitochondria by altering their membrane structure and enhancing their efficiency in energy production. A second change, occurring within 12 h, stimulated production of proteins in the mitochondria to further enhance energy production (e.g., by altering the lipid composition of the inner membrane of the mitochondria). The third effect, not evident for at least 24 h, was the stimulation of nuclear and mitochondrial DNA to produce necessary protein for mitochondriogenesis. Thus, there are fundamental interactions between mitochondrial performance and growth and  $T_3$  hormone. The action of  $T_3$  also seems to be involved in tissue differentiation so that there is a direct linkage of metabolic level and thyroid hormone and the energy demands of organism development. The relation of these factors in controlling human longevity and development/senescence needs extensive study.

Also significantly involved in senescence and longevity is the endocrine control of immune function during aging by the hypothalamic–pituitary–adrenal axis. Hormones play many roles in cytokine production, and cytokines perform a number of important tasks. It is thought that the level of interleukin-6 (IL-6) increases with age. IL-6 controls inflammatory processes in part by stimulating C-reactive protein production in the liver. However, this mechanism is controversial as a model of aging, disability, and longevity because IL-6 is not clearly elevated in centenarians. It appears that in the age range 85–96 years, mortality selection on genetic factors eliminates persons predisposed to certain types of immunosenescence. There is interest in IL-6 because it causes dysfunction of muscle tissue by catabolism, promoting a direct linkage of functional loss and mortality. IL-6 is also implicated in the neurodegenerative process of Alzheimer's disease. It is of interest that anti-inflammatory compounds such as NSAIDs (especially ibuprofen) and statins seem to be highly effective in reducing the risk of neurodegenerative processes (e.g., Alzheimer's disease) through multiple pathways, as may certain steroid hormones (e.g., testosterone).

## A MODEL LINKING HUMAN REPRODUCTION AND GROWTH WITH SENESCENCE AND MORTALITY

To adequately model the physiology of such complex hormonal effects and interactions, more sophisticated

quantitative population models of human survival and physiological function are necessary. Based on evolutionary arguments, much of the linkage of senescence and development and reproduction is due to the need to compromise between the energy demands of growth and reproduction, on the one hand, and somatic maintenance, on the other hand, both of which require considerable expenditure of energy. No current model of aging and longevity/human mortality effectively links the age-changing linkage of these processes over the life span, and it is likely in the complex linkage of these processes that endocrine control systems are involved in complex ways.

One candidate model uses a series of three random walk equations for an individual in a state variable space of  $j$  dimensions, where each state variable is denoted by  $x_j$ . The first describes the temporal trajectory of a person progressing/developing through  $j$  dimensions due to both deterministic forces and stochastic influences (modeled as the process  $w$ ):

$$d(x, t) = \beta(x, t)w(x) \quad (1)$$

where  $t$  is the age or time,  $d$  is change, and  $\beta$  is the position-specific rate of change.

The second describes the change in the probability of mortality conditional on the position in the state space and stochastic factors:

$$M(x, t) = P_m(x, t)w(x) \quad (2)$$

The third describes the probability of a birth for a female at that position in the state space:

$$B(x, t) = P_b(x, t)w(x) \quad (3)$$

The advantage of this model is that the energy devoted to reproduction is correlated with the risk of death for a person of a given age and physical state through the state dynamic process in  $x_t$  (Eq. 1). The stochastic process,  $x_1(t)$ , linking fertility and mortality is the system through which hormonal control factors may work as indexed by physiological measurements  $x$ .

To apply this model, we need to generate equations for the change in the state space distribution of a population of such individuals. For change in the state variables over time, we need a second-order dynamic equation, or

$$x_{t+1} = u_0 + R_{x1} + C(x_t^I \cdot x_t) + e \quad (4)$$

This includes not only the cross-temporal dependency of factors but also interaction terms, which suggests that the effects of risk factors over time are functions of the values on other variables. This reflects various

“field” effects in which trajectories are affected by multiple variables. The  $e$  term reflects stochasticity in these cross-temporal relations.

To model mortality and its dependence on state variables, we use an age-dependent quadratic form, or

$$\mu_T(x_1t) = (x_t^T M x_t) e^{\theta t} \quad (5)$$

The risk of death is a function of age,  $t$ , and risk factor values. The matrix  $M$  are hazard coefficients. The term  $e^{\theta t}$  reflects the effects of unobserved factors correlated with age of mortality. To the extent that measured factors reflect age dependence, the value of  $\theta$  will decrease.

From Eq. (5), we can estimate cause-specific functions,  $\mu_C$ , to reflect the different relation of risk factors to each of a set of  $C$  distinct causes of death. All cause-specific mortality functions, and the birth function, will be related through the stochastic dynamic process  $x_t$ .

The next component of the model is the birth function. This is modeled as a quadratic function of the same set of risk factors (where  $B$  are the birth probability coefficients),

$$b(x_1t) = (x_t^T B x_t) e^{\delta t}$$

where the values of  $x_t$  reflect the physiological state of the mother prior to birth. Thus, both mortality and birth are linked to a common multidimensional state variable process with age functions  $\delta$  and  $\theta$  that reflect different latent states affecting the potential for survival and reproduction. The measurable cross-age effects of endocrinological factors are represented in the stochastic process  $x_t$ .

To understand how this model could be applied, we use as an example our analysis of the 46-year follow-up of the Framingham Heart Study, in which every 2 years a number of factors were measured, including systolic and diastolic blood pressure, left ventricular hypertrophy, pulse rate, body mass index (BMI), vital capacity index, smoking or nonsmoking, serum cholesterol, and blood glucose, as well as demographic variables such as gender and chronological age.

These factors describe a wide range of physiological effects. Blood glucose is connected to diabetes and hence insulin function. This effect is modulated by BMI, which is related to thyroid function and leptin as well as other hormonal factors. Blood pressure indicates, among other things, how hormonal factors affect arterial endothelium and renal function. Implicit in the interrelation of the variables is the interaction of a wide range of endocrinological factors. For example, blood glucose, hypertension, and cholesterol have been jointly implicated in Reaven’s metabolic syndrome X.

Other patterns of change may reflect features of endocrine control of multiple variables.

Obviously, in the model it would be best to have direct sequential hormonal measures (e.g., corticosteroid, T<sub>3</sub>, T<sub>4</sub>, insulin, GH, and leptin), but using the data that are usually available on large, representative human populations, we must infer their effects on aging and longevity from their joint temporal pathways and evolution. The more detailed measures may be obtained in the future for large, longitudinally representative populations, such as the 2004 National Long Term Care Survey, as the technology for protein assays improves. One of the critical hypotheses suggested by the previously mentioned model is that longevity and fertility are correlated through the state variables. Such a model could be evaluated in, for example, the Framingham Heart Study, in which two offspring cohorts have been assessed in addition to the original cohort. The age, timing and number of births may be factors in determining longevity, as demonstrated by Promislow in British aristocratic families. Effective intervention should take into account how various profiles of risk factor relations are changed.

## CONCLUSION

As the biological complexity of organisms increases, hormonal factors increase in importance as determinants of the correlation of senescence and mortality. Since no single organ dominates survival, the coordination and communication between organ systems becomes more crucial to survival and successful functioning. Because of the increased biological complexity of humans, the linkage of GH, IGF-1, and insulin observed in many studies of simple organisms likely does not function in the same way in humans. The role of thyroid hormones, especially their ability to cause the cell nucleus to produce mitochondrial proteins, must be further studied as a fundamental factor in senescence and longevity. It is reasonable to assume hormonal factors significantly affect human longevity and related aging processes in a variety of complex ways, that these are altered by exogenous factors such as nutrition, physical activity, by genetic polymorphisms, and that, if understood, could indicate how to better control senescent loss of function and overall survival.

## See Also the Following Articles

Aging and the Male Reproductive System • Aging, Animal Models for • Aging, Immunology and • Autonomic Nervous System, Aging and • Body Weight, Body Composition, and Aging • Functional Genomics of Aging • Interleukin-6 • Neuroendocrine System and Aging • Oxidative Stress and Aging • Stress, Aging, and Central Nervous System Interactions • Thyroid, Aging and

## Further Reading

- Anson, R. M., Guo, Z., de Cabo, R., Lyun, T., Rios, M., Hagepanos, A., Ingram, D. K., Lane, M. A., and Mattson, M. P. (2003). Intermittent fasting dissociates beneficial effects of dietary restriction on glucose metabolism and neuronal resistance to injury from calorie intake. *Proc. Natl. Acad. Sci.* **100**(10), 6216–6220.
- Corona, M., Estrada, E., *et al.* (1999). Differential expression of mitochondrial genes between queens and workers during caste determination in the honeybee *Apis mellifera*. *J. Exp. Biol.* **202**(8), 929–938.
- Cristofalo, V., Allen, R., *et al.* (1998). Relationship between donor age and the replicative lifespan of human cells in culture: A reevaluation. *Proc. Natl. Acad. Sci. USA* **95**, 10614–10619.
- Hagen, T. M., Liu, J., Lykkesfeldt, J., Wehr, C. M., Ingersoll, R. T., Vinarsky, V., Bartholomew, J. C., and Ames, B. N. (2002). Feeding acetyl-L-carnitine and lipoic acid to old rats significantly improves metabolic function while decreasing oxidative stress. *Proc. Natl. Acad. Sci.* **99**(4), 1870–1875.
- Hayflick, L. (1965). The limited in vitro lifetime of human diploid cell strains. *Exp. Cell Res.* **37**, 614–636.
- Kramer, A., Hahn, S., *et al.* (1999). Aging, fitness and neurocognitive function. *Nature* **400**(6743), 418–419.
- Liu, J., Head, E., Gharib, A. M., Yuan, W., Ingersoll, R. T., Hagen, T. M., Cotman, C. W., and Ames, B. N. (2002). Memory loss in old rats is associated with brain mitochondrial decay and RNA/DNA oxidation: Partial reversal by feeding acetyl-L-carnitine and/or R- $\alpha$ -lipoic acid. *Proc. Natl. Acad. Sci. USA* **99**(4), 2356–2361.
- Rosenberg, B., Kemeny, G., *et al.* (1973). The kinetics and thermodynamics of death in multicellular organisms. *Mech. Ageing Dev.* **2**, 275–293.
- Roth, G. S., Lane, M. A., Ingram, D. K., Mattison, J. A., Elahi, D., Tobin, J. D., Muller, D., and Metter, E. J. (2002). Biomarkers of caloric restriction may predict longevity in humans. *Science* **297**(2), 811.
- Tatar, M., Bartke, A., and Antebi, A. (2003). The endocrine regulation of aging by insulin-like signals. *Science* **299**(5611), 1346–1351.
- Wrutniak-Cabello, C., Casas, F., and Cabello, G. (2001). Thyroid hormone action in mitochondria. *J. Mol. Endocrinol.* **26**, 67–77.



# Aging and the Male Reproductive System

David A. Gruenewald

University of Washington School of Medicine and VA Puget Sound Health Care System, Puget Sound, Washington, United States

## Glossary

**andropause** An age-related decline in serum testosterone in older men to levels below the normal range for young men; associated with symptoms and signs consistent with androgen deficiency.

**erectile dysfunction** Difficulty obtaining an erection of sufficient quality to allow for sexual intercourse.

**hypogonadism** Failure of testicular production of testosterone, spermatozoa, or both.

**libido** Sexual desire.

Compared to women, men experience a more variable, gradual, and progressive decline in reproductive function as they age. Male reproductive system aging is notable for reductions in testicular secretion of testosterone; alterations in hypothalamic/pituitary regulation of testicular function; structural changes in the testis, penis, and accessory sexual glands; and alterations in sexual function, spermatogenesis, and fertility. The significance of most of these age-related changes in male reproductive physiology is unclear, although erectile dysfunction and some consequences of the age-related decline in testosterone levels may have a major negative impact on quality of life in aging men.

## INTRODUCTION

Variability is a hallmark of male reproductive system aging and the aging process in general. Whereas some older men exhibit relatively normal reproductive function, others develop significant impairments in testosterone production, spermatogenesis, or erectile function with aging. Age-associated interindividual variability has important implications for the study and management of male reproductive aging.

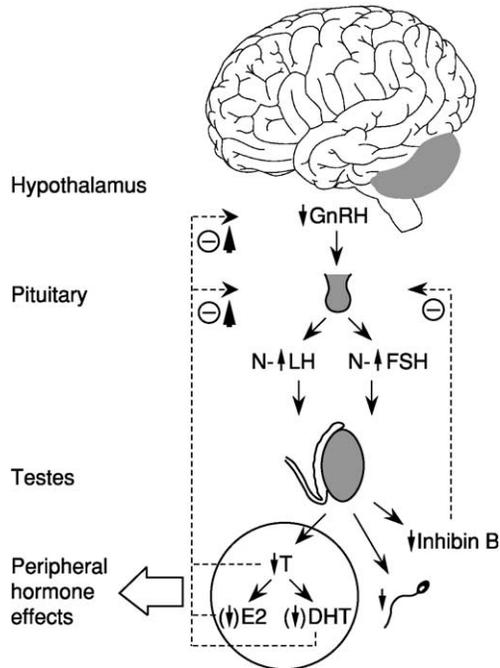
## PHYSIOLOGY OF AGING

### Decline in Serum Testosterone Levels with Aging

In longitudinal studies, serum testosterone levels decline progressively with aging in healthy men beginning in the third decade, and approximately 20% of men older than 60 years of age and half of men older than 80 years of age have total testosterone levels below the normal range for young men. However, even higher percentages of older men have bioavailable testosterone levels below normal levels for young men. This is because levels of sex hormone-binding globulin (SHBG; the major serum-binding protein for testosterone) increase with aging, and SHBG binds testosterone with high affinity. Therefore, bioavailable non-SHBG-bound testosterone, comprising free testosterone and testosterone weakly bound to serum albumin, declines more than total testosterone with aging.

Testosterone is metabolized to estradiol and 5 $\alpha$ -dihydrotestosterone (DHT) (Fig. 1). Like testosterone, bioavailable estradiol levels decline with aging in men, reflecting both the decline in its substrate, testosterone, and the increase in serum SHBG levels with aging. Serum total DHT levels are unchanged or slightly decreased with aging, but bioavailable DHT appears to decrease as a consequence of the age-related increase in SHBG.

In many older men, the age-related decline in testosterone levels is exacerbated by the effects of comorbid illnesses (e.g., diabetes mellitus, renal failure, and alcohol abuse), malnutrition, and medications (e.g., psychoactive medications and glucocorticoids) that further suppress testosterone levels. Elderly men with serious chronic illnesses, residents in nursing homes, and patients in inpatient rehabilitation units have significantly lower testosterone levels than healthy older men and a higher prevalence of testosterone levels below the normal range for young men.



**Figure 1** Schematic diagram of age-related changes in hypothalamic-pituitary-testicular axis function. Dashed lines denote negative feedback effects. Arrows denote aging effects on hormone levels or spermatogenesis, large arrowheads denote increased androgen feedback sensitivity, and arrows in parentheses denote age-related changes in bioavailable but not total hormone levels. Note that although luteinizing hormone (LH) and follicle-stimulating hormone (FSH) levels may be normal or increased with aging, they are inappropriately low in comparison to those of young men with similar reductions in testosterone levels. Also note that the effect of aging on gonadotropin-releasing hormone (GnRH) secretion is inferred from observation of peripheral hormone levels. N, unchanged; T, testosterone; E2, estradiol; DHT, 5 $\alpha$ -dihydrotestosterone.

## Decline in Testicular and Hypothalamic Regulation

The age-related decline in testosterone levels is a consequence of both primary testicular failure and impaired hypothalamic secretion of gonadotropin-releasing hormone (GnRH) that results in relatively suppressed pituitary secretion of luteinizing hormone (LH) (Fig. 1). Age-related primary testicular dysfunction is manifested by decreases in the number and volume of Leydig cells (testosterone-producing cells) in the aging testis, in basal testosterone secretion rates, and in maximal testosterone secretion after administration of human chorionic gonadotropin (an LH-like hormone). Metabolic clearance of testosterone is also decreased with aging, partially offsetting the decrease in testosterone production.

Spermatogenesis is also adversely affected by aging. Histologic studies show a decrease in spermatogenesis

with advancing age, although sperm concentration in ejaculated semen is unchanged or increased as a result of longer periods of abstinence compared with that of younger men. In addition, the number of Sertoli cells that support spermatogenesis in the seminiferous tubules is decreased with aging. There is a corresponding age-associated decrease in serum levels of inhibin B, which is secreted by Sertoli cells and mediates feedback inhibition of pituitary follicle-stimulating hormone (FSH) secretion. This decline in inhibin B is thought to reflect changes in spermatogenesis in the aging testis. However, most of the decline in inhibin B levels occurs by middle age, with little additional decline in late life. Functionally, sperm fertilizing capacity as determined by *in vitro* sperm penetration is well preserved with aging, although the percentage of sperm with normal morphology and motility is decreased. Fertility rates decrease with aging, but this is due primarily to a decrease in sexual activity.

Impairment of hypothalamic GnRH secretion with aging is thought to lead to a relative decrease in pituitary gonadotropin secretion and to contribute to impairment in testicular function. In turn, the age-related decline in serum testosterone levels is associated with a gradual increase in serum FSH and, to a lesser degree, LH levels as a result of diminishing testosterone negative feedback. However, gonadotropin levels often remain within normal limits for younger men. Even when levels of LH and FSH are elevated, they may remain inappropriately low in comparison to those of younger men with similar reductions in testosterone levels, suggesting hypothalamic-pituitary dysfunction (secondary hypogonadism).

Although hypothalamic GnRH secretion cannot be assessed directly in man, several lines of evidence indicate that age-related secondary hypogonadism is due to a defect at the level of the hypothalamus. First, although gonadotropin responses to acute GnRH administration are mildly impaired with aging, older men exhibit intact LH and FSH responses to prolonged pulsatile administration of GnRH, suggesting unimpaired pituitary gonadotropin secretion with aging. Second, the frequency of pulsatile LH secretion (an indicator of hypothalamic GnRH pulse generator activity) is decreased in healthy elderly men. Third, the circadian variation of serum testosterone levels with highest levels in the morning is attenuated in older compared to young men, suggesting an age-related impairment in the hypothalamic circadian pacemaker mechanism. Finally, the sensitivity of gonadotropin suppression to testosterone negative feedback is increased with aging. Taken together, these findings suggest that the decline in testicular

function with aging is due to both primary testicular failure and impaired hypothalamic GnRH regulation of gonadotropin secretion.

## POSSIBLE CONSEQUENCES OF AGING-ASSOCIATED ANDROGEN DEFICIENCY

The term andropause has been coined to denote an age-related decline in circulating testosterone to levels below normal for young men, associated with symptoms and signs consistent with androgen deficiency. Many of the physiological changes associated with advancing age are similar to the symptoms associated with hypogonadism in younger men, including decreases in muscle mass and strength, bone mass, libido, sexual activity, body hair, and hematopoiesis and increased fat mass. In young hypogonadal men, these symptoms improve with testosterone replacement. It has been hypothesized that in aging men, these problems are due at least in part to declining testosterone levels.

### Epidemiological Studies

Descriptive studies involving older men provide some support for this hypothesis. Testosterone levels and total or abdominal fat mass are inversely related in most studies of elderly men, and some studies report a positive correlation between testosterone levels and lean body or muscle mass and strength. In addition, low testosterone levels are associated with depressed mood and reduced performance on spatial cognition and some memory tasks in older men. Some longitudinal studies in healthy older men report a modest correlation between testosterone levels and sexual activity, whereas other studies find no association. Based on the observation that men have a higher incidence of coronary heart disease than women of similar age, it has been suggested that testosterone predisposes to coronary heart disease. However, most studies report a favorable or neutral correlation between testosterone levels and coronary heart disease in men.

Many of the peripheral effects of testosterone are mediated at least in part by its active metabolites estradiol and DHT (Fig. 1). As noted previously, bioavailable estradiol and DHT levels decline with aging. In older men, decreased bone mineral density and fracture risk are more strongly correlated with low estradiol than with low testosterone levels, suggesting an important role of estradiol in developing

and maintaining normal bone mass in men. Local production of the androgen DHT in the skin and prostate plays an important role in the development of male pattern baldness and prostate disorders, including benign prostatic hyperplasia with aging, but the significance of age-related changes in circulating DHT levels is unclear.

### Controlled Clinical Studies of Testosterone Treatment in Older Men

#### **Potentially Beneficial Effects**

Changes in lean body mass and fat mass are among the most consistently reported effects of testosterone supplementation. In trials of up to 3 years, lean body mass increased and total body and visceral fat mass decreased in healthy older men with low-normal or mildly decreased testosterone levels. Bone mineral density of the lumbar spine and hip increased in older men with low testosterone levels but not in men with testosterone levels in the normal range. However, fracture rates have not been reported to decrease in older men receiving testosterone.

Total and low-density lipoprotein cholesterol levels generally decreased with testosterone supplementation, whereas high-density lipoprotein cholesterol levels were unchanged. Most trials involving older men with coronary heart disease have reported decreased exercise-induced coronary ischemia with acute or chronic testosterone administration, although effects on angina were variable. The long-term effects of testosterone treatment on cardiovascular disease risk in older men are unknown.

Libido, sexual activity, energy, and subjective well-being improved with testosterone treatment in some studies but were unchanged in others. Studies in small numbers of subjects reported improvements in some aspects of cognition, including spatial ability and spatial, working, and verbal memory. However, in another trial testosterone appeared to impair verbal fluency. Testosterone therapy had no effect on clinical depression in one study.

Testosterone's effects on strength and functional status in older men were variable. In healthy older men, testosterone increased upper and lower extremity strength in some studies but not in others. Performance on functional tasks was unchanged in healthy men treated for 3 years, although testosterone appeared to prevent a decline in self-assessed physical function in these men. In frail older men undergoing inpatient rehabilitation, testosterone improved some measures of strength and functional status. However,

the effects of testosterone treatment in the prevention and treatment of frailty, and on health-related quality of life, remain largely unexplored.

### **Adverse Effects**

Testosterone supplementation was generally well tolerated by older men treated for up to 3 years. The most consistently reported adverse effect of testosterone was the development of hematocrit elevations above the normal range in 6–25% of older men. In most studies, the mean hematocrit increased 2.5–5% over baseline values during treatment. Erythrocytosis occurred during both parenteral and transdermal testosterone supplementation, although it was less common with transdermal administration.

Voiding symptoms and prostate examination abnormalities did not increase during treatment, although prostate size increased slightly in one study. Although most studies found no significant change in prostate-specific antigen (PSA), in a few studies PSA levels increased slightly during testosterone treatment, mostly within the normal range. However, the long-term risks of clinically significant prostatism or overt prostate cancer in men receiving testosterone therapy are unknown.

Sleep apnea has been reported as a potential complication of testosterone therapy in younger hypogonadal men, but it has not been observed in controlled studies of testosterone treatment.

## **CONCLUSION**

The foregoing evidence suggests that the age-related decline in testosterone levels contributes to some of the physiological changes associated with aging in men.

However, the available data are insufficient to assess the long-term benefits and risks of testosterone supplementation. Furthermore, age-associated physiological alterations such as altered body composition and sexual dysfunction are usually multifactorial in origin. Therefore, clinicians must identify and treat all potentially treatable or reversible contributory factors.

## **See Also the Following Articles**

Aging and Longevity of Human Populations • FSH (Follicle-Stimulating Hormone) • Gonadotropin-Releasing Hormone (GnRH) Actions • Gonadotropins and Testicular Function in Aging • Impotence and Aging • LH (Luteinizing Hormone) • Neuroendocrine System and Aging • Sexual Function and Androgens

## **Further Reading**

- Anonymous (2001). Report of the National Institute on Aging Advisory Panel on testosterone replacement in men. *J. Clin. Endocrinol. Metab.* **86**, 4611–4614.
- Bhasin, S., and Buckwalter, J. G. (2001). Testosterone supplementation in older men: A rational idea whose time has not yet come. *J. Androl.* **22**, 718–731.
- Gruenewald, D. A., and Matsumoto, A. M. (2003). Testosterone supplementation therapy for elderly men: Potential benefits and risks. *J. Am. Geriatr. Soc.* **51**, 101–115.
- Klee, G. G., and Hesser, D. W. (2000). Techniques to measure testosterone in the elderly. *Mayo Clinic Proc.* **75**(Suppl.), S19–S25.
- Matsumoto, A. M. (2002). Andropause: Clinical implications of the decline in serum testosterone levels with aging in men. *J. Gerontol. Med. Sci.* **57A**, M76–M99.
- Swerdloff, R. S., *et al.* (eds.) (2002). “Proceedings of the Endocrine Society Andropause Consensus Conferences.” Endocrine Society Continuing Medical Education Services.
- Vermeulen, A. (2001). Androgen replacement therapy in the aging male—A critical evaluation. *J. Clin. Endocrinol. Metab.* **86**, 2380–2390.



# Aging, Animal Models for

George A. Kuchel, Qing Zhu, Blanka Rogina, and Stephen H. Clark  
*University of Connecticut Health Center, Farmington, Connecticut, United States*

## Glossary

**DNA sequencing** Technique used to determine the sequence of nucleotide bases present in a DNA molecule or fragment.

**epithelial** Layers of cells forming the outer surface of the body and also lining its internal cavities.

**mutagenesis** Causing a permanent heritable change in the nucleotide sequence of a chromosome.

**phenotype** Observable characteristics of a cell or an organism.

**transgenesis** The stable introduction of a cloned gene into a plant or an animal, allowing its transmission to future generations.

Scientists interested in exploring fundamental mechanisms of aging need to judiciously choose animal models of aging most appropriate for their area of research. Similarly, great care must be taken when using animal models for the study of disease processes that tend to occur in late life. The establishment of funding agencies devoted to aging research (e.g., U.S. National Institute on Aging [NIA] in 1974) and a growing general interest in aging have led to an unprecedented increase in the use of aging models and in resulting publications. Although researchers are now able to choose from a broad variety of vertebrate and invertebrate models, most such research is still conducted using a very limited number of rat or mouse strains. Because no one single species or strain is optimal for all aging questions and studies, care must be given to the choice of the most appropriate animal aging model. Moreover, other animal-related issues, such as the presence of confounding disease, genetic background, diet, housing, husbandry, microbial status, and exercise, all can greatly influence experimental results. Thus, when designing aging studies, these factors must be given the same considerations as is the determination of an outcome measure of primary interest to investigators. Fortunately, although important issues still remain unresolved, researchers are now able to base their decisions on a reasonable body of experience and knowledge.

## GENERAL PRINCIPLES OF AGING RESEARCH

As described by Weindruch, animal models have been used in five general categories of aging studies. Many investigators have used cross-sectional and (less commonly) longitudinal study designs to examine the influence of normal aging on parameters that are thought to either cause aging or be the consequence of aging processes. One major pitfall of this type of study is attributing observed differences to normal aging without taking into account the presence of confounding disease processes. Other difficulties arise from an inappropriate choice of ages. For example, an investigator comparing very young (e.g., newborn) rats with older adult (e.g., middle-aged) rats may be studying maturational rather than aging processes, and a failure to include intermediate ages between young and old may cause the investigator to miss important changes that may herald the aging process. Interestingly, the absence of appropriate life tables for the determination of survival curves has been the limiting factor in the use of some potentially useful aging models.

Other, more comparative studies examine species with widely different longevity to identify factors associated with longevity and aging. Similarly, different strains within the same species (e.g., congenic mice) may possess major differences in longevity. Finally, animal models of accelerated aging (e.g., senescence accelerated mouse) and models of decelerated aging (e.g., *Drosophila* overexpressing catalase or superoxide dismutase) both have proved to be useful in aging studies.

## THE LABORATORY RAT

NIA's decision to provide the Fischer 344 (F344) rat strain to the scientific community has played a major

role in its great popularity for aging studies. Many investigators also welcomed this inbred strain because the commonly used outbred Sprague–Dawley strain tended to become grossly obese in old age and exhibited tremendous variability when purchased from different suppliers. Unfortunately, although inbred F344 rats have the advantage of being genetically identical, this also results in their frequent development of tumors (testicular interstitial cell and pituitary) as well as renal failure from nephropathy. Although these tumors are rarely metastatic or secretory, screening is important because both the tumors and renal failure could confound research data. Fortunately, the NIA also offers the inbred Brown Norway (BN) rat, as well as an F1 hybrid (F344/BN), through a contractual arrangement. Caloric restriction still remains the best validated and most robust means of extending longevity in rodents while also decreasing their burden of disease (including F344 nephropathy), and all three rat strains provided by the NIA are also available from a caloric-restricted colony. Many investigators have argued for the use of rodents that have undergone modest caloric restriction because such animals tend to be healthier in old age, minimizing the confounding effect of illness and disability. Interestingly, some investigators have used the heterogeneity present in other outbred strains (e.g., Wistar, Long–Evans) to great advantage in aging studies. For example, Long–Evans rats remain relatively healthy into old age and exhibit great variability in their spatial memory performance on a Morris water maze test.

The rat is a popular model for examining endocrine and reproductive issues in old age. Although this has resulted in the availability of considerable information regarding this model, great caution must be exercised when extrapolating from these studies to human conditions. For example, unlike women whose ovaries cease to produce estrogens past menopause, rat ovaries appear to be capable of normal or near-normal function throughout their life span. In contrast to human menopause, reproductive senescence in rats is associated with the development of irregular cycles. This is followed by a period constant estrus in nearly one-half of the animals, which then proceed to a state of persistent diestrus, whereas the other half proceeds immediately to persistent diestrus. As a result, approximately one-half of 2-year-old rats will be in a state of constant estrus with a minimum of three consecutive epithelial vaginal smear cycles, whereas the other half will proceed to a state of persistent diestrus with a minimum of three leucocytic smears. Interestingly, estrogen levels are only slightly higher in the former group. As discussed previously,

a high prevalence of pituitary and testicular tumors, particularly in aged F344 rats, also needs to be considered when using this strain in aging studies.

## THE LABORATORY MOUSE

Being smaller than rats, mice provide investigators with smaller amounts of tissue, and systemic physiological studies can sometimes be difficult. However, mouse genetics are extremely well known, and the ability to create genetically modified mice has become an important tool for studying aging and age-related processes. The presentation of a phenotype in any organism is a composite of the interactions between its genetic complement and the environment in which it is placed. Although the process of aging is a complex phenotype, it is no exception to this generally accepted biological principle. In contemporary biomedical research, the laboratory mouse has become a premier model organism to study processes important to human biology. This surge in mouse-related research has been driven largely by the ability to develop genetically engineered mice both by transgenesis via pronuclear injection and with gene targeting using embryonic stem (ES) cells. In addition, large mutagenesis programs have been initiated to isolate new mutations associated with a wide variety of biological processes. These activities have been complemented with other technological advances, including the initiation of a DNA sequencing effort to completely sequence the mouse genome and the development of microarray technology. Although the promise of these efforts in contributing to the genetic dissection of complex biological processes is extraordinary, having the appropriate model to which one can apply these emerging technologies is paramount to a successful and productive outcome.

Although the mouse has many advantages, it is not without its disadvantages. It must be recognized that not all discoveries or observations made with the mouse will have direct parallels in the human. Furthermore, with a life span of many months as compared with days in a species such as *Drosophila*, another powerful genetic model, data collection is inherently slower. Finally, housing costs and the operation of a barrier facility is a major consideration in the selection of this model.

The mouse model provides an opportunity for investigators to control both the nutrition and housing environments for their experiment animals. It has been well documented that caloric restriction can substantially lengthen the life span of laboratory mice. This

principle seems to be true not only for mice and other rodents but also for insects, worms, and nonhuman primates, with possible implications for humans. Clearly, continued work with genetic models such as *Drosophila* and the mouse will ultimately reveal the genetic basis for this phenomenon. In the design of an aging study, the program for housing animals should be evaluated carefully. Caging style, light–dark cycle, ambient temperature, and the number of mice housed per cage all are obvious concerns. Other factors for consideration include extraneous noise and consistency in caretaker handling of the animals. Careful training of caretakers for uniformity in handling of the mice during cage-changing activities should not be underestimated.

Infectious diseases have long been known to have a potential impact on experimental analyses in mice. It is important to be able to determine whether the onset of a disease state is a function of the aging process or the result of a pathogen infection in the colony. In this context, the development and maintenance of specific pathogen-free (SPF) colonies for aging studies is essential. A researcher is also aided in this area by the existence of a large literature base on diseases in mice and on variations in susceptibility to diseases in specific lines or strains. The SPF status of study animals should be defined and maintained by careful husbandry practices. Frequently, an SPF barrier facility equipped with micro-isolator caging is employed to minimize the risk of an unwanted pathogen entering the colony. Access to such barriers is limited, and strict protocols for entry and animal-handling practices are mandated for staff working in such areas. A carefully designed sentinel program is essential for monitoring the health status of the colony and should be associated with a careful plan to respond if a break in the SPF status should occur.

Many inbred strains of mice have been developed and maintained by brother–sister matings for numerous generations. One major advantage of an inbred strain is genetic uniformity from one generation to the next. This feature permits an experiment to be repeated in both time and space with an identical or nearly identical genetic environment. In addition, a wealth of information on individual strain characteristics has been accumulated through years of research with these lines. Several inbred strains have been used in aging studies and can be obtained from a variety of sources. Inbred mice currently available through the NIA include BALB/cBy, CBA, C57BL/6, and DBA/2. Hybrids and calorically restricted mice will be increasingly available in future years. However, genetic uniformity of the inbred strain does not come without a

price in that certain lines come with their own particular set of problems that can affect an aging study. In the selection of a particular strain for analysis, previous work with that strain, the nature of the aging question being addressed, and potential strain idiosyncrasies must be evaluated. Fortunately, databases that allow rapid electronic access to much of this information have been developed to assist investigators with strain selection.

One approach to reduce the potential problems associated with the use of inbred strains is the production of an F1 hybrid mouse that is created by mating individuals from two different inbred strains. These F1 animals have the virtue of being genetically identical as well as being potentially heterozygous at most loci (F1 mice will be homozygous only for loci that happened to be identical in their inbred parents), reducing the impact of certain idiosyncrasies found in their inbred parents.

The number of variants at individual genetic loci has grown precipitously as a function of genetic engineering and focused mutagenesis programs that have been initiated relatively recently. Because designer variants and many spontaneous mutations are often isolated on different or mixed genetic backgrounds, breeding programs are frequently conducted to place each variant in a defined genetic background. This is accomplished by a series of back-crosses of the variant locus to a selected inbred strain such as C57BL/6. Lines established in this fashion are referred to as congenic strains. To avoid the complications of studying a variant in an inbred setting, a second congenic line in a different inbred strain can be created. By crossing these two congenic strains, a particular variant can be analyzed in an F1 hybrid and compared with the phenotype in the inbred strain.

Although the inbred strains and F1 hybrids allow rigid control of the genetic environment in a set of experimental animals, not all researchers agree with this approach. Concerns about the lack of genetic variance in these experimental paradigms (inbred lines and F1 hybrid) has led some researchers to favor experimental systems that provide more genetic heterogeneity. One experimental paradigm involves F2 hybrids. In this model, experimental mice are obtained by mating two F1 mice of the same type (e.g., F1 mice produced by crossing C57BL/6 mice with C3H mice). Alternatively, F2 animals can be developed using the four-way breeding scheme. This breeding program involves the mating of two different F1 parents. In this type of mating system, the resulting F2 animals have four different grandparents, whereas they have only two different grandparents if the F1 parents are

of the same type. In either system, the F2 mice are genetically heterogeneous, yet the genetic makeup of the population is defined and reproducible from laboratory to laboratory.

## INVERTEBRATE MODEL SYSTEMS IN AGING

*Saccharomyces cerevisiae* (yeast), *Caenorhabditis elegans* (nematode), and *Drosophila melanogaster* (fruit fly) are the leading invertebrate model systems used in aging research. Their great advantages include ease of maintenance, short life spans, advanced molecular and genetic tools, and availability of their complete genomic sequences. Genetic screens have already identified a number of genes that appear to be involved in the aging process. Genetic and molecular characterization of these genes has provided valuable information about the physiological pathways that may be important in the aging process. Furthermore, the molecular conservation that has been so widely observed among yeast, *C. elegans*, *Drosophila*, and mammalian model systems such as mice and rats suggests that information obtained in these “simpler” systems is certain to facilitate our understanding of aging.

The budding yeast, *S. cerevisiae* (a single-cell organism), has been used in aging research since 1959, when Mortimer and Johnson described an increase in volume and number of scars as an aging phenotype. Aging of the yeast is measured as the average number of cell divisions that a mother cell can undergo or as the survival of nondividing cells during the stationary phase. The asymmetric cell divisions result in a larger mother cell and smaller daughter cells that can be easily separated, leaving behind the scar. Average cell divisions can range from 15 to 30 generations, depending on the strain. During the past decade, more than 16 longevity genes have been discovered in yeast, many of which are now being found to be important in the aging process of other invertebrates and mammals.

The nematode, *C. elegans*, is a very valuable tool in aging research. Some of the specific advantages of nematodes are their easy maintenance; small size (~1 mm as adult), short life span (~15 days at 20°C), and ability to be frozen indefinitely in liquid nitrogen and revived. Moreover, this relatively simple transparent organism is the only multicellular animal whose developmental anatomy has been described in great detail. The exact lineage of every cell in the body has been fully mapped out by direct observation of the developing animal from a single-cell egg to a

multicellular adult. Its genome has been completely sequenced. Recently developed microsurgical and genetic approaches can be applied to dissect the genetic control of specific developmental mechanisms.

*C. elegans* is a hermaphrodite, allowing for self-fertilization. As a result, a single heterozygous worm can produce homozygous progeny, an important feature that helps to make this organism exceptionally convenient for genetic studies. Meanwhile, genetic crosses are also possible given that the nematode can be mated with males. During normal development, *C. elegans* hatches from the egg as a first-stage larvae L1. After three more larval stages, it becomes sexually mature. Crowding, high level of pheromones, and food limitation induce entering of L2 to the dauer larval stage, an alternative L3 form that allows nematodes to survive difficult conditions. Dauer is a developmentally arrested form characterized by a lower metabolic rate, accumulations of fat, an increased level of antioxidant enzymes, and a longer life span. A nematode can survive during the dauer stage for up to 6 months. When conditions improve, the nematode develops into an L4 stage with a normal life span. Mutations in several genes extend the life span and affect dauer formation. Molecular and genetic characterizations of these genes reveal that some of them are components of the insulin-signaling pathway regulating both dauer formation and life span.

*D. melanogaster* or closely related species have been used to study aging since at least since 1915. There are numerous advantages such as the fact that fruit flies are not expensive to breed, are easy to maintain, and have a relatively short life span. *D. melanogaster* live about 2 months at 25°C. There are a number of environmental and genetic manipulations that can alter the life span of *Drosophila*. For instance, flies live about 3 months at 18°C and live only 1 month at 29°C. Also of interest is that a majority of the cells of adult fruit flies are postmitotic, with exceptions of some gut cells and gonads. Finally, the molecular genetic tools that have been developed over the past 90 years or so of *Drosophila* research have reached a point where experimenters have great control over the molecular genetic life of the fly. Through a combination of various already available techniques, it is possible to alter, increase, or decrease the level of expression of any gene or set of genes in any cell or group of cells at any time in life, from development to old age.

Reproduction has a strong negative effect on longevity in *D. melanogaster*. Both virgin female and male flies live significantly longer than do fully mated ones. The costs of reproduction in flies include energy

allocation for courtship and mating, egg production and egg laying, and direct toxic effects of mating. Although egg production does not play a role in aging in *C. elegans*, a signal from the germ line is involved in limiting life span such that removal of this signal results in an increase in life span. In addition to the reproductive system, other hormones, such as the insulin/insulin-like growth factor-1 (IGF-1) signaling pathways, have been shown to affect life span in nematodes and flies as well as in mice. Mutations in components of the highly conserved insulin pathway in *C. elegans* and *Drosophila* result in life-span extension.

## NONHUMAN PRIMATES

The high cost and small availability of aged nonhuman primates have limited their use to selected research centers. Further compounding the difficulties in conducting research with these animals is their long life span and a lack of good survival data. At the same time, the phylogenetic proximity of these primates to humans makes these highly attractive, and in some cases even indispensable, models for the study of human aging and disease. For example, true menstruation occurs only among some primate species, making some nonhuman primates invaluable models for the study of female human menopause. Another area is brain structure, where only nonhuman primates represent adequate models for the study of many aspects of higher cortical structure and physiology.

## See Also the Following Articles

Aging and Longevity of Human Populations • Aging and the Male Reproductive System • Aging, Immunology and

• Autonomic Nervous System, Aging and • Body Weight, Body Composition, and Aging • Functional Genomics of Aging • Neuroendocrine System and Aging • Oxidative Stress and Aging • Stress, Aging, and Central Nervous System Interactions • Thyroid, Aging and

## Further Reading

- Finch, C. E. (1990). "Longevity, Senescence, and the Genome." University of Chicago Press, Chicago.
- Finch, C. E. (1991). New models for new perspectives in the biology of senescence. *Neurobiol. Aging* **12**, 625–634.
- Hekimi, S., Burgess, J., Bussiere, F., Meng, Y., and Benard, C. (2001). Genetics of lifespan in *C. elegans*: Molecular diversity, physiological complexity, mechanistic simplicity. *Trends Genet.* **17**, 712–718.
- Helfand, S. L., and Rogina, B. (2003). From genes to aging in *Drosophila*. *Advances in Genetics* **49**, 67–109.
- LeFevre, J., and McClintock, M. K. (1988). Reproductive senescence in female rats: A longitudinal study of individual differences in estrous cycles and behavior. *Biol. Reprod.* **38**, 780–789.
- Lu, K. H., Hopper, B. R., Vargo, T. M., and Yen, S. S. (1979). Chronological changes in sex steroid, gonadotropin, and prolactin secretions in aging female rats displaying different reproductive states. *Biol. Reprod.* **21**, 193–203.
- Nelson, J. F. (1995). The potential role of selected endocrine systems in aging processes. In "Handbook of Physiology" (E. J. Masoro, ed.), pp. 377–394. Oxford University Press, New York.
- Sinclair, D. A. (2002). Paradigms and pitfalls of yeast longevity research. *Mechanisms Ageing Dev.* **123**, 857–867.
- vom Saal, F. S., Finch, C. E., and Nelson, J. F. (1994). Natural history and mechanisms of reproductive aging in humans, laboratory rodents, and other selected vertebrates. In "The Physiology of Reproduction" (E. Knobil and J. D. Neill, eds.). Raven Press, New York.
- Weindruch, R. (1995). Animal models. In "Handbook of Physiology" (E. J. Masoro, ed.), pp. 37–52. Oxford University Press, New York.



# Aging, Immunology and

Victor Appay

Institute of Molecular Medicine, John Radcliffe Hospital, London, United Kingdom

## Glossary

**antigen** A substance that can elicit an immune response following recognition by T cells or B cells. Antigens can be proteins, carbohydrates, nucleic acids, lipids, or hormones. The immune system usually responds only against foreign antigens (e.g., from pathogens), but it can also react against self-antigens as if they were foreign, which can result in autoimmune diseases.

**B cells** A population of white blood lymphocytes found in bone marrow, lymph nodes, and blood circulation. B cells are generated and mature in the bone marrow; their role is to produce antibodies to bind and neutralize pathogens and toxins. Each antibody is strictly specific to only one antigen.

**cytokines** Proteins synthesized by cells of the immune system. Cytokines include an extensive range of mediators (interleukins, interferons, chemokines, and growth factors) with various functions (e.g., immunostimulatory or immunosuppressive).

**immunosenescence** Deterioration of the immune system or its function with time.

**T cells** A population of white blood lymphocytes found in the thymus, lymph nodes, and blood circulation. T-cell precursors are generated in the bone marrow; following their maturation in the thymus, naive T cells are exposed to foreign antigens and produce memory T cells, which are long-lived antigen-experienced cells that respond rapidly to a subsequent exposure with the same antigen. T cells have many functions, including the elimination of foreign pathogens or infected cells and the regulation of the immune response by assisting or suppressing other cells of the immune system.

Immunology is the study of the immune system or immunity, a range of defenses developed by humans, for example, to protect against infection by surrounding microorganisms (viruses, bacteria, fungi, and other parasites) and related dangerous pathologies. Immunogerontology is the study of the immune system in the context of aging (i.e., in the elderly).

## INTRODUCTION

The study of our native immunity began in the late 18th century and has led to incredible discoveries regarding the role of our immune system and the mechanisms through which pathogens attempt to penetrate our defenses. In comparison, immunogerontology is a relatively new field that is gradually becoming more relevant because the older population is increasing at an unprecedented rate due to progress in health care, especially in developed countries, where life expectancy has nearly doubled during the past century. The aging of the human population represents a considerable challenge to public health authorities, raising both health care-related and socioeconomic issues.

The functional capacity of the immune system gradually declines with age so that the immune system cannot respond as quickly or as efficiently to stimuli, particularly new antigens. This deterioration, referred to as immunosenescence, is reflected by deregulation of the immune system, including changes in cellular phenotype and function, and alterations of whole organs. These changes are believed to contribute to the increased susceptibility to and severity of infectious disease, cancer, and autoimmune disease, which characterize individuals of advanced age. Research on the immunology of aging is needed to increase our understanding of the immunosenescence process and potentially to develop therapeutic intervention.

## IMMUNE ALTERATIONS ASSOCIATED WITH AGING

### Hayflick Limit

In 1961, Hayflick and Moorhead showed that cultures of fetal human fibroblasts could reach an irreversible state of growth arrest—the first evidence for replicative senescence. Accumulating evidence suggests that the so-called Hayflick limit is not restricted to cultured human fibroblasts but applies also to the cells of the immune system, which may have a limited replicative life span *in vivo*. The concept of limited replicative life

span is important in the context of aging and immunosenescence. The occurrence of replicative senescence is primarily related to the number of cell divisions, varying according to cell type and cellular environment. A commonly used marker of replicative history is the length of the telomeres (DNA repeats at the end of chromosomes), which is reduced after each cell division. Cells that have reached a state of replicative senescence have short telomeres.

## Hematopoiesis

All the cells that constitute our immune system originate from hematopoietic stem cells (in the bone marrow) that differentiate and commit themselves to a specific cellular lineage (e.g., myeloid or lymphoid) to continuously generate new granulocytes or naive lymphocytes. Aging appears to result in a deregulation of hematopoiesis: Progenitor cells in elderly individuals are present in lower numbers and exhibit a decline in their ability to generate new cells. Hematopoietic stem cell telomeres appear to be shorter in adults than in the cord blood of newborns. Moreover, granulocytes and naive T cells show a shortening in telomere length associated with age, suggesting that this also applies to stem cells. Increasing evidence supports the hypothesis that stem cells have a limited replicative potential so that their capacity to replenish the cell population of the immune system is restricted. Although there is debate regarding whether this phenomenon has a real consequence on the immune function in aging, it implies that the capacity of the immune system is not unlimited and can reach exhaustion over time.

## T Cells

During the past two decades, Graham Pawelec, Rita Effros, and others have provided strong evidence that aging results in important alterations of the T-cell function.

### Thymus Involution

The thymus is the organ in which T-cell precursors mature into naive T cells capable of mounting an immune response against foreign antigens. With aging, the thymus naturally atrophies, a process called thymic involution. At birth, the thymus is involved in massive generation of T cells to fill the immunological space and provide the elements necessary to protect the host against pathogens. T cells are produced continuously throughout life; however, over time progressive decay of the thymus results in decreased

production of T cells and a lower number of naive T cells in the elderly.

### Post-Thymic Development

When encountering a foreign antigen, naive T cells become activated and differentiate into antigen-experienced T cells that exhibit direct effector functions and can expand greatly to eliminate foreign antigens from the body. The ratio of naive to antigen-experienced T cells differs substantially in young and old subjects; it is high in newborns but decreases with age. Aging is also associated with a variety of changes in the characteristics and function of antigen-experienced T cells: (i) changes in T-cell receptor signal transduction and in the expression of cell surface receptors, including the loss of costimulatory receptors (e.g., CD27 and CD28) involved in activation; (ii) changes in the profile of cytokine expression (e.g., decreased production of interleukin-2, which is necessary for proliferation); (iii) reduced capacity to expand upon stimulation, associated with shorter telomere length, and increased susceptibility to undergo activation-induced cell death; and (iv) restricted T-cell repertoire so that T cells recognize a narrower range of antigens. All these functional changes have been postulated to play a role in the decreased immunocompetence associated with age. Interestingly, these characteristics may reflect a change in the T-cell subset composition, with a shift toward more differentiated antigen-experienced T cells that exhibit these characteristics (Fig. 1).

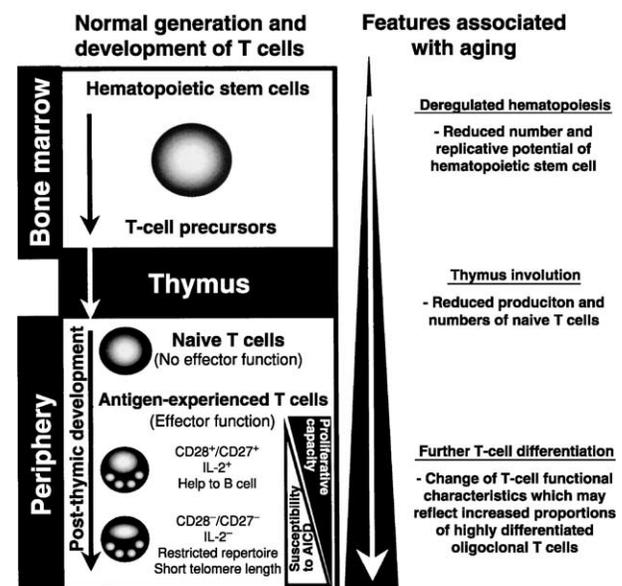


Figure 1 Alterations in the T-cell compartment associated with aging.

## Other Alterations

The T-cell compartment is clearly the most affected during the process of aging. However, aging also affects the quantity and quality of protective antibodies produced by B cells, or humoral immunity. These alterations reflect a change in the B-cell repertoire and include a decline in the production and activity of T-cell-dependent B<sub>2</sub> cells compared to T-cell-independent B<sub>1</sub> cells, associated with a reduction in the diversity and quantity of serum antibodies specific for foreign antigens and a shift in specificity toward autologous antigens (with production of auto-antibodies). Other cell types (including natural killer cells, macrophages, monocytes, neutrophils, and eosinophils) may also be affected by aging. However, there are few data available and results have been contradictory. More research is needed to resolve these discrepancies.

## IMPLICATIONS OF IMMUNOSENESCENCE

### Onset of Immunosenescence

The precise mechanisms involved in the development of immunosenescence are not completely understood; however, the following hypothesis has been developed from the available evidence. In adults, the ratio of naive to antigen-experienced T cells is progressively reversed as naive T cells are increasingly exposed to antigens, which also drives T cells to differentiate toward stages of replicative senescence. However, with age, the naive T-cell population is not replenished, likely a consequence of the gradual deterioration of the thymus, and highly differentiated oligoclonal senescent antigen-experienced cells accumulate. The very old individual may become practically devoid of naive T cells. Immunosenescence may therefore be the result of the continuous challenge by a variety of antigens (e.g., pathogens) to which the immune system is exposed throughout a lifetime. The maintenance of a protective immunity over time may cause the eventual exhaustion of the finite capacity of the adaptive immune system. The reduced function of T cells in the elderly may affect B-cell function because T cells play a role in the regulation of B cells and the production of antibodies. Although further investigation is required, the potential reduction in the replicative capacity of the hematopoietic stem cells might have a significant impact on thymus involution and the decline in the production of

naive T cells as well as on the overall alteration of homeostasis.

### Consequences of Immunosenescence

T cells play a prominent role in immunity, and alteration of their function has profound effects on the entire immune system and the host defense against pathogens. The age-associated decline in immunocompetence results in an increase in the incidence and/or severity of many infectious diseases (e.g., influenza, pneumonia, meningitis, sepsis, varicella zoster virus, and HIV), particularly in developing countries, in which health care is not optimum. Similarly, this decline may also be related to the prevalence of cancers in the elderly. In addition, deregulation of immune function may be involved in the onset of noninfectious diseases such as autoimmune disorders (e.g., diabetes, multiple sclerosis, and atherosclerosis) due to a decrease in the ability to discriminate between “self” and “nonself” and possibly neurodegenerative disorders such as Alzheimer’s disease. It is also well documented that the elderly do not respond to vaccinations as well as young individuals, as exemplified by influenza vaccination. In the elderly, T-cell and antibody responses to vaccines are slower and not as strong as in younger people, which makes the development of vaccines targeted for older adults a particular challenge. Interestingly, many of the immune characteristics associated with aging are also common to other conditions, independent of age, but result in a decline in immunocompetence (e.g., HIV infection and the genetic disorder ataxia telangiectasia). These observations link immunosenescence with the development of some degree of immunodeficiency. In the context of HIV infection, the virus is thought to cause elevated and chronic immune activation that may lead to premature exhaustion or aging of the immune resources, increasing the progression toward AIDS.

### See Also the Following Articles

Aging, Animal Models for • Autonomic Nervous System, Aging and • Immune System, Hormonal Effects on • Neuroendocrine System and Aging

### Further Reading

Appay, V., and Rowland-Jones, S. L. (2002). Premature ageing of the immune system: The cause of AIDS? *Trends Immunol.* **23**(12), 580–585.  
Dodet, B. (2000). *Vaccine* **18**(16), 1565–1724. [Special issue]

Effros, R. B., and Pawelec, G. (1997). Replicative senescence of T cells: Does the Hayflick limit lead to immune exhaustion? *Immunol. Today* **18**(9), 450–454.

National Institute on Aging and the National Institute of Allergy and Infectious Diseases of the National Institutes of Health (1996). Report of the Task Force on Immunology and Aging.

Available: NIAID Office of Communications, Bldg. 31, Room 7A50, 31 Center Drive, MSC 2520, Bethesda, MD 20892-2520.

Pawelec, G., Barnett, Y., Forsey, R., Frasca, D., Globerson, A., McLeod, J., Caruso, C., Franceschi, C., Fulop, T., Gupta, S., Mariani, E., Mocchegiani, E., and Solana, R. (2002). T cells and aging, January 2002 update. *Frontiers Biosci.* **1**(7), 1056–1183.



# Aging: Muscle

Blake B. Rasmussen and Elena Volpi

University of Southern California, Los Angeles, California, United States

## Glossary

**adrenopause** The gradual decrease in DHEA that normally occurs with aging.

**andropause** The decrease in androgens (primarily testosterone) that gradually occurs in men during the normal aging process.

**menopause** The termination of ovulation resulting in an abrupt decrease in circulating estrogens in women.

**muscle plasticity** An inherent property of skeletal muscle that allows muscle cells to adapt to external and internal stimuli.

**sarcopenia** The involuntary loss of muscle mass and function with aging that slowly develops over decades and becomes a significant contributor to disability in the older population.

**somatopause** The gradual decrease in circulating growth hormone and insulin-like growth factor-I that normally occurs with aging.

Aging is associated with a number of changes in muscle mass and function that may be caused at least in part by age-related endocrine changes and, conversely, may also be responsible for changes in the effects of some hormones. The most prominent effect of age is the involuntary loss of muscle mass, strength, and function called sarcopenia. Sarcopenia increases the risk of falls and vulnerability to injury and, consequently, can lead to functional dependence and disability. A decrease in muscle mass is also accompanied by changes in body composition and bone density, and it is associated with an increased incidence of insulin resistance in the elderly. A reduction in endocrine function, physical activity, and appropriate nutrition all play important roles in regulating muscle mass during aging.

## INTRODUCTION

Muscle is profoundly affected by age. Muscle mass decreases approximately 3–8% per decade after age 30, and this rate of decline is even higher after age 60. The involuntary loss of muscle mass, strength, and

function is a fundamental cause of and contributor to disability in older people. This has been called sarcopenia. Sarcopenia may lead to a reduction in physical activity, which may have possible metabolic effects, including decreased bone density, obesity, and impaired glucose tolerance.

The etiology of sarcopenia is not clearly understood, but several mechanisms have been proposed, including irreversible fiber damage or permanent denervation resulting in a loss of contact between the nerve and muscle, mitochondrial DNA deletion mutations subsequent to oxidative damage, changes in satellite cell recruitment, altered endocrine function (e.g., changes in hormone, growth factor, and/or cytokine release) and/or impaired muscle responsiveness to the hormonal stimuli, changes in muscle response to nutrients and/or malnutrition, and disuse (e.g., sedentary lifestyle). Most likely, sarcopenia is a multifactorial problem. This article examines the last three possibilities that are directly or indirectly related to the endocrine system: endocrine dysfunction, nutrition, and disuse.

## MUSCLE CHANGES WITH AGING

A variety of morphological and biochemical changes are detected in the aged human body and in skeletal muscle. Here, the alterations in body composition, skeletal muscle tissue, and muscle metabolism are discussed.

### Body Composition

There is a progressive decrease in fat-free mass and a progressive increase in fat mass as a person ages. Furthermore, bone density decreases, joint stiffness increases, and there is a small reduction in stature (kyphosis). These changes in body composition have probable implications for several conditions, including type 2 diabetes, obesity, heart disease, and osteoporosis.

## Muscle Tissue

The primary alteration is a reduction in total skeletal muscle mass. After age 30, the average person will lose approximately 3–8% of their muscle mass with each passing decade of life. Specific cellular alterations include reductions in muscle cell number, muscle twitch time and twitch force, sarcoplasmic reticulum volume, and calcium pumping capacity. Sarcomere spacing becomes disorganized, muscle nuclei become centralized along the muscle fiber, the plasma membrane of muscle becomes less excitable, and there is a significant increase in fat accumulation within and around the muscle cells. Neuromuscular alterations include a decrease in the nervous firing rate to muscle, the number of motor neurons, and the regenerative abilities of the nervous tissue. Motor unit size also increases.

## Muscle Metabolism

Biochemical changes are also associated with the aging process and include a reduction in glycolytic and oxidative enzyme activities, creatine phosphate and ATP stores within the muscle cell, mitochondrial volume, and a slight reduction in overall metabolic rate (~10%). These metabolic changes in muscle contribute to the overall physical fitness capacity of the elderly and are an important component of the ~30% reduction in ability to use oxygen during exercise (i.e.,  $\dot{V}O_{2max}$ ). Initial studies on a small number of elderly subjects suggested that aging is associated with a reduction in basal muscle protein synthesis, which might be responsible for the progressive reduction in muscle mass. However, recent data obtained by our group from the largest cohort of healthy older men did not confirm the results of earlier reports, leading to the conclusion that differences in basal muscle protein turnover between elderly and young men cannot explain muscle loss with age and suggesting that future research should focus on responses to specific stimuli, such as nutrition, exercise, or disease.

## AGE-RELATED ENDOCRINE CHANGES RELEVANT TO MUSCLE

A variety of hormonal changes are seen during the aging process. We have selected the most important ones in relation to their effect on skeletal muscle.

### Andropause

The primary and most potent androgen is testosterone. In approximately 60% of men older than age 65,

testosterone levels decrease below normal levels of younger men. Testosterone levels gradually decrease throughout the aging process (unlike the immediate decrease in estradiol seen with menopause). It has been proposed that the decrease in testosterone may cause a decrease in muscle protein synthesis and result in a loss of muscle mass. With this in mind, several studies have examined the effect of testosterone replacement therapy in men with overt hypogonadism or testosterone concentrations at the lower normal range. Testosterone was administered via injection, transdermal patch, or dermal gel. From these studies, it was shown that testosterone replacement to mid-normal levels resulted in a significant increase in muscle mass, muscle strength, muscle protein synthesis, and bone density. The results are positive and may lead to a reversal or attenuation of sarcopenia. However, testosterone is currently not recommended for the treatment of sarcopenia, and a careful evaluation of the potential benefits compared to the potential risks (e.g., increased prostate-specific antigen, hematocrit, and cardiovascular risk) should be performed before initiating testosterone replacement therapy for sarcopenia.

### Menopause

Estradiol levels abruptly decrease during menopause. Very little information is available regarding the role of menopause in sarcopenia. It appears that muscle mass is not affected by the decrease in estrogens. Cross-sectional studies evaluating the effects of age on lean body mass and appendicular muscle mass have shown that the rate of decline of muscle mass in women does not increase after menopause, suggesting a marginal role, if any, of this event in the development of sarcopenia in women. On the other hand, hormone replacement therapy can significantly increase serum steroid hormone-binding globulin, which leads to a significant decrease in serum free testosterone levels in women. Low serum free testosterone levels in women are associated with a lower muscle mass. Therefore, hormone replacement therapy may play a role in further reducing, rather than increasing, muscle mass in older women.

### Somatopause

The growth hormone (GH)/insulin-like growth factor-1 (IGF-1) axis also exhibits a gradual decline during normal aging. Although providing GH replacement therapy to GH-deficient adults resulted in

an increase in muscle mass and strength, recent studies in which GH replacement therapy was given to aged individuals have shown no effect on muscle strength. Therefore, GH replacement therapy in the elderly may be beneficial for lowering fat mass, improving blood lipid profiles, and increasing lean body mass, but these changes do not lead to an increase in muscle strength and function. In fact, muscle strength increased only when GH was given to elderly men undergoing a weight-training program compared to GH replacement therapy alone. It is also important to note that the methodologies used to measure body composition may be affected by water retention. Thus, an increase in muscle mass with no change in strength following GH therapy should be interpreted with caution because GH is notorious for increasing water retention, which can be misinterpreted as an increase in lean body mass. Regarding testosterone, GH replacement is not currently recommended for the treatment of sarcopenia due to both the unclear efficacy on muscle strength and function and the potentially serious side effects (arthralgia, edema, insulin resistance, cardiovascular risk, etc.).

### Adrenopause

The concentrations of DHEA in the blood also decrease gradually with normal aging. In fact, levels may be up to five times lower in very old men compared to younger men. Oral supplementation of DHEA restores DHEA levels in older persons to younger values, increases IGF-1 levels in men and women, increases estrogens in men, and increases testosterone in women. Unfortunately, no changes in lean body mass were detected and high-density lipoprotein-cholesterol levels were significantly decreased. However, in one study muscle strength was increased in older men (but not women) by DHEA supplementation. Recently, a very large study of older subjects showed that DHEA replacement therapy has no effect on muscle size, strength, or function.

### Insulin Resistance

The ability of muscle tissue to respond to insulin is an important aspect of overall insulin sensitivity. The incidence of insulin resistance and type 2 diabetes increases with aging, and sarcopenia may play an important role. It is unclear whether the increase in insulin resistance observed with aging is due to age per se or whether it is secondarily due to the reduction in metabolically active lean body mass. Some studies have

reported that when using the oral glucose tolerance test to compare insulin resistance between cohorts of young and older subjects, the prevalence of glucose intolerance is higher in older subjects when the data are reported per unit of body mass, but these differences disappear if the data are corrected by lean body mass. This suggests that the changes in body composition may drive the increase in insulin resistance with age. Although insulin is usually considered in the context of its ability to increase glucose uptake into cells, there is evidence that insulin resistance of muscle protein metabolism in the elderly may be an important contributor to sarcopenia. For example, when glucose is ingested with a regular meal, the subsequent increase in insulin concentrations has a negative effect on muscle protein synthesis. This implies that with normal aging, the ability of muscle cells to properly respond to circulating insulin (by increasing muscle protein synthesis) is impaired.

## AGING AND MUSCLE PLASTICITY

Human skeletal muscle is one of the most adaptable tissues in the body. The ability of muscle cells to respond to external and internal stimuli is often called muscle plasticity. We examine how two common physiological stimuli (exercise and nutrition) of skeletal muscle anabolism may be involved in the development of sarcopenia and may be used to counteract muscle loss with age.

### Exercise

Inactivity reduces muscle mass, even over a short period. Typical examples are the reduction of muscle mass observed in bed-ridden patients, following immobilization for fractures, or in weightlessness (space flight). Resistance exercise and aerobic exercise have a significant influence on skeletal muscle mass, strength, plasticity, and metabolism. Resistance exercise training results in hypertrophy of muscle cells and large increases in strength. Aerobic exercise training does not alter muscle cell size but has significant effects on the aerobic capacity of muscle cells (i.e., increases in oxidative enzymes and mitochondrial volume) with minimal effects on muscle strength. Resistance exercise acutely stimulates muscle protein synthesis in both young and older people. Both types of exercise reduce insulin resistance and increase glucose and fat oxidation. The elderly respond very well to resistance exercise training by increasing muscle mass and strength, and their aerobic capacity increases

following aerobic exercise training. Recent data also suggest that aerobic exercise may acutely increase muscle protein synthesis in older subjects. Thus, both types of exercise can be very useful to counteract sarcopenia and the associated metabolic alterations of the aging muscle.

## Nutrition

Malnutrition leads to muscle wasting. Several studies have shown that the caloric intake of many elderly individuals is too low and may be an important factor in the development of sarcopenia. Recent data also suggest that the recommended protein daily intake for the elderly (0.8 g/kg/day) might be insufficient to maintain a neutral nitrogen balance. Additionally, older people may voluntarily reduce their protein intake to comply with reduced fat and cholesterol diets. Amino acids from ingested protein directly stimulate muscle protein synthesis. Interestingly, healthy elderly subjects respond to an amino acid stimulus with an increase in muscle protein synthesis that is not significantly different from the effect observed in their younger counterparts. Unfortunately, high-protein diets given to older subjects have not proven to be effective on muscle size or strength. This may be explained by the fact that in these trials, the additional protein was given in combination with carbohydrate, which stimulated insulin secretion. Our recent data indicate that when carbohydrates are included in a protein meal for the elderly, the positive effect of amino acids on muscle protein synthesis is blunted, possibly due to the negative effect that insulin exerts on muscle proteins in the elderly. However, there are no long-term studies in which older subjects have been given only amino acids or proteins with no carbohydrate to stimulate muscle protein synthesis and provide additional energy.

## CONCLUSION

Sarcopenia is a multifactorial process. Reductions in endocrine function and physical activity and inadequate nutrition all play an important role in the reduction of muscle mass with normal aging. Testosterone replacement therapy may be a useful intervention in hypogonadal older men for increasing muscle mass and strength, although it is not currently recommended. Hormone replacement therapy for menopause, adrenopause, or somatopause appears to

have marginal or no positive effect on muscle mass and strength.

Exercise training and proper nutrition can have dramatic effects on muscle mass and strength. An optimal intervention program may include an exercise-training schedule that incorporates both resistance and aerobic exercise with adequate intake of total calories and protein. This would not only improve muscle mass and strength but also reduce insulin resistance, which is more prevalent in the elderly. Providing a nutritional supplement of only amino acids or protein might also be beneficial to promote muscle growth by stimulating muscle protein synthesis and increasing the total daily caloric intake; however, further investigation is needed.

Fortunately, aged muscle is still very plastic and can respond to anabolic stimuli by increasing its mass and strength. This knowledge is vital for designing interventions to reverse and/or attenuate the loss of muscle mass with aging and to improve metabolism and functional abilities in the elderly.

## See Also the Following Articles

Aging and Longevity of Human Populations • Aging and the Male Reproductive System • Body Weight, Body Composition, and Aging • Caloric Restriction, Aging and Oxidative Stress • DHEA and the Elderly • Growth Hormone (GH) • Neuroendocrine System and Aging • Osteoporosis in Older Men • Osteoporosis in Older Women

## Further Reading

- Baumgartner, R. N., Waters, D. L., Gallagher, D., Morley, J. E., and Garry, P. J. (1999). Predictors of skeletal muscle mass in elderly men and women. *Mech. Ageing Dev.* **107**, 123–136.
- Bhasin, S., and Buckwalter, J. G. (2001). Testosterone supplementation in older men: A rational idea whose time has not yet come. *J. Androl.* **22**(5), 718–731.
- Lamberts, S. W. J., van den Beld, A. W., and van der Lely, A.-J. (1997). The endocrinology of aging. *Science* **278**(5337), 419–424.
- Morley, J. E., Baumgartner, R. N., Roubenoff, R., Mayer, J., and Nair, K. S. (2001). Sarcopenia. *J. Lab. Clin. Med.* **137**(4), 231–243.
- Roubenoff, R., and Castenada, C. (2001). Sarcopenia: Understanding the dynamics of aging muscle. *J. Am. Med. Assoc.* **286**(10), 1230–1231.
- Volpi, E., and Rasmussen, B. B. (2000). Nutrition and muscle protein metabolism in the elderly. *Diabetes Nutr. Metab.* **13**, 99–107.
- Yarasheski, K. E. (2002). Managing sarcopenia with progressive resistance exercise training. *J. Nutr. Health Aging* **6**(5), 349–356.



# Agonadism, Male and Female

Marco A. Rivarola and Alicia Belgorosky  
*Garrahan Pediatric Hospital, Buenos Aires, Argentina*

## Glossary

**estradiol** Main female sex hormone secreted by the ovary; synthesized from ovarian male sex hormones by the action of the enzyme aromatase, localized in the granulosa cells.

**gonadotropin-releasing hormone (GnRH)** Small polypeptide hormone (10 amino acids) secreted by hypothalamic interconnected neurosecretory cells in a pulsatile fashion; reaches the pituitary gland via a local circulatory system to stimulate pituitary gonadotropins (gonadotrope cells).

**gonadotropins** Protein hormones synthesized and secreted into the blood by the pituitary gland or the placenta; stimulate the gonads.

**ovarian follicle** Main cellular structure of the ovary, containing one oocyte (germ cell) surrounded by many granulosa and theca cells; synthesizes female sex hormones.

**seminiferous tubules** Tubular structure of the testis containing the germ cells that are supported by a layer of Sertoli cells.

**testosterone** Main male sex hormone secreted by the testis; a small molecule belonging to the steroid family of chemicals.

**C**ongenital agonadism is the absence of both gonads in a newborn baby. The gonads (testes and ovaries) are formed during embryonic life under genetic control.

## INTRODUCTION

There is a sexual dimorphism in the genetic material that is evident under microscopic observation of the sex chromosome constitution. In humans, the two sexes have 46 chromosomes, but whereas females have two X chromosomes (46,XX), males have one

X chromosome and a Y chromosome (46,XY). The Y chromosome has fundamental genetic information necessary for the differentiation of the testis, even though genes located in other chromosomes (the X chromosome and the autosomes) are also necessary.

In the case of males, the function of the fetal testis is essential for the formation of male external genitalia (penis and scrotum); therefore, testes had to be present at the time of differentiation of the genitalia, during the first trimester of pregnancy. Indeed, in the absence of gonads, external genitalia (vulva and lower third of the vagina) as well as internal genitalia (fallopian tubes, uterus, and the upper third of the vagina) are female, regardless of genetic sex. For some reason, usually spontaneous testicular torsion that blocks vascular supply, testes might “vanish” after the time of genitalia differentiation and no longer be present at birth. This event is very rare; therefore, male agonadism is seen only infrequently. In a broader sense, partial loss of testicular function, properly called hypogonadism, is discussed in this article.

In the case of females, fetal ovaries are not necessary for the formation of female external genitalia. Girls with Turner’s syndrome, a condition secondary to an absence or abnormality of one chromosome X, have atrophic or hypotrophic ovaries but normal external and internal genitalia. These girls are short and usually have other physical problems.

The function of the gonads is under tropic control of the hypothalamo–pituitary axis by two pituitary hormones called gonadotropins. Even though gonadotropins are the same molecules in the two sexes, they take their name from their effects in females: follicle-stimulating hormone (FSH) and luteinizing hormone (LH). Hypogonadism is defined as “primary” when the defect is primarily of the gonads and as “secondary” when the defect is secondary to hypothalamo–pituitary deficiency.

## MALE AGONADISM

### The Testis: Two Compartments, Two Functions, Two Gonadotropic Controls

There are two distinct compartments in the testis with two different functions. First, the seminiferous tubules are multiple rolled-up microscopic tubules wrapped up in a capsule, within which millions of germ cells are produced daily (in adults) and released into a lumen along with a specific seminal fluid. These tubules end up in a common reservoir, and their contents leave the testis through a single duct. Germ cells are supported by Sertoli cells, which provide them with nutrition and functional modulation. Sertoli cells have receptors for the FSH gonadotropin in the cell membrane. Receptors for protein hormones, such as FSH, are specialized proteins that can recognize, bind, and retain a specific hormone arriving via the blood supply. After binding the hormone, receptors send a signal inside the cell that triggers a specific response. Sertoli cells also secrete a hormone, inhibin B, for pituitary FSH regulation. Second, the interstitial tissue surrounds the seminiferous tubules. This compartment contains capillary blood vessels and various specialized cells, among which the Leydig cells are in charge of synthesizing and secreting testosterone, the main male hormone. Testosterone is secreted into the blood for peripheral androgen stimulation, but it is also diffused locally to the seminiferous tubules in large concentrations. It is essential for proper development of spermatogenesis. Therefore, the close proximity of Leydig cells to seminiferous tubules serves a useful purpose. Leydig cells have receptors for the LH gonadotropin in the cell membrane. Testosterone also regulates pituitary LH secretion. Peripheral androgen stimulation induced by testosterone includes enlargement of the penis and contributes to sustaining penile erection and sexual potency. It also has many other actions, including central nervous system effects in sexual conduct.

Testes are located in the scrotum, outside the abdominal cavity, because spermatogenesis cannot proceed normally at 37° Celcius, the normal body temperature. This is so important that the scrotum has its own refrigeration system. The cooled venous blood returns from the scrotal tissues inside multiple small veins, which surround the arterial vessels, cooling off the incoming warmer arterial blood.

### Testicular Differentiation and Development

#### *From Embryonic Life to Late Prepuberty*

Because the underdeveloped seminiferous tubules do not have a lumen, they are named seminiferous cords. The reason why males do not have a uterus is that the embryonic and fetal Sertoli cells of these cords have the function of inducing regression of the embryonic Müllerian ducts, the precursors of fallopian tubes and uterus. For these purpose, they secrete anti-Müllerian hormone (AMH), also known as Müllerian-inhibiting substance (MIS). Secretion of AMH continues for a few years after birth for reasons that are not well understood. Seminiferous cords remained immature throughout childhood and become mature at adolescence along with maturation of spermatogenesis. The immature Sertoli cells have the capacity to proliferate (pre-Sertoli cells) during childhood. Maturation of germ cells during fetal life is halted at the stage of gonocytes and spermatogonia. Spermatogonia remain the most mature germ cells up to puberty.

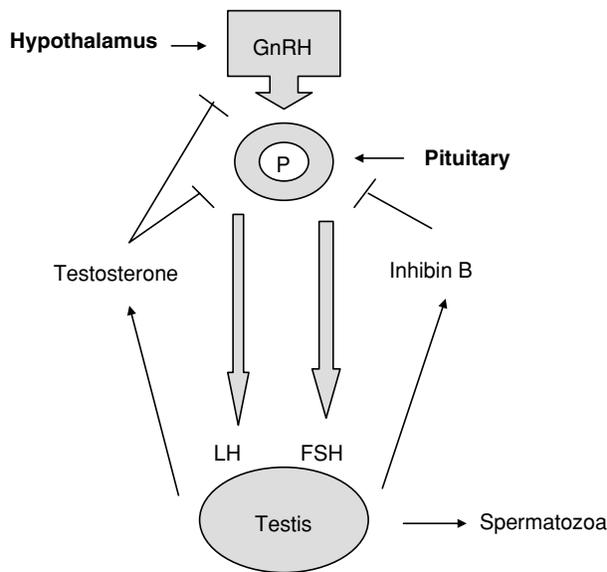
Germ cells are unique in their modes of division. They can divide by mitosis, like any other cell, but can also divide by meiosis, a complex process that includes chromatin interchange (gene mixing) and ends up with a reduction of gene material to one half (one allele). In males, the first meiosis takes place during adolescence, and it goes on throughout adult life.

Leydig cell development in humans is a triphasic phenomenon: testosterone production has peaks at 14 to 18 weeks of fetal life, 2 to 3 months after birth, and from puberty throughout adult life. Testosterone secretion by the fetal testes is essential for species existence because it induces differentiation of male external genitalia. At the critical period of male differentiation, it is stimulated by a placental gonadotropin called chorionic gonadotropin (CG). For final penile enlargement, the contribution of fetal pituitary LH is also necessary. The function of the early postnatal peak of testosterone secretion is unknown. It is a matter of controversy whether testosterone secretion during the fetal and postnatal phases may play a role on the central nervous system, programming hypothalamic gonadotropic modulation and/or adult male sexual orientation and behavior.

#### *Pubertal Maturation*

At puberty, dramatic changes take place in the testes. The initiation of testicular growth is the first sign of

male puberty taking place at approximately 11 to 13 years of age, but this event goes unnoticed by most boys. The process actually had started earlier by the activation of the gonadotropin-releasing hormone (GnRH) pulse generator, a neurosecretory structure located in the hypothalamus. The GnRH pulses reach the pituitary through a local portal vein system to induce a pulsatile secretion of LH and FSH, which in turn reach the gonads through the general blood circulation to stimulate Leydig and Sertoli cells, respectively. Initial testosterone production has a local effect on seminiferous tubules growth in combination with the direct FSH stimulation of Sertoli cells (Fig. 1). Secretion of testosterone increases gradually and, as mentioned previously, has effects in multiple organs. Those more obviously responsive are the growth of male external genitalia and pubic hair and a deepening of the voice, but many other tissues



**Figure 1** Interactions among the components of the hypothalamo-pituitary-testicular axis in males. The GnRH pulse generator of the hypothalamus exerts the overall control of the system. The GnRH hormone is delivered directly to the pituitary gland via a specialized local blood supply. The gonadotropic cells respond to GnRH by secreting the LH and FSH gonadotropins. They reach the testis through the general circulation. LH stimulates the secretion of testosterone by the Leydig cells of the interstitial compartment, whereas FSH stimulates the function of Sertoli cells to sustain spermatogenesis (along with testosterone). Testosterone, in turn, controls its own secretion by inhibiting an excess of gonadotropins (mainly LH) and also controls the GnRH pulse generator. This is known as the feedback system. Similarly, inhibin B, secreted by Sertoli cells, inhibits FSH secretion, a second feedback system.

respond as well. There is a simultaneous spurt in growth velocity, increases in muscle strength and bone mineral density, and important changes in the central nervous system function that affect general behavior and sexual interest.

Inside the seminiferous tubules, spermatogonia, the germ cell reserve, initiate a complex process of very active replication and maturation to produce spermatozoa. This process goes on continuously until senescence. Billions of spermatozoa are produced daily as a reassurance for the conservation of the species. Spermatozoa leaving the testes are immature, but they mature during their passage through the epididymis, a convoluted and unique tubule contributing to the seminal plasma and carrying the spermatozoa toward the deferent duct. Other exocrine glands, such as the prostate, also contribute to the seminal plasma. The semen is stored in the seminal vesicles, ready to be expelled during ejaculation.

Testosterone reaches most androgen-responsive organs via the blood supply. Cells of these organs have an intracellular receptor for androgens. Receptors for sex hormones do not need to be located in the external cell membrane because sex hormones have free access through the cell membrane. In some organs such as the prostate, testosterone is converted into a more potent androgen, dihydrotestosterone, before interacting with the androgen receptor. This is a protein with high affinity for androgens. The complex is then carried inside the cell nucleus to interact with androgen-responsive genes in the genomic DNA.

Puberty takes approximately 4 years to be completed. There are large variations in the onset of puberty in the two sexes. In males, age limits are conventionally set to diagnose early and late alterations in the onset of puberty. Normal variations are early puberty, when the onset is observed between 9 and 10 years of age, and delayed puberty, when it has not started by 14 years of age. An abnormality, and sometimes a serious problem, has to be suspected in boys when the onset of puberty is precocious (before 9 years of age) or very much delayed (after 15 years of age).

### Delayed Puberty

Those normal boys who undergo a delay in the initiation of puberty are frequently upset when they compare themselves with most other boys who are already in the process of sexual development, height velocity increments, and gains in physical strength. They have

difficulties in competing with their peers and might suffer from psychological problems. This is a relatively common complaint to pediatricians, who have to make the differential diagnosis between a normal variation of sexual development and hypogonadism, a disease affecting primarily the testis or, as already defined, secondary to hypothalamo–pituitary dysfunction.

The diagnosis of primary hypogonadism in males is much easier than in females because testes are outside the abdomen and, therefore, can be examined directly. In this respect, the position, size, and consistency of the pubertal testes should be evaluated. The progress of sexual maturation can be followed clinically; stages of sexual maturation for the two sexes were defined by James M. Tanner in 1962, and they are used universally in the follow-up of normal adolescence. Finally, an important clinical tool to evaluate the progress of body maturation is the determination of “bone age.” Normal maturational changes in the ossification of the hand and wrist in normal boys and girls are used to assess maturational bone age. In a single individual, bone age does not necessarily correspond to chronological age. It is a useful marker for whole body maturation as well as for sexual maturation. When necessary, imaging studies and hormonal tests can be used to establish a diagnosis.

In most instances, delayed puberty is associated with constitutional delay of growth, with the two expressing a delay in general body maturation. Frequently, there is a familiar trend. Because this is a variation of normality, the prognosis is good and no treatment is indicated. In exceptional instances, serious psychological problems might arise, and puberty can then be induced by transient administration of sex hormones.

### Undescended Testis (Cryptorchidism)

The embryonic differentiation and most of the fetal development of the testis take place in the abdomen. During the final trimester of pregnancy, the testis descends from the abdomen to the scrotum via the inguinal canal. Cryptorchidism is defined as the failure of this descent. It is observed quite often. Sometimes (2–5% of male newborns), and particularly in premature babies (up to 30%), testicular descent is not complete at birth, but the testis goes down spontaneously during the first 3 months of life. Cryptorchidism can be unilateral or bilateral. Bilateral cryptorchidism is more often associated with hormonal or genetic abnormalities than is unilateral cryptorchidism. Development of spermatogenesis, beginning at adolescence, requires that the testis

be located in the scrotum. However, the consensus among pediatricians and urologists is that the testis should be in the scrotum much earlier, around the first year of life.

Advances in the mechanism of testicular descent have been reported. However, this new information can explain only a few cases of undescended testis.

The incidence of testicular tumors is higher in adults who had had cryptorchidism during childhood than in the general population. However, early surgical correction does not change this incidence. In general, fertility during adult life is much higher in unilateral cryptorchidism than in bilateral cryptorchidism. A point of controversy is whether early correction of cryptorchidism is able to improve fertility in adult life. There is some weak evidence that this might be the case. Therefore, early correction of this condition (between 1 and 2 years of age) is recommended. This can be done with hormone stimulation (human CG [hCG]) for a few weeks followed by surgical correction in case of lack of descent.

### Primary (Hypergonadotropic) Hypogonadism

#### ***Klinefelter's Syndrome***

One of the frequent causes of hypergonadotropic hypogonadism is Klinefelter's syndrome. In 1942, H. F. Klinefelter described a syndrome of small testes, sterility, excessive gonadotropins, and bilateral breast development (gynecomastia). Different degrees of mental deficiency are frequently observed. In most patients, an extra X chromosome is found (47,XXY), but variants can be present. A defect in cell division of paternal or maternal germ cells, or during the first cell divisions after fertilization, is responsible for the sex chromosome excess. Frequency of the syndrome has been estimated to be between 1/500 and 1/1000 males—certainly a high proportion.

In most cases, these boys remain undiagnosed before puberty, even though school difficulties are common. Diagnosis is made at puberty because sexual development proceeds in the presence of small testes and gynecomastia becomes apparent. Soon after the initiation of spermatogenesis, germ cells start to die and the seminiferous tubules do not grow. Deficiencies in the secretion of inhibin B by Sertoli cells and testosterone by Leydig cells stimulate the secretion of FSH and LH, respectively. Even though testosterone secretion is subnormal, it is sufficient to induce sexual development; however, fertility is impaired. Mental deficiency and behavioral problems might be serious difficulties in these boys.

### **Anorchia**

Anorchia, or congenital agonadism, is the absence of testes in a newborn with male external genitalia and 46,XY chromosomal constitution. Androgen action is needed at a critical time of embryonic development to masculinize the external genitalia. Independently of the chromosomal constitution when there is a failure in the differentiation of the gonads, external genitalia are female. If there is partial differentiation of the testes, external genitalia undergo incomplete masculinization and the baby is born with ambiguous genitalia. The problem of human intersex is very complex and is not dealt with in this article. If, for some reason, testes become atrophic before birth but after formation of external genitalia, a male baby is born without palpable testes. True anorchia must be differentiated from intra-abdominal bilateral cryptorchidism. Monorchia is the presence of only one testis. In rare occasions, a 46,XX female with congenital adrenal hyperplasia has complete virilization of the external genitalia induced by an abnormal excess of adrenal androgens. These babies have a uterus and two ovaries that might go unnoticed. Therefore, it is important to establish the correct diagnosis at birth.

To detect the presence of nonpalpable intra-abdominal testes, some hormonal tests can be used. For many years, stimulation of testosterone secretion with hCG for 1 week has been the most useful test. Nowadays, a single blood sample determination of inhibin B or AHM, two products of Sertoli cells, can also detect the presence of intra-abdominal testes.

The lack of testes results in the absence of restraint to gonadotropin secretion. This is more marked than in Klinefelter's syndrome and becomes evident during late prepuberty, particularly for FSH.

Treatment of anorchia presents two problems. On the one hand, it is necessary to replace the function of the testes to induce puberty and to maintain appropriate androgen stimulation throughout adult life. On the other hand, the patient or parents often request that a testicular prosthesis be put in place to replenish the empty scrotum. If the boy is too small, the prosthesis will have to be replaced by a bigger one at adolescence. The induction and maintenance of testosterone function is relatively easy to accomplish using intramuscular long-acting preparations of testosterone (usually one shot per month) or dermal testosterone patches. The replacement of the second testicular function, the production of sperm, is not possible at this point. However, the astonishing advances in the techniques of human fertilization might bring solutions to this problem in the future.

### **Other Causes**

Other causes of hypergonadotropic hypogonadism include infections, trauma, and medications that are toxic for the testes. For instance, an unwanted side effect of the survival of boys with malignant diseases requiring chemotherapy is the destruction of the germinal epithelium. Radiation to the testes is also deleterious for germ cells. Before chemotherapy, there is the possibility of collecting sperm for storage for the purpose of using it to achieve fertility in the future.

### **Secondary (Hypogonadotropic) Hypogonadism**

In secondary hypogonadism, there is a deficiency in the secretion of the two gonadotropins: LH and FSH. In most instances, it is secondary to a failure in the function of the hypothalamic GnRH pulse generator.

It is interesting that the ontogeny of GnRH cells is closely associated with that of the olfactory neurons. Indeed, in the embryo, the two types of cells arise from the olfactory placode in the encephalon. They migrate together for some time and become separated later, with the GnRH neurons reaching the hypothalamus. This common biological origin might be associated with the known role of odors in sexual attraction in many animal species as well as in humans.

Kallman's syndrome is a relatively frequent type of hypogonadotropic hypogonadism that associates lack of sexual function and loss of sense of smell. The cause of Kallman's syndrome is heterogeneous. Loss of function mutations of at least two genes is responsible for the alteration. Because one of these genes is located in the X chromosome and the other one is present in an autosome, familiar transmission is different. Kallman's syndrome in boys might be suspected before puberty because of the association of small testes, cryptorchidism, and hyposmia or anosmia.

When lack of sexual development at puberty is the only symptom, differential diagnosis with pubertal delay might be difficult. Several hormonal tests have been devised for this purpose, but sometimes only the evolution of the signs and symptoms will define the diagnosis. In the meantime, patients can be treated temporarily.

Hypogonadotropic hypogonadism may be congenital or acquired, secondary to a lesion of the hypothalamo-pituitary axis. Such a lesion might be the consequence of the development of a tumor in this area. Even though these tumors are "benign,"

they might be life-threatening because of their location. Frequently, neurological or visual problems are the main concern. Other hormonal functions of the pituitary gland might be affected. Some of these tumors require surgery, but others respond to chemotherapy or hormonal treatment, depending on the tumor cell type. Unfortunately, surgery in the hypothalamic area frequently produces important invalidating sequelae. Paradoxically, tumors in this area can induce an opposite alteration: precocious puberty in small children.

Treatment of hypogonadotropic hypogonadism is needed for induction of sexual development at adolescence, maintenance of sexual function during adulthood, and stimulation of spermatogenesis when paternity is desired. When the pituitary gland is intact, the ideal treatment would be to replace the deficient secretion of the GnRH pulse generator. For this purpose, a portable pump that generates secretory pulses at the appropriate frequency (every 2 h) has been used. For practical reasons, this treatment is discontinued after completing sexual development in approximately 2 years. An alternative approach is to administer LH and FSH in adequate combinations. Finally, the most practical treatment is to administer replacement doses of testosterone as in boys with anorchia.

## FEMALE AGONADISM

### The Ovaries: Two Functions, Two Gonadotropic Controls

Similar to the testes but with multiple differences, ovaries have two separate functions (synthesis of female hormones [estrogens] and ovulation) and two gonadotropic controls (LH and FSH). However, compartmentalization is less evident. In the two sexes, during early embryogenesis, primitive germ cells migrate to the bisexual embryonic gonads. However, testes in males differentiate much earlier than do ovaries in females. Female germ cells, called oogonia, intermingle with surrounding somatic cells, called granulosa cells, to form the primordial ovarian follicles, each consisting of a single germ cell, now developed into an oocyte, wrapped in a layer of granulosa cells. Similar to Sertoli cells of the testis, granulosa cells have receptors for FSH and secrete inhibin for pituitary FSH regulation.

The growing follicle is encircled by a layer of steroidogenic cells called theca cells. Synthesis of estrogens results from cooperative action between theca

and granulosa cells in growing follicles. Theca cells have the steroidogenic machinery to synthesize androgens under LH stimulation. These androgens are then converted into estrogens by an enzyme of the granulosa cells under FSH stimulation. Curiously, female hormones are derived from male hormones. On the other hand, some of the actions of male hormones are mediated by their conversion into female hormones at target tissues.

## Ovarian Differentiation and Development

### *From Embryonic Life to Late Prepuberty*

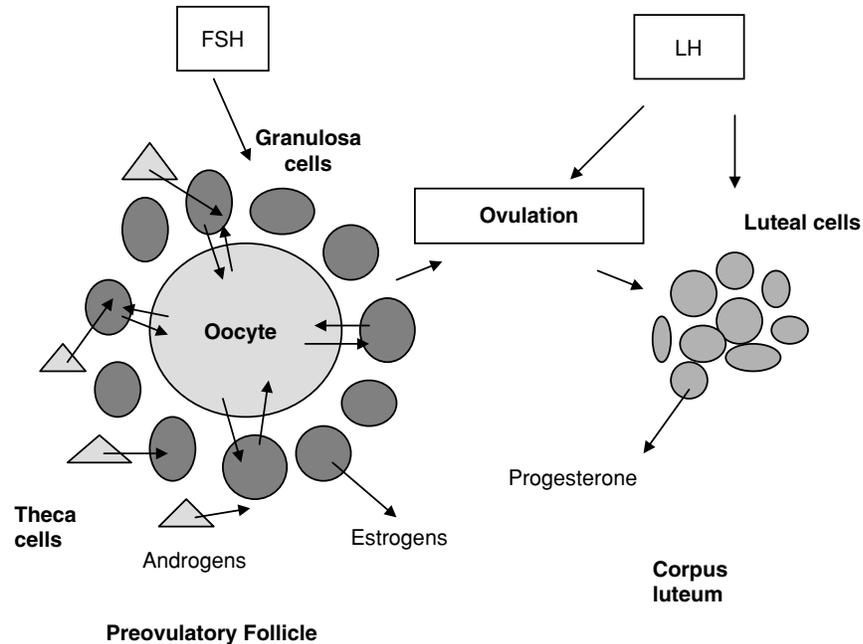
Different from testis, germ cells of the ovary initiate meiotic division during embryonic life. However, meiosis is already arrested at birth before completion. At birth, there are several hundred thousand primordial follicles, each one containing a single oocyte. This reserve of germ cells has to last a lifetime because there is no proliferation after birth. Ovaries are not necessary for differentiation of female internal genitalia (uterus and fallopian tubes) or external genitalia (vulva and vagina). Therefore, in the absence of ovaries, no clinical sign of the problem is evident in the external genitalia at birth.

Follicles develop continuously from the pool of primordial follicles throughout prepuberty and reproductive life. However, development during prepuberty is slow and restricted to partial growth. Therefore, estrogen secretion is minimal and there is no evidence of either secondary sexual characteristic development or ovulation.

Estrogen secretion starts to increase during late prepuberty. This minimal estrogen activity is often responsible for the increase in growth velocity observed at late prepuberty in girls. At the beginning of puberty, breast development becomes evident (thelarche). However, moderate and transient increases in breast development occasionally might be observed in girls at any time during prepuberty.

### *Pubertal Maturation*

As is the case in males, the process of pubertal development in females is controlled by the hypothalamic GnRH pulse generator. In the ovaries, follicle development, stimulated by pituitary FSH, is morphologically characterized by the proliferation of granulosa cells and a synchronous increase in the diameter of the oocytes (Fig. 2). During the first couple of years after the onset of puberty, growing follicles of the ovaries secrete increasing amounts of estrogens to develop secondary sexual characteristics, mainly breasts,



**Figure 2** Sequence of events in the cycling ovary showing interactions among the components of the preovulatory follicle. Pituitary FSH stimulates granulosa cell function. These cells, which surround the oocyte, control oocyte development; in turn, the oocyte controls granulosa cell function. Estrogen synthesis is also the result of cell cooperation. Indeed, theca cells synthesize androgens, which are converted to estrogens by granulosa cells. One of the growing follicles becomes the dominant follicle, accumulates fluid, and is the target for LH-induced ovulation. After ovulation, the oocyte is taken up by the fallopian tube. The remaining follicular cells undergo a process of change (luteinization) and form the corpus luteum. The corpus luteum secretes progesterone, an ovarian hormone that is necessary for preparing the endometrium of the uterus for egg nesting.

fallopian ducts, uterus, vagina, and vulva. Growing follicles undergo a process of regression, becoming atresic and being replaced by new growing follicles.

At one stage in the process of sexual maturation, the development of follicles (involving the recruitment of some primordial follicles from the resting pool and the continued growth of these follicles) reaches the point of selection of a dominant follicle in one of the ovaries. Active interaction between the oocyte and the granulosa cells, under LH stimulation, prepares the dominant follicle for ovulation. Two weeks later, at ovulation, the oocyte is expelled from the ovary and is taken up by the fallopian tube. Fertilization takes place in the fallopian tube after the arrival of mature spermatozoa several hours postcoitum. After ovulation, the dominant follicle undergoes an active process of cell reorganization and maturation, called luteinization, and is transformed into the corpus luteum (Fig. 2). This newly formed ovarian component secretes a different ovarian hormone called progesterone. One of the main functions of progesterone is to prepare the endometrium, the internal layer of the uterus, for possible nesting of the fertilized egg.

In the absence of fertilization, the corpus luteum regresses in approximately 2 weeks and the hormonal

activity of the ovary, both estrogens and progesterone, goes down.

These two hormones are necessary for maintaining the tropic development of the internal uterine layer. Hormonal withdrawal results in elimination of part of the endometrium along with its blood supply (menstruation). Afterward, the process of pituitary gonadotropic stimulation and follicle growth starts again. The cyclic nature of the ovarian function is in sharp contrast to the continuous functioning characteristic of the testis. The first menstruation is called menarche. As stated previously, it is usually preceded by approximately 2 years of acyclic ovarian activity that prepares the body for full sexual function. The age of menarche may vary greatly in normal girls, as much as between 10 and 15 years of age.

### Primary (Hypergonadotropic) Hypogonadism

Female hypergonadotropic hypogonadism can be congenital or acquired after birth. The most frequent cause of congenital primary hypogonadism in girls is Turner's syndrome. Other causes of congenital

hypergonadotropic hypogonadism in girls are very uncommon gene defects, which are of interest to medical specialists. Acquired hypergonadotropic hypogonadism can be secondary to trauma or surgical removal of the ovaries, chemotherapy or radiotherapy, ovarian infections, autoimmunity of the ovary, and/or ovarian resistance to gonadotropins of unknown cause.

### Secondary (Hypogonadotropic) Hypogonadism

Females can suffer from the same disorders of hypogonadotropic hypogonadism that have already been described for males. Complaints of delayed puberty are seen less frequently in girls than in boys. It is only a delay in the physiological activation of the GnRH pulse generator and is usually associated with a delay in stature growth. After some time, normal puberty should develop.

However, in females, a mild disorder may be limited to a loss of the cyclic function of the system. Menstrual irregularities or absent menstruation is the only complaint, with conservation of a reasonable degree of estrogen stimulation. Excessive exercise (e.g., sports, dancing) and anorexia or severe diets are well known and common causes of menstrual disorders. Finally, psychological problems can also disturb the adequate functioning of the GnRH pulse generator, showing the close linkage between the superior functions of the central nervous system and the hypothalamus. In extreme cases of anorexia, such as in the syndrome of anorexia nervosa (a profound psychological disorder), there might be a complete deficiency of gonadotropin secretion.

Hypogonadotropic hypogonadism can be isolated or associated with deficiency of other pituitary functions. Isolated hypogonadotropic hypogonadism is usually congenital and may or may not be present with olfactory dysfunction. Multiple pituitary deficiency might be congenital or acquired. Congenital multiple deficiencies are the consequence of deficiency of one of the various genes active during early pituitary embryogenesis. In these patients, symptoms vary and depend on the type of hormonal function affected.

Treatment of hypogonadotropic hypogonadism involves replacement of ovarian function with female sex hormones. In girls with sexual infantilism, estrogens are administered continuously, at increasing doses, over approximately 2 years. Then, estrogens are given in cycles of 3 weeks, with the addition of progesterone during the third week. During the week of resting, menses are expected. Estrogens should be administered under strict medical control. Periodic examination of the breasts and follow-up of uterine response by ultrasound are required. Occasionally, congenital absence of the uterus impairs menstruation.

Acquired pituitary deficiency in young girls is usually secondary to tumors of the pituitary gland or the hypothalamus. One relatively common benign tumor is a tumor of prolactin-secreting cells of the pituitary called prolactinoma. Prolactinomas produce hypogonadotropic hypogonadism (amenorrhea) and galactorrhea (abnormal milk secretion). They might grow outside the pituitary gland and produce lesions of the visual fields. When diagnosed early in their evolution, these tumors can be controlled with an oral medication and without the need for surgery.

### See Also the Following Articles

Anti-Müllerian Hormone • Delayed Puberty and Hypogonadism, Female • Delayed Puberty and Hypogonadism, Male • Genes and Gene Defects Affecting Gonadal Development and Sex Determination • Hypergonadotropic Hypogonadism • Kallmann's Syndrome and Idiopathic Hypogonadotropic Hypogonadism • Klinefelter's Syndrome • Ovarian-Follicular Apparatus • Testes, Embryology of • Undescended Testes

### Further Reading

- Bertrand, J., Rappaport, R., and Sizonenko, P. C. (1994). "Pediatric Endocrinology, Physiology, Pathophysiology, and Clinical Aspects." Lippincott Williams & Wilkins, Philadelphia.
- Grumbach, M. M., Sizonenko, P. C., and Aubert, M. L. (1990). "Control of the Onset of Puberty." Williams & Wilkins, Baltimore, MD.
- Lifshitz, F. (1996). "Pediatric Endocrinology," 3rd ed. Marcel Dekker, New York.
- Vanderhyden, B. (2002). Molecular basis of ovarian development and function. *Front. Biosci.* 7, D2006–D2022.



# Albright's Fibrous Dysplasia

Shaifali Kaushik and Michael Montileone  
*Virginia Commonwealth University, Richmond, Virginia, United States*

Atmaram Yarlagadda  
*University of Virginia Health System, Charlottesville, Virginia, United States*

## Glossary

**café au lait spots** Pigmented skin macules with irregular borders that are variable in shape, size, and age.

**fibrous dysplasia** A benign developmental disorder of the bone-forming mesenchymal cells resulting in replacement of spongiosa by fibrous bone.

**McCune–Albright syndrome** The triad of skin café au lait spots, polyostotic fibrous dysplasia, and peripheral precocious puberty.

**monostotic fibrous dysplasia** Involvement of a single bone by fibrous dysplasia.

**polyostotic fibrous dysplasia** Involvement of multiple bones by fibrous dysplasia.

**M**McCune–Albright syndrome is a sporadically occurring disorder caused by a somatic postzygotic missense mutation in the *GNAS1* gene encoding the subunit of the signal-transducing guanine nucleotide-binding protein (G protein). This abnormal Gs protein constitutively activates the receptors and the adenylyl cyclase system, resulting in autonomous cell proliferation and hormonal hypersecretion responsible for corresponding clinical features.

## INTRODUCTION

Fuller Albright described the triad of cutaneous café au lait spots, fibrous dysplasia and peripheral precocious puberty in 1937 after his observations of several such patients during the early 1930s. Around the same time, Donovan McCune reported a patient with the same triad and hyperthyroidism, revealing the associated endocrine abnormalities with this disease.

## CLASSIC FEATURES OF McCUNE–ALBRIGHT SYNDROME

The clinical features of the classical form of McCune–Albright syndrome (MAS) are skin café au lait spots, osseous fibrous dysplasia (FD), and precocious puberty that is caused by gonadotropin-independent sex steroid secretion. The classical form of the disease is more prevalent in females. Only two of the three features are encountered in the nonclassical form.

FD of bone is a benign osseous disorder of fibro-osseous tissue, accounting for approximately 2.5% of all bone neoplasms. The monostotic form affects one bone and accounts for approximately 70% of FD cases. The polyostotic form involves multiple bones and accounts for approximately 25% of cases. The polyostotic form occurring as part of MAS accounts for approximately 3 to 5% of the reported cases of FD. Clinical manifestations of MAS can occur as early as 0.7 years to as late as 11 years of age, with 50% of patients displaying bone dysplasia by 8 years of age.

Café au lait spots are thought to represent active proliferation of melanocytes and are variable in shape, size, and age. These pigmented skin macules are of varied color, are not raised above the skin, and typically have irregular borders (comparable to the coast of Maine) in contradistinction to neurofibromatosis café au lait spots, which have smooth margination (comparable to the coast of California).

Precocious puberty can affect both sexes; however, the majority of reported cases of precocious puberty are described in young girls. It can be evaluated based on the pubic hair development according to Tanner stages and breast enlargement, menarche, and estrogenization of external genitalia before 8 years of age. The patients develop ovarian cysts that are responsible for their elevated estrogen levels. Baseline luteinizing hormone (LH) and follicle-stimulating hormone (FSH) levels, as well as gonadotropin-releasing hormone (GnRH)-stimulated LH and FSH, are below

normal. MAS manifestations of skin, bone, and endocrine abnormalities with or without precocious puberty are occasionally seen in males. Precocious puberty in males may be considered on the basis of testicular volume greater than 4 ml according to Prader's orchidometer, androgenization of genitalia, and pubic hair development as per Tanner's standards before 9 years of age.

### POLYOSTOTIC FIBROUS DYSPLASIA

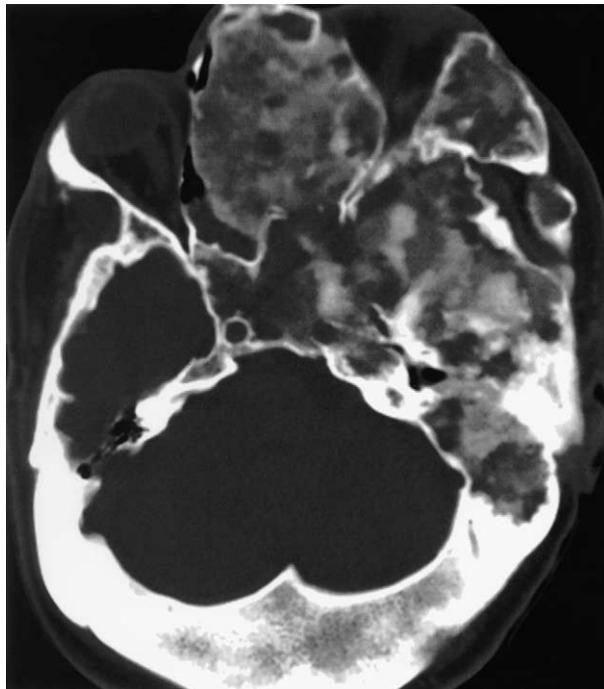
Polyostotic FD is the hallmark of MAS. Osseous lesions of FD are usually an incidental finding in a patient with an unrelated complaint. The radiographic appearance of FD is variable. Lesions containing larger proportions of fibrous content appear to be lucent, accounting for their typical ground glass appearance, whereas lesions with larger proportions of osseous material have a more sclerotic radiographic appearance. The disease has predilection for multiple bone involvement on one side of the body. The pelvis (Figs. 1 and 2), long bones, skull (Fig. 3), and ribs (Fig. 4) are involved in decreasing order of frequency. The expansile portion of the lesion may thin the cortex, and septations or cortical scalloping may be visualized (Fig. 2). The osseous radiographic features are a result of normal bone undergoing physiological resorption and replacement by fibrous tissue. There is an irregularity of the trabecular pattern. The osseous trabeculae in FD are composed of woven immature



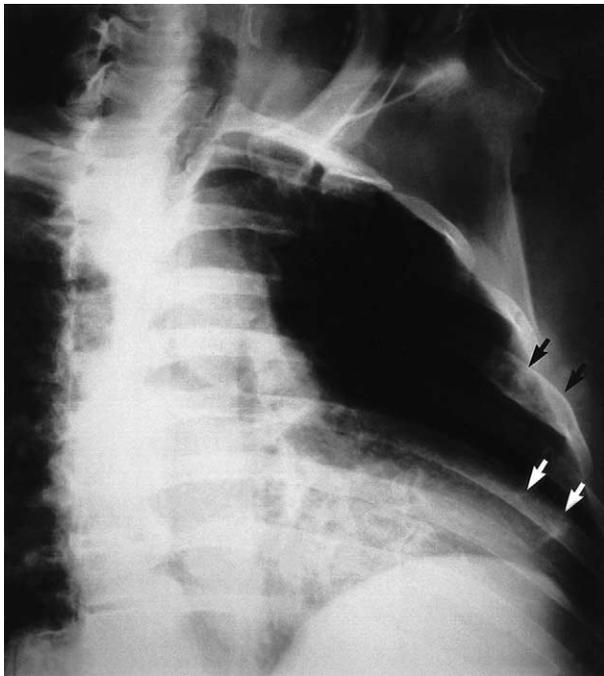
**Figure 1** Pelvis radiograph. Anteroposterior view shows an expansile lesion (arrows) with ground glass appearance involving the left superior pubic ramus and pubis symphysis.



**Figure 2** Pelvic computed tomography. Axial image shows expansile lytic lesion (arrows) with septations involving left superior pubic ramus. The low-density matrix of the involved bone is due to fibrous tissue replacement of the marrow. Compare the density of bony trabeculae of the normal right superior pubic ramus, not involved by fibrous dysplasia.



**Figure 3** Head computed tomography. Axial image shows significant involvement of the skull base and the left-sided and central facial bones. The involved bones are expanded with a mixed-density ground glass material. The nasal cavity and ethmoid sinuses are completely obliterated.



**Figure 4** Left rib series showing expansion of a left posterolateral rib (black arrows) containing mixed-density lesions of fibrous dysplasia. A normal rib (white arrows) is marked for width comparison.

bone and show no osteoblastic activity (i.e., “naked trabeculae”). The lesions increase in number and size until skeletal maturation. Pathological fracture is the most frequently encountered complication of polyostotic FD. Pathological fracture of femoral neck in these patients may lead to “shepherd’s crook” deformity. The other complications include bowing deformities and length discrepancies of the extremities. There is a tendency for polyostotic FD to be of the more symptomatic than monostotic type, resulting in a higher incidence of bowing deformities and pathological fractures. Craniofacial bone dysplasia may result in hearing and visual impairment due to cranial nerve compression. Involvement of cranial and facial bones may also lead to facial deformity and a “lion face” appearance. Surgical intervention in these areas can be associated with significant complications and morbidity; therefore, it is reserved mostly for palliative treatment.

## ENDOCRINOPATHIES ASSOCIATED WITH MAS

MAS has been reported in association with multiple endocrine abnormalities, including hyperthyroidism, hypercortisolism, hyperprolactinemia, and growth hormone (GH) hypersecretion. Concurrent

hypophosphatemic rickets may occur. Hepatocellular disease might not become evident without lab values. Dysfunction of several visceral organs has been reported. There is an increased risk for developing breast cancer in females with MAS, and this may be due partly to prolonged estrogen stimulation. Increased risk for osseous and thyroid malignancies has also been reported. Associated thymic dysplasia and intestinal polyps have been reported.

GH hypersecretion and acromegaly are rarely associated with MAS and result in elevated serum levels of insulin-like growth factor 1. When these individuals are subjected to an oral glucose load, their serum GH levels do not fall and may even increase. Autonomous adrenocortical hyperfunction resulting in Cushing’s syndrome has been reported as a rare complication of MAS. The majority of these cases are reported in young infants treated with bilateral adrenalectomy and subsequent hormone replacement therapy. Classic radiologic features of rickets, such as cupping, fraying, and widening of epiphysis, help to establish the diagnosis of rickets in MAS patients. These patients usually have normal calcium levels, normal or low-normal serum phosphorus levels, and normal or elevated serum alkaline phosphatase and urinary hydroxyproline levels attributable to polyostotic FD.

## DIAGNOSIS

Diagnostic evaluations in MAS include, but are not limited to, a GnRH stimulation test, estradiol and testosterone level determination, and liver and renal function tests. Thyroid function tests, including thyroid-stimulating hormone (TSH) levels, thyroid hormone levels, and thyroid ultrasound, are appropriate in patients with thyroid-related symptoms or abnormal clinical exams given that 30 to 40% of MAS patients are found to have thyroid abnormalities.

A skeletal survey is the study of choice to diagnose location, distribution, and laterality of FD and to follow up on disease progression and complications. Radionuclide bone scintigraphy is sensitive at determining the lesion distribution. However, other etiologies resulting in bone scan abnormalities should be considered in the differential diagnosis, and bone scintigraphy should not be considered a specific diagnostic test for MAS. Computed tomography is useful in defining the extent of disease involvement, particularly when surgical contemplation is anticipated. Although magnetic resonance imaging (MRI) signal characteristics of FD are variable, an MRI can be a

useful diagnostic test if there is spine involvement and information about the spinal canal encroachment is desired. An MRI can also achieve evaluation of soft tissues adjacent to the involved bone to differentiate from a soft tissue sarcoma or determine malignant transformation of a preexisting FD lesion.

The radiographic features of polyostotic FD show diaphyseal involvement of long bones with widening of the medullary cavity and endosteal scalloping. The involved area of bone has a ground glass appearance with sclerotic margins. The radiographic patterns of FD can be classified into three categories: predominantly sclerotic, predominantly cystic or cyst-like, and mixed patterns. The radiographic appearance of FD lesions depends on the amount of fibrous or osseous contents present. Lesions with higher fibrous content are relatively more radiolucent than are lesions with higher osseous content, which may appear to be sclerotic. In the ribs, the expansion of bone can result in an extrapleural sign. The rib lesions in FD can be long, usually greater than 4 cm in length. Involvement of pelvic bones results in typical radiographic features. If the acetabular region is involved, "protrusio acetabuli" may result. Bone softening from other causes may also result in acetabular protrusion and should be considered in the differential diagnosis.

## TREATMENT

Bone dysplasia treatment with bisphosphonates may reduce bone pain and fractures with probable retardation of bone disease progression. However, no medical treatment has been shown to alter the course of osseous disease. There is no "gold standard" treatment established for precocious puberty in MAS. The most commonly used therapies for precocious puberty include testolactone, tamoxifen, and cyproterone acetate (CPA). Testolactone is a competitive inhibitor of aromatase, the enzyme required to convert testosterone and androstenedione to estradiol and estrone. The effective treatment results in decreased frequency of menses as well as decreased growth rate and bone maturation in females with MAS. Males with MAS may occasionally demonstrate precocious puberty resulting from testicular hyperfunction. Precocious puberty in boys is treated with spironolactone, an androgen inhibitor, and testolactone. Patient monitoring, particularly for serum electrolyte levels,

is recommended for those taking spironolactone. The presence of endocrinopathies requires specific treatment aimed at the involved hormonal axis.

## RISK OF MALIGNANT TRANSFORMATION

Malignant transformation of FD is reported in approximately 0.4% of cases. Malignant transformation into osteosarcoma, fibrosarcoma, and chondrosarcoma has been described in decreasing order of frequency. Sarcomatous transformation is reported more frequently after radiation treatment.

## See Also the Following Articles

Bisphosphonates • Chondrodysplasias • Multiple Auto-immune Endocrinopathy

## Further Reading

- Albright, F., Butler, A. M., Hampton, A. O., *et al.* (1937). Syndrome characterized by osteitis fibrosa disseminata, areas of pigmentation and endocrine dysfunction, with precocious puberty in females: Report of five cases. *N. Engl. J. Med.* **216**, 727–746.
- Albright, F., and McCune, D. J. (1936). Osteitis fibrosa cystica: The case of a 9 year old girl who also exhibits precocious puberty, multiple pigmentation of the skin, and hyperthyroidism. *Am. J. Dis. Child.* **52**, 743.
- Benedict, P. H. (1962). Endocrine features in Albright's syndrome (fibrous dysplasia of bone). *Metabolism* **11**, 30–45.
- Feuillan, P. P. (1994). McCune–Albright syndrome. *Curr. Ther. Endocrinol. Metab.* **5**, 205–209.
- Feuillan, P. P., Shawker, T., Ross, S. R., *et al.* (1990). Thyroid abnormalities in the McCune–Albright syndrome: Ultrasonography and hormonal studies. *J. Clin. Endocrinol. Metab.* **71**, 1596–1601.
- Kaushik, S., Smoker, W. R. K., and Frable, W. J. (2002). Malignant transformation of fibrous dysplasia into chondroblastic osteosarcoma. *Skeletal Radiol.* **31**, 103–106.
- Kransdorf, M. J., Moser, R. P., Jr., and Gilkey, F. W. (1990). Fibrous dysplasia. *Radiographics* **10**, 519–537.
- Lee, P. A., Dop, C. V., and Migeon, C. J. (1986). McCune–Albright syndrome: Long-term follow-up. *J. Am. Med. Assoc.* **356**, 2980–2984.
- Ruggieri, P., Sim, F. H., Bond, J. R., and Unni, K. K. (1994). Malignancies in fibrous dysplasia. *Cancer* **73**, 1411–1424.
- Shenker, A., Weinstein, L. S., Moran, A., *et al.* (1991). Severe endocrine and nonendocrine manifestations of the McCune–Albright syndrome associated with activating mutations of the stimulatory G protein in the McCune–Albright syndrome. *N. Engl. J. Med.* **325**, 1688–1695.



# Aldosterone in Congestive Heart Failure

Karl T. Weber, Yao Sun, Ivan C. Gerling, and Robert A. Ahokas  
*University of Tennessee Health Science Center, Memphis, Tennessee, United States*

## Glossary

**aldosterone** A steroid hormone produced by zona glomerulosa cells of the adrenal cortex that is integral to the salt and water retention that begets the congestive heart failure syndrome with its characteristic signs and symptoms.

**endothelial dysfunction** An abnormality of endothelial cells that form the endothelium or internal lining of blood vessels and whose metabolic function (e.g., release of nitric oxide) is involved in regulating the vasomotor reactivity (dilatation and constriction) of these vessels.

**heart failure** The circumstance in which the heart does not provide for the delivery of oxygen and other nutrients to the body's tissues at a rate in keeping with their metabolic requirements.

**Mg<sup>2+</sup> homeostasis** The balance that exists within the body's various stores of this predominantly intracellular divalent cation and that is involved in hundreds of enzymatic reactions.

**neuroendocrine-immune interface** The role played by various neurohormonal system effector hormones (e.g., aldosterone, angiotensin, catecholamines) in regulating the behavior of the immune system and its peripheral blood mononuclear cells (monocytes and lymphocytes).

**A**ldosterone (ALDO) was discovered more than 50 years ago. It is a mineralocorticoid hormone with well-known endocrine properties in epithelial cells of kidneys, colon, and sweat and salivary glands that contribute to the pathophysiology of congestive heart failure. This includes sodium (Na<sup>+</sup>) resorption at the expense of potassium (K<sup>+</sup>) excretion. Magnesium (Mg<sup>2+</sup>) excretion is likewise enhanced by ALDO, whereas adrenal ALDO secretion is regulated by extracellular Mg<sup>2+</sup>. An emerging body of information continues to identify other endocrine actions of ALDO receptor–ligand binding, including promoting an efflux of cytosolic-free Mg<sup>2+</sup>, in exchange for Na<sup>+</sup>, in nonepithelial cells such as peripheral blood mononuclear cells; endothelial cell function; and its central nervous system actions

involving regulation of cerebrospinal fluid composition produced by epithelial cells of the choroid plexus, activity of the hypothalamic paraventricular nucleus involved in Na<sup>+</sup> appetite, Na<sup>+</sup> and water (H<sub>2</sub>O) excretion and sympathetic nerve activity, and the regulation of tumor necrosis factor- $\alpha$  (TNF- $\alpha$ ) production from central and/or peripheral sources. The past decade or so has witnessed a revival of interest in this steroid molecule, and in years to come an even broader understanding of ALDO's contribution to the pathophysiology of congestive heart failure will undoubtedly emerge.

## INTRODUCTION

Initially termed electrocortin, aldosterone (ALDO) was discovered some 50 years ago. Identification of its 18-aldehyde steroid structure and adrenal origin led Tait and Simpson to rename this steroid hormone aldosterone. Its ability to alter sodium (Na<sup>+</sup>) uptake in exchange for potassium (K<sup>+</sup>) prompted Selye to refer to ALDO as a mineralocorticoid. Sites of ALDO's action and receptor–ligand binding were found to reside within epithelial cells of the kidney, colon, and sweat and salivary glands. The importance of this circulating hormone in promoting salt and water retention at each of these target tissues was demonstrated by J. O. Davis and included edema-forming states such as congestive heart failure (CHF), nephrosis, and cirrhosis. Further confirmation of its biological importance was the efficacy of spironolactone, an ALDO receptor antagonist developed by Kagawa and introduced into clinical practice more than 40 years ago, in counteracting Na<sup>+</sup> retention and alleviating edema.

The past decade or so has witnessed a resurgence of interest in this steroid molecule in CHF, prompted by several observations reviewed by Weber. These included a recognition of ALDO's broader range of actions that extended beyond classic target tissues and were associated with an adverse structural remodeling of the cardiovascular system that contributes to pathophysiological expressions and the progressive nature of CHF. In addition, the inability of angiotensin-converting enzyme (ACE) inhibition to provide

for sustained suppression of plasma ALDO in patients with CHF indicated that the regulation of its secretion by zona glomerulosa cells of the adrenal cortex is more complex than that governed by angiotensin II alone. Finally, Pitt and colleagues reported that in a controlled clinical trial (RALES), conducted in 19 countries on five continents and involving more than 1600 patients with CHF, spironolactone (*vis-à-vis* placebo), in combination with an ACE inhibitor and loop diuretic, reduced the risk of all-cause and cardiac-related mortality and cardiovascular morbidity by 30%. Pitt and colleagues also reported on results of another controlled clinical trial (EPHESUS), conducted in 674 centers in 37 countries, that further underscore the importance of ALDO and the efficacy of ALDO receptor antagonism in a large cohort of patients (>6000) with left ventricular dysfunction following myocardial infarction (MI). As contrasted with placebo, the addition of eplerenone to standard therapy with ACE inhibitors or AT<sub>1</sub> receptor blocker, beta blocker, aspirin, and diuretics led to a significant reduction in overall mortality and in cardiovascular morbidity and mortality.

This article provides a broader perspective of ALDO's properties that contribute to the pathophysiology of CHF and are mediated by receptor–ligand binding. They include ALDO's influence on magnesium (Mg<sup>2+</sup>) homeostasis and the role of extracellular Mg<sup>2+</sup> ([Mg<sup>2+</sup>]<sub>o</sub>) in regulating adrenal ALDO secretion, vascular remodeling and immune cell activation, endothelial cell dysfunction, and the central nervous system.

## Mg<sup>2+</sup> HOMEOSTASIS AND ALDOSTERONE SECRETION

The distribution of Mg<sup>2+</sup> within body tissues is as follows: 53% in bone, 27% in skeletal muscle, and 19% in other soft tissues such as the heart. Less than 1% of total body Mg<sup>2+</sup> is present in blood, and cytosolic-free Mg<sup>2+</sup> ([Mg<sup>2+</sup>]<sub>i</sub>) represents only 0.5 to 5% of total cellular Mg<sup>2+</sup>, with approximately 80% bound to adenosine triphosphate (ATP) and other phosphometabolites sequestered within organelles such as mitochondria and endoplasmic reticulum. [Mg<sup>2+</sup>]<sub>i</sub> homeostasis is maintained by exchange with these intracellular stores, whereas [Mg<sup>2+</sup>]<sub>o</sub> is held within narrow limits by Mg<sup>2+</sup> efflux from tissue stores.

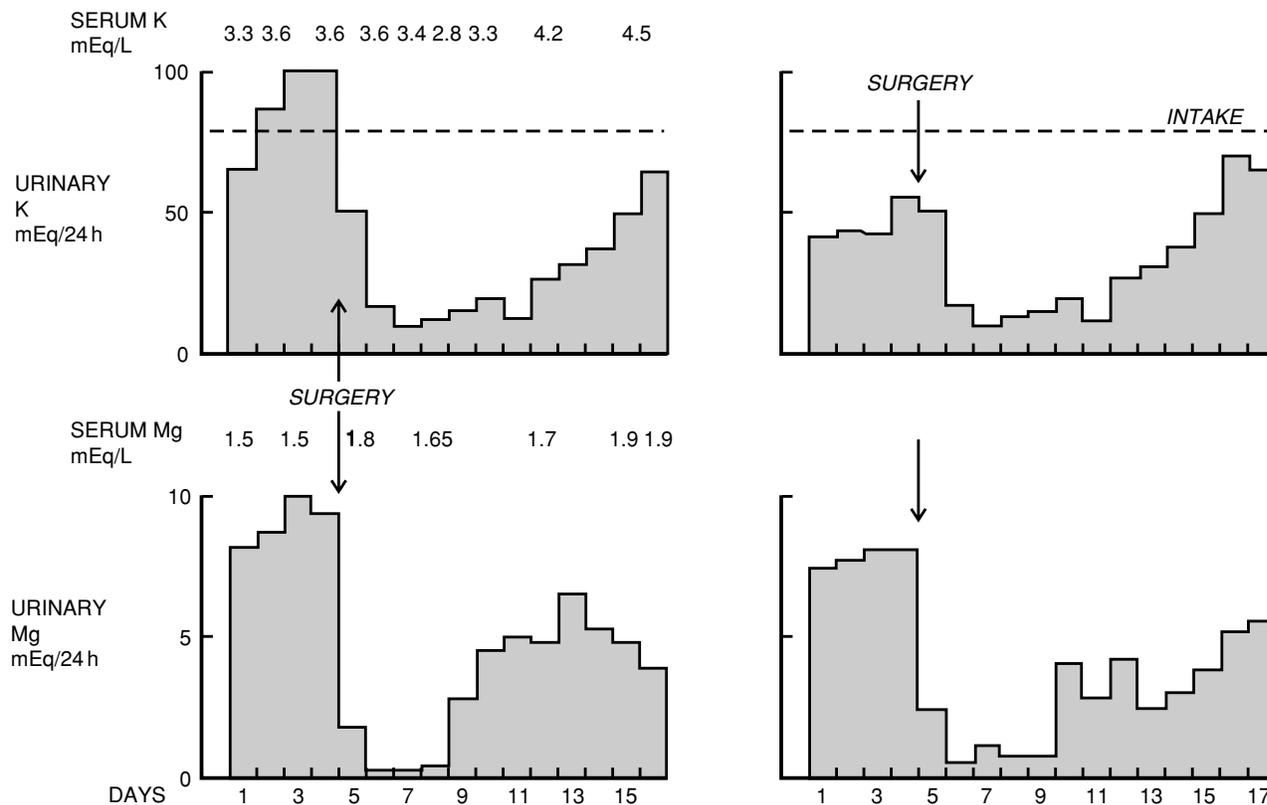
### Epithelial Cells

In 1955, Mader and Iseri reported that a patient with adrenal adenoma had experienced spontaneous

episodes of hypomagnesemia together with enhanced Mg<sup>2+</sup> excretion in urine and stool. Others would also note the presence of hypomagnesemia in patients with primary aldosteronism, suggesting that Mg<sup>2+</sup> deficiency accompanies long-standing, ALDO-induced Mg<sup>2+</sup> excretion. These clinical findings implicated ALDO in regulating both K<sup>+</sup> and Mg<sup>2+</sup> excretion in classic target tissues. In 1962, Horton and Biglieri addressed urinary K<sup>+</sup> and Mg<sup>2+</sup> excretion in five patients with PAL, where each patient had low normal or reduced serum Mg<sup>2+</sup> levels. The influence of surgical resection of adrenal adenoma on urinary K<sup>+</sup> and Mg<sup>2+</sup> excretion was assessed in two patients (Fig. 1). In both patients, there was an immediate and marked fall in excretion of these monovalent and divalent cations after surgery together with a gradual normalization of plasma Mg<sup>2+</sup>. A lower basal excretion of Mg<sup>2+</sup>, comparable to values seen for normal controls on the same Mg<sup>2+</sup> diet, was seen postoperatively. In another patient, spironolactone was shown to reduce both urinary K<sup>+</sup> and Mg<sup>2+</sup> excretion, which returned to previous increased basal levels following discontinuation of spironolactone (Fig. 2). Thus, the importance of ALDO in promoting urinary Mg<sup>2+</sup> excretion was evident. In 1963, Conn, who had earlier coined the term PAL to connote autonomous adrenal ALDO production independent of plasma renin activity, concluded that hypomagnesemia, hypokalemia, hypernatremia, hypochloremia, and metabolic alkalosis were cardinal metabolic abnormalities of PAL.

The importance of the adrenal cortex in regulating Mg<sup>2+</sup> and K<sup>+</sup> excretion was further underscored in adrenal insufficiency or following adrenalectomy. ALDO treatment of dogs or rats with surgically induced bilateral adrenalectomy was shown to increase fecal K<sup>+</sup> and Mg<sup>2+</sup> excretion and to normalize their plasma concentrations. Horton and Biglieri treated an adrenalectomized patient with d-ALDO for 8 days (Fig. 3). On day 5 of this regimen, spironolactone cotreatment was initiated. Exogenous ALDO promoted a prompt elevation in urinary K<sup>+</sup> and Mg<sup>+</sup> excretion that was abrogated by receptor antagonist. Thus, the body of evidence is compelling that ALDO promotes both K<sup>+</sup> and Mg<sup>2+</sup> excretion at classic target tissue sites.

Mg<sup>2+</sup> excretion is likely increased in patients with CHF, where elevated plasma ALDO levels are expected. In patients with CHF treated with a loop diuretic, there exists an independent stimulus to urinary Mg<sup>2+</sup> excretion and the potential for exaggerated Mg<sup>2+</sup> loss. However, its serum concentration does not reflect intracellular Mg<sup>2+</sup>, and methods to detect



**Figure 1** Urinary  $K^+$  and  $Mg^{2+}$  excretion in two patients with primary aldosteronism before and after surgical removal of adrenal adenoma. Serum  $K^+$  and  $Mg^{2+}$  levels are shown for one of these individuals. Reproduced from Horton and Biglieri (1962). Effects of aldosterone on the metabolism of magnesium. *J. Clin. Endocrinol. Metab.* 22, 1187–1192. © The Endocrine Society.

biologically active, cytosolic-free  $[Mg^{2+}]_i$  are not widely used; therefore, the assessment of this important clinical problem remains to be defined. Intracellular  $Mg^{2+}$  deficiency may contribute to morbid and mortal events such as sudden cardiac death, which occurs in 50% of patients with CHF. The 30% reduction in risk of sudden cardiac death observed in the RALES trial may be related, in part, to spironolactone's ability to restore and preserve  $Mg^{2+}$  homeostasis.

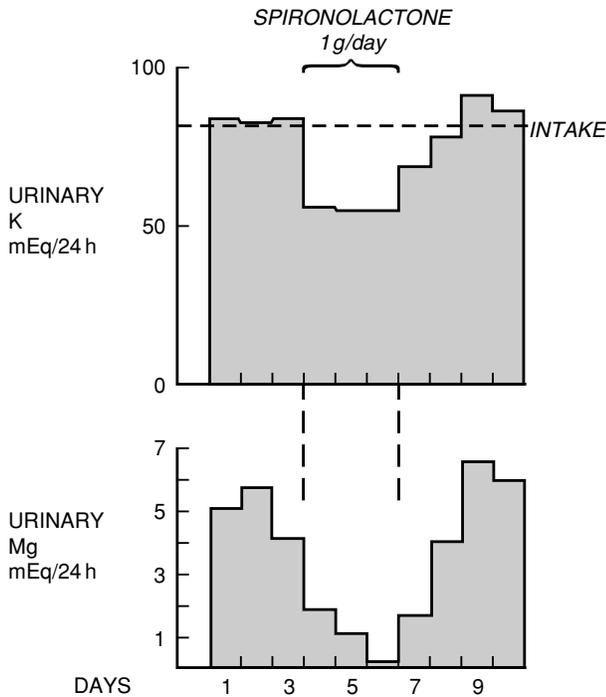
### Nonepithelial Cells

ALDO regulates  $Mg^{2+}$  exchange by nonepithelial cells, such as peripheral blood mononuclear cells (PBMCs), where it binds to a single class of cytosolic receptors. In patients with either PAL or renin-dependent secondary aldosteronism (SAL), ALDO-binding sites are reduced by 50%. Following surgical removal of adenomatous adrenal tissue, ALDO receptor binding is normalized in these cells. ALDO promotes the efflux of  $Mg^{2+}$  from cultured human

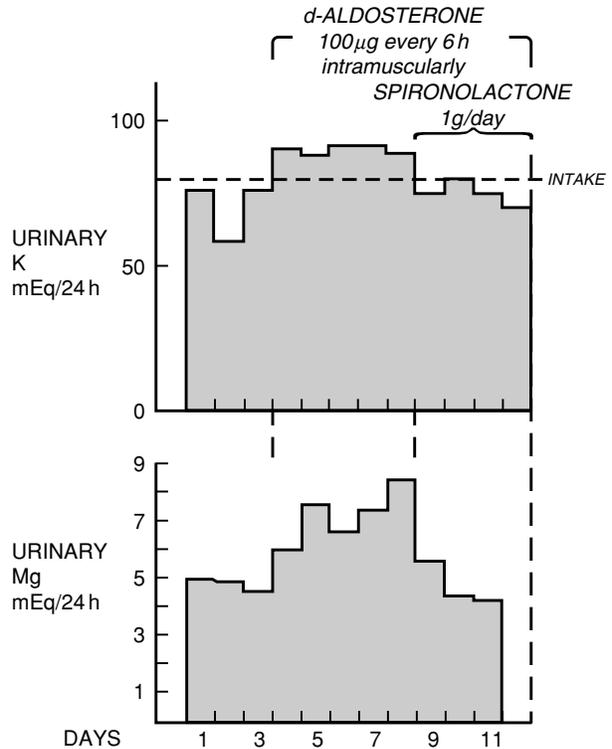
lymphocytes in exchange for  $Na^+$  via receptor–ligand binding. This  $Na^+$ -dependent response, measured by mag-fura-2, a fluorescent probe, involves both transcription and protein synthesis, as demonstrated by its respective abrogation by cycloheximide and actinomycin D. Lymphocyte-ionized  $[Mg^{2+}]_i$  is reduced in patients with PAL secondary to either adrenal adenoma or hyperplasia. In uninephrectomized rats treated with ALDO by implanted mini-pump, PBMC  $[Mg^{2+}]_i$  is significantly reduced and accompanied by immune cell activation (vide infra).

### Aldosterone Secretion

$[Mg^{2+}]_o$  participates in the regulation of adrenal ALDO secretion to create a pathway of reciprocal regulation in  $Mg^{2+}$  homeostasis (Fig. 4). In healthy normotensive men and women, a 3-h intravenous infusion of magnesium sulfate ( $MgSO_4$ ) suppresses plasma ALDO levels. On the other hand, dietary-induced  $Mg^{2+}$  deficiency with reduced  $[Mg^{2+}]_o$  is accompanied by an expanded width to the adrenal



**Figure 2** Urinary  $K^+$  and  $Mg^{2+}$  excretion in a patient with primary aldosteronism before, during, and after oral spironolactone treatment. Reproduced from Horton and Biglieri (1962). Effects of aldosterone on the metabolism of magnesium. *J. Clin. Endocrinol. Metab.* 22, 1187–1192. © The Endocrine Society.

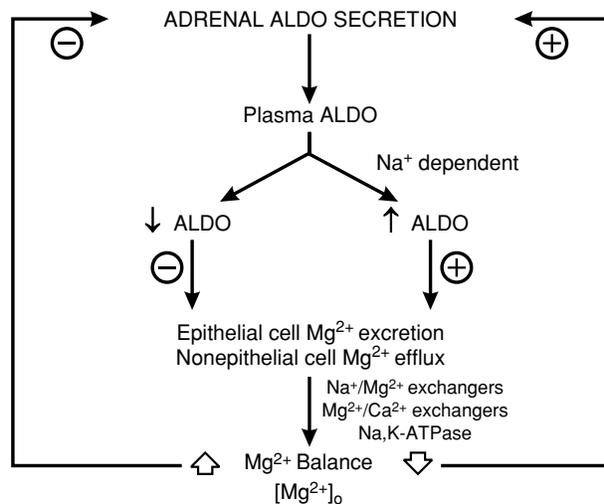


**Figure 3** Urinary  $K^+$  and  $Mg^{2+}$  excretion in a patient with bilateral adrenalectomy before and during aldosterone treatment (without and then with oral spironolactone coadministration). Adapted from Horton and Biglieri (1962). Effects of aldosterone on the metabolism of magnesium. *J. Clin. Endocrinol. Metab.* 22, 1187–1192. © The Endocrine Society.

zona glomerulosa together with hyperplasia of the renal juxtaglomerulosa cells, increased adrenal ALDO secretion, increased plasma ALDO, and a reduction in urinary  $Na^+/K^+$  ratio. Dietary  $Mg^{2+}$  deficiency combined with a high- $Na^+$  diet attenuates, but does not abrogate, heightened ALDO secretion. In cultured zona glomerulosa cells, superfusate  $[Mg^{2+}]_o$  regulates ALDO production; high  $[Mg^{2+}]_o$  suppresses, whereas an  $[Mg^{2+}]_o$ -free media augments, their elaboration of ALDO. The SAL that accompanies dietary  $Mg^{2+}$  deficiency is associated with a time-dependent rise in  $[Na^+]_i$  and  $[Ca^{2+}]_i$  in heart, skeletal muscle, kidney, and bone and is suggestive of an inhibition of  $Na,K$ -ATPase, an  $Mg^{2+}$ -dependent pump, and increased  $Na^+/Ca^{2+}$  exchange at these sites.

### NEUROHORMONE-IMMUNE INTERFACE

A structural remodeling of the cardiovascular by fibrous tissue accompanies aldosteronism derived from either endogenous or exogenous sources. This fibrogenic phenotype includes intramural arteries of the heart, kidney, pancreas, mesentery, and vaso



**Figure 4**  $Mg^{2+}$  balance and reciprocal regulation of adrenal ALDO secretion. Reproduced with permission from Weber, K. T. (2003). Aldosterone revisited: Perspectives on less well recognized actions of aldosterone. *J. Lab. Clin. Med.* (in press).

vasorum of the aorta and pulmonary artery. Cotreatment with a receptor antagonist (e.g., spironolactone, eplerenone), in either nondepressor or depressor doses, prevents this remodeling, indicating its independence of elevations in blood pressure. In a substudy of the RALES trial, survival benefit was associated with a reduction in circulating markers of collagen synthesis that presumably reflected an attenuation in ongoing vascular fibrosis. In this connection, urinary excretion of hydroxyproline, a marker of collagen turnover, is increased in adrenalectomized rats treated with ALDO, 1% dietary sodium chloride (NaCl), and cortisone. On the other hand, glucocorticoids reduce urinary hydroxyproline excretion, and their inhibition of collagen formation in bone is associated with osteoporosis.

In 1995, Campbell and colleagues found that the perivascular fibrosis of the coronary vasculature that ultimately appears in aldosteronism is preceded by a proinflammatory vascular phenotype that features invading monocytes/macrophages and lymphocytes and adhesion molecule expression. These findings, including their independence of blood pressure, were confirmed by others more recently. An interrogation of molecular responses involved in the invasion of coronary vessels by these inflammatory cells was addressed in animal models of PAL and SAL. In rats receiving ALDO/salt treatment (ALDOST), where plasma renin and angiotensin II both are suppressed, Sun and colleagues more recently tested the hypothesis that oxidative stress was involved in the appearance of the proinflammatory/fibrogenic cardiac phenotype. At week 3 of ALDOST, there was no evidence of cardiac pathology. However, at weeks 4 and 5, inflammatory cells (monocytes/macrophages and lymphocytes) were found to have invaded intramural coronary vessels in both ventricles. By immunohistochemistry, invading PMBCs were found to express the gp91<sup>phox</sup> subunit of nicotinamide adenine dinucleotide phosphate (NADPH) oxidase, whose activation is a major source of superoxide in leukocytes, and 3-nitrotyrosine, a product of peroxynitrite with stable protein tyrosine residues and where peroxynitrite is formed by the reaction between superoxide and nitric oxide. The RelA subunit of nuclear factor- $\kappa$ B (NF- $\kappa$ B), a redox-sensitive transcription factor integral to inflammatory responses, was likewise activated in these cells. *In situ* hybridization localized increased mRNA expression of intercellular adhesion molecule-1 (ICAM-1), monocyte chemoattractant protein-1 (MCP-1), and proinflammatory cytokine tumor necrosis factor- $\alpha$  (TNF- $\alpha$ ) at vascular sites that involved the normotensive nonhypertrophied right atrium and ventricle and left

atrium as well as the hypertensive hypertrophied left ventricle. Cotreatment with either spironolactone or an antioxidant (either pyrrolidine dithiocarbamate or N-acetylcysteine) prevented the appearance of these cells and associated molecular responses as well as the subsequent perivascular fibrosis. Thus, ALDOST induces oxi/nitrosative stress within inflammatory cells invading the intramural coronary vasculature, and it is this proinflammatory vascular phenotype that leads to intramural coronary artery pathology and the subsequent perivascular fibrosis.

What is responsible for the induction of oxi/nitrosative stress in PMBCs? Gerling and colleagues identified an ALDO-mediated reduction in  $[Mg^{2+}]_i$  that appears in PMBCs at week 1 of ALDOST, long before the appearance of cardiac pathology at week 4. The early activation of these cells is evident in their transcriptome (expressed genes) and proteome (expressed proteins). The interrogation of complex molecular events that account for immune cell activation and subsequent homing of these cells to the heart may explain why cardiac pathology does not appear until week 4 of ALDOST and calls into question the prospect of an autoimmune response similar to that which can follow MI.

In rodents treated with a  $Mg^{2+}$ -deficient diet, a putative state of exaggerated aldosteronism, lymphocyte  $Mg^{2+}$  is reduced to an extent comparable to the  $Mg^{2+}$  depletion that appears in skeletal muscle and cardiac tissue. Weglicki and colleagues identified an early (week 1) induction of oxi/nitrosative stress and depletion of antioxidant defenses in PMBCs and endothelial cells. Lymphocyte activation that appears includes their production of proinflammatory cytokines and a neurogenic peptide, substance P, together with the expression of its receptors. Cardiac lesions are first seen in this model at week 3 and can be prevented by a substance P receptor antagonist.

The potential for the SAL that accompanies human CHF to likewise be accompanied by reduced  $[Mg^{2+}]_i$  and immune cell activation remains unknown. The prospect does exist for an immune cell origin to the "cytokine storm" characteristic of CHF that features elevations in circulating proinflammatory cytokines such as TNF- $\alpha$  and interleukin-6 (IL-6). Cells of the monocyte-phagocyte system are a potent source of these cytokines. Another source of cytokine production in heart failure is the central nervous system. Irrespective of their origin, prolonged elevations in these proinflammatory cytokines contribute to the progressive systemic illness that accompanies CHF and features tissue wasting to eventuate in cardiac cachexia.

## DYSFUNCTION OF THE ENDOTHELIUM

In patients with PAL or renal artery stenosis with SAL, forearm vasomotor reactivity to endothelial cell-dependent acetylcholine is diminished compared with that in normotensive controls, whereas nonendothelial cell-dependent sodium nitroprusside-induced vasodilatation is preserved. Following surgical removal of adrenal adenoma, the impairment in endothelial cell-dependent vasodilation is restored. In the SAL that accompanies CHF, diminished forearm vasomotor reactivity to acetylcholine is normalized by spironolactone treatment. Acetylcholine-induced, nitric oxide-dependent vasorelaxation of aortic rings is reduced in rats following MI. This vasomotor dysfunction, together with increased superoxide formation by aortic tissue, is normalized by spironolactone alone or in combination with an ACE inhibitor. In cultured aortic endothelial cells, reduced  $Mg^{2+}$  concentration of culture medium is associated with increased oxidant production and reduced intracellular glutathione, an antioxidant reserve consumed in neutralizing oxidative stress. Abnormal  $Mg^{2+}$  homeostasis may account for the underlying pathophysiological basis of endothelial dysfunction seen in either PAL or SAL.

## CENTRAL NERVOUS SYSTEM

ALDO receptors are found at diverse sites in the central nervous system. These include epithelial cells of the choroid plexus. A role for ALDO in the genesis of idiopathic intracranial hypertension (IIH) has been proposed given the reported association between IIH and PAL and SAL as well as the prevalence of headaches among patients with PAL. That this proposition does not apply to all patients with IIH is underscored by its appearance in patients with adrenal insufficiency unless ALDO is produced *in situ* within the central nervous system, as is now recognized to be the case for the cardiovascular system.

The choroid plexus, a site of high-affinity ALDO receptor binding, is involved in the production of cerebrospinal fluid (CSF) and is a target site for ALDO, spironolactone, and ouabain, an endogenous, digitalis-like substance released by the adrenals and the hypothalamic-pituitary axis. ALDO exerts its biological actions on epithelial cells by enhancing the activity and number of Na,K-ATPase pumps in their apical membrane. An ouabain-sensitive Na,K-ATPase is present in the microvilli of the plexus and is involved in the regulation of CSF formation and electrolyte composition (e.g., ouabain reduces CSF

production). ALDO is present in CSF, where its concentration correlates with plasma levels. Either the systemic or intracerebroventricular administration of a mineralocorticoid (ALDO or deoxycorticosterone [DOC]) is accompanied by a fall in CSF  $K^+$ , together with a rise in arterial pressure, without changes in blood volume, cardiac output, plasma catecholamines, or vasopressin. This hypertensive response is abrogated by intracerebroventricular infusion of  $K^+$  or a mineralocorticoid receptor antagonist. Therefore, ALDO has a central action involved in the regulation of blood pressure as well as CSF volume and composition. Produced locally within the brain, ALDO's paracrine properties may likewise contribute to blood pressure regulation.

ALDO's central actions, which may contribute to the pathophysiology of CHF, were reviewed and expanded by Felder and colleagues. The hypothalamic paraventricular nucleus (PVN), a forebrain site involved in the regulation of extracellular volume and sympathetic nerve activity, is governed by circulating neurohormones and effector signals originating from the brainstem. In rats with MI induced by coronary artery ligation, the activity of the PVN is increased. Systemic or intracerebroventricular administration of spironolactone reduces this activity, improves baroreflex regulation of renal sympathetic nerve activity (albeit in a time-dependent manner), and prevents the increase in  $Na^+$  appetite and decline in urinary  $Na^+$  and  $H_2O$  excretion that appear in this model. Plasma levels of TNF- $\alpha$  rise progressively over weeks 1 to 3 following MI, a response abrogated by intracerebroventricular infusion of spironolactone started 24 h after coronary ligation, suggesting that central ALDO receptor activation is involved in regulating the release of this proinflammatory cytokine. However, the cellular source of TNF- $\alpha$  remains uncertain and may include central and/or peripheral tissues.

## FUTURE DIRECTIONS

Since its discovery more than 50 years ago, ALDO has had a well-established importance in clinical medicine, including its role in CHF. The past decade or so has witnessed a resurgence of interest in the adrenal's most potent mineralocorticoid as well as its *de novo* production through steroidogenesis within the cardiovascular system and brain. An ever-expanding role for this steroid molecule in the metabolism of monovalent and divalent cations by epithelial and nonepithelial cells has warranted an even broader perspective of its portfolio of actions. The many peripheral and central actions of ALDO that can contribute to the

pathophysiology of CHF syndrome remain to be defined. This is no more evident than in the adverse structural remodeling of the heart and systemic organs that accompanies chronic elevations in plasma ALDO (inappropriate relative to dietary  $\text{Na}^+$ ) and that may be secondary to PBMC activation induced by  $[\text{Mg}^{2+}]_i$  depletion and transduced by oxi/nitrosative stress. Molecular mechanisms involved in immune cell responses remain to be elucidated. Today's technologies will permit an assessment of ALDO's role in altering the molecular phenotype of these immune cells, specifically their transcriptome and proteome. Such insights may provide for the development of serologic biomarkers that address the risk, onset, and progression of vascular injury in CHF and could lead the way toward refined and even newer drug targets.

### Acknowledgments

The editorial assistance of Richard A. Parkinson is appreciated. This work was supported, in part, by NIH R01-DK62403 and R24-RR-15373 (I.C.G.), NHLBI R01-HL67888 (Y.S.), NHLBI R01-HL62229 (K.T.W.), and grants from the Center of Excellence in Connective Tissue Diseases (Y.S. and K.T.W.) at the University of Tennessee Health Science Center.

### See Also the Following Articles

Aldosterone Receptors • Atrial Natriuretic Factor and Family of Natriuretic Peptides • Hypertension, Endocrine • Immune System, Hormonal Effects on • Mineralocorticoids and Mineralocorticoid Excess Syndromes • Primary Aldosteronism (PAL) • Tissue Renin-Angiotensin-Aldosterone System

### Further Reading

Campbell, S. E., Janicki, J. S., and Weber, K. T. (1995). Temporal differences in fibroblast proliferation and phenotype expression

- in response to chronic administration of angiotensin II or aldosterone. *J. Mol. Cell. Cardiol.* **27**, 1545–1560.
- Conn, J. W. (1963). Aldosteronism in man: Some clinical and climatological aspects. *J. Am. Med. Assoc.* **183**, 871–878.
- Felder, R. B., Francis, J., Zhang, Z. H., Wei, S. G., Weiss, R. M., and Johnson, A. K. (2003). Heart failure and the brain: New perspectives. *Am. J. Physiol.* **284**, R259–R276.
- Horton, R., and Biglieri, E. G. (1962). Effect of aldosterone on the metabolism of magnesium. *J. Clin. Endocrinol. Metab.* **22**, 1187–1192.
- Kagawa, C. M., Cella, J. A., and Van Arman, C. G. (1957). Action of new steroids in blocking effects of aldosterone and deoxycorticosterone on salt. *Science* **126**, 1015–1016.
- Lösel, R. M., Feuring, M., Falkenstein, E., and Wehling, M. (2002). Nongenomic effects of aldosterone: Cellular aspects and clinical implications. *Steroids* **67**, 493–498.
- Mader, I. J., and Iseri, L. T. (1955). Spontaneous hypopotassemia, hypomagnesemia, alkalosis, and tetany due to hypersecretion of corticosterone-like mineralocorticoid. *Am. J. Med.* **19**, 976–988.
- Pitt, B., Remme, W., Zannad, F., Neaton, J., Martinez, F., Roniker, B., Bittman, R., Hurley, S., Kleiman, J., and Gatlin, M. (2003). Eplerenone, a selective aldosterone blocker, in patients with left ventricular dysfunction after myocardial infarction. *N. Engl. J. Med.* **348**, 1309–1321.
- Pitt, B., Zannad, F., Remme, W. J., Cody, R., Castaigne, A., Perez, A., Palensky, J., and Wittes, W. (1999). The effect of spironolactone on morbidity and mortality in patients with severe heart failure. *N. Engl. J. Med.* **341**, 709–717.
- Selye, H. (1946). The general adaptation syndrome and the diseases of adaptation. *J. Clin. Endocrinol.* **6**, 117–230.
- Sun, Y., Zhang, J., Lu, L., Chen, S. S., Quinn, M. T., and Weber, K. T. (2002). Aldosterone-induced inflammation in the rat heart: Role of oxidative stress. *Am. J. Pathol.* **161**, 1773–1781.
- Tait, J. F., and Tait, S. A. S. (1990). A decade (and even more) of aldosterone and other adrenal steroids. In "Endocrine Hypertension" (E. G. Biglieri and J. C. Melby, eds.), pp. 5–27. Raven Press, New York.
- Weber, K. T. (2001). Aldosterone in congestive heart failure. *N. Engl. J. Med.* **345**, 1689–1697.
- Weglicki, W. B., Mak, I. T., Stafford, R. E., Dickens, B. F., Cassidy, M. M., and Phillips, T. M. (1994). Neurogenic peptides and the cardiomyopathy of magnesium-deficiency: Effects of substance P-receptor inhibition. *Mol. Cell. Biochem.* **130**, 103–109.



# Aldosterone Receptors

Marc Lombès and Maria-Christina Zennaro  
INSERM U 478, Xavier Bichat, Paris, France

## Glossary

**aldosterone** A steroid hormone synthesized in the zona glomerulosa of the adrenal cortex; the major mineralocorticoid hormone in humans.

**mineralocorticoid** Corticosteroid hormone secreted by the adrenal gland and exerting its function through the mineralocorticoid receptor; also referred to as the hormone that affects water and electrolyte homeostasis.

**A**ldosterone receptors, or mineralocorticoid receptors, are defined as intracellular proteins that are able to bind aldosterone and mediate hormone action within target cells.

## INTRODUCTION

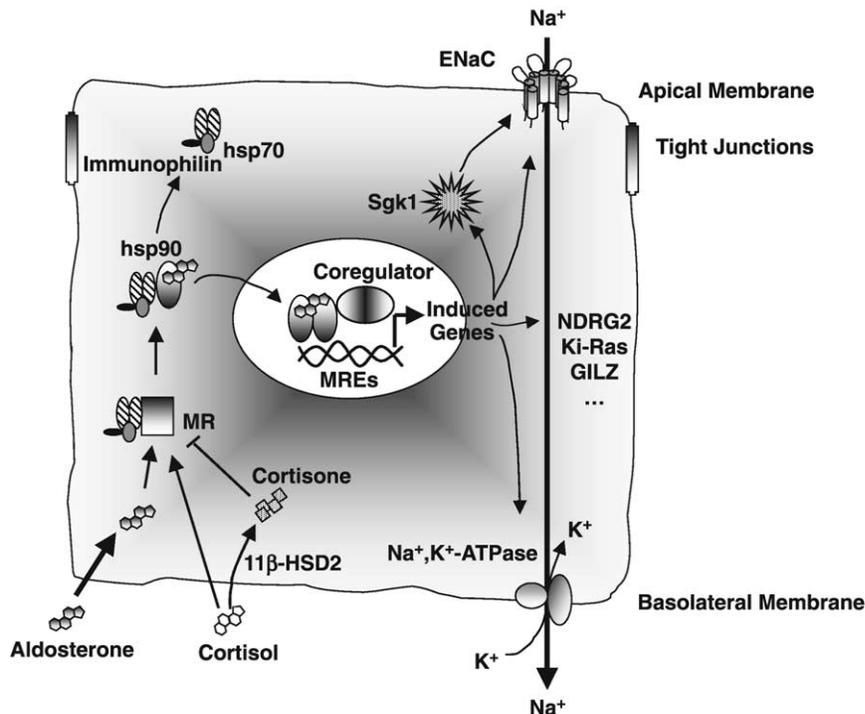
The aldosterone receptor, also referred to as the mineralocorticoid receptor (MR), is a member of the nuclear receptor superfamily that acts as a ligand-dependent transcription factor mediating aldosterone effects on a variety of target tissues. These include epithelial cells in the kidney and colon but also nonepithelial cells in the cardiovascular and central nervous systems.

## MECHANISM OF ALDOSTERONE ACTION

The classical model of aldosterone action is illustrated schematically in Fig. 1. Aldosterone penetrates a target cell, typically a polarized epithelial cell in the distal nephron and presumably by passive diffusion, and specifically binds to the MR. In its unliganded state, the MR resides predominantly in the cytoplasm and is complexed with various receptor-associated proteins, including a dimer of heat shock protein 90 (hsp90), heat shock protein 70 (hsp70), and other proteins such as immunophilins, cyclophilin (Cyp40), and FKBP52, which are known to bind immunosuppressive agents.

On aldosterone binding, the MR undergoes conformational change that leads to the dissociation of receptor-associated proteins. Aldosterone receptor complexes then translocate to the nuclear compartment and bind as homodimers to specific DNA sequences that are known as mineralocorticoid response elements (MREs) located in the regulatory regions of aldosterone-sensitive genes. The consensus sequence of MREs generally consists of an inverted hexameric palindrome separated by three nucleotides (AGAA-CAnnnTGGTCT). The MR is then able to recruit specific coactivators, in a sequential and/or combinatorial manner, that subsequently enhance transcriptional activation through direct interaction with the basal transcription factors and chromatin remodeling involving histone acetylation/methylation. Several aldosterone-regulated genes have been identified, including the serum- and glucocorticoid-inducible kinase (sgk1), a serine threonine kinase that is able to phosphorylate the ubiquitin ligase Nedd4-2, which in turn controls the retrieval of the subunits of the epithelial sodium channel from the apical membrane of the cell. Other aldosterone-induced proteins, such as the small monomeric GTP-binding protein Kirsten Ras (Ki-Ras), the glucocorticoid-induced leucine zipper protein (GILZ), and the N-myc down-regulated gene 2 (NDRG2), also seem to play an important role during the early phase of aldosterone responses in the renal tubule. Collectively, aldosterone stimulates the biosynthesis and activity of sodium channels and pumps, leading to an enhanced vectorial transepithelial sodium transport from the tubular lumen to the basolateral space.

Another important issue in the mechanism of aldosterone action is the apparent nonselectivity of the MR. Indeed, aldosterone and glucocorticoids such as cortisol are equally able to bind to the MR with high affinity ( $K_d$  in the nanomolar range). However, the enzyme 11 $\beta$ -hydroxysteroid dehydrogenase type 2 (11 $\beta$ -HSD2), by metabolizing 11 $\beta$ -hydroxysteroids into 11 keto-derivatives such as cortisone, which exhibits no affinity for the MR, plays a pivotal role in



**Figure 1** Schematic representation of the molecular mechanism of aldosterone action in a polarized epithelial cell. MR, mineralocorticoid receptor or aldosterone receptor; hsp90 and hsp70, heat shock protein 90 and 70, respectively; MREs, mineralocorticoid response elements; ENaC, epithelial sodium channel; Sgk1, serum- and glucocorticoid-inducible kinase; NDRG2, N-myc down-regulated gene 2; Ki-Ras, small monomeric GTP-binding protein Kirsten Ras; GILZ, glucocorticoid-induced leucine zipper protein; Na<sup>+</sup>,K<sup>+</sup>-ATPase, sodium/potassium ATPase; 11β-HSD2, 11β-hydroxysteroid dehydrogenase.

mineralocorticoid selectivity in epithelial cells, preventing permanent occupancy of the MR by the more prevalent glucocorticoid hormones.

## STRUCTURE OF THE MR

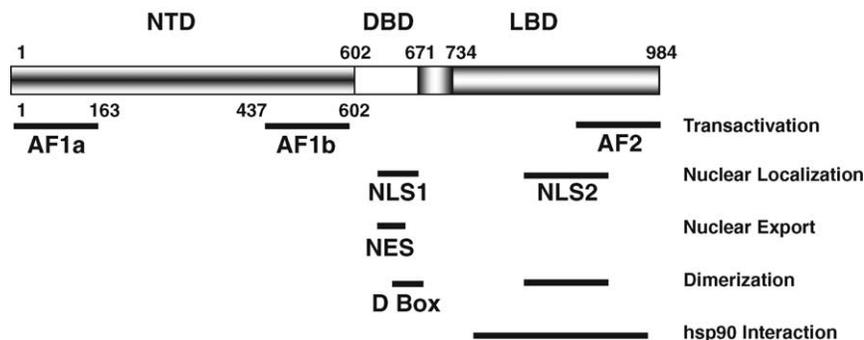
The cloning of the human MR cDNA in 1987 by Ron Evans's laboratory facilitated the deduction of its primary structure and the definition of different functional domains. Like other members of the nuclear receptor superfamily of which it is a member, the MR displays a common modular structure and is composed of three distinct functional domains. A schematic representation of the MR is given in Fig. 2.

The amino terminal region, also referred to as the transactivation domain, is 602 amino acids long and so constitutes the longest domain among nuclear receptor family members. Its primary sequence bears little resemblance to that of other family members, sharing less than 15% homology with its closely related receptor, the glucocorticoid receptor. However, among all known mammalian species, more than 85%

of this MR domain is highly conserved, suggesting that it contains specific and important functions. This domain harbors two ligand-independent activation functions: AF1a located in aa 1 to 167 and AF1b spanning aa 445 to 602 and presumably an inhibitory transactivation region between aa 167 and aa 437.

The centrally located, highly conserved DNA-binding domain is responsible for the specific interaction with hormone response elements located in the promoters of aldosterone target genes. This domain has a rigid structure, is highly hydrophilic, is very rich in cysteine residues, and is composed of two zinc finger structures. It contains a P box that constitutes the interacting contact with the half-site of the response element and a D box that is responsible for weak dimerization. A nuclear export signal has also been identified between the two zinc fingers, and a weak ligand-independent nuclear localization signal, NSL1, has been shown to be located next to the C-terminal site of the DBD.

The ligand-binding domain at the carboxy-terminus part of the receptor is 250 amino acids



**Figure 2** Different functional domains of the human mineralocorticoid receptor. Numbering represents the amino acids. NTD, N-terminal domain of the receptor; DBD, DNA-binding domain; LBD, ligand-binding domain; AF, activating function; NLS, nuclear localization signal; NES, nuclear export signal; D Box, dimerization box; hsp90, heat shock protein 90.

long and is separated from the DBD by a hydrophilic proline-rich hinge region. The complex C-terminal domain is responsible for ligand binding and contains a ligand-dependent nuclear localization signal (NLS2), multiple contact sites for hsp90 interaction, and a ligand-dependent activating function AF2 domain. This domain is highly structured and is composed of 12  $\alpha$ -helices (H1–H12) and one  $\beta$  antiparallel sheet on which the steroid hormone lies.

Aldosterone binds to the MR with high affinity, but glucocorticoids such as cortisol bind to the receptor with equivalent affinity. Although similar dissociation constants ( $K_d$ ) have been calculated by Scatchard plot analysis, dissociation rates ( $k - 1$ ) are much faster for glucocorticoids than for aldosterone. This intrinsic property of the MR to discriminate between aldosterone and glucocorticoids constitutes an additional molecular mechanism that ensures the mineralocorticoid selectivity of aldosterone action within target cells from a dynamic point of view. Finally, it has been shown that the aldosterone–MR complex presumably adopts conformation different from that of the glucocorticoid–MR complex, whereby distinct interaction between the N-terminal domain and the LBD occurs. This leads to the recruitment of particular coactivators resulting in a highly specific transcriptional response.

The three-dimensional structure of the MR was deduced by using an analogy of the crystal structures of the ligand-binding domain of other steroid hormone receptors. This facilitated the precise definition of the amino acid contacts with the functional groups of the steroid. Thus, the 3-ketone function of aldosterone interacts with the glutamine residue at position 776 and the arginine residue at position 817. On the other hand, the 20-ketone function contacts the cysteine residue at position 942, and the

21-hydroxyl and 18-hydroxyl groups are anchored by the asparagine residue at position 770. On ligand binding, the helix 12 within the AF2 domain rotates tightly against the LBD, and this, together with the changes in helices 3 to 5, facilitates the interaction of the receptor with coactivators of the steroid receptor coactivator (SRC) family.

### CELL-SPECIFIC EXPRESSION AND SUBCELLULAR LOCALIZATION OF THE MR

The tissue-specific expression of the MR is presented in [Table I](#). The MR is essentially expressed at relatively high levels in polarized cells of sodium-transporting epithelia in the distal parts of the nephron (from the cortical part of the thick ascending limb of Henle's loop, distal tubule, and connecting tubule to the cortical and medullary collecting tubule), colon, pneumocytes, and salivary and sweat glands. It is now well established that the MR is also present in “non-classical” aldosterone target tissues, most notably neurons of the hippocampus, cardiomyocytes, adipocytes, vascular smooth muscle cells, and (presumably) other cell types. With respect to intracellular localization, initial immunocytochemical studies demonstrated that the MR is predominantly located in the cytoplasmic compartment in the absence of ligand and is translocated into the nucleus on aldosterone exposure. Experiments using green fluorescent protein–MR chimeras allowed the subcellular localization of the MR and its kinetics in living cells to be examined. As illustrated in [Fig. 3](#), on aldosterone binding, the MR rapidly translocates in the nucleus within minutes and is sequestered in specific areas within this

**Table 1** Tissue-Specific Expression of Mineralocorticoid Receptor

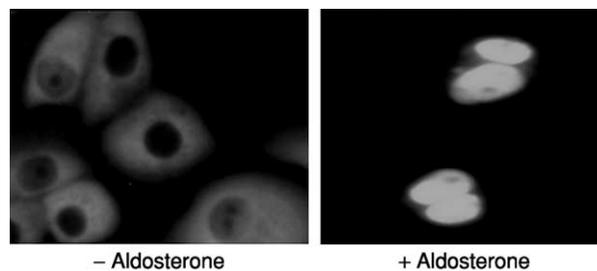
Epithelial cells
Distal nephron
Colon
Salivary glands
Sweat glands
Pneumocytes
Nonepithelial cells
Neurons
Cardiomyocytes
Adipocytes
Smooth muscle cells

compartment, where transcriptionally active regions of the chromatin presumably exist.

## REGULATION OF MR EXPRESSION

The human MR (hMR) gene is localized to the q31.2 region of chromosome 4 and spans approximately 450 kb. The hMR gene is composed of 10 exons, including 2 untranslated first exons referred to as exon 1 $\alpha$  and exon 1 $\beta$  (Fig. 4). Alternative transcription of these two 5'-untranslated exons generates two mRNA isoforms: hMR $\alpha$  and hMR $\beta$ . Given that the hMR translation initiation site, as defined by the start codon ATG, is located 2 bp downstream from the beginning of exon 2, these two isoforms give rise to the same translation product. The last 8 exons, from 2 to 9, encode the various functional domains of the protein. Exon 2 codes for the N-terminal domain of the receptor. The two small exons 3 and 4 encode each of the two zinc fingers of the DNA-binding domain, whereas the last 5 exons encode the ligand-binding domain of the receptor. However, the existence of other hMR splice variants has been demonstrated, and this seems to play a major role in modulating receptor function.

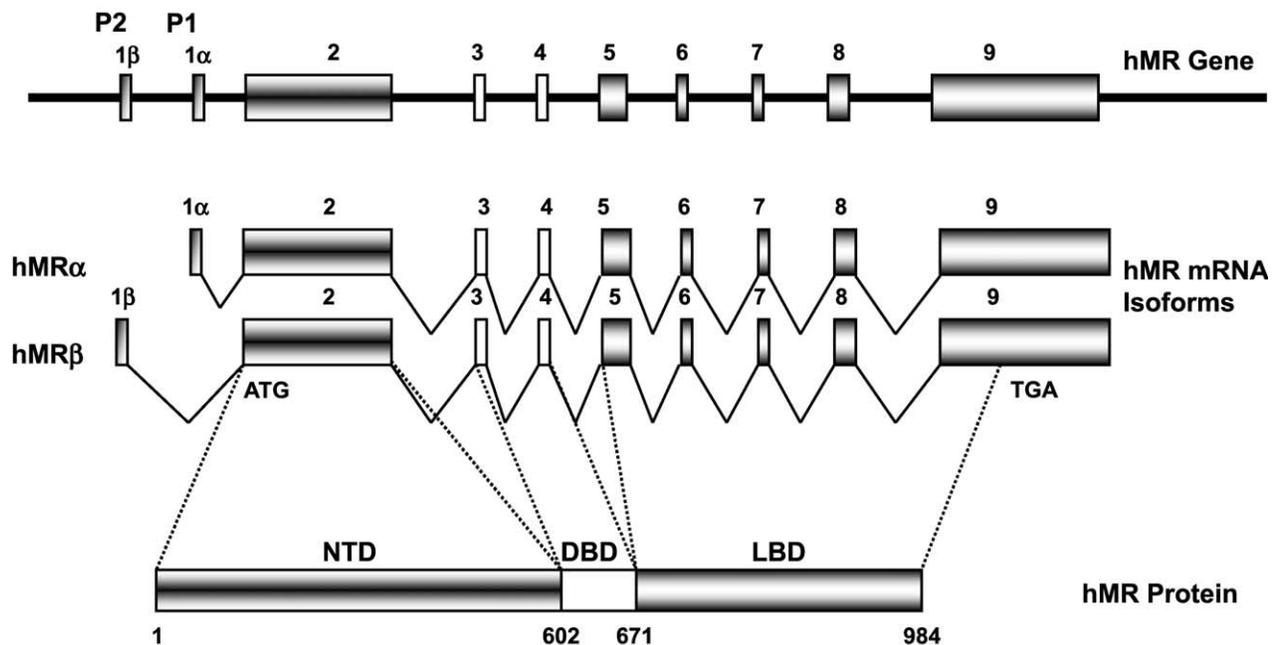
The hMR gene expression is controlled by two different promoters that differ in terms of their basal activity as well as their hormonal regulation. Experiments in transgenic mice have shown distinct tissue-specific use and activity of these two hMR regulatory regions *in vivo*. The proximal P1 promoter corresponding to the 5'-flanking region of exon 1 $\alpha$  is a relatively strong promoter that is transcriptionally active in all aldosterone target tissues, whereas the distal P2 promoter flanking exon 1 $\beta$  is weaker and has a more restricted pattern of expression; thus, it is presumably used during specific developmental stages or physiological situations.



**Figure 3** Aldosterone-dependent nuclear translocation of the hMR. RCSV3 rabbit renal cells were transfected with the expression vector encoding for a chimera protein consisting of human mineralocorticoid receptor and enhanced green fluorescent protein EGFP. The intracellular distribution of EGFP-hMR in living cells was observed directly on an inverted IRB microscope. In the absence of aldosterone, the receptor is predominantly in the cytoplasm, whereas 15 min after addition of  $10^{-8}$  M aldosterone, the receptor is localized in the nucleus.

## PATHOPHYSIOLOGICAL EFFECTS OF ALDOSTERONE

Aldosterone is primarily implicated in the maintenance of water and salt homeostasis by regulating sodium reabsorption and potassium excretion across tight epithelia. As such, aldosterone plays a key role in the regulation of blood pressure, and in turn, the dysfunctional regulation of aldosterone secretion and action is implicated in many human diseases such as hypertension and heart failure. In addition to the renal effects of aldosterone, it has become evident that aldosterone exerts direct effects on the cardiovascular system. Aldosterone excess leads to the development of cardiac hypertrophy and fibrosis, which are involved in cardiac remodeling and heart failure. Even though the molecular mechanisms remain obscure, the detrimental effects of aldosterone on cardiovascular function led to major clinical trials aimed at demonstrating beneficial effects of antimineralocorticoid compounds. Initially, the RALES study demonstrated the efficiency of spironolactone treatment by showing that it significantly reduces the morbidity and mortality (by 30%) of patients with severe congestive heart failure. A fairly new selective aldosterone antagonist, eplerenone, has been approved by the Food and Drug Administration for the treatment of high blood pressure. This antimineralocorticoid has reduced progestagenic and antiandrogenic activities more than has spironolactone. The EPHESUS study also clearly demonstrated the beneficial effects of eplerenone in patients with acute myocardial infarction complicated by left ventricular dysfunction and heart failure. It is likely that aldosterone receptor



**Figure 4** Structure of human mineralocorticoid receptor gene, mRNA isoforms, and hMR protein. P1 and P2 are the alternative promoters of the hMR gene. Numbering represents the exons. ATG and TGA are the starting and stop codons, respectively.

blockade preventing renal, cerebral, and vascular injuries will become increasingly important in managing hypertension, heart failure, and atherosclerosis as well as for patients during the postmyocardial infarction period.

## GENETIC ALTERATIONS OF THE MR

There is now clear evidence that genetic alterations of the MR are associated with human diseases. The first MR mutations were found in patients with autosomal dominant or sporadic pseudohypoaldosteronism type I, an inherited disorder characterized by renal salt wasting during infancy and associated with failure to thrive, hyponatremia, hyperkalemia, and high plasma aldosterone levels. Therefore, these clinical and biological features are consistent with aldosterone resistance. These heterozygous frameshift, nonsense, or missense mutations occur within different functional domains of the MR receptor that, in turn, affect receptor function in different ways, generally resulting in receptor inactivity.

Conversely, a gain of function mutation in the MR has been described in a family with severe early-onset hypertension that is exacerbated by pregnancy. The mutation, a substitution of leucine for serine at codon 810, lies within the ligand-binding domain and has

been shown to drastically modify the receptor steroid specificity. Indeed, further experiments demonstrated a constitutive MR activation in the absence of ligand. In addition, progesterone, spironolactone, and even cortisone were able to fully activate the mutant receptor, consistent with the clinical presentation of gestational hypertension. Interestingly, various genetically engineered animals mimicking human disease have proven to be useful in terms of analyzing the *in vivo* function of the MR. MR gene inactivation achieved by homologous recombination leads to knockout mice developing symptoms of pseudohypoaldosteronism. In contrast, hMR overexpressing transgenic mice exhibit specific alteration in renal and cardiac function. Altogether, these animal models constitute attractive new experimental systems to further explore the widespread and pleiotropic function of aldosterone receptors *in vivo* and to decipher the molecular and cellular events underlying the aldosterone signaling pathway.

## See Also the Following Articles

Aldosterone in Congestive Heart Failure • Hypertension, Overview • Mineralocorticoids and Mineralocorticoid Excess Syndromes • Primary Aldosteronism (PAL) • Tissue Renin-Angiotensin-Aldosterone System

## Further Reading

- Arriza, J. L., Weinberger, C., Cerelli, G., Glaser, T. M., Handelin, B. L., Housman, D. E., and Evans, R. M. (1987). Cloning of human mineralocorticoid receptor complementary DNA: Structural and functional kinship with the glucocorticoid receptor. *Science* **237**, 268–275.
- Berger, S., Bleich, M., Schmid, W., Cole, T. J., Peters, J., Watanabe, H., Kriz, W., Warth, R., Greger, R., and Schutz, G. (1998). Mineralocorticoid receptor knockout mice: Pathophysiology of  $\text{Na}^+$  metabolism. *Proc. Natl. Acad. Sci. USA* **95**, 9424–9429.
- Fajart, J., Wurtz, J. M., Souque, A., Hellal-Levy, C., Moras, D., and Rafestin-Oblin, M. E. (1998). Antagonism in the human mineralocorticoid receptor. *EMBO J.* **17**, 3317–3325.
- Farman, N., and Rafestin-Oblin, M. E. (2001). Multiple aspects of mineralocorticoid selectivity. *Am. J. Physiol. Renal Physiol.* **280**, F181–F192.
- Geller, D. S., Farhi, A., Pinkerton, N., Fradley, M., Moritz, M., Spitzer, A., Meinke, G., Tsai, F. T., Sigler, P. B., and Lifton, R. P. (2000). Activating mineralocorticoid receptor mutation in hypertension exacerbated by pregnancy. *Science* **289**, 119–123.
- Geller, D. S., Rodriguez-Soriano, J., Vallo Boado, A., Schifter, S., Bayer, M., Chang, S. S., and Lifton, R. P. (1998). Mutations in the mineralocorticoid receptor gene cause autosomal dominant pseudohypoaldosteronism type I. *Nat. Genet.* **19**, 279–281.
- Le Menuet, D., Isnard, R., Bichara, M., Viengchareun, S., Muffat-Joly, M., Walker, F., Zennaro, M.-C., and Lombès, M. (2001). Alteration of cardiac and renal functions in transgenic mice overexpressing human mineralocorticoid receptor. *J. Biol. Chem.* **276**, 38911–38920.
- Le Menuet, D., Viengchareun, S., Penfornis, P., Walker, F., Zennaro, M.-C., and Lombès, M. (2000). Targeted oncogenesis reveals a distinct tissue-specific utilization of alternative promoters of the human mineralocorticoid receptor gene in transgenic mice. *J. Biol. Chem.* **275**, 7878–7886.
- Pitt, B., Remme, W., Zannad, F., Neaton, J., Martinez, F., Roniker, B., Bittman, R., Hurley, S., Kleiman, J., and Gatlin, M. (2003). Eplerenone, a selective aldosterone blocker, in patients with left ventricular dysfunction after myocardial infarction. *N. Engl. J. Med.* **348**, 1309–1321.
- Stockland, J. D. (2002). News ideas about aldosterone signaling in epithelia. *Am. J. Physiol.* **282**, F559–F576.
- Zennaro, M.-C., Keightley, M.-C., Kotelevtsev, Y., Conway, G., Soubrier, F., and Fuller, P. J. (1995). Human mineralocorticoid receptor genomic structure and identification of expressed isoforms. *J. Biol. Chem.* **270**, 21016–21020.



# Alkaline Phosphatase

David Goltzman and Dengshun Miao  
McGill University, Montreal, Quebec, Canada

## Glossary

**alkaline phosphatase** A membrane-bound metalloenzyme that consists of a group of isoenzymes, all glycoproteins, encoded by at least four different gene loci: tissue nonspecific, intestinal, placental, and germ cell alkaline phosphatase; catalyzes the hydrolysis of a wide range of phosphomonoesters at alkaline pH.

**hypophosphatasia** An inheritable disorder characterized by defective bone mineralization and a deficiency of tissue nonspecific alkaline phosphatase activity.

**isoenzymes** Enzymes that have the same catalytic activity but differ slightly in amino acid sequence and/or posttranslational modifications.

Since its first description by Suzuki and colleagues in 1907, alkaline phosphatase (ALP) has been investigated continuously and extensively. For most of the past century, there has been widespread use of ALP activity in serum as an enzymatic signal for a variety of disease states involving, in particular, the liver and bone. Investigations directed at the molecular properties of the enzyme have been relatively recent. Quantitation of serum ALP activity has been a routine in hospital laboratories since the 1930s, and this test is perhaps the most frequently performed enzyme assay. Bone ALP became the clinically most relevant enzyme in the diagnosis of bone disease. In spite of its broad use as a clinical marker, the physiological function of this protein, which is ubiquitous in nature, is largely unknown. Identification of ALP gene mutations in hypophosphatasia, a rare heritable form of rickets, has confirmed that ALP functions importantly in skeletal mineralization in humans. Still, many questions remain.

## GENETICS AND EXPRESSION

Alkaline phosphatase (ALP) is a membrane-bound metalloenzyme that consists of a group of isoenzymes. Each isoenzyme is a glycoprotein encoded by different gene loci. At least four loci have been identified: tissue nonspecific, intestinal, placental, and germ cell ALP.

It is believed that the evolution of the ALP gene family has involved the duplication of a primordial tissue nonspecific ALP (TN-ALP) gene to create the TN-ALP gene and an intermediate intestinal ALP (IAP) gene, followed by additional duplications of the latter to create intestinal, placental, and germ cell ALP genes. Only humans and great apes have placental ALP; all other mammals have IAP.

The gene encoding TN-ALP maps to the short arm of chromosome 1, bands p36.1–p34. It is expressed at its highest levels in liver, bone, and kidney (hence its alternate name L/B/K ALP) as well as at lower levels in various other tissues. Differential processing of the TN-ALP gene product occurs within the cell or during passage of the protein molecule out of its cell of origin. In this way, differential glycosylation of TN-ALP gives rise to tissue-specific isoforms.

The gene encoding IAP is a member of the gene family mapping to the long arm of chromosome 2 (q34–q37). IAP is present at high levels in intestinal tissue and at trace levels in the kidney. In contrast to the other ALP isoenzymes, the carbohydrate side chains of IAP are not terminated by sialic acid. Distinct IAPs can be isolated from fetal and adult intestinal tissue, with the fetus forming a sialylated isoenzyme in contrast to the adult. The fetal and adult forms differ not only in the carbohydrate content but also in the protein moiety itself, suggesting that a separate ALP gene locus may exist in humans during fetal development. Fetal IAP is present in amniotic fluid and in the meconium. This fetal/embryonic gene has also been identified in cancer cells and is designated the Kasahara isoenzyme.

The human placental ALP gene was also mapped to chromosome 2. It exhibits 87% homology with the IAP gene. There are, however, amino acid differences at their carboxyl-terminals. Placental ALP is a heat-stable enzyme present at high levels in the placenta. A trace amount of this isoenzyme can be detected in normal sera. Part of the serum placental-type activity originates from neutrophils. Placental ALP activity has also been detected in normal type I pneumocytes

and in the human ovary and cervix. The placental ALP gene can be reexpressed by cancer cells as the Regan isoenzyme. Placental ALP is a very polymorphic enzyme, with up to 18 identified allelozymes resulting from point mutations, in contrast to the other ALP isoenzymes.

The gene encoding germ cell ALP (GCAP) was also mapped to chromosome 2. It encodes testis/thymus ALP and can be expressed in the placenta at low levels. GCAP in testis appears to be localized to the cell membrane of immature germ cells and, like the other ALP isoenzymes, is attached to the cell membrane by means of a glycan-phosphatidylinositol (GPI) anchor. Like the intestinal and placental ALP genes, it can be reexpressed by cancer cells.

## LOCALIZATION OF ALP

Localization of ALP in the cell has been performed by fractionating subcellular components by ultracentrifugation and then performing electron microscopic cytochemistry. From these studies, one can conclude that most of the ALP activity in the majority of cell types is located on the plasma membrane. Still, a small amount of enzyme activity has been shown to be present at other intracellular sites such as the nucleus, endoplasmic reticulum, and Golgi apparatus.

Histochemical techniques at the level of the electron microscope showed reaction product for ALP on the microvillar membranes, multivesicular bodies, Golgi complex, lysosomes, endoplasmic reticulum, and nuclear membranes of the absorbing epithelial cells of the intestinal mucosa. ALP in the kidneys is localized on the microvilli of the proximal tubular cells. In liver, the reaction product of ALP is present on the microvilli of the bile canaliculi. In bone, ALP activity is present on the outer surface of plasma membranes of maturing and hypertrophic chondrocytes and of preosteoblasts and osteoblasts. The reaction product is also present in chondrocyte lacunae, in matrix vesicles in cartilage matrix, as well as among uncalcified collagen fibrils of osteoid in bone. In placenta, ALP activity is found along the plasma membrane lining the microvilli and in vesicles of the syncytiotrophoblast. However, in eosinophil leukocytes, the ALP reaction product was found in only minimal amounts in the plasma membranes but was prominent on the nuclear membranes and outer compartment of mitochondria. The Golgi membranes and endoplasmic reticulum reacted but did so less intensely. The positive cytochemical reactions for ALP are confined to the cytoplasm of the more mature

neutrophilic granulocyte. ALP activity was also localized to fibroblasts, vascular smooth muscle, and endothelial cells in brain, heart, and endocrine glands.

## STRUCTURE AND ANCHORING OF ALP TO THE CELL MEMBRANE

Human ALP enzymes have not yet been obtained in crystalline forms suitable for X-ray analysis. However, their active sites, and the active site of *Escherichia coli* ALP, show a high degree of homology, so that key regions of the human isoenzymes can be interpreted with reference to the corresponding regions of the bacterial enzyme. *E. coli* ALP exists as a dimer of identical subunits, each of which contains 429 amino acids. Crystallographic observations of the molecule from *E. coli* have revealed the three-dimensional structure of dimeric ALP: a bat-like figure with a metal ion triplet in each active site region. Zinc can bind to all of these sites but binds particularly strongly to four sites per dimer, with magnesium occupying two sites of the dimer.

ALP belongs to the large group of proteins attached to the outer surfaces of cells by a C-terminal GPI anchor. As such, it is an ectoenzyme expressed on the outside of the cell. The enzyme is a tetramer when it is membrane bound, but it circulates as a dimer. Phospholipase C or D, which is abundant in plasma, potentially converts the membrane-bound form to a soluble form.

## IDENTIFICATION OF ALP ISOENZYMES

Total serum ALP activity remains one of the most frequently measured enzyme activities in clinical medicine. Several different substrates have been introduced for its assay. Of these, paranitrophenylphosphate (PNPP) is probably the most widely used. Many methods have been proposed to separate the various ALP isoenzymes, including heat denaturation, chemical inhibition of selective activity, gel electrophoresis, precipitation by wheat germ lectin, and immunoassay.

The heat denaturation method is based on the gradation in heat stability at 56°C of the ALP enzymes found in serum. This heat stability ranges from placental ALP, which is completely heat stable; to liver ALP, which has intermediate stability; to bone ALP, which is very labile. The mean remaining enzyme activities after 15 min at 56°C are 11, 21, 90, and 87% of the original activity for the bone,

liver, intestine, and placenta isoenzymes, respectively. Determining the tissue origin of the enzymes by employing differential heat inactivation demands precise control of the temperature during the assay.

Selective chemical inhibitors have also been used to separate ALP isoenzymes. L-phenylalanine and L-tryptophane inhibit intestinal and placental ALP, whereas levamisole and L-homoarginine inhibit TN-ALP. Each of these inhibitors is stereospecific and noncompetitive.

Another common method for distinguishing among ALP isoenzymes is polyacrylamide gel electrophoresis (PAGE). Liver ALP carries the highest net negative charge, followed by the placental, bone, and intestinal forms. Liver and bone ALP can be separated sufficiently to allow visual assessment of their relative proportions, but these methods are quite tedious and there is often overlap between the two isoenzymes, making precise quantification difficult.

Wheat germ lectin binds to N-acetylglucosamine and sialic acid residues, and it provides a method by which to separate liver and bone ALP. Based on the differing glycosylation patterns of liver and bone, wheat germ lectin selectively binds the bone form. However, proper standards and lectin concentrations are necessary for accurate resolution.

Attempts to produce tissue-specific monoclonal antibodies have resulted in antibodies with preferential recognition of the liver, intestinal, and placental ALP. A two-site immunoradiometric assay that relies on the use of two monoclonal antibodies has been developed for the bone isoform.

## FUNCTION OF ALP

### Biochemical Function

Three *in vitro* functions have been attributed to ALP: (1) phosphohydrolysis of organic phosphomonoesters of low molecular mass, (2) phosphotransferase activity, and (3) protein phosphatase activity. Whether any of these relate to the physiological role of the enzyme is as yet unknown. ALP has little preference for a particular substrate and will hydrolyze all phosphomonoesters (diesters are not substrates). Catalysis includes phosphorylation of a serine residue at the active site, followed by transfer of the phosphoryl group to either water (phosphohydrolysis) or an organic acceptor alcohol (phosphotransferase). However, phosphoester cleavage is faster if the transfer of phosphate is to an acceptor rather than to water, and the hydrophobic nature of plasma membrane-bound ALP might lead to a preference for organic

acceptors and, thus, principally to a transphosphorylation reaction.

### Physiological Role of ALP

ALPs are widely distributed in nature; they are present in all species, from bacteria to humans. This is an indication that the enzymes are involved in fundamental biochemical processes.

#### **Embryonal Development**

Studies in amphibian embryos led to predictions concerning the expected distribution, on the embryonic flank mesoderm, of a cell surface molecule involved in guidance information for the embryonic cell migration of pronephric duct cells. An investigation of effects of disruption of the embryonic ALP gene on mouse preimplantation development revealed that the absence of the embryonic ALP gene resulted in fewer blastocysts *in vitro*, delayed parturition, and reduced litter size *in vivo*. This observation indicated that the presence of an active ALP gene is beneficial for preimplantation development.

#### **Regulation of Lipid Transport and Intestinal and Renal Phosphate Transport**

The local abundance of ALP on the membranes of cells involved in the transport of many substances (e.g., duodenal cell and small intestinal enterocytes, renal tubular cell, type I pneumocyte in the lung) suggests a role for the enzyme in complex active transport.

The evidence has been accumulating that IAP may play a role in lipid transport. Elevation of serum IAP activity has been demonstrated in rats following feeding of a high-fat diet. The magnitude of this response is dependent on fatty acid chain length. The correlation between lipid concentration and IAP activity in human lymph has led to the speculation that IAP might be involved in lipid transport.

The views concerning the role of ALP in renal inorganic phosphate (Pi) transport differ considerably. The administration of ALP inhibitors decreases Pi reabsorption in the kidney *in vivo*. Studies *in vitro*, however, showed that compounds that inhibit ALP activity did not block the uptake of Pi, suggesting that ALP is not a Pi-transporting enzyme analogous to ATPase in Na<sup>+</sup> transport but rather that the role of ALP in Pi transport is indirect. In the kidney, ALP activity is highest in the early proximal tubule, which coincides with the region of the proximal luminal brush border membrane transport of Pi. Positive

correlations are found between the brush border membrane transport of Pi and the activity of ALP but not of other brush border membrane enzymes. The function of ALP in the intestinal epithelial transport of Pi (and calcium) is also a controversial subject in that inhibitors of ALP are unable to inhibit Pi transport by the intestinal cell.

### **Bone Formation**

Another important role that has been assigned to ALP is its role in skeletal mineralization. Robison was the first to recognize the role of ALP in skeletal mineralization. The strongest evidence that TN-ALP functions in skeletal formation come with the delineation of inherited hypophosphatasia. This disorder is characterized by defective bone mineralization. The severity of hypophosphatasia is highly variable, ranging from stillbirth with almost no mineralized bone to pathological fractures first presenting during adulthood. The clinical heterogeneity of hypophosphatasia probably reflects the numerous mutations that have been described in the TN-ALP gene and that may give rise to various degrees of clinical severity.

The following roles have been postulated for bone ALP in the mineralization process: (1) hydrolysis of organic phosphate esters, resulting in high local Pi concentration and facilitating precipitation of calcium phosphate; (2) destruction of physiological crystal growth inhibitors such as inorganic pyrophosphate and adenosine triphosphate (ATP) through its hydrolyase activity; (3) action as a Pi transporter; and (4) active transport of  $\text{Ca}^{2+}$  or Pi via its ATPase activity. In bone matrix vesicles (functionally active shed plasma membrane fragments of osteoblasts) and the initial site of hydroxyapatite crystal formation, bone ALP activity can be as much as 20 times more than that on the plasma membrane surface of intact osteoblasts. ALP may be involved here in initiation of the calcification process by raising the local concentration of phosphate ions. The extracellular matrix-binding domain of TN-ALP may also be important in directing the migration of matrix vesicles along collagen fibers during the process of bone mineralization.

### **ALP in Liver**

Although liver ALP generally represents at least half of total serum ALP activity in healthy adults, there is little evidence to support a critical hepatic function for liver ALP. It is generally assumed from its location on the sinusoidal membrane that liver ALP acts as a transport protein. Nevertheless it is doubtful that ALP fulfills a key function in liver given that gross deficiency of the liver isoenzyme in congenital

hypophosphatasia does not appear to give rise to any obvious clinical manifestation. However, one report has suggested that liver ALP may protect liver function from immunological injury by a mechanism involving neutralization of endotoxin.

### **ALP in Microvessels**

ALP is one of the main enzymes present in brain and heart microvessels. ALP may have a role in the metabolism of pyridoxal 5'-phosphate, a cofactor of enzymes such as glutamate decarboxylase and glutamate transaminase involved in the metabolism of neural tissue, and might be a key enzyme of the blood-brain barrier regulated by insulin. Thus, insulin has been found to significantly inhibit brain ALP activity. ALP activity associated with capillary endothelial cells is also clearly affected by a hypoxic environment. ALP activity was significantly reduced or absent in areas of hypoxic skeletal and cardiac muscle, and a similar reduction was found in brain emboli associated with cardiopulmonary bypass. Focal loss of ALP activity demonstrated by histochemical methods appears to be a useful probe in identifying ischemic or hypoxic loci in these tissues. It has also been proposed that the expression of ALP in endothelia in brain and heart microvessels may contribute to the vascular hardening and calcification observed in humans. This, in turn, could be related to vascular aging, vascular disease, and the resultant weakening and/or rupture of vessel walls.

## **PATHOLOGICAL SIGNIFICANCE OF ALP**

The major factors that affect ALP activity are age, sex, and hormonal status (puberty or menopause). From birth to 6 weeks, both bone ALP and intestinal ALP increase. No liver ALP is observed until 6 months of age. In children, a wide range of ALP activity exists and correlates with height and weight, and until puberty the bone isoenzyme represents 77 to 87% of the total. Activity increases in children around the age of puberty, with the maximum being earlier in girls than in boys, and corresponds temporally with growth spurts in both sexes. Bone ALP has been reported to increase during pregnancy; however, a gradual increase in total ALP activity is observed during the first 6 months of pregnancy, followed by a rapid increase during the final trimester. This increase is due primarily to the placental enzyme.

In healthy adults, the ratio of bone activity to liver activity is approximately 1:1. After 50 years of age,

total ALP again increases. Bone ALP activity is generally found to be higher in postmenopausal women than in premenopausal women. There is a great deal of interindividual variation in adult ALP levels, but for any one individual, values change little with time. ALP is cleared from the blood very slowly; the half-life varies from 40 h for bone to 7 days for placental isoforms. Biological daily variation of total ALP is estimated to be less than 4%.

## Bone Disease

### **Rickets and Osteomalacia**

Rickets in growing children is characterized by alterations in chondrocyte differentiation and reduced matrix mineralization in the cartilaginous growth plate, by increased osteoblastic activity, and by defective mineralization of bone matrix. The most common cause is vitamin D deficiency. Primary disorders of phosphate homeostasis and renal tubular disorders also can cause rickets in children.

A number of conditions result in vitamin D-deficient rickets, including inadequate exposure of the skin to ultraviolet radiation with inadequate dietary intake or intestinal malabsorption of this vitamin, chronic impaired renal function with insufficient production of 1,25-dihydroxy vitamin D (the metabolically active form of this vitamin), vitamin D receptor abnormalities, and an inherited deficiency of the enzyme 1 $\alpha$ -hydroxylase that produces 1,25-dihydroxy vitamin D. Bone ALP in these children is moderately to greatly elevated. Bone ALP activity can be further increased shortly after starting vitamin D treatment when healing begins, and it can decline progressively when therapy is effective.

X-linked hypophosphatemia due to a mutation in the enzyme PHEX and autosomal dominant hypophosphatemia due to a mutation in FGF23 are characterized by progressively severe skeletal deformities and dwarfism. Bone ALP activity is also high in these children.

Rickets can occur in a variety of disorders with impaired proximal tubular function that produce increased renal clearance of inorganic phosphate and hypophosphatemia. Glomerular filtration can be entirely normal or near normal. Because the serum concentrations of inorganic phosphate and calcium are critical for the formation of hydroxyapatite crystals and mineralization of bone, hypophosphatemia results in defective mineralization.

Osteomalacia is the adult equivalent of rickets in children. It is characterized by a defect in the

mineralization of the osteoid matrix synthesized by the osteoblasts, but because the growth plate is closed, disorders of epiphyseal cartilage do not occur and so growth abnormalities are not seen. The result is an accumulation of unmineralized bone, which in turn results in decreased bone density, as shown by X-ray or other techniques. As in children, disorders of vitamin D metabolism in adults are the most common causes of osteomalacia, and bone ALP is significantly increased in these cases.

### **Paget's Disease of Bone**

Paget's disease of bone is a common disorder in the elderly where excessive bone turnover occurs, leading to the production of structurally abnormal bone. It is characterized during its initial phase by an increased resorption of bone, followed by an intense osteoblastic response. ALP is a very sensitive biochemical sign manifestation of Paget's disease of bone, as are indexes of bone resorption. ALP activity in Paget's disease is higher than in any other bone disease, excluding primary tumors of bone (osteosarcomas) and osteoblastic metastases from extraskeletal tumors (notably prostate cancer). Bone ALP may eventually decrease, possibly as the disease moves into a sclerotic phase. Serum ALP and indexes of bone resorption correlate well with the extent of skeletal involvement and with the response to treatment.

### **Congenital Hypophosphatasia**

Hypophosphatasia is an inherited disorder characterized by defective bone mineralization and a deficiency of TN-ALP activity. Other ALP isoenzyme activity is unaffected. The severe homozygous form is lethal, whereas less severely affected children show generalized bone deformities due to defective mineralization of osteoid. The milder adult form is characterized by pathological fractures and precocious loss of teeth. Three phosphocompounds accumulate endogenously in hypophosphatasia—phosphoethanolamine (PEA), inorganic pyrophosphate (PPi), and pyridoxal 5'-phosphate (PLP)—indicating that these are natural substrates for TN-ALP.

## Hepatobiliary Diseases

Hepatic diseases, such as acute and chronic hepatitis, cirrhosis, carcinoma of the liver, metastatic carcinoma of the liver, and acute and chronic biliary obstruction, are associated with increases in liver ALP activity.

Since the first report of elevated ALP activity in human serum in connection with obstructive jaundice,

much of the current knowledge concerning ALP in cholestasis has been obtained from experiments with bile duct-ligated rats. Studies in rats have shown that bile duct ligation or exposure of cultured explants of rat liver to bile from bile duct-ligated rats causes a marked increase in hepatic ALP synthesis in the liver, followed by an increase in serum ALP activity in the bile duct-ligated rats. Several bile acids have the capacity to stimulate ALP activity in a dose-dependent fashion in rat liver cell cultures.

In patients with liver cirrhosis, the catabolism of ALP may be reduced. Thoracic lymph flow is increased several-fold in cirrhotic patients, and the intestinal isoenzyme enriches the thoracic lymph and the serum, especially after a fatty meal. It is conceivable that these two factors (decreased clearance by the cirrhotic liver and enrichment of thoracic lymph with intestinal isoenzyme) are responsible for the frequent occurrence of the intestinal isoenzyme in the serum of cirrhotic patients.

### Tumor Forms of ALP

Since the first description of ectopic production of ALP by tumor tissue, three different isoenzymes from independent gene loci have been detected in cancer patients: (1) Regan isoenzyme or term placental ALP; (2) Nagao, testicular, or placental ALP-like; and (3) Kasahara or fetal intestinal ALP.

Total serum ALP has been shown to be an effective indicator for metastatic breast cancer, with a sensitivity for metastases of 30 to 40%. Total ALP has also been suggested to be a significant prognostic factor for prostate cancer. Using various statistical methods, sensitivity, specificity, and accuracy for liver metastases in lung cancer were 71, 89, and 86%, respectively. Serum bone ALP was twice as sensitive as total enzyme activity in the diagnosis of the presence of bone metastases. Serum placental ALP has been

described as a potential marker in seminoma patients, but environmental influences have been shown to affect placental ALP levels significantly (e.g., smoking leads to a reexpression of placental ALP). Despite these environmental influences, placental ALP has been used successfully for monitoring testicular and ovarian carcinomas.

Therefore, ALP forms appear to play an important role in the diagnosis of a variety of benign and malignant clinical disorders, especially of bone and liver, and can be a useful index for monitoring therapy in these diseases.

### Further Reading

- Bai, X. Y., Miao, D., Goltzman, D., and Karaplis, A. C. (2003). The autosomal dominant hypophosphatemic rickets R176Q mutation in fibroblast growth factor 23 resists proteolytic cleavage and enhances *in vivo* biological potency. *J. Biol. Chem.* **278**, 9843–9849.
- Dehghani, H., Narisawa, S., Millan, J. L., and Hahnel, A. C. (2000). Effects of disruption of the embryonic alkaline phosphatase gene on preimplantation development of the mouse. *Dev. Dynamics* **217**, 440–448.
- Hui, M., and Tenenbaum, H. C. (1998). New face of an old enzyme: Alkaline phosphatase may contribute to human tissue aging by inducing tissue hardening and calcification. *Anatomical Rec.* **253**, 91–94.
- Miao, D., and Scutt, A. (2002). Histochemical localization of alkaline phosphatase activity in decalcified bone and cartilage. *J. Histochem. Cytochem.* **50**, 333–340.
- Narisawa, S., Wennberg, C., and Millan, J. L. (2001). Abnormal vitamin B6 metabolism in alkaline phosphatase knock-out mice causes multiple abnormalities, but not the impaired bone mineralization. *J. Pathol.* **193**, 125–133.
- Van Hoof, V. O., and De Broe, M. E. (1994). Interpretation and clinical significance of alkaline phosphatase isoenzyme patterns. *Crit. Rev. Clin. Lab. Sci.* **31**, 197–293.
- Whyte, M. P. (1994). Hypophosphatasia and the role of alkaline phosphatase in skeletal mineralization. *Endocrinol. Rev.* **15**, 439–461.
- Xu, Q., Lu, Z., and Zhang, X. (2002). A novel role of alkaline phosphatase in protection from immunological liver injury in mice. *Liver* **22**, 8–14.



# Alternative Promoters

Erik Jansen, Torik A. Y. Ayoubi, Wim J. M. Van de Ven, and John W. M. Creemers  
*University of Leuven and Flanders Interuniversity Institute for Biotechnology, Belgium, The Netherlands*

## Glossary

**promoter** DNA structures containing *cis*-acting regulatory elements required for efficient initiation of transcription and for controlling expression of a gene.

**transcription units** Templates for RNA polymerases that encode information for the production of a single protein product (simple transcription units) or templates that produce multiple mature mRNAs that can give rise to multiple, albeit related, proteins (complex transcription units).

Alternative promoters can be defined as two or more promoters used to generate the same primary transcript or at least partially overlapping primary transcripts. Alternative promoters will always give rise to differential transcription initiation, resulting in differences in the 5' region of the mRNA isoforms.

## INTRODUCTION

Probably the biggest surprise coming from the human genome sequence is the much lower than anticipated gene number. Therefore, it might be expected that complexity of the human proteome is achieved by transcriptional, translational, and posttranslational diversity. The use of alternative promoters is a frequently used way in which to generate multiple protein isoforms from a single gene. In addition, alternative promoters play an essential role in the control of transcription in a spatial and temporal fashion necessary for the proper development and differentiation of specialized cell types that define multicellular organisms.

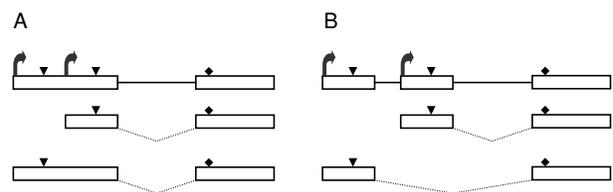
## THE NEED FOR ALTERNATIVE PROMOTERS

Expression of genes in more than one tissue or developmental stage often requires distinct combinations

of transcription factors. One single promoter might not always be sufficient to accommodate all of the required responses. Another role for alternative promoters is to be able to respond in different cell types to the same extracellular signals or to respond in the same cell to different signals. Finally, alternative transcripts can generate diversity by influencing mRNA stability, translation efficiency, and amino terminus of the encoded protein. A different amino terminus, in turn, can lead to alterations in protein levels, functions, or subcellular localization.

## STRUCTURAL ORGANIZATION OF ALTERNATIVE PROMOTERS

Although there are various patterns of alternative promoter use, the two basic mechanisms are shown schematically in Fig. 1. In the first case (panel A), two tandemly arranged promoters are positioned within the same exon. In the second case (panel B), alternative promoter use will result in alternative first exons. If the start codon is located within the second exon (indicated with a black diamond), there will be no difference between the transcripts at the protein level. However, if start codons are located in the first exons (indicated with arrowheads), different protein isoforms will be generated. Another



**Figure 1** Schematic representation of the organization of genes containing alternative promoters. Exons are depicted as boxes, and intervening sequences are depicted by solid lines. Dotted lines connecting exons indicate splicing patterns. Arrows indicate transcription initiation sites. Diamonds and arrowheads represent start codons, as discussed in the text.

layer of complexity can be added by additional promoters and/or exons upstream of the second promoter.

## MODES OF REGULATION OF ALTERNATIVE PROMOTERS

Promoters respond to specific stimuli. In most cases, this is a direct result of the presence or absence of specific response elements in a given promoter. Alternative promoters provide the possibility of responding differentially to ubiquitous transcription activators and repressors, tissue- and developmental stage-specific regulators, and maternal or paternal allele-specific (imprinted) regulators. In combination with the generation of different protein isoforms, this will result in the differential expression and presence of, for instance, hormone receptor  $X_A$  in tissue A and its variant  $X_B$  in tissue B. This variant could display altered ligand specificity and/or altered intracellular signaling. In addition, feedback mechanisms exist in which the gene product (in)directly up- or down-regulates one or several of its alternative promoters.

## SELECTED EXAMPLES

The genes encoding transcription factors CREB (cyclic AMP response element-binding protein) and CREM (cyclic AMP response element-modulatory protein) are an interesting example of regulatory mechanisms. The peculiar aspect resides in the fact that they can encode different isoforms, either activating or inhibiting gene expression, by mechanisms of alternative exon splicing, alternative promoter use, and auto-regulation of promoters. In particular, an internal promoter of the CREM gene directs the expression of a repressor isoform that auto-regulates the alternative promoter, thereby generating a negative feedback loop.

The specific regulation of expression of the cytochrome P450 aromatase gene is a good example of alternative promoter use in the context of endocrine disorders. P450 aromatase is expressed in several normal human tissues such as ovary, placenta, testis, brain, adipose, bone, and skin; in some pathological tissues such as breast and endometrial tumors; and tissues from endometriosis and myofibroma of the uterus. The regulation of expression of P450 aromatase is quite different in these tissues; follicle-stimulating hormone (FSH) stimulates the expression of P450 aromatase in ovary, phorbol esters and

ligands of retinoic acid receptors stimulate the expression in choriocarcinoma cells, and androgens stimulate the expression in the hypothalamus. It has been shown that the tissue-specific and hormonal milieu-specific expression is achieved mainly by tissue-specific alternative use of different promoters. At least nine alternative promoters and 5'-untranslated first exons have been identified. FSH increases the expression of aromatase in ovary by the use and activation of the proximal promoter II. In placenta, phorbol esters regulate the expression of the most upstream promoter I.1. And in adipose tissue, tumor necrosis factor- $\alpha$  stimulates the use of promoter I.4 through activation of transcription factor AP-1. The potential role of P450 aromatase promoter switching in various physiological and pathological processes is illustrated by observations that distinct transcripts are expressed in fetal versus adult human liver and in healthy versus cancerous breast adipose tissue.

A summit of complexity is represented by the GNAS1 gene encoding a stimulatory G protein that is essential for activation of intracellular signaling in response to specific hormone-receptor interactions. This gene has at least four alternative promoters. In addition, it is an imprinted gene that produces different gene products from the maternal and paternal allele through the use of oppositely imprinted alternative promoters. Moreover, abnormal imprinting of the GNAS1 gene promoters can lead to disease states such as pseudo-hypoparathyroidism and somatotrophs adenomas.

## CONCLUSION

Heterogeneity in the 5' ends of mRNAs generated by alternative promoter use in a tissue- or developmental stage-specific manner is common to a large group of genes. It is obvious that the increasing complexity of the mRNA population is an appropriate means of achieving a differential and spatiotemporal expression of the corresponding gene products and their potential pleiotropic actions. There is increasing evidence for several genes that switching of gene expression from one mRNA variant to another may be a key regulatory mechanism in several physiological and pathological processes.

## See Also the Following Articles

Alternative Splicing • Peptide Hormones, Regulation and Gene Expression

## Further Reading

- Ayoubi, T. A. Y., and Van de Ven, W. J. M. (1996). Regulation of gene expression by alternative promoters. *FASEB J.* **10**, 453–460.
- Bulun, S. E., Noble, L. S., Takayama, K., Michael, M. D., Agarwal, V., Fisher, C., Zhao, Y., Hinshelwood, M. M., Ito, Y., and Simpson, E. R. (2001). Endocrine disorders associated with inappropriately high aromatase expression. *J. Steroid Biochem. Molec. Biol.* **61**, 133–139.
- De Cesare, D., and Sassone-Corsi, P. (2000). Transcriptional regulation by cyclic AMP-responsive factors. *Prog. Nucleic Acid Res. Mol. Biol.* **64**, 343–369.
- Lalande, M. (2001). Imprints of disease at GNAS1. *J. Clin. Invest.* **107**, 793–794.



# Alternative Splicing

Scott A. Akker, Bernard C. E. Khoo, and Shern L. Chew  
St. Bartholomew's Hospital, London, United Kingdom

## Glossary

**exon** A nucleotide sequence, separated by introns, that is included in the final mRNA.

**intron** A transcribed nucleotide sequence that is excised in the production of mRNA.

**splicing** The process by which introns are excised and exons united to form mRNA.

Alternative splicing is the process by which different mRNAs are generated from the same pre-mRNA by the selection of alternative exons.

## THE SPLIT GENE, ALTERNATIVE SPLICING, AND PROTEIN DIVERSITY

“Split” genes were first recognized in 1977. It is now estimated that 95% of human genes are split, with coding exons split by noncoding introns. Splicing is the process by which the introns are excised and the exons ligated to form a translatable message. Thus, DNA is transcribed to form premessenger RNA (pre-mRNA), spliced to form messenger RNA (mRNA), and then exported from the nucleus for translation to protein.

Alternative splicing is now recognized as the major contributor to protein diversity in man. Analysis of the human genome suggests that only 1.1% of its 2.91 billion base pairs code for exons. These exons are distributed over an estimated 32,000 genes, and alternative splicing is important in generating diversity and complexity from this limited gene pool. Current estimates suggest that at least 60% of the genes can produce two or more mRNAs by alternative splicing. The potential for generating diversity is demonstrated by the *Slo* gene, which has the potential to generate more than 500 protein isoforms.

The regulation of alternative splicing provides an essential control step in post-transcriptional gene

expression, and the endocrine system is likely to be one of the key systems that directs this in a tissue- and cell-specific manner. Next, we describe the current understanding of the role, mechanisms, and control of alternative splicing, with particular reference to endocrinology.

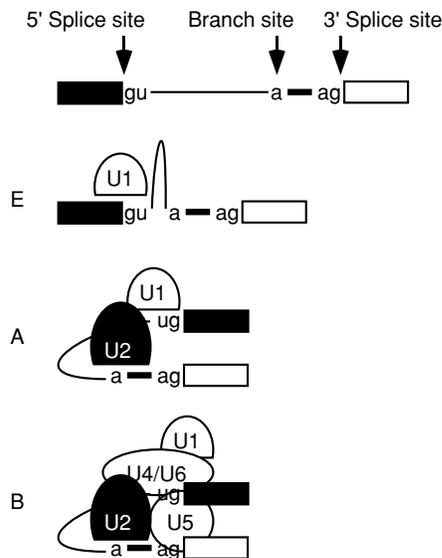
## THE SPLICING MECHANISM

Splicing is performed by a large complex of proteins and small nuclear ribonucleoproteins (snRNPs) called the spliceosome. The spliceosome must be able to recognize the exon/intron boundaries in a precise and reproducible manner. At least part of the exon/intron definition is provided by certain key intronic elements, including a 3' and 5' splice site and a branch site and polypyrimidine tract upstream of each spliced exon (Fig. 1).

Splicing occurs in close physical and temporal relationship to pre-mRNA transcription, with the C-terminal domain of RNA polymerase II helping direct splicing factors to the pre-mRNA. The snRNPs are key players in the spliceosome, and each one has a RNA component that is capable of interacting with the pre-mRNA through Watson–Crick base pairing. U1 snRNP binds the 5' splice site of the exon. U2 snRNP then associates with the polypyrimidine tract of the upstream 3' splice site and the U4/U6.U5 tri-snRNP is recruited. The spliceosome then catalyzes cleavage of the pre-mRNA at the 5' splice site, and a lariat structure with the branch site is formed (Fig. 1). Cleavage at the 3' splice site precedes exon ligation and lariat release. Splicing is followed by the addition of a poly(A) tail, which is also subject to regulation.

## PATTERNS OF ALTERNATIVE SPLICING

The most primitive form of alternative splicing is intron retention; in man, however, the most commonly described form of alternative splicing is exon skipping. Several other mechanisms exist, including the use of alternative 3' and/or alternative 5' splice



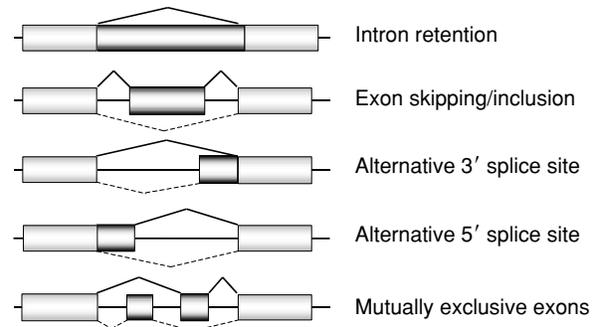
**Figure 1** Basic mechanisms of splice site selection and spliceosome assembly. The exons are shown as boxes and the intron as a line. The bold line in the intron, between the branch site and the 3' splice site, represents the polypyrimidine tract. The stepwise assembly of the spliceosome complexes, early (E) and subsequent (A and B), is shown. Note that only the U1, U2, and U4/U6.U5 snRNPs are depicted, and the many non-snRNP proteins are omitted for clarity. From Akker *et al.* (2001), reproduced by permission of the Society for Endocrinology.

sites (Fig. 2). Many genes in the endocrine system undergo a combination of different alternative splicing events, employing various mechanisms.

## REGULATION OF ALTERNATIVE SPLICING

### Splice Sites

The splice sites that flank exons are necessary but not sufficient to explain exon recognition. Splice sites have a varying degree of similarity to highly degenerate consensus sequences. In fact, splice site-like sequences that also match the consensus occur with great frequency throughout the genome and define a set of pseudosites and pseudoexons that never undergo splicing. Such pseudoexons may outnumber true exons 10 to 1. Thus, there is a spectrum of exons: those that are always spliced (constitutive exons), those that are spliced only under certain conditions (alternatively spliced exons), and those that never undergo splicing (pseudoexons). As a rule, alternatively spliced exons tend to have splice sites with a weaker match to the consensus than constitutive exons, and this allows them to be more amenable to the effects of other sequence elements and factors. Experimentally



**Figure 2** Common patterns of alternative splicing. The central exons undergo alternative splicing, with one pathway represented by solid lines and the alternative with dashed lines. As seen in IGF-1, the same exon within a gene may undergo more than one form of alternative splicing.

enhancing the splice site strength of alternatively spliced exons usually causes them to be included constitutively. The splice site strength of the exons neighboring an alternatively spliced exon is also important. Strengthening the 5' splice site of the upstream exon or the 3' splice site of the downstream exon can lead to an increased degree of skipping of the middle exon.

### Other RNA Sequence Elements and Their Interactions

Major advances in understanding regulated alternative splicing have occurred with the identification of primary sequence elements involved in promoting exon selection (enhancers) or repressing splicing (silencers). The factors functioning through several such regulatory sequences have been isolated. Many important non-snRNP proteins are involved in spliceosome assembly and function. A number of these also regulate splice site selection. In addition to the elements described previously, examples of elements that may function through secondary structure have been described, such as in the insulin receptor described later. Alternative splicing could be subject to regulation via RNA helicases, which function to unwind the secondary structure to allow access of the spliceosome to splice sites and regulatory elements.

#### Splicing Enhancers and Serine-Arginine-Rich Proteins

There are many examples of splicing enhancer elements. These are most commonly sited within exons but may also be intronic. A family of splicing factors has been shown to interact with these elements—the serine-arginine-rich (SR) proteins. SR proteins are characterized by RNA-recognition motifs and domains

containing serine and arginine repeats, which are important for protein–protein interactions. SR proteins are required for constitutive splicing and are also crucial mediators of regulated alternative splicing. SR proteins may regulate alternative splicing through other means. The SR protein SF2/ASF has been proposed to increase the stability of certain mRNAs by direct binding. Some proteins act by modulating SR protein function; for example, SRp86 directly interacts with two other SR proteins to enhance (SRp20) and repress (SC35) their actions in a dose-dependent manner.

### Splicing Silencers and Heterogeneous Nuclear Ribonucleoproteins

Less well documented are examples of splicing silencers. These have also been described in both exonic and intronic contexts. It has been proposed that silencers occur frequently throughout human introns and act to repress pseudo-splice sites. The major family of proteins shown to interact with these elements is the heterogeneous nuclear ribonucleoproteins (hnRNP) family. Only a few have been characterized in terms of their function. Perhaps the best characterized hnRNP is polypyrimidine tract-binding protein (PTB; also known as hnRNP I). PTB exists as several isoforms through alternative splicing and has been associated with several types of exon skipping. PTB binding sites are often found on both upstream and downstream introns, and it has been proposed that an interaction occurs between PTBs, which can form multimers, and together they promote exon skipping. In the case of hnRNP A1, cooperative binding between several molecules occurs, starting from a high-affinity binding site. This results in the “spreading” of hnRNP binding along the pre-mRNA.

The overall picture that is emerging is that each alternatively spliced exon lies in a balance of recognition versus nonrecognition. The spliceosome makes its decision based on the splicing signal strength and the degree of signal enhancement or repression present. A major model of alternative splicing holds that the binding of SF2/ASF to a pre-mRNA is in competition with that of hnRNP A1. The binding of SF2/ASF recruits spliceosomal factors via protein–protein interactions. SF2/ASF may also function to block the propagation of hnRNP A1 binding along the pre-mRNA. Any change in the ratio of enhancing factors to silencing factors can swing the balance toward or away from splicing.

The following mechanisms may regulate enhancing/silencing factors: phosphorylation, demonstrated by STREX and insulin action on SRp40; cellular

localization/sequestration (e.g., the stress-induced cytoplasmic sequestration of hnRNP A1); and temporal regulation of factor concentrations in tissues (e.g., insulin up-regulates the expression of SRp40). A combination of all three of these mechanisms is likely to occur *in vivo*: for example, cytoplasmic sequestration of hnRNP A1 is associated with phosphorylation mediated by the MKK<sub>3/6</sub>-p38 kinase pathway.

## ENDOCRINE EXAMPLES OF ALTERNATIVE SPLICING

Table I provides examples of endocrine stimuli that affect alternative splicing. Here, we focus on endocrine examples that demonstrate some principles of alternative splicing regulation.

Table I Examples of Hormonally Regulated Alternative Splicing Events<sup>a</sup>

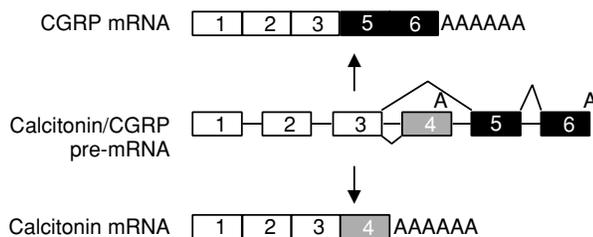
Alternatively spliced mRNA	Stimulus
Insulin receptor	Dexamethasone, glucose, insulin
Calcitonin/CGRP	Dexamethasone
Protein kinase C $\beta$	Insulin
IGF-1	Growth hormone
Fibroblast growth factor receptor	Cytokines
Phosphotyrosine-1B	PDGF, EGF, bFGF
hPMCA2	Calcium
CD44	PDGF, IGF-1, via hnRNP A1
Fibronectin EIIIB (rat)	Insulin, via HRS
Fibronectin ED (human)	TGF- $\beta$ <sub>1</sub> , vitamin D, retinoic acid
Kv3.1 channel	bFGF/depolarization
Agrin	Nerve growth factor
MHC-B	Nerve growth factor
SRp20	Serum/cell cycle
Slo BK channel	Hypophysectomy/ACTH
Thyroid hormone receptor- $\beta$	Tri-iodothyronine
Tau	Tri-iodothyronine
Activating transcription factor-3	TNF- $\alpha$
SERCA3	Retinoic acid, “hypertension”

<sup>a</sup>Abbreviations used: bFGF, basic fibroblast growth factor; EGF, epidermal growth factor; hPMCA2, human plasma membrane Ca-ATPase; IGF-1, insulin-like growth factor-1; Kv3.1, potassium voltage-gated channel; MHC-B, myosin heavy chain II-B; PDGF, platelet-derived growth factor; SERCA3, sarco/endoplasmic reticulum Ca-ATPase 3; TGF- $\beta$ <sub>1</sub>, transforming growth factor- $\beta$ <sub>1</sub>; TNF- $\alpha$ , tumor necrosis factor- $\alpha$ .

### The Calcitonin/ $\alpha$ -Calcitonin Gene-Related Peptide Gene: Tissue-Specific Alternative Splicing Altered by a Hormonal Stimulus

The first described example of a hormonal stimulus altering splicing was in the calcitonin/ $\alpha$ -calcitonin gene-related peptide ( $\alpha$ -CGRP) gene in 1986. The gene contains six exons and two polyadenylation signals. Alternative splicing leads to two proteins, distinct in both structure and function. In thyroid C cells, exon 4 is recognized and spliced with use of the exon 4 polyadenylation signal to generate calcitonin. In neuronal cells, exon 4 is skipped, and exons 5 and 6 are spliced with use of the exon 6 polyadenylation signal to generate  $\alpha$ -CGRP (Fig. 3). Exon 4 is the key because its inclusion leads to cleavage and polyadenylation at its poly(A) site, thus preventing recognition of the downstream exons. The “constitutive” pathway is that of exon 4 inclusion, and it is only skipped in neuronal and cardiac tissue.

The important elements in and around exon 4 have been characterized. It is not surprising that exon 4 undergoes skipping because it contains weak splicing signals (non-canonical branch site and polypyrimidine tract) as well as a weak polyadenylation signal. Mutation of these signals to canonical sequences leads to exon 4 inclusion in all circumstances. Given the weak signals, it is perhaps more surprising that exon 4 is generally included. However, no fewer than three exonic splicing enhancers, one intronic splicing enhancer, and five pentanucleotide repeats, which also act as enhancers, have been described. These appear to compensate for the weak splicing signals. The intronic splicing enhancer is particularly interesting because it contains consensus sequences for both 3' and 5' splice sites that interact with U1 snRNP, PTB, and SF2/ASF. The enhancer and the factors binding it regulate both polyadenylation and terminal exon splicing.



**Figure 3** Alternative splicing of the calcitonin/CGRP gene. The two potential mRNAs are schematically shown: CGRP mRNA (top) and calcitonin mRNA (bottom). A, the polyadenylation signals on the pre-mRNA.

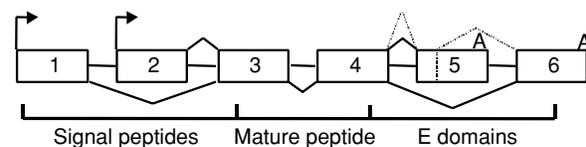
In a medullary carcinoma-derived cell line, dexamethasone provides a stimulus that alters alternative splicing, up-regulating calcitonin production over  $\alpha$ -CGRP production. The intermediaries between dexamethasone and the elements and factors described previously are not known.

### Growth Hormone/Insulin-like Growth Factor-1 Alternative Splicing: Generation of Complexity

This system provides insight into how levels of complexity in gene expression can be established because growth hormone (GH), insulin-like growth factor-1 (IGF-1), their binding proteins, and their receptors all undergo alternative splicing.

GH is coded for by the *GH-1* gene, a five-exon gene that produces two major transcripts and at least three minor transcripts. The predominant isoform of GH is the 22-kDa form resulting from inclusion of all exons. Approximately 5–10% of human GH in the pituitary and in lymphocytes is the 20-kDa form, resulting from the use of an alternative splice site within exon 3. Use of this site is dependent on secondary structure elements and leads to a 15-amino acid deletion from the final protein. The role of this isoform is not known; somatotroph adenomas produce proportionally more 20-kDa isoform than 22-kDa isoform, and GH is carried by two GH-binding proteins that have different affinities for the 20- and 22-kDa isoforms. A small proportion of GH pre-mRNAs undergo exon 3 or exons 3 and 4 skipping, and these isoforms are believed to have a dominant negative effect, inhibiting the action of normal GH. Exon 3 inclusion has been shown to be dependent on an upstream intronic splicing enhancer. Mutations within this intronic splicing enhancer lead to exon skipping and cause isolated GH deficiency.

IGF-1 is derived from a six-exon pre-mRNA. Four of the six exons undergo alternative splicing and lead to the production of different precursor peptides but an identical mature peptide (Fig. 4). The exact roles of these different precursors are not well understood.



**Figure 4** Alternative splicing of the IGF-1 gene. Alternative promoter sites are represented by the arrows of exon 1 and exon 2. Alternative splicing pathways are represented by the sloped lines.

Exon 5 splicing is dependent on tissue type and the hormonal milieu, and an exonic splicing enhancer located within exon 5 is responsive to the SR protein SF2/ASF. Increasing the activity/amount of SF2/ASF enhances exon 5 inclusion. The presence of exon 5 is linked to nuclear and nucleolar localization. The role of the precursors at different sites remains to be determined.

### Insulin Receptor Alternative Splicing: A Role in Insulin Resistance?

The insulin receptor (IR) gene is a 22-exon gene. As in the calcitonin/ $\alpha$ -CGRP gene, there are two alternatively spliced isoforms that are generated by the inclusion or skipping of a single exon—exon 11. Exon 11 is only 36 base pairs and codes for 12 amino acids that lie in the C terminus of the  $\alpha$ -subunit. The two receptor isoforms have been shown to differ in their tissue distribution, dimerization properties, substrate binding, and function (Table II).

In a human hepatoblastoma cell line, dexamethasone increases insulin sensitivity, and this is associated with up-regulation of the IR-B isoform. Glucose also increases IR-B splicing, and insulin levels have been linked with different IR isoform expression, independent of glucose.

Closer analysis has revealed some of the elements involved in alternative splicing of exon 11 and that ultimately mediate hormonal regulation. The weak splice sites of exons 10–12 predispose to regulation. These splice sites appear to compete for limiting splicing factors. Changing the splice sites effects the level of splicing of exon 11. Furthermore, four regulatory elements have been identified. Two elements lie in intron 10; a GA-rich region acts as an intronic

splicing enhancer that may enhance the neighboring weak splice sites. Downstream of this element lies an intronic splicing silencer, which is predicted to form a stem-loop structure with exon 11. The final two elements lie in close proximity to exon 11—one acts as an enhancer and the other as a silencer.

It is generally accepted that hyperinsulinemia is associated with relatively increased IR-A expression in muscle and that this precedes hyperglycemia. This contributes to insulin resistance, but it is not established whether hyperinsulinemia precedes the change in IR alternative splicing or vice versa. There is an additional level of complexity because both IR isoforms form heterodimers with the homologous IGF-1 receptor. Heterodimers of the IGF-1 receptor with the IR-A isoform bind IGF-1, IGF-2, and insulin, and signaling occurs along the IGF pathway. In contrast, heterodimers of the IGF-1 receptor with the IR-B isoform bind only IGF-1. Therefore, hyperinsulinemia can cause activation of both insulin and IGF pathways via IR-A homodimers and IR-A/IGF-1 receptor heterodimers, respectively. The clinical relevance is clear because heterodimers are found at higher levels in skeletal muscle and adipose tissue in type 2 diabetes.

### CaRRE: The First Hormonally Responsive Splicing Regulation Element

The *Slo* gene encodes BK calcium and voltage-activated potassium channels. These channels exhibit functional diversity, partially due to alternative splicing. Two alternatively spliced stress axis-regulated exons (STREXs) have been identified that enhance repetitive firing when included. The skipping of these exons in adrenal chromaffin cells is increased

Table II Characteristics of Insulin Receptor Isoforms

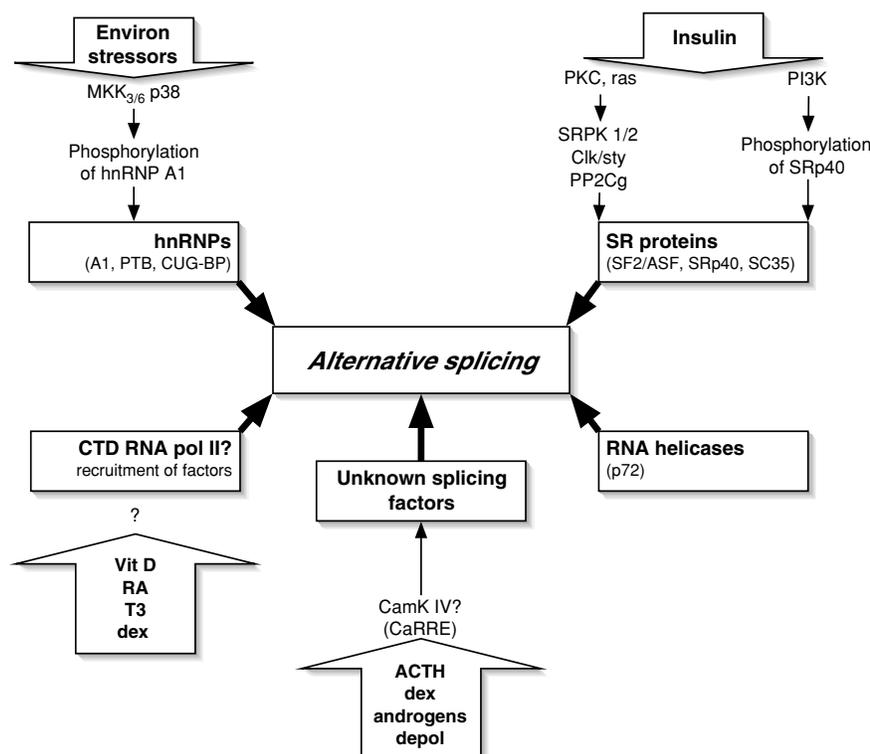
Insulin receptor isoform	IR-A (exon 11 skipping)	IR-B (exon 11 inclusion)
Tissue distribution	Ubiquitous; predominant during embryonic development	Insulin-sensitive tissues: liver, muscle, adipocytes, kidney
Insulin binding affinity	++	+
Sensitivity to insulin action	+	++
Insulin action in $\beta$ cells	Increased insulin transcription	Increased glucokinase transcription
IGF-1 binding affinity	++	$\pm$
IGF-2 binding affinity	+	$\pm$
IGF-1 receptor heterodimer	Binds insulin, IGF-1, IGF-2	Binds IGF-1
Expression in type 2 diabetes	Increased	Decreased
Expression in cancer cells	Increased	Decreased

by hypophysectomy in rats. This effect is prevented by ACTH injections. Treatment of these cells with glucocorticoids increases skipping, whereas androgens decrease skipping. It is proposed that stress hormones regulate alternative splicing of STREX in adrenaline-secreting cells.

One of the sequence elements involved in this alternative splicing has been identified. Depolarization of GH<sub>3</sub> pituitary cells decreases STREX inclusion, and this is dependent on the action of a Ca<sup>2+</sup>/calmodulin-dependent protein kinase (CaMK IV). CaMK IV in turn acts via an element [CaMK IV-responsive RNA element (CaRRE)] upstream of the STREX. Transfer of this element to another exon confers CaMK IV repressibility on that exon. STREX skipping reduces the BK channel Ca<sup>2+</sup> sensitivity, and this provides a feedback pathway by which chronic depolarization will reduce the Ca<sup>2+</sup> sensitivity of the cell.

## Hormones as Regulators of Alternative Splicing

Hormones regulate the alternative splicing of their target genes via multiple mechanisms, and many different second messenger systems mediate their effects (Fig. 5). Many of the downstream splicing factors have been identified, but the intermediaries remain elusive. Many kinases act to regulate alternative splicing, and the phosphorylation of splicing proteins is likely to be a key process. For example, terminal exon inclusion of the protein kinase C (PKC)  $\beta$ II pre-mRNA in skeletal muscle is up-regulated by insulin, which enhances insulin-stimulated glucose uptake. Insulin regulates the alternative splicing via the PI3K signaling pathway by phosphorylation of the SR protein, SRp40. Hormones acting via nuclear receptors may act on the C-terminal domain of RNA polymerase II to regulate the recruitment of splicing factors.



**Figure 5** Possible pathways involved in the regulation of alternative splicing. Primary stimuli are indicated by the large arrows. dex, dexamethasone; depol, depolarization; environ, environmental; T3, tri-iodothyronine; RA, retinoic acid. Possible effector mechanisms are shown in boxes (examples of molecules). Second messenger molecules: Clk/sty, clk/sty kinase; MKK<sub>3/6</sub> p38, p38 mitogen-activated protein kinase kinase; PP2Cg, protein phosphatase 2C gamma; CTD RNA pol II, C-terminal domain of RNA polymerase II; SRPK, SR protein kinases.

## SPlicing AND ENDOCRINE DISEASE

Splicing defects are associated with an increasing array of disease processes and are particularly well represented in inherited endocrinopathies, such as congenital adrenal hyperplasia, multiple endocrine neoplasia, and neurofibromatosis type 1. Mutations that directly affect splicing may be classified into those that disrupt the splice sites and those that change non-splice site sequences. In the former class, mutations at 5' splice sites may cause activation of nearby cryptic 5' splice sites or, more commonly, skipping of the entire adjacent upstream exon. In the latter class, mutations of non-splice site sequences may disrupt regulatory elements for nearby splice sites, as previously described for isolated GH deficiency. An unusual endocrine splicing mutation occurs in the GH receptor gene, in which a deep intronic point mutation directly activates a pseudoexon, resulting in an additional 36 amino acids being included in the GH receptor, which in turn leads to GH insensitivity (Laron's syndrome). Thus, point mutations may cause disease by disrupting splicing regulatory elements.

The best described example of a disease caused by disruption of alternative splicing is myotonic dystrophy (DM1) and insulin resistance. DM1 is caused by CTG trinucleotide repeats at the 3' untranslated end of the DM protein kinase gene. In DM1 skeletal muscle, there are high levels of the hnRNP CUG-binding protein (CUG-BP), causing aberrant regulation of IR alternative splicing and a higher proportion of the IR-A isoform. Overexpression of this protein in normal cells also induces a switch to the IR-A isoform. This has been shown to occur via an intronic CUG-BP binding element upstream of the crucial exon 11. The same binding protein was previously shown to alter alternative splicing of the cardiac troponin T gene in both cardiac and skeletal muscle. Overexpression of CUG repeat RNA in normal cells appears to increase the half-life and steady-state levels of CUG-BP. Thus, a specific pathway has been proposed that leads to insulin resistance and possibly some of the other effects of this multisystem disorder.

Interestingly, higher levels of the IR-A isoform have also been demonstrated in poorly differentiated cancer cells. An autocrine loop has been proposed in anaplastic thyroid cancer in which high levels of IGF-2 have been associated with high levels of IR-A and IR-A autophosphorylation. Because IGF-2 binds IR-A and not IR-B, it may be acting as a mitogen. The presence of heterodimerization with signaling through the IGF system enhances this effect.

## THERAPIES

Most attempts to alter splicing have focused on the disrupted splicing present in the cancer process, and several novel approaches to chemotherapy have been proposed. Complementary antisense oligonucleotides can switch alternative splicing but have problems with delivery and toxicity. Although suitable for cancer therapy, such approaches are less promising as long-term solutions to genetic disease caused by splicing dysregulation. However, two examples demonstrate the potential for novel therapies as our understanding increases. First, dystrophin synthesis is rescued in a Duchenne muscular dystrophy cell line by chimeric snRNAs with complementarity to exon 51 splice sites. Second, work is focusing on the development of compounds to treat human spinal muscular atrophy (SMA). Mutations in the *SMN1* gene cause SMA, but all SMA patients can produce functional SMN protein from the *SMN2* gene. This gene is identical to the *SMN1* gene except for one nucleotide in exon 7. This single nucleotide disrupts an SF2/ASF-dependent exonic splicing enhancer and leads to alternative splicing of exon 7. Exon 7 is therefore excluded from 80% of transcripts derived from the *SMN2* gene, and the 20% full SMN protein produced is not sufficient to prevent disease. Several novel splicing compounds have been identified as enhancers of exon 7 inclusion, and the search continues for a way to up-regulate exon 7 alternative splicing; however, several candidates have been found from cell culture work.

## CONCLUSION

Alternative splicing is an essential means of generating diversity from a limited number of genes. Genes coding for hormones, their binding proteins, and their receptors are particularly well represented. Alternative splicing is an integral component of the endocrine system because it both regulates and is regulated by hormones.

Recent insights have shown that pre-mRNA has many sequence elements within introns and exons that regulate alternative splicing. Many factors acting directly on such elements have been described. The intermediates between a given hormone and its potential effect on a splicing factor are not known. Neither is it understood how splicing factors, many of which are ubiquitously and generously expressed, are regulated to produce subtle changes in spliceosome action.

As further knowledge of the alternative splicing process and its relation to disease processes is obtained, it is hoped that novel therapies can be

designed that divert alternative splicing to the appropriate pathway.

### See Also the Following Articles

Alternative Promoters • Insulin-like Growth Factors

### Further Reading

- Akker, S. A., Smith, P. J., and Chew, S. L. (2001). Nuclear post-transcriptional control of gene expression. *J. Mol. Endocrinol.* **27**, 123–131.
- Caceres, J. F., and Kornblihtt, A. R. (2002). Alternative splicing: Multiple control mechanisms and involvement in human disease. *Trends Genet.* **18**, 186–193.
- Cartegni, L., Chew, S. L., and Krainer, A. R. (2002). Listening to silence and understanding nonsense: Exonic mutations that affect splicing. *Nat. Rev. Genet.* **3**, 285–298.
- Krainer, A. R. (1997). “Eukaryotic mRNA Processing.” IRL Press, Oxford.
- Maniatis, T., and Tasic, B. (2002). Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature* **418**, 236–243.
- Stoss, O., Stoilov, P., Daoud, R., Hartmann, A. M., Olbrich, M., and Stamm, S. (2000). Misregulation of pre-mRNA splicing that causes human diseases. Concepts and therapeutic strategies. *Gene Ther. Mol. Biol.* **5**, 9–30.
- Webster, N. J. G., and Huang, Z. (1999). Hormonal regulation of alternative splicing. In “Post-transcriptional Processing and the Endocrine System” (S. L. Chew, ed.), pp. 1–17. Karger, Basel, Switzerland.



# Alzheimer's Disease and Hormones

Sigbritt Rasmuson and Tommy Olsson  
Umeå University Hospital, Umeå, Sweden

## Glossary

**Alzheimer's disease** The most common cause of cognitive disturbance and dementia; a progressive neurodegenerative disease.

**apolipoprotein E (apoE)** A protein that is important for the metabolism of lipids (e.g., cholesterol); there are three different alleles coding for three isoforms of apoE, and an increased frequency of the apoE  $\epsilon$ 4 allele is seen in Alzheimer's disease.

**cognition** The process of knowing or perceiving.

**declarative memory** Consists of items that are easy to verbalize and generally accessible to conscious recall; this includes both episodic and semantic memory.

**entorhinal cortex** A structure within the hippocampal region that provides the primary pathway by which sensory information enters and leaves the hippocampal region.

**hippocampus** A region of the brain that plays an important role in memory and learning and that has a high concentration of glucocorticoid receptors.

**neuroendangerment** Exposure that makes neurons less likely to survive neurological insults.

**plaques and tangles** Hallmark lesions in the brain in Alzheimer's disease; plaques are extracellular deposits consisting of the protein  $\beta$ -amyloid, whereas tangles are abnormalities of the neural cytoskeleton.

Alzheimer's disease is the most common form of dementia. The hallmark lesions are neurofibrillary tangles and neuritic plaques. The hippocampal formation, a key area for learning and memory processes, is affected early during the course of the disease. A number of hormones may affect the development and progression of this disease. These include hormones readily entering the brain through the blood-brain barrier as well as hormones locally produced within the brain. Potential prevention of dementia as well as treatment of this devastating disease may be achieved in the future through manipulation of hormone effects on brain cells.

## ALZHEIMER'S DISEASE

In 1907, Alois Alzheimer described the first case of the most common cause of dementia. Alzheimer's disease (AD) accounts for at least 50% of all cases of dementia today. The prevalence increases with age, from approximately 2 to 3% at 70 to 74 years to 12 to 16% at 85 to 90 years. The most common cognitive symptoms in AD are impairments in explicit memory, visuospatial problems, inability to calculate, loss of judgment, and progressive loss of language. Furthermore, there are various noncognitive, behavioral, and psychological symptoms such as depression, anxiety, and aggressiveness. The definitive diagnosis of AD is based on a combination of an appropriate clinical history and a histopathological investigation. There are two hallmark pathological brain lesions: neurofibrillary tangles and neuritic plaques. The hippocampus seems to be the brain region that shows the earliest neuron loss in AD, and the pattern of neurodegeneration results in a functional isolation of the hippocampus from the cortical association areas. There are disturbances in multiple neurotransmitter systems in AD, including a profound loss of cortical and hippocampal cholinergic innervation, reduced central somatostatin levels, and changes in glutamatergic and serotonergic activity. Neurotransmitters and circulating hormones play key roles in modulating the complex neurobiological interactions that occur within the brain. Many hormones are also produced locally in the brain. There are important bidirectional links between the central nervous system and the peripheral endocrine organs. Thus, lesions in the brain of different origins may influence endocrine function.

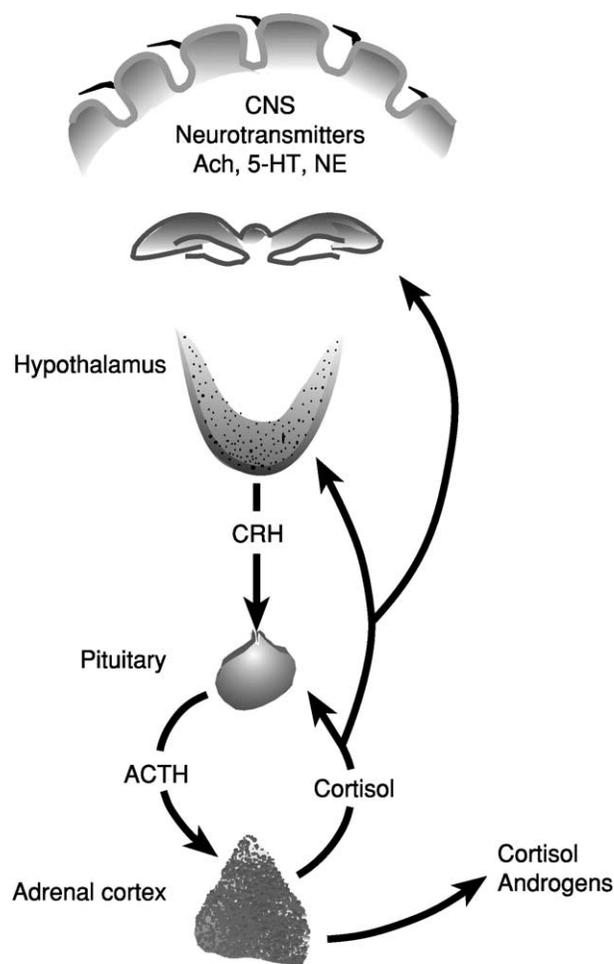
## THE LIMBIC-HYPOTHALAMIC-PITUITARY-ADRENAL AXIS

### Components of the LHPA Axis

The limbic-hypothalamic-pituitary-adrenal (LHPA) axis plays an important role in maintaining homeostasis

under basal and challenge conditions. The components of this axis function as an integrated feedback system in a hierarchic way (Fig. 1). The hippocampus, a limbic forebrain region, is an important neuroanatomical target for interactions between neurotransmitter systems and circulating hormones. The hippocampus has a high density of glucocorticoid receptors (GRs) and an inhibitory effect on cortisol release and is also important for negative feedback function.

Steroid hormones have profound effects on the central nervous system (CNS). These include effects



**Figure 1** Schematic diagram of the limbic-hypothalamic-pituitary-adrenal axis. In Alzheimer's disease, there may be increased glucocorticoid production related to increased corticotropin-releasing hormone (CRH) secretion and disturbances in negative feedback. Circulating adrenal androgen levels are increased during the early phase of the disease, and this may indicate increased androgen production in the zona reticularis of the adrenal cortex or an altered metabolism of androgens. CNS, central nervous system; Ach, acetylcholine; 5-HT, serotonin; NE, norepinephrine; ACTH, adrenocorticotropic.

on mood, memory, and learning. These effects are exerted by glucocorticoids as well as by other steroid hormones such as androgens and estrogens.

### Experimental Studies

Glucocorticoids have potent effects on the CNS, affecting cellular metabolism, receptor and ion channel density, neurotransmission, cell division, maturation, and cell survival. It was originally suggested that prolonged supraphysiological glucocorticoid levels with increasing age and/or stress down-regulated central GR expression. Reduced GRs in key brain areas, notably the hippocampus, might decrease the capacity for feedback, leading to further glucocorticoid hypersecretion. This would then lead to a vicious cycle in which hippocampal neuronal cell death might ensue, leading to permanently attenuated feedback and glucocorticoid hypersecretion. There are several caveats to this hypothesis. High glucocorticoid levels have only a moderate influence on GR expression. Instead, neurotransmitter influx, notably by serotonin and noradrenaline, is of major importance. Furthermore, it is not clear whether excessive glucocorticoid exposure *in vivo* per se might lead to neuron death. It seems more likely that glucocorticoids may lead to neuroendangerment. Thus, glucocorticoids make neurons less likely to survive a brain insult. The aging brain, especially the hippocampus, appears to be more susceptible than other parts of the brain to adverse effects of glucocorticoids. In combination with other insults associated with advanced age such as brain ischemia, various pathological processes might be accelerated by hypercortisolism. On a cellular level, glucocorticoids have a clear impact on neurons. Thus, increased levels of glucocorticoids seem to reduce neurogenesis within the hippocampus in close connection to excitatory amino acid efflux. Second, reversible stress-induced modeling of dendrites in hippocampal neurons is mediated by glucocorticoids along with excitatory amino acids. Finally, excitability of hippocampal neurons is influenced to a major extent. Thus, chronic excess of glucocorticoids might cause or accelerate some aspects of degenerative brain aging, notably hippocampus-associated cognitive dysfunction in rodents and humans.

### Glucocorticoids and Cognition

In accordance with experimental data, several studies suggest an association between hypercortisolism, on the one hand, and hippocampal function and volume,

on the other. Healthy adults treated with low/medium doses of synthetic glucocorticoids develop a reversible decrease in both immediate and delayed recall. In patients with Cushing's syndrome with very high circulating cortisol levels, impaired declarative memory performance and mood changes are common. This is associated with reduced hippocampal volume. Notably, this volume loss seems to be a reversible phenomenon. In long-standing major depression, often associated with a moderate increase in glucocorticoid production, hippocampal volume reduction seems to be common. Interestingly, increasing serum cortisol levels in healthy elderly persons over years has been associated with decreasing hippocampal volume, with concomitant deficits in hippocampus-dependent memory tasks. In individuals with mild cognitive impairment (MCI), often preceding a later development of AD, salivary cortisol levels are inversely correlated with the result in an immediate recall task. From these data, it is not clear whether hypercortisolism is related to reversible hippocampal dysfunction or whether any association exists between hypercortisolism and future neurodegenerative disease.

### Cortisol and Alzheimer's Disease

Disturbances on several different levels of the LHPA axis are present in AD. Some studies have been performed on patients in advanced stages of the disease, but it seems reasonable to assume that it is important to consider studies on mild to moderate AD, where these disturbances are early events in the disease rather than a late response to extensive CNS damage or medications.

There are several indicators of an increased central drive of the LHPA axis in AD, contributing to hypercortisolism. Postmortem studies have demonstrated increased corticotropin-releasing hormone (CRH) mRNA levels in the paraventricular nucleus of the hypothalamus. This putatively increased CRH secretion is associated with an increased cortisol but blunted adrenocorticotropin (ACTH) responsiveness to exogenous CRH in AD patients. An insensitivity to glucocorticoid feedback in AD probably also contributes to this increased activity through an inability to "shut off" temporary increases in LHPA axis activity. In line with this, increased glucocorticoid production has been reported in women with mild to moderate AD. These results show that increased glucocorticoid production is an early event in AD rather than a late response to chronic illness and/or medications. In addition, an altered cortisol metabolism is

present in mild to moderate AD, with increased excretion of A-ring-reduced metabolites in urine. This indicates that LHPA activation can be secondary to changes in cortisol clearance in AD. Alternatively, altered cortisol metabolism may be a protective mechanism that, together with down-regulation of GR expression in the periphery, can explain the absence of clinical features of hypercortisolism in peripheral tissues in AD (i.e., no increased prevalence of hypertension, diabetes mellitus, abdominal obesity, etc., as seen in Cushing's syndrome). Importantly, because of tissue-specific differences in cortisol metabolism and receptor expression, the brain may be excessively exposed to glucocorticoids. Thus, increased cortisol levels in cerebrospinal fluid (CSF) have been found in AD. Adverse effects on the brain through increased glucocorticoid exposure may be worsened via an inability of GRs in the brain to down-regulate their expression when exposed to increased glucocorticoid levels.

Thus, hypercortisolism may contribute to neuropsychiatric symptoms and accelerate neuronal damage in AD via tissue-specific alterations in pre-receptor glucocorticoid metabolism and receptor sensitivity/reactivity.

### ANDROGENS

Adrenal androgens include dehydroepiandrosterone (DHEA), its sulfate (DHEAS), and androstenedione. DHEA enhances neuronal and glial survival and enhances memory retention in rodents, and placebo-controlled DHEA supplementation studies in humans have suggested that administration of DHEA may improve physical and physiological well-being.

During the more advanced phases of dementia, levels of DHEAS are clearly decreased. The latter finding has been suggested to be of importance for progression of neurodegeneration, partly through loss of a glucocorticoid-antagonistic effect by DHEA, but this is a matter of controversy. In contrast, we have found increased levels of androgens in mild to moderate AD. This indicates an increased androgen production in the zona reticularis of the adrenal cortex or an altered metabolism of androgens. Furthermore, dynamic studies show an enhanced response of androgens, including DHEA, in AD after ACTH stimulation. In one longitudinal study on AD, lower levels of DHEA were associated with superior cognitive performance.

Importantly, DHEA has no known receptor; instead, local brain metabolism of this hormone

may govern its biological activity. A cytochrome p450 enzyme, CYP7B, converts DHEA to 7 $\alpha$ -hydroxy-DHEA and may be responsible for putative antigluco-corticoid effects of DHEA. CYP7B mRNA is highly expressed in the human hippocampus, and this expression is significantly decreased in hippocampal neurons in patients with AD. Thus, alterations in androgen metabolism in AD might be important for tissue effects of adrenal androgens, with key interactions with glucocorticoids on a cellular level.

## ESTROGEN AND OTHER FEMALE GONADAL HORMONES

Estrogens have numerous effects on the brain, including influences on development and adult brain plasticity. Beneficial effects on neuronal plasticity and blockade of neurotoxic effects may prevent or retard the development of neurodegenerative disease, including AD.

In favor of this hypothesis, estrogens regulate synaptogenesis in the rat hippocampus, notably in the CA1 subregion, which seems crucial for learning and memory functions. Synaptic spine density is related to circulating estradiol levels, and this is linked to memory function. These effects are associated with an increase in N-methyl-D-aspartate (NMDA) receptors in hippocampal neurons that relates to increased efficacy in long-term potentiation, that is, a proposed neurophysiological correlate to memory. Estrogens may also work in collaboration with neurotrophins, mainly nerve growth factor (NGF), to stimulate neuronal development, differentiation, and growth via colocalized receptors on neurons in the rodent fore-brain hippocampus and cerebral cortex. Estrogens also seem to promote neurogenesis.

In addition, estrogens influence the function in several neurotransmitter systems, including the cholinergic, serotonergic, dopaminergic, and noradrenergic systems. Estrogen-induced enhancement of cholinergic functions via increased activity of choline acetyl transferase (ChAT) in the basal fore-brain, hippocampus, and frontal cortex may be of particular relevance for AD. Related to this, the growth hormone response to pyridostigmine is increased in postmenopausal women taking estrogen replacement therapy, indicating an increased central cholinergic tone.

As evident from studies of, for example, serotonin receptor expression after manipulation of estradiol and progesterone levels, it is important to notice that estradiol may have very different effects when given

alone versus given in combination with progesterone regarding site and type of effects. This is also clear for effects on the noradrenergic system; estradiol given alone inhibits noradrenaline uptake, whereas estradiol followed by progesterone increases reuptake of this neurotransmitter. This influences the interpretation of hormone replacement studies in postmenopausal women that mainly has been done with estrogen in combination with a progestin.

*In vitro*, estrogens have neuroprotective effects, including inhibition of  $\beta$ -amyloid formation from its precursor protein. This may be due partly to reduced accumulation of reactive oxygen and nitrogen species. Importantly, estrogens seem to inhibit the production of proinflammatory cytokines, notably interleukin-6 (IL-6). Glucocorticoid-induced hippocampal neuronal damage also seems to be reduced. *In vivo* estrogens reduce damage from ischemic stroke insults, and this may be relevant for protection against development and/or progression of AD. Thus, cerebrovascular disease may affect the clinical expression of AD. In contrast, high levels of estradiol might be neurotoxic following injury, and this can be relevant for early stages of AD.

Neuroimaging studies in humans have implicated that estrogen influences the pattern of brain activation during memory processing, with regional increments in cerebral blood flow and glucose metabolism and with modulation of activity in specific brain regions affected during the early stages of AD. This may be partly related to direct effects on cerebral blood flow by estrogens.

Epidemiological studies have suggested that estrogen is protective against the development and/or progression of neurodegenerative disorders. Thus, low circulating levels of estrogen have been associated with an increased risk of AD. Some, but not all, case control studies, as well as a couple of prospective studies, have suggested that estrogen treatment may be protective against the development of AD. Importantly, there may be a protective effect of estrogen in long-term users of estrogen and/or when treatment is given during the latent preclinical stage of AD that may extend a decade or more before the onset of diagnosable dementia. Obviously, prospective randomized double-blind studies are needed for verification of these observations. Notably, circulating levels of estradiol have been reported to be slightly increased (and so not decreased) with concomitant elevated adrenal androgen hormone levels (DHEA and androstenedione) in patients with mild to moderate AD. This may reflect increased secretion and/or alterations in metabolism of these hormones during

the early phases of AD. These alterations of endogenous gonadal hormone levels during the early stages of AD imply that there may be no need for increasing estradiol levels further by supplementation in mild to moderate dementia. Indeed, increased estradiol levels may even be neurotoxic, and it might be worth exploring beneficial effects on reducing the production of adrenal androgens during the early phases of neurodegeneration/AD, thereby also reducing an excessive increase in estrogens.

The finding of a novel estrogen receptor, the estrogen  $\beta$ -receptor (ER $\beta$ ), has intensified the possibility of finding drugs that may act as neuroprotectants specifically through this receptor. The ER $\beta$  has a clear role in the development of the cerebral cortex and also in survival of hippocampal neurons after exposure to excitatory neurotoxins. In contrast, ER $\alpha$  is the major receptor subtype expressed in basal forebrain cholinergic neurons. Thus, estrogen probably acts via ER $\alpha$  to enhance cognitive functions through the production of acetylcholine. On the other hand, ER $\beta$  is the only estrogen receptor expressed in the dorsal raphe nucleus suggesting important effects on the serotonin system, indirectly affecting neuronal plasticity. In ER $\beta$  knockout mice, spatial learning is impaired and treatment with cytotoxins causes marked apoptosis in hippocampus at doses that do not affect wild-type litter mates. Eight exon ER $\beta$  knockout mice also show an increase in astroglia numbers, with a concomitant decrease in neuron number. This cell loss affects the limbic system, as well as the substantia nigra, to a major extent. Complex interactions among the subtypes of estrogen receptors is important to elucidate further, and it is not surprising that administration of estrogen compounds may induce very complex responses. Development of selective ligands for ER $\alpha$  and ER $\beta$  may have profound effects on cognitive function and neuronal survival, notably in relationship to  $\beta$ -amyloid-induced neurotoxicity.

## THE ROLE OF METABOLIC DYSFUNCTION

A number of recent studies have suggested an association between AD and risk markers for cardiovascular diseases. These include hypertension, diabetes mellitus, lipid abnormalities, and the presence of the apolipoprotein E (apoE)  $\epsilon$ 4 allele. Related to this, antihypertensive treatment and treatment with statins may reduce the incidence of dementia. There are several explanations for these associations, including

overlapping pathophysiology and clinical features, notably metabolic dysfunction. In the "metabolic syndrome," associated with increased risk of type 2 diabetes and cardiovascular disease, hyperinsulinemia is a key element.

Insulin receptors are widely distributed in the brain, mainly in the cerebral cortex and hippocampus. Insulin receptors are localized at the synapse, where they regulate neurotransmitter release and receptor recruitment. This indicates a role for insulin in synaptic plasticity. Disruption of cerebral insulin receptor functions leads to progressive cognitive impairments in rodents, and high insulin levels may directly influence the development of neuropathological changes in AD. Thus, insulin may be involved in the formation of neurofibrillary tangles via its regulatory activity on tau phosphorylation. Insulin also seems to affect amyloid metabolism, inhibiting  $\beta$ -amyloid degradation.

Desensitization of the neural insulin receptor reduces transport of glucose, the major nutrient for brain cells, and so might be a crucial link between metabolic dysfunction/hyperinsulinemia and cognitive dysfunction in AD. This could be worsened through secretion of  $\beta$ -amyloid because this protein may influence insulin binding and action through competitive binding to the insulin receptor. Furthermore, glucocorticoid overexposure to the brain, due to either a primary increase in glucocorticoid secretion or local changes in prereceptor metabolism, can decrease insulin sensitivity.

Epidemiological studies indeed suggest a link among insulin resistance, diabetes mellitus, and AD. Ethnic factors may be of importance given that longitudinal studies of Caucasians and Japanese Americans have generated conflicting results, with clear associations between these factors found in a population-based study from Rotterdam, Netherlands. Interestingly, antidiabetic drugs may influence AD pathology. Troglitazone, a thiazolidindione antidiabetic agent acting as an "insulin sensitizer" via activation of peroxisome proliferator-activated receptor- $\gamma$  (PPAR- $\gamma$ ) receptors, antagonizes an amyloid-stimulated proinflammatory response and neurotoxicity. This indicates a link among an inflammatory component of the metabolic syndrome, development of atherosclerosis, and pathology in AD.

Proinflammatory cytokines produced by activated microglia, especially IL-1, seem to trigger enhanced synthesis of amyloid precursor protein and production of  $\beta$ -amyloid. Amyloid deposits per se stimulate further cytokine production by activated microglia, leading to a vicious cycle with continuous production

of amyloid precursor protein and  $\beta$ -amyloid. The link between inflammatory mechanisms and AD pathology has been strengthened by the association between head trauma and later development of AD. Epidemiological studies also suggest that nonsteroidal anti-inflammatory drugs (NSAIDs) can retard the development of AD. Notably, the positive epidemiological findings with NSAIDs are reported for drugs that activate PPAR- $\gamma$  receptors, lowering  $\beta$ -amyloid production (A $\beta$ 42). This strengthens a possible link between insulin resistance and the development of  $\beta$ -amyloid-associated neurodegeneration.

Another potential link between insulin resistance and a later development of cardiovascular disease might be the adipocyte-derived hormone leptin. Leptin is an important regulator of satiety and energy expenditure. Based on the fact that overweight is associated with high circulating levels of leptin with associated "leptin resistance," a basic physiological function for leptin as protective against neuroendocrine consequences of starvation has been proposed. Increased leptin, strongly related to increased fat depot size, has been reported to predict the later development of myocardial infarction and stroke, independent of other risk factors. In patients with early AD, a physiological link between circulating cortisol and leptin levels over 24h seems to be lost. This suggests that leptin regulation is disturbed during the early phase of neurodegeneration, and this may be linked to abnormalities in cardiovascular risk factors as well as weight loss in patients with AD.

In summary, epidemiological and experimental data implicate that metabolic dysfunction precedes and may influence the development, progression, and symptomatology of AD.

## GROWTH HORMONE

The activity of the growth hormone (GH)–insulin-like growth factor-1 (IGF-1) axis declines during aging. In humans, GH secretion takes place in a pulsatile manner, regulated by stimulatory effects of growth hormone-releasing hormone (GHRH) and inhibitory input from somatostatin. Reduced central somatostatin levels have been found in autopsy studies in AD patients. In the periphery, circulating basal GH levels have been reported to be elevated in a few studies, but the diurnal pattern of GH secretion in AD seems to be unaltered. There also seems to be a high variability in reported responsiveness to a GHRH challenge. Related to these findings, it is not clear whether peripheral IGF-1 levels are altered in AD and, if so,

whether this is a trait or a stage-dependent change in GH–IGF-1 axis function.

## THYROID FUNCTION

In the Rotterdam study of aged individuals, the relative risk of AD at follow-up was increased more than three-fold for participants with reduced thyroid-stimulating hormone (TSH) concentrations at baseline. In contrast, no association was found between an increased TSH level and incident AD. Thus, the findings in the Rotterdam study are the first to suggest that subclinical hyperthyroidism among the elderly may increase the risk of AD.

## CATECHOLAMINES

Noradrenergic axons arising from the locus ceruleus (LC) project to several cortical areas, including the prefrontal and entorhinal cortexes. In the brain, the levels of norepinephrine (NE) are highest in the hypothalamus, and NE plays an important role in attention, arousal, and stress reactions as well as in cognition. The level of NE in plasma is widely used as a marker of the activity of the sympathetic nervous system (SNS), and a decline in NE levels is reported in aging individuals. In AD, studies assessing the SNS are relatively few, with small numbers of patients and varying severities of the disease. Decreases in cortical NE levels are described in various degenerative diseases of the brain, including AD, and a marked LC neuronal loss is considered as a classic postmortem pathological hallmark of AD. Despite these findings, increased concentrations of NE in CSF of AD patients have been described, possibly due to a compensatory activation of remaining LC neurons in this disorder or to increased turnover. Taking into consideration the great variation between studies concerning both severity of the disease and number of patients included, the overall result of the studies on SNS in AD points to a noradrenergic dysfunction that may contribute to cognitive impairment and behavioral symptoms.

## See Also the Following Articles

Acetylation • ACTH (Adrenocorticotropic Hormone) • Aging and Longevity of Human Populations • Aging, Immunology and • Brain, Effects of Steroid Hormones • Catecholamines • DHEA and the Elderly • Functional Genomics of Aging • Growth Hormone (GH) • Insulin-like Growth Factors • Leptin • Neuroendocrine System

and Aging • Somatostatin Analogs • Stress, Aging, and Central Nervous System Interactions

## Further Reading

- Alderson, A. L., and Novack, T. A. (2002). Neurophysiological and clinical aspects of glucocorticoids and memory: A review. *J. Clin. Exp. Neuropsychol.* **24**, 335–355.
- Ferrari, E., Cravello, R., Muzzoni, B., Casarotti, D., Paltro, M., Solerte, S. B., Fioravanti, M., Cuzzoni, G., Pontiggia, B., and Magri, F. (2001). Age-related changes of the hypothalamic-pituitary-adrenal axis: Pathophysiological correlates. *Eur. J. Endocrinol.* **144**, 319–329.
- Gasparin, I. L., Netzer, W. J., Greengard, P., and Xu, H. (2002). Does insulin dysfunction play a role in Alzheimer's disease? *Trends Pharmacol. Sci.* **23**, 288–293.
- Kalmijn, S., Mehta, K. M., Pols, H. A., Hofman, A., Drexhage, H. A., and Breteler, M. M. (2000). Subclinical hyperthyroidism and the risk of dementia: The Rotterdam study. *Clin. Endocrinol. (Oxf.)* **53**, 733–737.
- Lupien, S. J., deLeon, M., deSanti, S., Convit, A., Tarshish, C., Nair, N. P. V., Thakur, M., McEwen, B. S., Hauger, R. L., and Meaney, M. J. (1998). Cortisol levels during human aging predict hippocampal atrophy and memory deficits. *Nature Neurosci.* **1**, 69–73.
- McEwen, B. S., and Sapolsky, R. M. (1995). Stress and cognitive function. *Curr. Opin. Neurobiol.* **5**, 205–216.
- Rasmuson, S., Andrew, R., Näsman, B., Seckl, J. R., Walker, B. R., and Olsson, T. (2001). Increased glucocorticoid production and altered cortisol metabolism in women with mild to moderate Alzheimer's disease. *Biol. Psychiatry* **49**, 547–552.
- Rasmuson, S., Nasman, B., Eriksson, S., Carlstrom, K., and Olsson, T. (1998). Adrenal responsivity in normal aging and mild to moderate Alzheimer's disease. *Biol. Psychiatry* **43**, 401–407.
- Seckl, J. R., French, K. L., O'Donnell, D., Meaney, M. J., Nair, N. P., Yates, C. M., and Fink, G. (1993). Glucocorticoid receptor gene expression is unaltered in hippocampal neurons in Alzheimer's disease. *Brain Res. Mol. Brain Res.* **18**, 239–245.
- Seckl, J. R., and Olsson, T. (1995). Glucocorticoid hypersecretion and the age-impaired hippocampus: Cause or effect? *J. Endocrinol.* **145**, 201–211.
- Wooley, C. S. (1998). Estrogen-mediated structural and functional synaptic plasticity in the female rat hippocampus. *Horm. Behav.* **34**, 140–148.



# Amidation

Joseph Bell, Betty A. Eipper, and Richard E. Mains  
*University of Connecticut Health Center, Farmington, Connecticut, United States*

## Glossary

**$\alpha$ -amidated peptide** A peptide with a carboxy-terminal amide group (-CONH<sub>2</sub>) in place of a free -COOH group.

**$\alpha$ -amidation** The enzymatic reaction that results in the presence of a carboxy-terminal  $\alpha$ -amide group.

**bioactive peptide** Any peptide with a biological activity. They typically bind to receptors at nanomolar or picomolar concentrations and show very high amino acid sequence specificity. Most bioactive peptides work through G protein-coupled metabotropic receptors with seven transmembrane domains.

**large dense core vesicle (LDCV)** Also referred to as secretory granule. A membrane-delimited organelle formed from the trans-Golgi network that stores concentrated peptides and proteins for regulated secretion (secretion-on-demand). Standard electron microscopy preparations show an electron-dense core surrounded by a membrane.

Amidation is the process by which the -CONH<sub>2</sub> structure is created. The side chains of asparagine and glutamine contain this moiety, but this article discusses the process by which the carboxy terminus of a peptide acquires an  $\alpha$ -amide.

## INTRODUCTION

Endocrine cells and neurons communicate to other cells in the body (neurons, endocrine cells, muscles, glands, etc.) in part by secreting a variety of small molecules (e.g., glutamate, glycine, norepinephrine, and ATP), steroids, and peptides. Peptides are by far the most numerous class of secreted communication molecules. Bioactive peptides are normally synthesized as part of a larger inactive protein that is cotranslationally inserted into the lumen of the endoplasmic reticulum. The active peptide regions are then liberated by a small group of highly selective endo- and exoproteases. The majority of these cleavages and

other covalent modifications occur in specialized storage organelles called large dense core vesicles (LDCV). More than half of the known biologically active peptides have an  $\alpha$ -amide group instead of a free carboxyl group on their COOH terminus (Fig. 1). Amidated fatty acids (e.g., oleamide) and other molecules have also been reported, and the formation of these  $\alpha$ -amide groups may follow the same process established for  $\alpha$ -amidated peptides.

For peptides, the immediate precursor of the  $\alpha$ -amidated peptide is the corresponding peptide with an additional glycine residue at its COOH terminus. This peptidyl-glycine intermediate is formed from the initial larger precursor by the combined actions of endoproteases and carboxypeptidases; occasionally, the Gly residue is the final residue in the larger precursor. The  $\alpha$ -amidation reaction begins with peptidyl-glycine and finishes with peptidyl-NH<sub>2</sub>, the bioactive peptide. A single amidating enzyme can produce peptides terminating with all 20 amino acid amides.

## WHY IS PEPTIDE AMIDATION IMPORTANT?

For most amidated peptides, biological potency decreases two or three orders of magnitude when the  $\alpha$ -amide group is absent, with either the intact glycine residue remaining or the free acid exposed after the  $\alpha$ -amide group is removed (Fig. 2). The  $\alpha$ -amide is important because its presence means that the COOH terminus of the peptide does not change charge as a function of physiological changes in pH. Thus, the uncharged COOH terminus of the peptide can associate closely with its membrane receptor, binding to a site on the receptor at or within the transmembrane domains. This leaves the NH<sub>2</sub> terminus of the peptide exposed to the extracellular aqueous solution, enabling it to bind to additional portions of the peptide receptor and thus giving the peptide receptor pair more specificity and tighter binding than usually seen with smaller ligands, such as acetylcholine or glutamate.

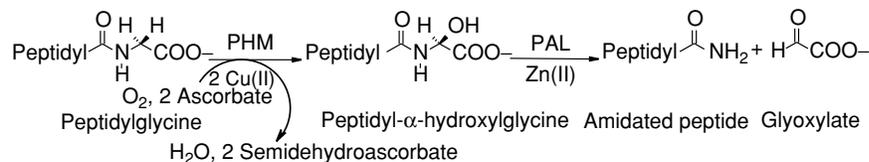


Figure 1 The two-step  $\alpha$ -amidation reaction.

## THE $\alpha$ -AMIDATION REACTION

The peptide  $\alpha$ -amidation reaction is a two-step process (Fig. 1). The first step is performed by peptidylglycine- $\alpha$ -hydroxylating monooxygenase (PHM; E.C. 1.14.17.3; also called peptidylglycine- $\alpha$ -monooxygenase). The second step is performed by peptidyl- $\alpha$ -hydroxyglycine- $\alpha$ -amidating lyase (PAL; E.C. 4.3.2.5; also known as peptidylamidoglycolate lyase).

In the  $\alpha$ -hydroxylation reaction, ascorbate reduces the two copper ions that are bound to PHM, yielding two semidehydroascorbate molecules that leave the enzyme. Cytosolic reducing equivalents, delivered by cytochrome  $b_{561}$ , an integral membrane protein, are used to return the semidehydroascorbate to its fully reduced state. Only after the peptidyl-Gly substrate binds to PHM is molecular oxygen split, yielding peptidyl- $\alpha$ -hydroxy-Gly and water—soluble products that leave the enzyme. The two atoms of molecular oxygen are incorporated into the hydroxylated peptide and water. The  $\alpha$ -hydroxylation reaction is closely analogous to the  $\beta$ -hydroxylation reaction performed by dopamine  $\beta$ -monooxygenase, which converts dopamine to norepinephrine and tyramine to octopamine. The most distinguishing features of this

reaction are its absolute dependence on copper and its use of ascorbate and molecular oxygen.

Although the second, or lyase, reaction can occur spontaneously as the pH is increased above 7.5, in cells it must be catalyzed by a separate enzyme. In LDCVs, with an internal pH of 5.0–5.5, the peptidyl- $\alpha$ -hydroxy-Gly product created by PHM is quite stable unless the lyase reaction is catalyzed by an enzyme. The lyase reaction, the stereospecificity of which matches that of PHM, requires enzyme-bound divalent metal, and bound zinc has been detected. It is not clear whether the zinc is catalytic or structural or both. A structural zinc could aid in the folding and stabilization of PAL, just as  $\text{Ca}^{2+}$  is thought to stabilize the prohormone convertases involved in propeptide endoproteolysis. A catalytic zinc could play a role like the one it plays in alcohol dehydrogenase, just as  $\text{Zn}^{2+}$  plays a crucial catalytic role in the carboxypeptidases.

## THE PEPTIDE AMIDATING ENZYMES

The two enzymatic activities necessary to perform the peptide  $\alpha$ -amidation reaction are initially synthesized as a single bifunctional precursor protein called peptidylglycine- $\alpha$ -amidating monooxygenase (PAM) (Fig. 3). Mammals have a single gene encoding PAM. Alternative splicing generates several forms of PAM that differ in important ways. Elimination of the exon encoding the transmembrane domain yields a soluble, secreted enzyme. Elimination of the flexible linker region between PHM and PAL (exon 16) greatly reduces the ability of cells to separate the two catalytic activities. In LDCVs, the same prohormone convertases that cleave propeptide precursors also cleave the PAM precursor, producing soluble PHM, soluble and membrane-anchored PAL, and soluble bifunctional PAM. In invertebrates such as *Drosophila* and *Cnidarians*, PHM and PAL are encoded by separate genes. Elimination of the *Drosophila* PHM gene is lethal, causing death in the very late embryo and young larvae stages. In snails, both catalytic functions are encoded in the same gene, but there are four copies of the PHM domain, each with a similar

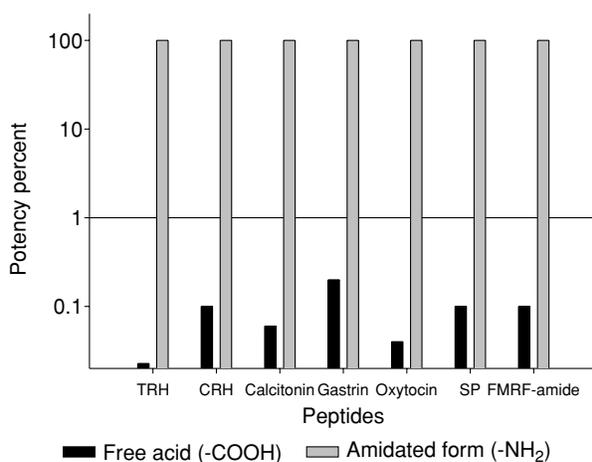
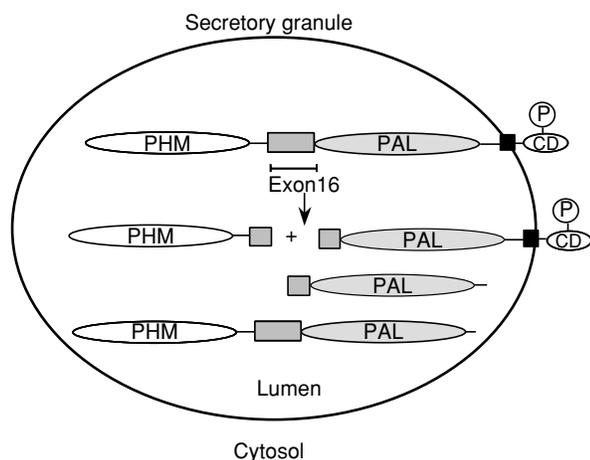


Figure 2 Relative biological potency of  $\alpha$ -amided peptides (each normalized to 100%) compared to the corresponding free acid form of each peptide.



**Figure 3** Forms of the amidating enzyme as found in mammalian LDCVs.

dependence on copper and ascorbate but a unique peptide substrate specificity.

The catalytic core of PHM was defined using controlled protease digestion, and its structure was explored by assigning disulfides, examining site-directed mutants, and employing spectroscopy and X-ray crystallography. The catalytic core of PHM consists of two  $\beta$ -clamshell or sandwich domains. Each approximately 150-amino acid domain contains a single copper binding site. The  $\text{NH}_2$ -terminal domain, with its three disulfide bonds, uses three His residues to bind Cu (the  $\text{Cu}_A$  or  $\text{Cu}_H$  site). The  $\text{COOH}$ -terminal domain, with its two disulfide bonds, uses two His residues and one Met residue to bind Cu ( $\text{Cu}_B$  or  $\text{Cu}_M$ ). The two domains are held together by a single hydrophilic linker strand, whereas the interiors of the domains are very hydrophobic. All of the histidine and methionine residues involved in coordinating the two catalytic copper ions are conserved in all known PHM sequences. Interestingly, all the spectral and crystallography data indicate that the two copper ions are farther apart than expected for a reaction requiring both copper ions to undergo a reduction-oxidation cycle. The X-ray structure places the two copper ions 11 Å apart and separated by a solvent-filled cleft; the Gly-extended peptide substrate binds closer to  $\text{Cu}_B$ . Dopamine  $\beta$ -monooxygenase shares many conserved disulfide bonds and contains histidine and methionine residues that may bind to copper in a similar manner. Based on sequence similarity, two additional potential family members, monooxygenase X and dopamine- $\beta$ -hydroxylase-L, have been identified; their substrate specificity has not been determined.

PAL has not been studied as extensively. However, of the amino acid residues that are conserved among species, site-directed mutagenesis has identified a subset that are candidates for involvement in maintaining the structure of PAL and additional residues that are candidates for involvement in the PAL catalytic mechanism. Unlike the copper ions in the PHM reaction, which cycle from  $\text{Cu}^{1+}$  to  $\text{Cu}^{2+}$  during each reaction cycle, transferring reducing equivalents to molecular oxygen, the catalytic zinc presumably interacts directly with the  $\alpha$ -hydroxylated substrate. PAL remains unique, with no close homologues identified in database screens. The yeast genome lacks enzymes homologous to either PHM or PAL.

## FUTURE STUDIES ON PEPTIDE AMIDATION

Future work should focus on the peculiar copper/oxygen chemistry employed by PHM and its relatives and on the unique structure of PAL and the divalent metals it binds. Cell biological studies should focus on how membrane PAM enters and leaves LDCVs and how it traverses the endocytic pathway. The role of copper transporters (e.g., ATP7A and the Menkes protein), cytochrome  $b_{561}$ , and ascorbate transporters in providing the right vesicular milieu must be elucidated. The biological consequences of disrupting amidation by interfering with copper metabolism (mottled/brindled mice) or the PAM protein (PAM knockout mice) must be explored.

## Further Reading

- Bell, J., Ash, D. E., Snyder, L. M., Kulathila, R., Blackburn, N. J., and Merkler, D. J. (1997). Structural and functional investigations on the role of zinc in bifunctional rat peptidylglycine  $\alpha$ -amidating enzyme. *Biochemistry* **36**, 16239–16246.
- Eipper, B. A., Stoffers, D. A., and Mains, R. E. (1992). The biosynthesis of neuropeptides: Peptide  $\alpha$ -amidation. *Annu. Rev. Neurosci.* **15**, 57–85.
- Eipper, B. A., Milgram, S. L., Husten, E. J., Yun, H.-Y., and Mains, R. E. (1993). Peptidylglycine  $\alpha$ -amidating monooxygenase: A multifunctional protein with catalytic, processing and routing domains. *Protein Sci.* **2**, 489–497.
- Feng, J., and May, S. W. (2001). High-performance liquid chromatographic enantiomeric separation of an enzyme inhibitor which possesses both a chiral center and tautomeric moieties. *J. Chromatogr. A* **905**, 103–109.
- Francisco, W. A., Merkler, D. J., Blackburn, N. J., and Klinman, J. P. (1998). Kinetic mechanism and intrinsic isotope effects for the peptidylglycine  $\alpha$ -amidating enzyme reaction. *Biochemistry* **37**, 8244–8252.
- Kolhekar, A. S., Mains, R. E., and Eipper, B. A. (1996). Peptidylglycine  $\alpha$ -amidating monooxygenase (PAM): An ascorbate requiring enzyme. *Methods Enzymol.* **279**, 35–43.

- Kulathila, R., Merkle, K. A., and Merkle, D. J. (1999). Enzymatic formation of C-terminal amides. *Nat. Prod. Rep.* **16**, 145–154.
- Mains, R. E., and Eipper, B. A. (1999). Peptides. In “Basic Neurochemistry” (G. R. Siegel, *et al.*, eds.), pp. 363–382. Lippincott-Raven, Philadelphia.
- Merkle, D. J. (1994). C-terminal amidated peptides: Production by the in vitro enzymatic amidation of glycine-extended peptides and the importance of the amide to bioactivity. *Enzyme Microb. Technol.* **16**, 450–456.
- Mounier, C. E., Shi, J., Sirimanne, S. R., Chen, B. H., Moore, A. B., Gill-Woznichak, M. M., Ping, D., and May, S. W. (1997). Pyruvate-extended amino acid derivatives as highly potent inhibitors of carboxyl-terminal peptide amidation. *J. Biol. Chem.* **272**, 5016–5023.
- Prigge, S. T., Mains, R. E., Eipper, B. A., and Amzel, L. M. (2000). New insights into copper monooxygenases and peptide amidation: Structure, mechanism and function. *Cell Mol. Life Sci.* **57**, 1236–1259.



# Amiodarone and Thyroid

Wilmar M. Wiersinga

Academic Medical Center and University of Amsterdam, Amsterdam, The Netherlands

## Glossary

**action potential** Changes in the plasma membrane during depolarization and the subsequent repolarization of cardiomyocytes. The duration of the action potential is reflected by the QT time of the electrocardiogram (interval between the start of the QRS-complex and the end of T-top). Class III anti-arrhythmic drugs (such as amiodarone) act via prolongation of the action potential.

**colloid** Proteinaceous substance in thyroid follicles, containing sizable quantities of iodine, thyroglobulin, and thyroid hormone.

**deiodination** Enzymatic removal of iodine atoms from organic compounds.

**euthyroidism** Normal thyroid function.

**IC<sub>50</sub> value** Concentration of a compound at which it inhibits a particular phenomenon by 50%.

**lysosome** A membranous bag of hydrolytic enzymes; this organelle is used for the intracellular digestion of macromolecules.

**organification** The incorporation of iodine atoms in organic compounds.

**thionamides** Class of anti-thyroid drugs (such as carbimazole, methimazole, and propylthiouracil) that inhibit thyroid hormone synthesis by interfering with organification.

**thyroid hormone receptor** Protein that, after binding of triiodothyronine (T<sub>3</sub>), binds to specific DNA sequences in the promoter region of T<sub>3</sub>-dependent genes, thereby modulating the transcription of these genes.

**Wolff-Chaikoff effect** Decreasing yield of organic iodine from increasing doses of inorganic iodide.

Amiodarone is a very potent anti-arrhythmic drug, which is successfully used in the treatment of atrial fibrillation and life-threatening ventricular arrhythmias. The drug, however, has many side effects. Amiodarone influences thyroid hormone secretion and metabolism in all patients taking the drug. In a

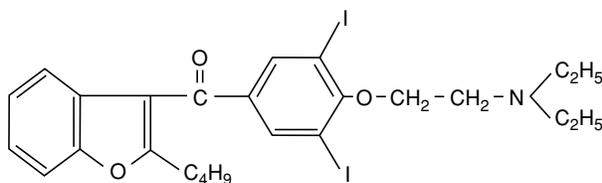
subset of patients, this results in amiodarone-induced hypothyroidism or thyrotoxicosis. Amiodarone also acts as a thyroid hormone receptor antagonist.

## PHARMACOLOGY

The structural features of amiodarone are its high iodine content and its resemblance to thyroxine (Fig. 1). The drug is prescribed as amiodarone hydrochloride (MW 681.82), which contains 37.25% iodine by weight. It was originally introduced in 1962 in clinical medicine for the treatment of angina pectoris, but later was found to be very efficacious in the treatment of cardiac arrhythmias. Amiodarone is classified as a class III anti-arrhythmic agent; it lengthens the duration of the action potential and repolarization time in cardiac tissues. It also has weak anti-adrenergic effects, it causes smooth muscle relaxation resulting in a dilation of coronary arteries and an increase in coronary blood flow, and it induces peripheral arterial vasodilation and a decrease in systemic blood pressure and afterload.

Dosage forms of amiodarone (trade name Cordarone) are tablets or injections. Because of the large distribution volume, the onset of the drug's action is delayed; consequently, a loading dose to saturate the large body stores is frequently required. The highest levels of amiodarone and its major metabolite desethylamiodarone (DEA) are found in adipose tissue, liver, and lung (Table I). The slow turnover of amiodarone from deep compartments such as adipose tissue explains its exceptionally long terminal half-life of  $\pm 10$  days (half-life for DEA is  $57 \pm 27$  days).

The major metabolic pathways of amiodarone are N-dealkylation, resulting in its main (and biologically active) metabolite DEA, and deiodination, resulting in monoiodo-, desdiiodo-, and desethyl-desdiiodo-amiodarone. A daily oral dose of 200 mg amiodarone results in a urinary iodide excretion of approximately 14,000  $\mu\text{g}/24$  liter, which is approximately 45 times higher than the optimal daily iodine intake of 150–300  $\mu\text{g}$  recommended by the World Health



**Figure 1** Molecular structure of amiodarone (chemical name: 2-butyl-3-[3,5-diiodo-4-( $\beta$ -diethylamino-ethoxy)-benzoyl] benzofuran).

**Organization.** Amiodarone medication thus causes chronic iodine excess. Extensive glucuro-conjugation of amiodarone occurs and biliary excretion and fecal elimination account for 65–75% of the ingested drug. Transplacental transfer of amiodarone and DEA varies from 10 to 20%. Plasma amiodarone and DEA concentrations in the newborn are approximately fourfold lower than those in the mother. Amiodarone and DEA concentrations in breast milk are higher than in the plasma of the mother, due to their high lipid solubility.

Side effects are numerous, occurring in ~80% of patients. Most prevalent are corneal microdeposits (almost 100%), gastrointestinal symptoms such as anorexia and nausea (80%), photosensitivity and unusual blue-gray skin discoloration of exposed areas (55–75%), and neurologic symptoms such as tremor, ataxia, and peripheral neuropathy (48%). Uncommon but severe adverse effects include pulmonary toxicity, liver failure, and proarrhythmias. Most side effects develop slowly and are related to the cumulative dose of amiodarone. The mechanism of amiodarone toxicity is multifactorial, but a drug-induced phospholipidosis with disturbed lysosomal function explains many side effects. Amiodarone as an amphiphilic drug binds strongly to intralysosomal phospholipids, rendering them indigestible by phospholipases. The bound complexes form the intralysosomal multilamellar inclusion bodies, which have been found in lung,

liver, heart, skin, corneal epithelium, and nerve fibers of amiodarone-treated patients.

## EFFECTS ON THYROID HORMONE SECRETION AND METABOLISM

Amiodarone treatment invariably results in changes in the plasma concentrations of thyrotropin [thyroid-stimulating hormone (TSH)] and thyroid hormones. There is an initial rise in plasma TSH starting in the first week of treatment, with a return to normal values after 3 months. This is due to a chronic iodine excess generated during biotransformation of the drug, which transiently inhibits the synthesis and release of thyroid hormones (the so-called Wolff-Chaikoff effect), explaining the rise in TSH. The thyroid usually escapes from these inhibitory effects and plasma TSH returns to normal values.

Amiodarone simultaneously affects extrathyroidal thyroid hormone metabolism; it strongly inhibits type I iodothyronine-5'-deiodinase, which catalyzes the deiodination of thyroxine (T4) into triiodothyronine (T3) and that of reverse triiodothyronine (rT3) into 3,3'-diiodothyronine. Consequently, plasma T3 decreases and plasma rT3 increases.

Plasma T4 and free thyroxine (FT4) concentrations also increase, predominantly due to a reduced metabolic clearance rate of T4 related to inhibition of T4 uptake in the liver. Inhibition of T4 entry into tissues decreases the availability of the substrate T4 for 5'-deiodination, thereby contributing to decreased production of T3 and decreased clearance of rT3. Chronic administration of amiodarone thus results in elevated plasma T4 and FT4 concentrations in the presence of a normal plasma TSH: a remarkable combination of test results.

## AMIODARONE-INDUCED THYROID DISEASES

The clinical diagnosis of amiodarone-induced hypothyroidism (AIH) and amiodarone-induced thyrotoxicosis (AIT) can be made very easily if the classical symptoms and signs of thyroid hormone deficiency or excess are present, but this is not always the case. Worsening of cardiac arrhythmia can be an important clue for the diagnosis of AIT. A TSH value within the normal reference range reliably excludes AIH and AIT. An elevated TSH with a low T4 or FT4 in plasma indicates AIH. When plasma TSH is suppressed and plasma T3 is elevated, the diagnosis of AIT is straightforward. However, a normal plasma T3

**Table 1** Tissue Distribution of Amiodarone (A) and Desethylamiodarone (DEA) in Human Autopsies

Tissue	A ( $\mu\text{g/g}$ )	DEA ( $\mu\text{g/g}$ )	A/DEA
Adipose tissue	316	76	4.2
Liver	391	2354	0.12
Lung	198	952	0.21
Kidneys	57	262	0.22
Heart	40	169	0.24
Muscle	22	51	0.43
Thyroid	14	64	0.22

**Table II** Incidence of Amiodarone-Induced Hypothyroidism (AIH) and Amiodarone-Induced Thyrotoxicosis (AIT) in Relation to Environmental Iodine Intake

Iodine intake	AIH	AIT	AIH + AIT	Country
High	13.2%	1.7%	14.9%	USA, UK
Intermediate	5.7%	7.9%	13.6%	Spain, Australia, The Netherlands
Low	6.4%	11.9%	18.4%	Italy, Belgium

does not exclude AIT (in view of the decreased T<sub>3</sub> production due to inhibited deiodination of T<sub>4</sub> into T<sub>3</sub>) and AIT may present as T<sub>4</sub> toxicosis.

AIH or AIT develops in 16% of amiodarone-treated patients (Table II). The incidence of AIH is higher in patients residing in areas with a high iodine intake and AIT occurs more often in regions with a low iodine intake. Amiodarone may also give rise to a small, firm goiter (induced by the iodine excess) in the presence of a normal level of TSH, but this occurs less frequently than AIH and AIT.

### Amiodarone-Induced Hypothyroidism

AIH occurs more often in females than in males and most cases are seen in the first 18 months of treatment. It is caused by a failure of the thyroid gland to escape from the Wolff-Chaikoff effect, resulting in permanent inhibition of organification. This is more likely to occur in subjects with preexisting autoimmune thyroiditis. Consequently, AIH is, to a certain extent, predictable and women with preexisting thyroid peroxidase antibodies are at risk of developing AIH. Thyroidal radioiodine uptake is preserved in AIH despite the increased stable iodide pool; it is explained by the drop in organification, which normally also produces compounds (presumably iodinated lipids) that inhibit iodide uptake.

Discontinuation of amiodarone usually restores euthyroidism in 3–4 months, but permanent hypothyroidism may ensue in patients with preexistent autoimmune thyroiditis. Potassium perchlorate (which acutely inhibits thyroidal iodine uptake via the sodium–iodide symporter) may shorten the time to reach euthyroidism. Thyroxine medication is effective and allows continuation of amiodarone. Fetal hypothyroidism occurs in 11% of patients treated with amiodarone during pregnancy and should be treated at once with thyroxine.

### Amiodarone-Induced Thyrotoxicosis

AIT occurs more often in males than in females and new cases continue to occur throughout the duration of treatment. Two types of AIT with a different pathogenesis have been distinguished (Table III). Type I occurs in patients with preexisting thyroid disease (Graves' disease, nodular goiter), obviously caused by increased thyroid hormone synthesis due to overrepletion of intrathyroidal iodine stores by the iodine excess. Type II is the result of destructive thyroiditis caused by the cytotoxic effects of amiodarone and DEA on thyrocytes by interference with lysosomes. Disruption of the normal thyroidal architecture allows the release of colloid contents (very rich in thyroid hormone) into the circulation, causing thyrotoxicosis. The stores of thyroid hormone in the colloid are finite, which explains the often self-limited nature of type II.

Treatment for AIT depends on its severity, which varies from mild to very severe, and on the cardiac condition, which may or may not allow discontinuation of amiodarone. In AIT type I, discontinuation of amiodarone is recommended, but patients are still thyrotoxic 6–9 months thereafter. Combination therapy using thionamides (which are less effective during iodine excess) with potassium perchlorate shortens

**Table III** Characteristics of Amiodarone-Induced Thyrotoxicosis Types I and II

	Type I	Type II
Underlying thyroid abnormality	Yes	No
Pathogenetic mechanism	Excessive thyroid hormone synthesis due to iodine excess	Excessive thyroid hormone release due to destructive thyroiditis
Goiter	Usually diffuse or nodular goiter	Occasionally small diffuse goiter
Thyroid radioiodine uptake	Low, normal, or high	Low
Serum interleukin-6	Normal or slightly elevated	Markedly elevated
Spontaneous remission	Less likely	More likely
Preferred drug treatment	Potassium perchlorate plus thionamides	Glucocorticoids plus thionamides
Subsequent hypothyroidism	Unlikely	Possible

Table IV Hypothyroid-like Effects of Amiodarone in Various Tissues

Tissue effect	Hypothyroidism	Amiodarone	Amiodarone + T3
Heart			
QT interval	↑	↑	N
Heart rate	↓	↓	N
β-adrenoceptor density	↓	↓	N
Ca <sup>2+</sup> ATPase activity of myosin	↓	↓	
Liver			
LDL receptor density	↓	↓	N
Triglyceride lipase activity	↓	↓	↓
Adipose tissue			
Lipoprotein lipase activity	(†)	↑	N

Note. ↑, increase; ↓, decrease; N, return to normal; (†), increase not significant.

the period until euthyroidism to less than 3 months. <sup>131</sup>I therapy is seldom feasible in view of the low radioiodine uptake. Total thyroidectomy has been performed successfully in resistant cases. In AIT type II, spontaneous recovery to euthyroidism is the rule within 3–5 months after stopping amiodarone. Faster improvement is usually obtained with prednisone given for 7–12 weeks in combination with thionamides. Although most authors still favor discontinuation of amiodarone in AIT type II, a favorable outcome under continuation of amiodarone is certainly possible. Despite all efforts, some patients do not respond to multidrug treatment with thionamides, potassium perchlorate, and steroids. There have been reports of fatal cases in which AIT patients have died of thyroid storm.

### AMIODARONE AS A THYROID HORMONE RECEPTOR ANTAGONIST

Amiodarone is prescribed for cardiac arrhythmias and angina pectoris, the rationale being induction of bradycardia, lengthening of the cardiac action potential, and depression of myocardial oxygen consumption. Essentially similar phenomena are observed in hypothyroidism. The hypothesis has thus been put forward that one of the main mechanisms of action of amiodarone is via induction of a local hypothyroid-like condition in extrathyroidal tissues, notably the heart (Table IV). The hypothesis is quite attractive, in view of the markedly decreased tissue concentrations of T3 induced by amiodarone. However, other drugs, such as iopanoic acid (equally potent in inhibiting type I 5'-deiodination), do not induce hypothyroid-like effects. In support of the hypothesis is the finding that DEA inhibits the binding of T3 to its nuclear receptors, resulting in a dose-dependent

decrease in the expression of several T3-dependent genes. Interestingly, DEA is a competitive inhibitor of T3 binding to thyroid hormone receptor  $\alpha_1$  (TR $\alpha_1$ ) (IC<sub>50</sub> value 47  $\mu$ M) but a noncompetitive inhibitor of T3 binding to TR $\beta_1$  (IC<sub>50</sub> value 27  $\mu$ M). The intracellular concentrations of DEA reached *in vivo* are high enough (50–500  $\mu$ M) for the drug to be able to interfere with T3 binding. Protein-protein binding studies with TR $\beta_1$  and the coactivator glucocorticoid receptor-interacting protein showed an inhibitory effect of DEA on the T3-dependent binding of the coactivator to TR $\beta_1$ . Further studies have indicated that residues on the outside of the TR ligand-binding domain are involved in the binding of DEA.

The available studies provide good evidence that DEA rather than amiodarone itself is responsible for the hypothyroid-like actions. The drug appears to meet the criteria of a thyroid hormone receptor antagonist.

### See Also the Following Articles

Hypothyroidism, Causes of • Hypothyroidism, Systemic Manifestations of • Hypothyroidism, Treatment of • Iodine • Thyroid Hormone Action • Thyroid Hormone Metabolism • Thyroid Hormone Receptors • Thyrotoxicosis, Overview of Causes • Thyrotoxicosis, Systemic Manifestations • Thyrotoxicosis, Treatment

### Further Reading

- Daniels, G. H. (2001). Amiodarone-induced thyrotoxicosis. *J. Clin. Endocrinol. Metab.* **86**, 3–8.
- Martino, E., Bartalena, L., Bogazzi, F., and Braverman, L. E. (2001). The effects of amiodarone in the thyroid. *Endocr. Rev.* **22**, 240–254.
- Newman, C. M., Price, A., Daview, D. W., Gray, T. A., and Weetman, A. P. (1998). Amiodarone and the thyroid: A practical

- guide to the management of thyroid dysfunction by amiodarone therapy. *Br. Med. J.* **79**, 121–127.
- Trip, M. D., Düren, D. R., and Wiersinga, W. M. (1994). Two cases of amiodarone-induced thyrotoxicosis successfully treated with a short course of antithyroid drugs while amiodarone was continued. *Br. Heart J.* **72**, 266–268.
- Trip, M. D., Wiersinga, W. M., and Plomp, T. A. (1991). Incidence, predictability and pathogenesis of amiodarone-induced thyrotoxicosis and hypothyroidism. *Am. J. Med.* **91**, 507–511.
- Van Beeren, H. C., Jong, W. M., Kaptein, E., Visser, T. J., Bakker, O., and Wiersinga, W. M. (2003). Dronedarone acts as a selective inhibitor of 3,5,3'-triiodothyronine binding to thyroid hormone receptor- $\alpha_1$ : *In vitro* and *in vivo* evidence. *Endocrinology* **144**, 552–558.
- Wiersinga, W. M. (1997). Pharmacotherapeutics of the thyroid gland. In "Handbook of Experimental Pharmacology" (A. P. Weetman and A. Grossman, eds.), Vol. 128, pp. 225–287. Springer-Verlag, Berlin/Heidelberg, Germany.



# Anderson's Disease (Chylomicron Retention Disease)

Nathalie Berriot-Varoqueaux and Marie-Elisabeth Samson-Bouma

Xavier Bichat, Paris, France

Lawrence P. Aggerbeck

University Pierre and Marie Curie and University of Paris-Sud, Paris, France

## Glossary

**apolipoproteins (apo)** Protein components of plasma lipoproteins, such as apoA1 and apoB. ApoB-48 corresponds to the amino-terminal 48% of apoB-100, which is produced in the intestine following RNA editing.

**chylomicrons** Triglyceride-rich lipoproteins secreted by the intestine after a meal.

**hypocholesterolemia (hypocholesterolemic disorders)** Syndromes characterized by low levels of plasma cholesterol.

**lipoproteins** Particles in the plasma composed of lipids (phospholipids, free and esterified cholesterol, and triglycerides) and apoproteins (e.g., apoA1, apoB, and apoC) that are responsible for lipid and lipid-soluble vitamin transport. The major lipoprotein classes are very low-density lipoproteins (VLDLs), low-density lipoproteins (LDLs), and high-density lipoproteins (HDLs).

**microsomal triglyceride transfer protein (MTP)** Protein heterodimer in the endoplasmic reticulum that is necessary for the formation of apo B-containing lipoproteins, such as chylomicrons or VLDLs.

Anderson's disease and chylomicron retention disease are terms used for very similar, if not identical, rare hereditary (most likely autosomal recessive) hypocholesterolemic disorders characterized by a lipid malabsorption syndrome with steatorrhea (chronic diarrhea) and growth retardation.

## INTRODUCTION

The term Anderson's disease has been used since 1970 by French medical scientists as the diagnosis for patients having manifestations similar to those of a patient described by the Australian physician

Charlotte Anderson. North American medical scientists used the term chylomicron retention disease to describe a cohort of patients with similar signs and symptoms in whom it appeared that chylomicrons were assembled but not secreted. Both terms have been used subsequently by physicians and scientists in other areas of the world as diagnoses for patients having the clinical and laboratory manifestations described in this article.

The diagnosis is made on the basis of the clinical findings (diarrhea, steatorrhea, and lipid malabsorption), endoscopy (the presence of a "white hoary" layer or *gelée blanche* on the small intestinal mucosa), the intestinal biopsy (vacuolated enterocytes that stain positively with Oil red), the levels of plasma cholesterol (hypocholesterolemia), the presence of apolipoprotein (apo)B-100 and the absence of apoB-48 in the plasma, and the absence of chylomicrons and apoB-48 in the plasma after a fat feeding. The parents of patients are asymptomatic.

## CLINICAL DESCRIPTION

Three well-described, inherited hypocholesterolemic states, characterized by lipoprotein deficiency, affect apoB-containing lipoproteins: abetalipoproteinemia, familial hypobetalipoproteinemia, and Anderson's disease or chylomicron retention disease. Abetalipoproteinemia is a rare autosomal recessive disease that manifests in infancy. It is characterized by profound hypocholesterolemia, hypotriglyceridemia, lipid malabsorption, diarrhea, retinitis pigmentosa, acanthocytosis, spinocerebellar degeneration, and the complete absence of apoB-containing lipoproteins. The molecular basis of the disease is a mutation in the gene encoding the large subunit of the microsomal triglyceride transfer protein (MTP), which results in a defect in lipoprotein assembly. Familial hypobetalipoproteinemia is an autosomal codominant

disorder. Two major phenotypes have been described. Individuals who are homozygous for null alleles, in which plasma apoB is absent, are phenotypically similar to individuals who have abetalipoproteinemia. Individuals in whom truncated forms of apoB are found in the plasma, whether homozygous or heterozygous, are generally asymptomatic clinically; however, in both cases, there are decreased plasma and low-density lipoprotein (LDL) cholesterol levels. Mutations in the apoB gene form the molecular basis of the disease.

Anderson's disease is clinically distinguishable from abetalipoproteinemia and null alleles of homozygous hypobetalipoproteinemia by the absence of acanthocytosis, retinitis pigmentosa, and severe neurological symptoms and by the presence of apoB-100-containing lipoproteins. Anderson's disease is distinguishable from heterozygous familial hypobetalipoproteinemia and homozygous familial hypobetalipoproteinemia (with truncated apoB) by the presence of diarrhea, malabsorption, and steatorrhea and an autosomal recessive mode of inheritance.

At least 35 cases of Anderson's disease in 26 families have been described. Sixteen cases (10 different families) are of North African origin (Algeria, Morocco, and Tunisia). Ten cases are from 9 Canadian families. The 9 remaining cases are from six different countries (Spain, Pakistan, Turkey, England, Lebanon, and the United States). Consanguinity has been described in 9 families, no consanguinity is present in 12 families, and no information regarding consanguinity is available for the remaining 5 families.

Individuals with these disorders exhibit a malabsorption syndrome with steatorrhea and growth retardation under a normolipidemic alimentary regime. The mucosal surface of the small intestine, as observed by endoscopy, is covered with a whitish layer ("a white stippling-like hoar frosting" or *gelée blanche*). Bloating of the stomach, osteomalacia, and rickets have been observed in several cases. Hepatic steatosis has been noted in four cases but without evolution to cirrhosis. Neuroretinal manifestations occur later and are less severe than in abetalipoproteinemia. Neurological signs most frequently consist of a loss of deep tendon reflexes. There is occasional alteration of position and vibratory senses, nerve conduction velocities, and evoked auditory and visual potentials. For patients diagnosed as adults, the neurological signs are more severe and may also include areflexia, ataxia, alteration in deep and vibratory senses, myopathy (with lipofuscin deposits on muscle biopsy), and polyneuropathy. Although nystagmus and delayed dark adaptation may occur, there is no visual loss

or retinitis pigmentosa. Acanthocytosis is typically absent.

Institution of a low-fat diet supplemented with lipid-soluble vitamins (A and E) and essential fatty acids results in the resumption of normal growth and abatement of the gastrointestinal symptomatology. Departure from the low-fat diet results in recurrence of symptoms.

## PLASMA LIPID AND LIPOPROTEIN AND BIOCHEMICAL ANALYSES

Plasma cholesterol levels are decreased but remain higher than 50 mg/dl. Fasting plasma triglyceride levels are normal. Postprandially, there is no increase in plasma triglycerides, and chylomicrons are not detected. However, the absorption of luminal fatty acids and their consecutive esterification by epithelial cells appear normal. Although apoB-48-containing lipoproteins are absent from the plasma, lipoproteins containing apoB-100 are present but in decreased amounts. The plasma levels of high-density lipoproteins (HDLs) and apoAI, -AIV, -E, and -C are also decreased, and there are low levels of total lipids, phospholipids, carotenoids, and lipid-soluble vitamins (particularly vitamin E) as well as vitamins A, K, and D. Lipoprotein composition is abnormal in that it has decreased amounts of cholesterol and increased amounts of phospholipid and triglyceride. Very low-density lipoproteins (VLDLs) are increased in size, whereas LDLs and HDLs are decreased in size. Analysis of mRNA synthesis in the human intestine has shown the presence of mRNA for apoAI, -AIV, -E, -CII, and -CIII. ApoB messenger RNA appeared to be correctly edited in the intestine in two cases of chylomicron retention disease in one family. Normal MTP protein and activity were detected in intestinal biopsies of several patients.

## IMMUNOCHEMICAL ANALYSES

Despite the absence of apoB-48 in the plasma in this disease, this apolipoprotein as well as apoAI, -AIV, -CII, and -CIII have been detected by immunochemical techniques in the enterocytes of patients, along with the lipid components that are normally assembled into triglyceride-rich lipoproteins. Immunoprecipitation with polyclonal antibodies to apoB or apoAIV of the homogenates of organ cultures of intestinal biopsies from patients shows the presence of normal-sized apoB-48 and apoAIV in amounts three- to fivefold more abundant compared to those

found in normal individuals. Analysis of homogenates of organ cultures and of the culture media shows that the patients assemble and secrete some normal-sized apoB-48 and apoAIV, which are coimmunoprecipitated and which float like chylomicrons. No apoB-100 is detected in the biopsy or culture medium.

## ULTRASTRUCTURAL ANALYSES BY LIGHT AND ELECTRON MICROSCOPY

Studies of the intestine by light microscopy have shown that villi are present in normal number and length but that the enterocytes are overloaded with fat droplets. The pattern and extent of lipid loading are variable among patients and also among biopsy sections for a given patient. In the regions of the villi that contain lipid-laden enterocytes there are always a few morphologically normal-appearing cells. The cells in the inferior approximately one-third of the villus characteristically show no accumulation of lipids.

When examined by electron microscopy, the enterocytes in some regions of biopsies have an intracellular architecture like that found in normal fasted individuals in whom intra- and intercellular lipoprotein-like particles are not readily apparent and the Golgi apparatus is flat and nondistended. In other regions, the enterocytes contain large amounts of lipid particles. Many of these are chylomicron- and VLDL-sized particles (approximately 300 nm in diameter; range, 169–580 nm) in membrane-bound compartments. When clearly identifiable, the Golgi apparatus is frequently distended and empty, although membrane-bound compartments containing particles are in close juxtaposition. These membrane-bound particles resemble lipoprotein particles seen in normal fed individuals that are situated in a membrane-bound compartment and are seen budding from the lateral aspect or located near the Golgi apparatus. The identity of the membrane-bound compartment that contains the lipoprotein-sized particles is not entirely clear, and the composition of the particles has not been carefully defined.

Other, larger particles (368–2127 nm mean diameter) appear to be lipid droplets that are free in the cytoplasm. These lipid droplets may derive from the breakdown of membrane-bound compartments that contain lipid particles unassembled with protein (putative second step triglyceride-rich particles). Large non-membrane-bound lipid droplets (presumably not assembled with protein) predominate in abetalipoproteinemia, whereas smaller membrane-bound

particles predominate in Anderson's disease. Finally, smaller particles (63 nm mean diameter) are rarely found in the intercellular spaces of affected individuals, suggesting that secretion can occur.

Even after treatment with a low-fat, lipid-soluble, vitamin-supplemented diet for at least 6 months and abatement of gastrointestinal symptoms, biopsies performed in patients after 12–15 h of fasting remained lipid laden with lipoprotein- and lipid-like particles with densities exceeding those found in the fed normal individual.

## LINKAGE ANALYSES

Using an autosomal recessive mode of transmission and highly polymorphic microsatellite markers [most frequently the  $(CA)_n$  type], segregation analyses of four families, excluded as a cause of the disease significant regions of the genome surrounding the genes for apoAI, -CIII, and -AIV (15 cm on chromosome 11); the apoCII gene, which includes the apoCI and apoE genes (24 cm on chromosome 19); and the genes encoding three proteins involved in intracellular lipid transport—MTP (30 cm on chromosome 4) and fatty acid-binding proteins 1 (20 cm on chromosome 2) and 2 (30 cm on chromosome 4). No evidence of linkage was found for a distance of at least 5 cm on either side of the reported location of the apoB gene (results exclude the apoB gene in 13 cases from seven families with Anderson's disease).

## INTRACELLULAR PROCESSING OF APOLIPOPROTEIN B

Two basic types of asparagine-linked glycans (N-glycosylation) are found on apoB-48 present in normal enterocytes. One form contains only high-mannose oligosaccharides and represents the newly glycosylated protein in the endoplasmic reticulum. The other form contains, in addition to high-mannose glycans, complex oligosaccharides that arise by processing of some of the high-mannose glycans in the trans-Golgi apparatus. Five of six potential asparagine glycosylation sites may be used in apoB-48 based on results obtained with apoB-100, and there are probably one high-mannose chain and four complex-type oligosaccharide chains on the same molecule. The mixed glycosylation pattern (high-mannose and complex oligosaccharides on the same molecule) is apparently due to the masking of some of the oligosaccharide chains on apoB-48 to the action of Golgi

glycosyltransferases. The addition and modification of N-linked carbohydrates occur in distinct intracellular compartments. The advancement of a protein along the secretory pathway can thus be assessed by evaluating the differences in the sensitivity of the N-linked oligosaccharides to endoglycosidases. In Anderson's disease, there is a time-dependent transformation of high-mannose endoglycosidase H-sensitive oligosaccharides of endoplasmic reticulum origin to complex endoglycosidase H-resistant oligosaccharides, added in the Golgi network, as in normal individuals. The apoB-48 containing oligosaccharides only partially sensitive to endo H (high-mannose plus complex oligosaccharides) has reached the trans-Golgi and is certainly already assembled with lipids and is probably destined to be secreted. In contrast, in abetalipoproteinemic patients, there is a single intracellular population of apoB-48 containing only high-mannose and hybrid glycans, indicating that apoB-48 does not reach the Golgi apparatus since the complex glycans that are added in the medial and trans-Golgi network are absent.

## CONCLUSION

Biochemical and ultrastructural analyses suggest that triglyceride-rich lipoprotein assembly takes place in enterocytes. ApoB-48 oligosaccharide processing indicates that the defect is not located between the endoplasmic reticulum and the Golgi apparatus but rather is distal to the trans-Golgi apparatus. Given that apo-AIV and apo-AI seem to be well secreted in patients, this implies a cargo-specific defect in transport between the trans-Golgi apparatus and the basolateral surface. Two independent genome-wide linkage analyses have shown that a locus on chromosome 5q31.1 segregates with affected status in several families affected with Anderson's Disease, chylomicron retention disease, or chylomicron retention disease with the neuromuscular disorder Marinesco-Sjogren syndrome. In 10 affected individuals, coding sequence variants (including two frameshift, one splice site, and five missense mutations) were identified in both alleles of the SARA2 gene, which is located in the region of apparent homozygosity. The SARA2 gene product (sar1b) belongs to the Sar1-ADP ribosylation factor family of small GTPases which are involved in COP-coated vesicle mediated intracellular transport.

## Acknowledgments

This study was supported by the Centre National de la Recherche Scientifique; the Institut National de la Santé et de la Recherche Médicale (PROGRES, Grant 4P016D); the Ministère de l'Éducation, de la Recherche et de la Technologie; the Fondation pour la Recherche Médicale (N.B.-V.); the Caisse Nationale de l'Assurance Maladie des Travailleurs Salariés (Grant 4AIC01); and the Société Nationale Française de Gastroentérologie.

## See Also the Following Articles

Abetalipoproteinemia • Familial Low Cholesterol Syndromes, Hypobetalipoproteinemia

## Further Reading

- Anderson, C., Townley, R. R. W., Freeman, M., and Johansen, P. (1961). Unusual causes of steatorrhea in infancy and childhood. *Med. J. Aust.* **11**, 617–622.
- Berriot-Varoqueaux, N., Dannoura, A. H., Moreau, A., Verthier, N., Sassolas, A., Cadiot, G., Lachaux, A., Munck, A., Schmitz, J., Aggerbeck, L. P., and Samson-Bouma, M. E. (2001). Hereditary malabsorptive and hypocholesterolemic syndromes: Apolipoprotein B48 glycosylation in abetalipoproteinemia and Anderson's disease. *Gastroenterology* **121**, 1101–1108.
- Bouma, M. E., Beucler, I., Aggerbeck, L. P., Infante, R., and Schmitz, J. (1986). Hypobetalipoproteinemia with accumulation of an apoprotein B-like protein in intestinal cells. Immunoenzymatic and biochemical characterization of seven cases of Anderson's disease. *J. Clin. Invest.* **78**, 398–410.
- Dannoura, A. H., Berriot-Varoqueaux, N., Amati, P., Abadie, V., Verthier, N., Schmitz, J., Wetterau, J. R., Samson-Bouma, M. E., and Aggerbeck, L. P. (1999). Anderson's disease: Exclusion of apolipoprotein and intracellular lipid transport genes. *Arterioscler. Thromb. Vasc. Biol.* **19**, 2494–2508.
- Jones, B., Jones, E. L., Bonney, S. A., Patel, H. N., Mesenkamp, A. R., Eichen-Baum-Voline, S., Rudling, M., Myrdal, U., Annesi, G., Naik, S., Meadows, N., Quattrone, A., Islam, S. A., Naoumova, R. P., Angelin, B., Infante, R., Levy, E., Roy, C. C., Freemont, P. S., Scott, J., and Shoulders, C. C. (2003). Mutations in a Sar1 GTPase of COPII vesicles are associated with lipid absorption disorders. *Nat. Genet.* **34**, 29–31.
- Kane, J. P., and Havel, R. J., (1995). Disorders of the biogenesis and secretion of lipoproteins containing the B apolipoproteins. In "The Metabolic and Molecular Basis of Inherited Disease" (C. R. Scriver, A. L., Beaudet, W. S., Sly, D., Valle, eds.), Vol. 2, pp. 1853–1885. McGraw-Hill, New York.
- Levy, E., Marcel, Y. L., Deckelbaum, R. J., Milne, R., Lepage, G., Seidman, E., Bendayan, M., and Roy, C. C. (1987). Intestinal apo B synthesis, lipids and lipoproteins in chylomicron retention disease. *J. Lipid Res.* **28**, 1263–1274.
- Roy, C. C., Levy, E., Green, P. H. R., Sniderman, A., Letarte, J., Buts, J. P., Orquin, J., Brochu, P., Weber, A. M., Morin, C. L., Marcel, Y. L., and Deckelbaum, R. J. (1987). Malabsorption, hypocholesterolemia and fat filled enterocytes with increased intestinal apoprotein B. Chylomicron retention disease. *Gastroenterology* **92**, 390–399.



# Androgen Biosynthesis and Gene Defects

Mark R. Vanderwel and Richard J. Auchus  
University of Texas Southwestern Medical Center, Dallas, Texas, United States

## Glossary

**cryptorchidism** The condition in which the testes fail to descend into the scrotum and are retained within the abdomen or inguinal canal.

**follicle-stimulating hormone (FSH)** A gonadotropin secreted and released by the anterior pituitary. FSH stimulates the ripening of the follicles in the ovary and formation of sperm in the testes by acting on granulosa and Sertoli cells, respectively.

**haploinsufficiency** A mutation in only one of the two alleles for an autosomal gene, resulting in a reduced gene dosage for the protein encoded on the normal allele.

**human chorionic gonadotropin (hCG)** A hormone similar to the pituitary gonadotropin luteinizing hormone; it is produced by the placenta during pregnancy.

**hypospadias** A congenital abnormality in which the labioscrotal folds have not completely fused, so that the opening of the urethra is on the underside of the penis.

**Leydig cells** The cells interspersed between the seminiferous tubules of the testis. They secrete androgens in response to luteinizing hormone.

**luteinizing hormone (LH)** A gonadotropin secreted by the anterior pituitary. LH stimulates androgen synthesis by the Leydig cells of the testis and the theca cells of the ovary; it also stimulates the “luteinization” of ovarian cells after ovulation, forming the corpus luteum, which makes progesterone.

**male pseudo-hermaphroditism** A congenital abnormality in which the genitalia of a 46,XY infant are not completely masculinized, despite the presence of testes.

**Müllerian structures** Structures that develop from the paramesonephric duct in normal females. These include the fallopian tubes, uterus, and upper part of the vagina.

**Sertoli cells** Cells found in the walls of the seminiferous tubules of the testis. They anchor and nourish the developing germ cells.

**Wolffian structures** Structures that develop from the mesonephric duct in normal males. These include the epididymis and vas deferens.

**F**ormation of the male external genitalia in human beings requires the production of dihydrotestosterone during the critical period of sexual differentiation at 8–12 weeks of gestation. At this time, chorionic gonadotropin (hCG) stimulates responsive Leydig cells to convert cholesterol to testosterone, which is transformed to dihydrotestosterone in target tissues. Mutations in the genes encoding the enzymes, cofactor proteins, receptors, and stimulatory hormones may disrupt this process at various steps, leading to congenital defects in androgen production and male pseudo-hermaphroditism. Defects in gonadotropin-releasing hormone will not interfere with fetal androgen production but will prevent the production of luteinizing hormone (LH) and not allow pubertal progression. This article describes the syndromes caused by mutations in these genes, focusing on the clinical and biochemical features of disorders that impair enzymatic steps in dihydrotestosterone production.

## NORMAL ANDROGEN BIOSYNTHESIS

All steroid hormone production begins with the conversion of cholesterol to pregnenolone. A “mobile pool” of free cholesterol in the outer mitochondrial membrane (OMM) is physically inaccessible to the side chain cleavage enzyme (CYP11A, P450<sub>scc</sub>), which resides in the inner mitochondrial membrane (IMM). Stimulation of Leydig cells with hCG or LH, both of which bind to the same LH/hCG receptor, increases intracellular cyclic AMP (cAMP). The rise in cAMP induces the expression and activation of the labile steroidogenic acute regulatory (StAR) protein, which allows the cholesterol to flow from the OMM to the IMM, where CYP11A converts the cholesterol to pregnenolone. The microsomal enzyme CYP17

(P450c17) sequentially oxygenates pregnenolone (the 17 $\alpha$ -hydroxylase reaction) and cleaves the C17—C20 bond (the 17,20-lyase reaction), yielding dehydroepiandrosterone (DHEA). Although human CYP17 hydroxylates pregnenolone and progesterone with comparable efficiencies, its 17,20-lyase activity is much more efficient for the  $\Delta^5$ -steroid 17 $\alpha$ -hydroxypregnenolone (17Preg) than for its  $\Delta^4$ -congenitor 17 $\alpha$ -hydroxyprogesterone (17OHP). Consequently, DHEA is the substrate for 17 $\beta$ -hydroxysteroid dehydrogenase type 3 (17 $\beta$ -HSD3), which reduces DHEA to  $\Delta^5$ -androstenediol, and 3 $\beta$ -hydroxysteroid dehydrogenase/isomerase type 2 (3 $\beta$ -HSD2) converts this  $\Delta^5$  steroid to testosterone (T). The sequence of these latter two reactions (17 $\beta$ -HSD3 and 3 $\beta$ -HSD2) may also proceed in reverse, with androstenedione (AD) as the intermediate. T produced by the testis diffuses into peripheral tissues and those tissues that contain the enzyme 5 $\alpha$ -reductase type 2 (SRD5A2) (i.e., prostate, genital skin) metabolize T to the potent androgen dihydrotestosterone (DHT). The importance of all of these steps and the lack of adequate redundancy are demonstrated by the clinical disorders caused by mutations in the genes encoding the key proteins described herein.

## DISORDERS OF LEYDIG CELL STIMULATION

### Hypothalamic Hypogonadism

Because fetal T synthesis during weeks 8–12 of gestation is driven primarily by placental hCG, 46,XY children with defects in gonadotropin-releasing hormone (GnRH) or gonadotropin production are born with relatively normal male external genitalia. However, T production in late gestation is driven by LH, so micropenis is often present, and these individuals fail to experience puberty. Mutations in the genes for the orphan nuclear receptors DAX1 and SF1, the homeodomain transcription factors HESX1, LH3, and PROP1, the  $\beta$ -subunits of LH and follicle-stimulating hormone, the GnRH receptor, or the extracellular matrix protein anosmin-1 have been identified in patients with defects in androgen production and/or infertility due to abnormalities in hypothalamus and/or pituitary development. Most of these disorders are extremely rare and variable in their manifestations, but the more common syndromes whose genetic basis has been at least partially defined are discussed below.

### Kallman's Syndrome and Variants

Kallman's syndrome refers to the combination of hypogonadotropic hypogonadism and anosmia. Consequently, 46,XY patients are born with normal male genitalia except for micropenis, and they fail to initiate puberty. The diagnosis is made when LH does not rise in response to a bolus of GnRH or GnRH analogue. Magnetic resonance imaging (MRI) with attention to the olfactory bulbs and midline structures is helpful but not essential, and olfactory testing should be performed. At the time of expected puberty, sex steroid replacement appropriate for gender is commenced. Treatment with pulsatile GnRH using a programmable pump will not only induce sexual maturation but may restore fertility in individuals of both sexes.

Kallman's syndrome can be sporadic, autosomal dominant or recessive, or X-linked. The best understood form of Kallman's syndrome is the X-linked variety, which accounts for nearly half of cases and most often results from mutations in the *KAL* gene. This gene encodes an extracellular matrix protein called anosmin-1 that guides the migration of both the olfactory and GnRH-producing neurons from the olfactory placode to their proper location in the head and brain. Forms of hypothalamic hypogonadism without anosmia may result from mutations in the gene for the GnRH receptor, but these conditions are rare. In approximately half the cases of Kallman's syndrome and its variants, the molecular basis is unknown.

### Septo-optic Dysplasia

Developmental defects in midline structures often ablate hypothalamic–pituitary axes, and the growth hormone and GnRH–LH axes are particularly vulnerable. Severe defects, such as holoprosencephaly, characteristically involve large portions of the brain, but milder developmental defects can involve few structures. Septo-optic dysplasia refers to the combination of optic nerve hypoplasia and hypothalamic–pituitary maldevelopment. Evaluation and management are similar to those for Kallman's syndrome, although vision testing and MRI evaluation are important for guiding follow-up and for prognosis. Mutations in the homeobox gene *HESX1* have been identified in patients with septo-optic dysplasia, and the clinical severity roughly correlates with the impairment in DNA binding by the mutant protein.

### Leydig Cell Agenesis or Hypoplasia (Testicular Unresponsiveness to hCG/LH)

When a 46,XY fetus has a mutation in the LH/hCG receptor, Leydig cells fail to develop appropriately,

and T production is impaired from conception. This defect leads to varying degrees of genital anomalies at birth, depending on the amount of hCG-independent T production prior to the 10th week of gestation and the severity of the hCG/LH receptor dysfunction. Because secretion of anti-Müllerian hormone by the Sertoli cells is intact, 46,XY children with Leydig cell hypoplasia do not retain Müllerian structures. The testes lack distinct Leydig cells on biopsy, and Sertoli cells may appear at puberty. However, the seminiferous tubules, if present, often show spermatogenic arrest, and the testes degenerate progressively.

The 46,XY infants born with completely female genitalia may not be identified until puberty, when they present with failure to develop breasts and to undergo menarche. Milder forms in 46,XY infants cause undervirilization, including hypospadias, micropenis, and cryptorchidism. 46,XX females homozygous for LH receptor defects will have normal female genitalia and may experience some breast development at puberty, but with amenorrhea and infertility. The diagnosis is confirmed by low or absent T, AD, and 17OHP production in response to hCG stimulation testing. Basal and GnRH-stimulated gonadotropin values are elevated in pubertal subjects.

Management depends on the age of diagnosis and the degree of virilization. When the defects are severe enough to produce phenotypically female genitalia, assignment of the female gender, with gonadectomy and estrogen replacement therapy at the time of expected puberty, is usually recommended. For less severely affected individuals with undervirilized male genitalia, surgery may be necessary to correct hypospadias, and testosterone therapy is used to stimulate phallic development and to virilize the patient at puberty.

Leydig cell hypoplasia is an autosomal recessive condition due to mutations in the hCG/LH receptor. Several genetic defects have been reported, including missense, nonsense, and null mutations. The null mutations, such as Arg554Stop, are associated with the most severe clinical phenotypes.

## VARIANTS OF CONGENITAL ADRENAL HYPERPLASIA

The most common form of congenital adrenal hyperplasia (CAH) is 21-hydroxylase (CYP21, P450c21) deficiency, but CYP21 is not expressed in the gonads and does not participate in T biosynthesis. Other, less common forms of CAH that involve enzymes or proteins expressed both in the adrenals and the gonads are discussed below.

## Lipoid CAH

### StAR Deficiency

Because StAR facilitates the transport of cholesterol from the OMM to the IMM, inactivating mutations in StAR block the production of pregnenolone and thus impair all steroidogenesis, both in the adrenals and in the gonads. Under the stimulation of adrenocorticotrophic hormone (ACTH) and LH, cholesterol esters massively accumulate in the adrenal glands and testes, respectively, affording the characteristic enlarged, lipid-laden adrenals from which the name lipoid CAH derives. Secondarily, sterol auto-oxidation products accumulate in the adrenals and Leydig cells, altering cell structure and ultimately provoking cell destruction.

Both 46,XY and 46,XX individuals will have female external genitalia at birth. Affected 46,XY individuals have abdominal, inguinal, or intralabial testes, a blind vaginal pouch, and no uterus or fallopian tubes. Wolffian duct remnants may be preserved in 46,XY individuals secondary to low levels of StAR-independent steroidogenesis. All reported patients are diffusely hyperpigmented from pro-opiomelanocortin excess. Because the theca cells of the ovary do not normally make steroids during fetal and neonatal life, the ovaries of 46,XX subjects do not suffer lipid accumulation and cell death during childhood. Consequently, 46,XX subjects may produce enough estrogens in early puberty to undergo some breast development and may even menstruate until lipid accumulation and auto-oxidation obliterate ovarian function as well.

The diagnosis of lipoid CAH is confirmed by low or absent glucocorticoids, mineralocorticoids, gonadal steroids, their precursors, and their metabolites in plasma and/or urine, even after stimulation. In particular,  $3\beta$ -HSD2 deficiency is excluded by documenting low concentrations of not only the active  $\Delta^4$  steroids but of the  $\Delta^5$  precursors pregnenolone and 17Preg. On abdominal computed tomography scan or MRI, the lipid-laden adrenals are strikingly enlarged, displacing the kidneys caudad.

Treatment requires replacement doses of glucocorticoids and mineralocorticoids in the newborn period, which must be continued throughout life. All affected 46,XY males have been reared as females, and orchidectomy is advised. Estrogen replacement therapy for individuals of both genotypes is required at puberty to initiate female secondary sexual characteristics, and low-dose testosterone may be used to elicit a female pattern of sexual hair growth. Lipoid CAH is an autosomal recessive disease with a male/female ratio of approximately 3/1; however, this ratio may be skewed

by ascertainment bias. This condition is rare in the United States and Europe, but is the second most common form of CAH in Japan and Korea. In one series, mutation Gln258Stop accounted for 80% of the affected alleles from Japanese and Korean subjects, suggesting a founder effect that causes the relatively high incidence of lipoid CAH in these countries. Mutation Arg182Leu was found in 78% of affected alleles from Palestinian subjects in the same series.

### **Side Chain Cleavage Enzyme (CYP11A or P450scc) Deficiency**

For many years, it was hypothesized that homozygous CYP11A deficiency caused lipoid CAH, but an absence of CYP11A would preclude placental progesterone synthesis and promote spontaneous abortion after the 8th to 10th week. However, haploinsufficiency of CYP11A has been shown to produce a milder clinical picture of lipoid CAH than in StAR deficiency. Tajima and associates described a 46,XY patient with clitoromegaly, a blind vaginal pouch, hyperpigmentation, and absent Müllerian structures. This patient was raised as a female, and testes were removed from the inguinal region. Adrenal insufficiency with hyperplasia did not occur until the child was 4 years old, and no mutation was found in the gene for StAR. Instead, one allele of the gene for CYP11A had a 6 bp in-frame insertion, adding Gly-Asp between Asp-271 and Val-272. This mutant enzyme had no activity and appeared to impair the function of wild-type CYP11A when expressed in the same cells, suggesting a partial, dominant-negative mode of action, leading to less severe disease in early childhood rather than infancy.

Treatment is similar to that for StAR protein deficiency, with gluco- and mineralocorticoid replacement to prevent life-threatening adrenal insufficiency at the time of diagnosis, plus estrogen replacement therapy at the time of puberty.

### **3 $\beta$ -HSD2 Deficiency**

The 3 $\beta$ -HSD enzymes catalyze the conversion of the  $\Delta^5$  steroids pregnenolone, 17Preg, DHEA, and  $\Delta^5$ -androstenediol to their corresponding  $\Delta^4$  steroids progesterone, 17OHP, AD, and T, respectively. One of these conversions is required in the biosynthesis of all active steroid hormones, so severe 3 $\beta$ -HSD deficiency will also result in a form of CAH with impaired androgen production.

46,XY individuals with 3 $\beta$ -HSD deficiency most frequently exhibit male pseudo-hermaphroditism with a small phallus, hypospadias, partial labioscrotal fusion, and possibly a urogenital sinus with a blind vaginal pouch. Testes usually lie in the lower inguinal region, and Müllerian structures are absent. Paradoxically, 46, XX individuals often have trace clitoral enlargement and progressive masculinization if undertreated. Severe 3 $\beta$ -HSD deficiency can present with salt-wasting crisis from glucocorticoid and mineralocorticoid insufficiency within the first week of life. Less severe forms of 3 $\beta$ -HSD deficiency are usually diagnosed in genetic males because of genital abnormalities, but may be difficult to diagnose in females. Androgen production increases at puberty in both sexes but at a rate that is intermediate for males and females; consequently, girls show signs of androgen excess, but boys often develop gynecomastia. Fertility has been reported in affected individuals of both sexes.

The diagnosis of 3 $\beta$ -HSD deficiency hinges on elevated ratios of  $\Delta^5$  steroids to their  $\Delta^4$  congeners. These ratios, which are already increased at baseline, are accentuated by cosyntropin stimulation and should reach >12 SD above normal. Adult females with hirsutism often have high circulating DHEA-S concentrations with high ratios of  $\Delta^5$  to  $\Delta^4$  steroids, so extremely elevated ratios are required to confidently diagnose 3 $\beta$ -HSD deficiency.

Therapy includes early gluco- and mineralocorticoid replacement in salt-wasting individuals to prevent life-threatening adrenal insufficiency, and similar replacement is used in non-salt-wasting individuals to limit sexual precocity caused by increased synthesis of adrenal DHEA-S. Females require estrogen replacement, and males require testosterone supplementation to achieve full development of secondary sexual characteristics.

Two functional 3 $\beta$ -HSD genes are encoded on chromosome 1p13. The type 2 enzyme is the dominant isoform expressed in the adrenals and gonads, and its gene is mutated in 3 $\beta$ -HSD deficiency. Adult females who present with hirsutism, infertility, and relatively high DHEA-S concentrations do not have mutations in the 3 $\beta$ -HSD genes. The type 1 enzyme is expressed in the placenta but also in liver and skin, and this enzyme accounts for the peripheral conversion of  $\Delta^5$  precursors to  $\Delta^4$  steroids in 3 $\beta$ -HSD deficiency, leading to the paradoxical androgen excess in females. Mutations in the gene for the type 1 enzyme have not been reported and are probably lethal because, as with homozygous CYP11A deficiency, this disorder would compromise progesterone synthesis

by the placenta. Mutations in the type 2 gene have been found from residues Ala-10 to Gly-294, with no mutations dominating large series. Salt-wasting subjects have completely inactive alleles, and non-salt-wasting individuals can harbor mutations that retain partial activity when expressed in heterologous systems.

### CYP17 (P450c17) Deficiencies

CYP17 catalyzes both the 17 $\alpha$ -hydroxylase and 17,20-lyase reactions. Severe mutations will ablate both activities, but milder mutations may either partially impair both activities or preferentially impair 17,20-lyase activity with 17 $\alpha$ -hydroxylase activity remaining relatively normal. Furthermore, the 17,20-lyase activity is particularly dependent on the interaction of CYP17 with its redox partners, cytochrome P450 oxidoreductase (CPR) and cytochrome b<sub>5</sub>. In particular, the presence of cytochrome b<sub>5</sub> increases the 17,20-lyase activity 10-fold but minimally influences 17 $\alpha$ -hydroxylase activity. Thus, alterations in redox partners and/or their interactions with CYP17 may preferentially impair 17,20-lyase activity.

#### **Combined 17 $\alpha$ -Hydroxylase/17,20-Lyase Deficiency**

The classical description of severe, combined, 17 $\alpha$ -hydroxylase/17,20-lyase deficiency is sexual infantilism and hypokalemic hypertension in both genetic sexes. The presentation in 46,XY individuals varies from completely female external genitalia with a blind vaginal pouch to an undervirilized male genital phenotype with hypospadias and a small phallus. Testes in these subjects may be intra-abdominal, in the inguinal canal, or in labioscrotal folds. Müllerian structures are absent, and Wolffian derivatives are usually hypoplastic. Severely affected patients will fail to develop secondary sexual characteristics, including pubic and axillary hair. With milder disease, males may develop gynecomastia at puberty and females may progress to Tanner stage 5 with ovulatory menses. Based on studies with recombinant enzyme, it has been estimated that at least 25% of normal fetal CYP17 activities are required for masculinization of the external genitalia. Hypertension with hypokalemia occurs in childhood and is often severe, with myopathy and hypertensive complications at a young age.

The diagnosis should be entertained not only in male pseudo-hermaphrodites, but also in any individual with hyporeninemic hypertension, hypokalemic alkalosis, and a suppressed aldosterone. This diagnosis may be confirmed by obtaining elevated serum levels

of ACTH and the precursors that accumulate proximal to the block in 17-hydroxylation: progesterone, 11-deoxycorticosterone (DOC), corticosterone (B), 18-hydroxy-DOC, and 18-hydroxy-B. Both DOC and B have mineralocorticoid activity, leading to hypertension and hypokalemia. However, signs of adrenal insufficiency rarely develop, because the weak glucocorticoid B is present in abundance. Gonadotropins are elevated at puberty, and serum concentrations of aldosterone, 17OHP, cortisol, and sex steroids are low or absent.

Treatment includes replacement of glucocorticoids to suppress DOC and B secretion and thereby to normalize potassium homeostasis and blood pressure. Mineralocorticoid receptor antagonists, such as spironolactone, can be added to reduce the doses of glucocorticoids and to prevent iatrogenic Cushing syndrome. At puberty, sex steroid replacement is indicated, and gonadectomy should be performed in 46,XY patients assigned a female sex of rearing.

Mutations in CYP17 have been identified throughout the protein, and some missense mutations retain partial activity when expressed in heterologous systems. Most mutations that yield a completely inactive enzyme also destabilize the enzyme structure and ablate heme binding. A deletion of Phe-53 has been found in several Japanese subjects, and a CATC duplication following Ile-479 has been described in both Dutch Frieslanders and Canadian Mennonites. Recent reports suggest that 17-hydroxylase deficiency is most common in Brazil, where mutations W406R and R362C dominate.

#### **Isolated 17,20-Lyase Deficiency**

When only the 17,20-lyase activity is deficient, as has been suggested in 18 case reports, adrenal gluco- and mineralocorticoid synthesis is normal, but testosterone synthesis is impaired. Therefore, serum potassium and blood pressure are normal, but sexual development may be hampered. In 46,XY patients, the external genitalia are that of an undervirilized male, due to some residual 17,20-lyase activity. Müllerian structures are absent, Wolffian derivatives are either hypoplastic or normal, and testes may be intra-abdominal, inguinal, or in the scrotum. In 46,XX females, this condition is believed to lessen adrenarchal and pubertal development, but no confirmed cases have been studied in detail.

Cosyntropin stimulation tests yields normal or elevated 17-hydroxysteroid values, including cortisol and 17OHP, but DHEA and AD do not rise proportionately. Similarly, hCG stimulation testing will produce an increase in 17OHP concentrations, but AD and

T do not rise normally. The ratio of the rise in the C<sub>19</sub> steroids to their C<sub>21</sub>, 17-hydroxy precursors is the most discriminatory parameter for diagnosing isolated 17,20-lyase deficiency.

Only six cases of isolated 17,20-lyase activity have been confirmed with molecular genetic and biochemical studies. These subjects were all 46,XY individuals homozygous for mutations Arg358Gln (one) or Arg347His (three) or heterozygous for a completely inactive allele and one copy of Arg347Cys (two). These residues, Arg-347 and Arg-358, are both located on the redox-partner-binding surface of the CYP17 enzyme. Consistent with the known dependence of the 17,20-lyase activity on the interaction of CYP17 with redox partners CPR and cytochrome b<sub>5</sub>, these mutations at this binding surface appear to preferentially impair 17,20-lyase activity.

Gonadectomy is recommended in 46,XY subjects raised as females and gender-appropriate sex steroid replacement will be necessary at the time of puberty. Optimal management in 46,XX subjects has not been established.

## DEFECTS AFFECTING ONLY TESTOSTERONE AND DIHYDROTESTOSTERONE PRODUCTION

The final two genetic disorders discussed involve the terminal steps of T and DHT biosynthesis. These two conditions are unique in that only males (46,XY) experience clinical manifestations that are solely due to androgen deficiency *in utero*. Furthermore, other enzymes partially compensate for the genetic deficiencies, but only at puberty.

### 17 $\beta$ -Hydroxysteroid Dehydrogenase Type 3 (17 $\beta$ -HSD3) Deficiency

The human genome contains several 17 $\beta$ -HSD isoforms, but 17 $\beta$ -HSD3 is the enzyme that is defective in the clinical entity "17 $\beta$ -HSD deficiency," also known as "17-ketosteroid reductase deficiency." The 17 $\beta$ -HSD3 enzyme catalyzes the conversion of C<sub>19</sub>, 17-ketosteroids to 17 $\beta$ -hydroxysteroids using NADPH as cofactor: AD to T, DHEA to  $\Delta^5$ -androstenediol, 5 $\alpha$ -androstanedione to DHT, and 5 $\alpha$ -androsterone to 5 $\alpha$ -androstenediol. Because 17 $\beta$ -HSD3 is expressed exclusively in the testes, the loss of this enzyme impairs androgen biosynthesis only in males.

Most affected 46,XY individuals with 17 $\beta$ -HSD3 deficiency have predominantly female external

genitalia with a blind vaginal pouch. Surprisingly, Wolffian derivatives, such as the epididymis, vas deferens, seminal vesicles, and ejaculatory duct, are present, suggesting that an alternate pathway in these tissues enables some testosterone production, perhaps mediated by the 17 $\beta$ -HSD type 5 isoform. Testes are usually located in the inguinal canal, and Müllerian structures are absent.

Most of these children are raised as females. At puberty, testicular AD production increases and significant extraglandular conversion of this AD to T elicits marked physical changes. The phallus enlarges and can reach lengths of 4 to 8 cm; the voice may deepen, male body hair develops, and muscle mass increases. Several affected individuals have changed gender role from female to male in adolescence because of the prominent physical and psychological masculinization they experience. In contrast, 17 $\beta$ -HSD3 is not expressed in the human ovary, so 46,XX patients with this disorder are asymptomatic.

The diagnosis is based on markedly elevated AD concentrations in the face of low T in the neonatal period or in adolescence. The discrepancy in the AD/T ratio is accentuated with hCG stimulation. In the past, affected 46,XY males were frequently raised as females and underwent castration followed by estrogen substitution therapy at puberty, but infants with adequate phallic structures and mild hypospadias may be reared as males and undergo genitoplasty. This approach anticipates the tendency for gender reversal associated with virilization at puberty. However, even within members of a kindred with identical genotypes, affected individuals vary in their decisions about gender reversal at puberty when reared initially as females. If the patient is reared as a male, T replacement at puberty is necessary to achieve full masculinization and to prevent the development of gynecomastia. Spermatogenesis is absent because intratesticular T synthesis is blocked, and postpubertal elevations in gonadotropins may increase the risk for testicular neoplasms.

Most mutations in the gene for 17 $\beta$ -HSD are located on exon 9 and impair all enzyme functions. One common mutation, identified in both Brazilian and Palestinian subjects, is R80Q, which lies in the Rossman fold area and primarily disrupts the binding of cofactor, but not of steroid.

### 5 $\alpha$ -Reductase Type 2 (SRD5A2) Deficiency

The disorder 5 $\alpha$ -reductase deficiency (also known as pseudo-vaginal perineoscrotal hypospadias)

Table I Clinical Manifestations in 46,XY Individuals

	Leydig cell hypoplasia	Lipoid CAH	3 $\beta$ -HSD2 deficiency	17-OHase deficiency	Isolated 17,20-lyase deficiency	17 $\beta$ -HSD3 deficiency	5 $\alpha$ -Reductase 2 deficiency
Possible appearance of genitalia	Female/ambiguous/hypoplastic male	Female	Ambiguous/hypospadiac male	Female/ambiguous/hypospadiac male	Female/hypoplastic male $\pm$ hypospadias	Female/ambiguous	Ambiguous hypospadias, small phallus
Wolffian duct derivatives	Absent/hypoplastic	Absent/hypoplastic	Normal	Absent/hypoplastic	Hypoplastic/normal	Hypoplastic	Normal
Müllerian duct derivatives	Absent	Absent	Absent	Absent	Absent	Absent	Absent
Gonads in 46,XY	Testes, no Leydig cells <sup>a</sup>	Testes lipid laden	Testes	Testes	Testes	Testes	Normal testes
Adrenal insufficiency?	No	Severe	Severe–mild	No	No	No	No
Virilization at puberty?	None to poor	None	Poor to mild	None to poor	Poor	Yes	Yes
Increased hormone concentrations	LH, FSH	LH, FSH, renin	LH, FSH, pregnenolone, <b>17Preg, DHEA</b>	LH, FSH, <b>DOC, B, Progesterone</b>	LH, FSH	LH, FSH, <b>Estrone, AD</b>	<b>T</b>
Decreased hormone concentrations	T, DHT	All adrenal & gonadal steroids	Progesterone, <b>17OHP, AD, T, DHT</b>	Renin, cortisol, <b>17OHP, DHEA, AD, T, DHT</b>	<b>DHEA, AD, T, DHT</b>	<b>T, DHT</b>	<b>DHT</b>
Chromosomal location	2p21	StAR: 8p11.2 CYP11A: 15q23–q24	1p13	10q24.3	10q24.3	9q22	2p23

<sup>a</sup>Small undescended testes with absent or decreased numbers of Leydig cells to descended testes of normal size with decreased numbers of Leydig cells.

provides strong genetic evidence that DHT is required for complete formation of the male genitalia in human beings. The  $5\alpha$ -reductases catalyze the conversion of T to its  $5\alpha$ -reduced metabolite DHT, and the type 2 isoform executes this transformation in the prostate and genital skin. Consequently, 46,XY male infants with a deficiency in the type 2 enzyme are born with hypospadias and a phallic structure that resembles an enlarged clitoris, often bound in chordee. The urogenital sinus with a blind vaginal pouch opens on the perineum, and the scrotum is bifid, with testes located in the inguinal canal or labioscrotal folds. With the abundance of T, Wolffian structures are well differentiated, and Müllerian structures are absent. The ejaculatory ducts terminate in the blind vaginal pouch or onto the perineum next to the urethra, and the prostate is hypoplastic.

As is the case in  $17\beta$ -HSD3 deficiency, masculinizing changes occur at puberty as circulating T concentrations rise into the normal adult male range. The voice deepens, muscle mass increases, the phallus grows to 4 to 8 cm, and the subject may experience erections. The testes enlarge, the scrotal structure becomes rugated and pigmented, and spermatogenesis may occur, but is often impaired from cryptorchidism. Facial hair and body hair are sparse and acne and temporal hair recession do not occur, presumably because DHT production is low. As in 46,XY infants with  $17\beta$ -HSD3 deficiency, most 46,XY individuals with  $5\alpha$ -reductase deficiency are reared as females but reverse gender role with the masculinizing changes of puberty, and in some cultures where the disorder is endemic, this process has achieved a socially acceptable status. However, unlike  $17\beta$ -HSD3 deficiency, gynecomastia does not develop. The 46,XX females with  $5\alpha$ -reductase deficiency are phenotypically normal at birth, but at puberty they have decreased body and sexual hair and delayed menarche yet normal fertility.

Diagnostic testing includes measurement of serum T and DHT, and a T/DHT ratio  $>30$  confirms the diagnosis. One pitfall of testing is that after puberty, the activity of the type 1 isozyme may provide measurable levels of DHT, emphasizing the importance of the T/DHT ratio.

Treatment of  $5\alpha$ -reductase deficiency is DHT therapy, often applied as a cream to the genitalia, to increase phallic length and to facilitate hypospadias repair. Supraphysiologic dosing of T in adults may generate adequate DHT via a partially functional type 2 enzyme and via the type 1 isoform.

Approximately 40% of children with  $5\alpha$ -reductase deficiency are born to consanguineous parents; uniparental disomy has also been described. The founder mutation Arg246Trp is prevalent in the Dominican Republic, where the disease is common, but other mutations are found in other kindreds there. A 20 kb deletion that includes the *SRD5A2* gene is prevalent in Papua New Guinea, and an A insertion into amino acid 251 causes  $5\alpha$ -reductase deficiency in a Turkish cluster. Several mutations are found in Brazil, and Gln126Arg is the most common. The type 1 isozyme is expressed in the liver up to 2 to 3 years of age and in nongenital skin. Mutations in the type 1 isozyme have not been described.

Table I compares clinical manifestations in 46,XY children born with these disorders. Hypothalamic hypogonadism is not included, because boys with this disorder generally have normal genitalia, except for micropenis.

## See Also the Following Articles

Adrenal Androgens • Androgen Biosynthesis Inhibitors and Androgen Receptor Antagonists • Androgens, Gender and Brain Differentiation • Androgen Insensitivity Syndrome • Estrogen and the Male • Pseudohermaphroditism, Male, Due to  $5\alpha$ -Reductase-2 Deficiency • Sexual Function and Androgens • Undescended Testes

## Further Reading

- Achermann, J. C., Weiss, J., Lee, E. J., and Jameson, J. L. (2001). Inherited disorders of the gonadotropin hormones. *Mol. Cell. Endocrinol.* **179**, 89–96.
- Andersson, S., Geissler, W. M., Wu, L., Davis, D. L., Grumbach, M. M., New, M. I., Schwarz, H. P., Blethen, S. L., Mendonça, B. B., Bloise, W., Witchel, S. F., Cutler, G. B., Jr., Griffin, J. E., Wilson, J. D., and Russell, D. W. (1996). Molecular genetics and pathophysiology of  $17\beta$ -hydroxysteroid dehydrogenase 3 deficiency. *J. Clin. Endocrinol. Metab.* **81**, 130–136.
- Auchus, R. J. (2001). The genetics, pathophysiology, and management of human deficiencies of P450c17. *Endocrinol. Metab. Clin. N. Am.* **30**, 101–119.
- Bose, H. S., Sugawara, T., Strauss, J. F., III, and Miller, W. L. (1996). The pathophysiology and genetics of congenital lipid adrenal hyperplasia. *N. Engl. J. Med.* **335**, 1870–1878.
- Geller, D. H., Auchus, R. J., Mendonça, B. B., and Miller, W. L. (1997). The genetic and functional basis of isolated  $17,20$ -lyase deficiency. *Nat. Genet.* **17**, 201–205.
- Imperato-McGinley, J., and Zhu, Y. S. (2002). Androgens and male physiology in the syndrome of  $5\alpha$ -reductase-2 deficiency. *Mol. Cell. Endocrinol.* **198**, 51–59.
- Moisan, A. M., Ricketts, M. L., Tardy, V., Desrochers, M., Mebarki, F., Chaussain, J. L., Cabrol, S., Raux-Demay, M. C., Forest, M. G., Sippell, W. G., Peter, M., Morel, Y., and Simard, J. (1999). New insight into the molecular basis of  $3\beta$ -hydroxysteroid

- dehydrogenase deficiency. *J. Clin. Endocrinol. Metab.* **84**, 4410–4425.
- Richter-Unruh, A., Martens, J. W., Verhoef-Post, M., Wessels, H. T., Kors, W. A., Sinnecker, G. H., Boehmer, A., Drop, S. L., Toledo, S. P., Brunner, H. G., and Themmen, A. P. (2002). Leydig cell hypoplasia: Cases with new mutations, new polymorphisms and cases without mutations in the luteinizing hormone receptor gene. *Clin. Endocrinol.* **56**, 103–112.
- Tajima, T., Fujieda, K., Kouda, N., Nakae, J., and Miller, W. L. (2001). Heterozygous mutation in the cholesterol side chain cleavage enzyme (p450scc) gene in a patient with 46-XY sex reversal and adrenal insufficiency. *J. Clin. Endocrinol. Metab.* **86**, 3820–3825.



# Androgen Biosynthesis Inhibitors and Androgen Receptor Antagonists

Stephen E. Borst, David T. Lowenthal, and George Zavros  
*University of Florida and Malcom Randall VA Medical Center, Gainesville, Florida, United States*

## Glossary

**androgenetic alopecia** Male-pattern baldness, requiring both a genetic predisposition and sufficient action of dihydrotestosterone.

**clinical flare** Temporary, but serious, side effects resulting from administration of gonadotropin-releasing hormone agonists. Includes bone pain, nerve compression, and blockage of the ureters.

**gynecomastia** Excessive development of male mammary glands.

**hirsutism** The presence of excessive facial and body hair, especially in women.

**paraphilias** Sexual practices that are socially prohibited.

The androgens testosterone and dihydrotestosterone (DHT) play a central role in a number of disease states, including the progression of prostate cancer, benign prostatic hyperplasia, male pattern baldness, hirsutism, acne, and virilizing syndromes in women. In adults, most of the undesirable effects of androgens are mediated specifically by DHT. This article discusses the use of drugs that produce anti-androgenic effects by the following mechanisms of action: (1) androgen receptor antagonism, (2) inhibition of the conversion of testosterone to DHT, and (3) inhibition of testosterone synthesis.

## ANDROGEN RECEPTOR ANTAGONISTS

The androgen receptor, the gene of which is located near the centromere on the long arm of the X chromosome, was first described in 1969 and was cloned in 1988. It is present in highest concentrations in androgen target tissues such as the accessory organs of male reproduction. Tissues such as skeletal muscle, heart, and placenta have smaller amounts of androgen

receptor. Both testosterone and DHT bind to the androgen receptor in the cell, leading to the transcription of certain genes. The receptor's affinity is four times greater for DHT than for testosterone.

### Flutamide

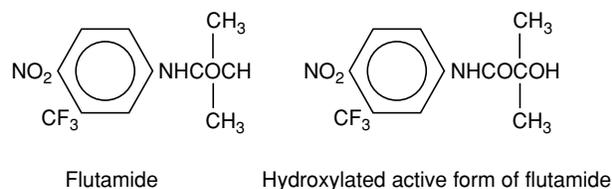
Flutamide is a nonsteroidal androgen receptor antagonist used in the management of metastatic prostate cancer and in the treatment of hirsutism in women. It is a pure androgen antagonist and produces no androgenic or other steroidal effects. It is metabolized in the liver to hydroxyflutamide, which is the active anti-androgen (see Fig. 1). It has a relatively short half-life of approximately 5.2 h. Side effects of flutamide treatment are gynecomastia, abnormal liver function, diarrhea, and gastrointestinal complaints. It is contraindicated in patients with severe hepatic impairment.

### Nilutamide

Nilutamide is structurally related to flutamide and binds to the androgen receptor with an affinity similar to that of hydroxyflutamide. It is used in combination with surgical castration for the treatment of metastatic prostate cancer. Nilutamide has a longer half-life than flutamide, approximately 40 h, allowing once a day oral administration. The major side effects of nilutamide are diarrhea, impaired adaptation to darkness, alcohol intolerance, and the occasional serious side effect of acute interstitial pneumonitis. It is contraindicated in patients with severe respiratory insufficiency or severe hepatic impairment.

### Casodex

Casodex is another nonsteroidal androgen receptor inhibitor used in the management of prostate cancer. Casodex is indicated for use in combination therapy



**Figure 1** Chemical structure of flutamide and its hydroxylated active form.

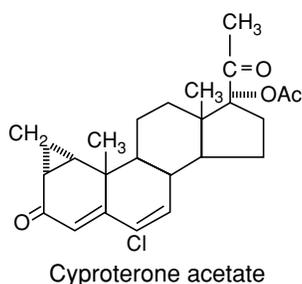
with a luteinizing hormone-releasing hormone analogue for the treatment of stage D2 metastatic carcinoma of the prostate. It is generally well tolerated and the most common side effects include gynecomastia and breast pain. It should be used with caution in patients with moderate to severe hepatic impairment as hepatotoxicity has been reported during the first 3 to 4 months of treatment.

### Cimetidine

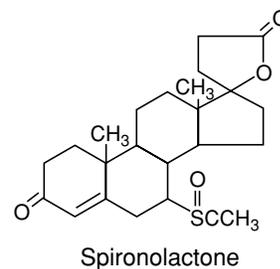
Cimetidine was the first histamine-2 blocker introduced for general clinical use in the treatment of duodenal ulcers and other gastric hypersecretory conditions. It exerts anti-androgenic properties by binding to the androgen receptor, causing loss of libido, impotence, and gynecomastia. It has been used as an orphan drug for the treatment of androgenetic alopecia, hirsutism, and warts.

### Cyproterone

Cyproterone acetate, a potent androgen antagonist, has been used for the treatment of acne, male pattern baldness, and hirsutism in men and for virilizing syndromes in women as well as in the treatment of the paraphilias (see Fig. 2). Its usefulness is limited by the fact that it can cross the placenta and cause male pseudo-hermaphroditism in male embryos and it has been associated with severe liver damage. It is the most widely used anti-androgen internationally. It has orphan drug status in the United States.



**Figure 2** Chemical structure of cyproterone acetate.



**Figure 3** Chemical structure of spironolactone.

### Spironolactone

Spironolactone is a steroidal androgen receptor blocker that has been shown to be beneficial in the management of acne, hirsutism, and androgenetic alopecia (see Fig. 3). Spironolactone is a weak anti-androgen in blocking the androgen receptor but it also acts by inhibiting androgen biosynthesis. It is best known as an aldosterone receptor antagonist, it has been used traditionally as a potassium-sparing diuretic, and hyperkalemia is one of its side effects. Other side effects include menstrual irregularities, breast tenderness, fatigue, and headache.

### Organophosphates

Organophosphate insecticides are widely used in agricultural and residential settings. The organophosphate insecticide fenestration has structural similarity to flutamide, a known anti-androgen, and has been identified as an androgen receptor antagonist.

## ANDROGEN BIOSYNTHESIS INHIBITORS

### Androgen Biosynthesis

Gonadotropin-releasing hormone (GnRH) is released from the hypothalamus in a pulsatile (on-off) manner, stimulating the anterior pituitary to produce luteinizing hormone (LH) and follicle-stimulating hormone (FSH). LH stimulates the Leydig cells of the testes to secrete testosterone. FSH stimulates the Sertoli cells of the testes, increasing spermatogenesis. Testosterone is the major circulating androgen and is converted to DHT in tissues expressing the  $5\alpha$ -reductase enzyme. Muscle and bone express very low levels of  $5\alpha$ -reductase and the anabolic actions of testosterone in these tissues are the direct result of binding of testosterone to androgen receptors. The prostate highly expresses  $5\alpha$ -reductase and the effects of testosterone in promoting prostate enlargement and

cancer are both mediated by DHT. Testosterone and DHT bind to the same androgen receptors, but DHT has greater potency. In addition, binding of DHT dimerizes the androgen receptor, causing additional effects. The following tissues express 5 $\alpha$ -reductase and exhibit androgenic effects that are dependent on DHT: hair follicles (chest, scalp, and beard), liver, kidney, adrenals, seminal vesicles, prostate, testes, foreskin, and scrotum. Virilization of the external genitalia is dependent on the conversion of testosterone to DHT and a deficiency in 5 $\alpha$ -reductase results in an incomplete form of male pseudohermaphroditism. However, in adults the effects of DHT are generally considered to be undesirable. The latter include increased body hair, acne, male-pattern baldness, and prostate enlargement. In contrast, the direct effects of testosterone are generally considered to be desirable. The latter include increased muscle and bone mass, deepening of the voice, increased libido and sense of well-being, spermatogenesis, penile and scrotal enlargement, and increased hematocrit (desirable in the absence of polycythemia). Blocking the effects of DHT, while preserving the direct effects of testosterone, presents an attractive therapeutic target.

### Finasteride

Finasteride is a selective inhibitor of 5 $\alpha$ -reductase type 1 (see Fig. 4). Finasteride is effective for treating male-pattern baldness and is marketed for this

purpose as Propecia (1 mg/day). Finasteride is also effective for the treatment of benign prostatic hyperplasia (BPH) and is marketed for this purpose as Proscar (5 mg/day). Finasteride causes an approximate 20% reduction in prostate size and brings symptomatic relief of BPH within several months. In contrast,  $\alpha$ -adrenergic receptor antagonists bring more rapid relief by decreasing urinary hesitation and spasm in the bladder neck. However,  $\alpha$ -antagonists are less effective on a long-term basis, because they do not reduce or prevent an increase in the size of the prostate. Finasteride is less effective than other treatments for treating prostate cancer and is not indicated for this purpose. The ineffectiveness of finasteride in prostate cancer may be due to incomplete suppression of DHT or to elevation of testosterone within the prostate. Whereas 5 $\alpha$ -reductase type 1 predominates in the prostate, 5 $\alpha$ -reductase type 2 is also present and finasteride is a weak inhibitor of the latter enzyme. Finasteride reduces prostate tissue levels of DHT by approximately 70%, but the remaining DHT may be sufficient to produce a permissive effect in prostate cancer. Several compounds that are dual inhibitors of both type 1 and type 2 5 $\alpha$ -reductase are under evaluation. These compounds cause a near-total suppression of prostate DHT levels. Whether they have greater efficacy than finasteride in treating BPH and prostate cancer remains to be determined. By blocking the conversion of testosterone to DHT, finasteride causes an increase in prostate testosterone levels. Although testosterone promotes

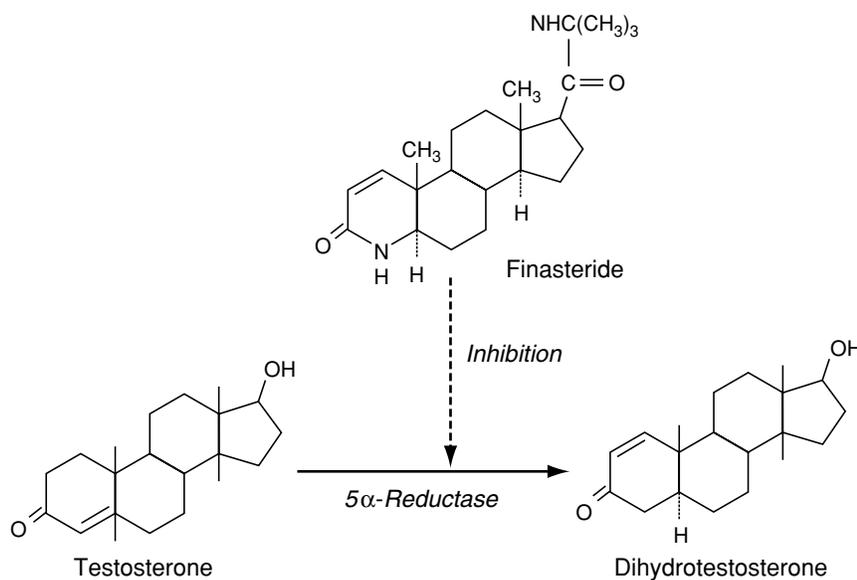


Figure 4 Finasteride inhibits 5 $\alpha$ -reductase and thus blocks the conversion of testosterone to DHT.

5-O-Pro-His-Trp-Ser-Tyr-Gly-Leu-Arg-Gly-Pro-Gly-NH<sub>2</sub>  
Endogenous GnRH

5-O-Pro-His-Trp-Ser-Tyr-Gly-D-Leu-Arg-Gly-Pro-NHCH<sub>2</sub>CH<sub>3</sub>  
Leuprolide (Leupron)

**Figure 5** Peptide sequence of endogenous GnRH and that of a GnRH analogue, leuprolide (Leupron), with a D-amino acid substituted at the 6 position and with ethylamide substituted at the 10 position.

prostate cancer to a much lesser degree than does DHT, 5 $\alpha$ -reductase inhibitors may elevate prostate testosterone levels sufficiently to promote or to maintain a permissive effect on prostate cancer.

### GnRH Agonists

GnRH is a decapeptide secreted by the hypothalamus in a pulsatile manner, causing the anterior pituitary to release LH and FSH. Endogenous GnRH has a short half-life. Analogues with a D-amino acid substituted at the 6 position and with ethylamide substituted at the 10 position are GnRH agonists with greater potency and longer duration of action than endogenous GnRH (see Fig. 5). Leuprolide (Leupron) and goserelin (Zoladex) are GnRH agonists that have been used successfully to treat prostate cancer. Leupron is formulated in microspheres and Zoladex is formulated in a polymer matrix. In both cases, the slow release of the drug from an intramuscular injection site suppresses androgens for several months. GnRH agonists continuously stimulate pituitary GnRH receptors, as opposed to endogenous GnRH, which stimulates them intermittently. Continuous stimulation of GnRH receptors produces a transient "clinical flare reaction," which is caused by a surge of androgen production and which lasts 5 to 12 days. Clinical flare occurs in approximately 10% of patients with prostate cancer who are treated with GnRH agonists. The condition is dangerous and is associated with bone pain, nerve compression, and blockage of the ureters. These symptoms may be controlled with androgen receptor antagonists, such as flutamide. However, androgen antagonists also produce serious side effects. Eventually, continuous stimulation causes GnRH receptors to be down-regulated, resulting in a marked and long-lasting inhibition of LH and FSH secretion

and producing a hypogonadal state in both men and women. The GnRH agonist nafarelin (Synarel) is administered as a nasal spray for the treatment of endometriosis. Suppression of ovarian steroidogenesis and clinical effectiveness are similar to those of danazol.

### GnRH Antagonists

In an effort to avoid the transient androgen surge associated with GnRH agonists, a number of GnRH antagonists have been developed. Abarelix is the first to reach clinical trials. One study has shown that aberelix causes a near-complete suppression of serum testosterone, without a transient increase. Evaluation of the effectiveness of aberelix in treating prostate cancer is under way.

### See Also the Following Articles

Androgen Biosynthesis and Gene Defects • Androgen Insensitivity Syndrome • Gonadotropin-Releasing Hormone (GnRH) Actions • Gynecomastia • Sexual Function and Androgens

### Further Reading

- Bartsch, G., Rittmaster, R. S., and Klocker, D. (2000). Dihydrotestosterone and the concept of 5 $\alpha$ -reductase inhibition in human benign hyperplasia. *Eur. Urol.* **37**, 367–380.
- Chrousos, G. P., Zoumakis, E., and Gravanis, A. (2001). In "Basic and Clinical Pharmacology," (Katzung, B. G., ed.), 8th ed., pp. 634–636 and 704–706. Lange Medical Books/McGraw-Hill, New York.
- Cook, T., and Sheridan, W. P. (2000). Development of GnRH antagonists for prostate cancer: New approaches to treatment. *Oncologist* **5**, 162–168.
- Curtis, L. C. (2001). Organophosphate antagonism of the androgen receptor. *Toxicol. Sci.* **60**, 1–2.
- Grumbach, M. M., and Conte, F. A. (1998). In "Williams' Textbook of Endocrinology," (Wilson, J. D. and Foster, D. W., eds.), 9th ed., p. 859. W. B. Saunders, Philadelphia, PA.
- Mandell, J. (1998). Sexual differentiation: Normal and abnormal. In "Campbell's Urology," (Walsh, P. C., et al., eds.), 7th ed., p. 2147. W. B. Saunders, Philadelphia, PA.
- McLeod, D. G. (1993). Antiandrogenic drugs. *Cancer* **71**, 1046–1049.
- Miller, J. W. (2000). In "Melmon and Morrelli's Clinical Pharmacology," (Carruthers, S. G., Hoffman, B. B., Melmon, K. L., Nierenberg, D. W., eds.), 4th ed., pp. 630–640. McGraw-Hill, New York.
- Taplin, M. E., and Ho, S. M. (2001). The endocrinology of prostate cancer. *J. Clin. Endocrinol. Metab.* **86**, 3467–3477.



# Androgen Insensitivity Syndrome

Yuan-Shan Zhu and Julianne Imperato-McGinley  
Weill Medical College of Cornell University, New York, New York, United States

## Glossary

**androgen receptor** An intracellular protein that specifically binds to androgens and mediates androgen action.

**coregulator** A protein that interacts with nuclear receptors to enhance (coactivators) or reduce transactivation (corepressors) of target genes but does not significantly affect the basal transcription rate.

**dihydrotestosterone** A derivative of testosterone formed by the conversion of testosterone in the action of 5 $\alpha$ -reductase isozymes.

**testosterone** A male hormone that is produced by the testes and induces and maintains male secondary sexual characteristics.

Androgen receptor (AR), a member of the nuclear receptor superfamily and a ligand-dependent nuclear transcription factor, mediates the action of androgens. AR mutations that inactivate its function result in androgen insensitivity, classified as either complete androgen insensitivity syndrome (CAIS) or partial androgen insensitivity syndrome (PAIS). Affected 46,XY individuals with CAIS have female external genitalia, normal female breast development, absent or scanty axillary and pubic hair, and absent female internal genitalia. Testes are present, with levels of plasma testosterone, estradiol, and luteinizing hormone that are high-normal or elevated relative to those of normal males. The plasma dihydrotestosterone (DHT) level is normal or low-normal, which can sometimes result in an elevated testosterone:DHT ratio. Individuals with PAIS have a wide spectrum of phenotypic features, ranging from decreased body hair, infertility, and/or gynecomastia to severe ambiguity of the genitalia. Studies of androgen insensitivity highlight the importance of androgen-AR function in male sexual differentiation and physiology.

## MALE SEXUAL DETERMINATION AND DIFFERENTIATION

Male sexual determination and differentiation is a complex process involving multiple steps, including

testes formation (sex determination) under the control of the *SRY* (sex-determining region of the Y chromosome) gene on the short arm of the Y chromosome. Other autosomal and X-chromosomal genes are also known to be involved in gonadal or testes development.

Before 6 weeks of gestation, human embryos with either a 46,XY (genetic male) or 46,XX (genetic female) karyotype develop identically, and an undifferentiated gonad is present in both genetic male and female fetuses. At approximately 6 or 7 weeks of gestation, testicular cords evolve from the primary sex cords of the indifferent gonad. The Sertoli cells within the cords enlarge, become contiguous, and engulf the germ cells. The seminiferous cords interconnect to form a network of solid cords, which connect with the mesonephric tubules and ultimately to the ductuli efferentes. Leydig cells are apparent by 8 weeks of fetal life and completely fill the interstitial spaces of the developing testes at 3 months of gestation.

Two major hormones, testosterone and anti-Müllerian hormone, are synthesized in the testes and involved in the translation of gonadal sex to phenotypic sex. Anti-Müllerian hormone (also called Müllerian-inhibiting substance or Müllerian-inhibiting factor), a glycoprotein and a member of the transforming growth factor- $\beta$  family, secreted from the Sertoli cells of the testes, promotes regression of the Müllerian ducts, resulting in lack of development of female internal structures (uterus, fallopian tubes, and upper vagina). The Müllerian ducts, which appear at 40–48 days of gestation, regress at approximately 8½ weeks of gestation in the male fetus.

Testosterone secreted by the Leydig cells of the testes beginning at 8 weeks of gestation mediates the differentiation of the Wolffian ducts to the epididymides, vasa deferentia, and seminal vesicles, or the internal masculinization. The process of Wolffian duct differentiation is mediated by testosterone, probably via a paracrine action, and completed by 12 weeks of gestation.

Male external genital differentiation begins soon after Wolffian ductal differentiation. The development of the prostate is dependent on the local conversion of testosterone to the more potent androgen dihydrotestosterone (DHT) by the  $5\alpha$ -reductase-2 isozyme present in these tissues. The urogenital tubercle becomes the glans penis, the urogenital folds become the shaft of the penis, and the urogenital swellings become the scrotum. The urogenital sinus forms the prostate, bulbourethral glands, and the prostatic and membranous portion of the urethra. The entire process of male external sexual differentiation is completed by 14–16 weeks of gestation. Descent of the testes and growth of genitalia occur during the last two trimesters of pregnancy.

The actions of both testosterone and DHT on internal and external genital masculinization are mediated via androgen receptor (AR). Thus, a functional AR is required for normal sexual differentiation. Any defect in the production or action of androgens during these critical periods can result in disorders in male sexual differentiation.

## COMPLETE ANDROGEN INSENSITIVITY SYNDROME

### Clinical Syndrome

Androgen unresponsiveness *in utero* causes a 46,XY fetus with testes and normal androgen secretion to be born with female genitalia and an absent or severely hypoplastic Wolffian ductal system. The labia, especially the labia minora, may be underdeveloped. The clitoris is normal or small. The vagina ends blindly. Due to normal secretion of anti-Müllerian hormone by the testes *in utero*, Müllerian-derived structures are absent or rudimentary, and thus the uterus and cervix are absent or rudimentary.

During puberty, there is normal or augmented breast development due to the unopposed estrogenic action by androgens. Pubic and axillary hair is scant or absent.

The testes of patients with complete androgen insensitivity syndrome (CAIS) are usually located in the abdomen or inguinal canal. They cannot be distinguished histologically from those of normal males before puberty. However, postpubertal histologic studies reveal immature tubular development with Sertoli cells, spermatogonia, and no spermatogenesis. There is frequent clumping of tubules with formation of tubular adenomas. Leydig cells are hyperplastic and electron microscopy reveals ample smooth endoplasmic reticulum and mitochondria with tubular cristae.

This correlates well with the usually elevated plasma testosterone levels, although in some respects Leydig cells have been reported to resemble fetal Leydig cells with absent crystals of Reinke.

### Biochemical Characteristics

In postpubertal individuals with CAIS, plasma luteinizing hormone (LH) is frequently increased with high-normal or elevated testosterone and correlates well with the histological findings of Leydig cell hyperplasia. The elevation of LH apparently results from androgen unresponsiveness in the hypothalamus and/or pituitary. However, LH levels are not in the castrate range due to the negative feedback effect of estrogen on the hypothalamus and/or pituitary. Follicle-stimulating hormone is normal or elevated.

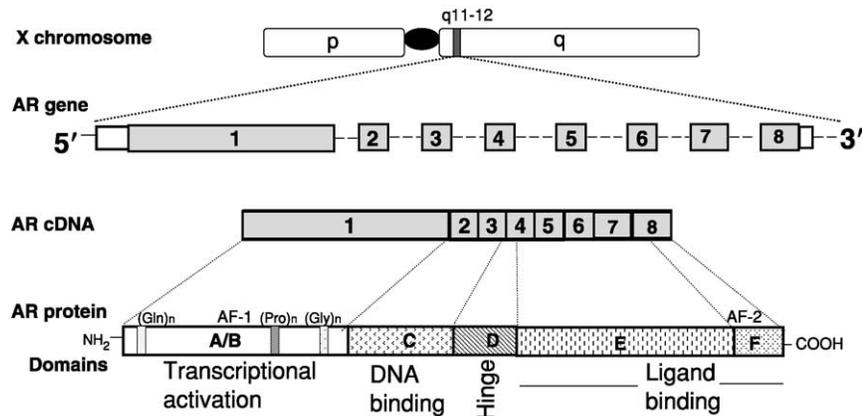
Although the plasma testosterone level is in high-normal or elevated, plasma DHT levels may be low-normal, resulting in an elevated testosterone:DHT ratio. This may be due to a secondary deficiency of  $5\alpha$ -reductase activity since DHT exhibits positive feedforward control of  $5\alpha$ -reductase activity.

Plasma and urinary estrogens range from high male to low female levels. The estrogens originate mainly from the testes and, to a lesser extent, peripheral aromatization of androstenedione and testosterone by aromatase. An unopposed estrogen effect, due to increased estrogen and androgen unresponsiveness, is the likely explanation for breast development during puberty.

The sex hormone binding globulin (SHBG) levels in individuals with CAIS are higher than those of normal males and similar to those of normal females. Castrated CAIS patients not receiving estrogen have SHBG levels similar to those found in normal males.

### Molecular Biology of the Androgen Receptor and Genetic Basis of Androgen Insensitivity Syndrome

The AR, a member of the nuclear steroid receptor superfamily and a ligand-dependent nuclear transcription factor, was cloned in 1988. It has approximately 910–919 amino acids encoded by the AR gene located on Xq11–12 (Fig. 1). The AR gene is a single-copy X-chromosomal gene that spans approximately 90 kilobases of genomic DNA. The encoding region of the AR gene comprises eight exons separated by seven introns. Like other steroid receptors, the AR is a single polypeptide composed of relatively distinct domains: an amino-terminal domain, a DNA-binding



**Figure 1** AR gene location, gene structure, and protein domains. (Top) The location of the AR gene at the q11–12 of the X chromosome. (Middle) The AR gene consists of eight exons (boxes) and seven introns (dashed line). (Bottom) The AR complementary DNA and AR protein, for which the AR domains and the exons encoding each domain are illustrated. Relative positions of glutamine (Gln), proline (Pro), and glycine (Gly) repeats within the N-terminal domain are shown. The transactivation functions AF-1 and AF-2 are located within the N-terminal domain and ligand-binding domain, respectively.

domain, a hinge region, and a steroid-binding domain in the carboxyl terminal.

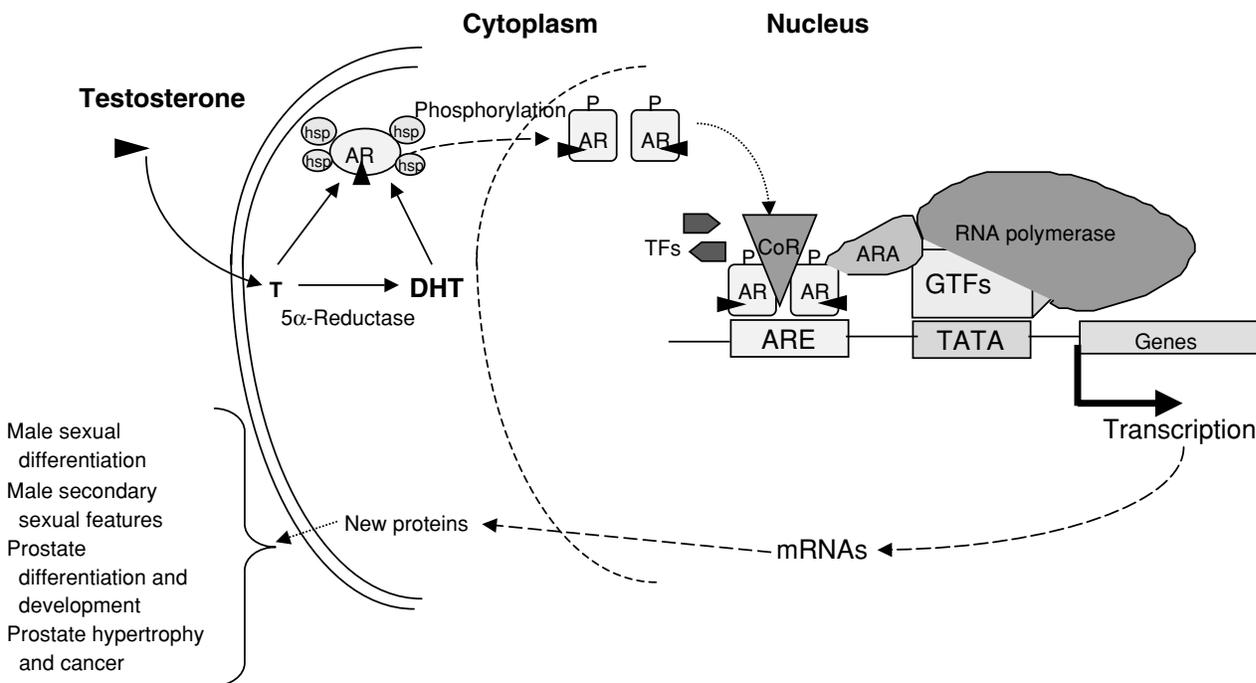
The large amino-terminal domain that comprises approximately half of the AR molecule is encoded by exon 1. It is involved in the transcriptional activation of target genes and contains a transactivation domain, activation function 1 (AF-1). This domain plays a role in AR functions by intramolecular and/or intermolecular interaction with other factors. There are three highly polymorphic direct repeats of amino acid residues: one each containing glutamine, proline, and glycine residues. The increase in size of the glutamine homopolymeric segment is related to the pathogenesis of the spinal and bulbar muscular atrophy (Kennedy's disease).

The DNA-binding domain encoded by exons 2 and 3 contains two “zinc finger” motifs that are hallmarks of all nuclear steroid receptors, and it is the most highly conserved region among steroid receptors. The formation of the two zinc fingers involves eight cysteine residues. These two zinc-coordinated stem-loop structures are responsible for the specific interaction with the cognate DNA of target genes by interacting with the major groove of the DNA duplex. The first zinc finger (proximal to the N-terminal domain) is associated with the determination of the sequence specificity of DNA binding, whereas the second finger helps stabilize the DNA–receptor complex. Despite their exquisite functional specificity in the physiological context, receptors for androgens, glucocorticoids, progesterone, and mineralocorticoids can recognize the same DNA response element both in an *in vitro* binding assay and in functional analysis

using transiently transfection analysis. This paradox remains to be solved.

The carboxyl terminal of the AR is the ligand-binding domain, encoded by the 3' portion of exon 4 and exons 5–8, and is responsible for the specific high-affinity ligand binding. Studies indicate that androgens interact with the ligand-binding domain mostly through hydrophobic interaction as well as some hydrogen bonding. The carboxyl terminal also contains subdomains involved in dimerization and transcriptional activation. The second transactivation function (AF-2) domain of AR resides within the ligand-binding domain. Upon ligand binding, the AF-2 domain can interact with coregulators, such as coactivators, to affect AR function. Between the DNA-binding domain and the steroid-binding domain is the hinge region, which contains the nuclear translocation signal.

Both testosterone and dihydrotestosterone, potent natural androgens, bind to the same AR at the ligand-binding domain to regulate androgen target gene expression. The binding of androgen on the AR results in an AR conformational change that promotes the dissociation of chaperone proteins and facilitates receptor dimerization, nuclear transportation, phosphorylation, and DNA binding. Upon the recruitment of coregulators and general transcription factors, the transcription of a target gene is either induced or inhibited and ultimately leads to a change in androgen target proteins and cellular or biological structures and functions (Fig. 2). A number of coregulators have been identified that can interact with AR to either enhance or reduce androgen–AR action on target



**Figure 2** The molecular events of androgen-AR action in a target cell. When testosterone enters the cell, it can be converted to DHT by the action of 5 $\alpha$ -reductase. Both testosterone (T) and DHT bind to the same AR. The androgen-AR complex translocates to the nucleus, interacting with other factors and target genes to regulate gene transcription. The changes in androgen-target proteins in the cell eventually affect cellular structure and function related to male sexual differentiation, physiology, and pathophysiology. ARE, androgen response element; GTFs, general transcription factors; ARA, androgen receptor-associated proteins; CoR, coregulator; TFs, transcription factors; hsp, heat shock protein.

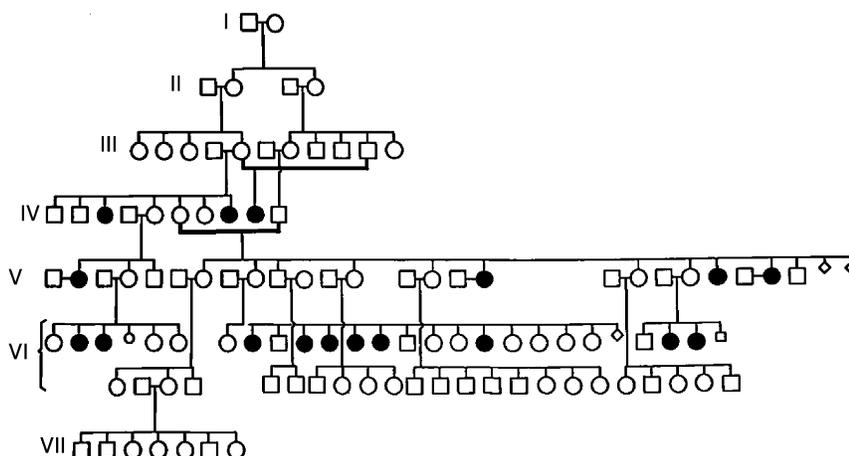
gene transcription. Although significant progress has been made in the past decade in understanding androgen-AR action, the detailed process from androgen binding on AR to the alteration of target gene transcription remains to be elucidated.

The mutations of the AR that cause loss of function result in androgen insensitivity, which is inherited as an X-linked recessive condition, with genetic males expressing the condition (Fig. 3). To date, more than 490 different mutations in all eight exons in the AR gene have been reported, including the mutation for the largest pedigree of CAIS (see the AR gene mutation database at [www.mcgill.ca/androgendb](http://www.mcgill.ca/androgendb)). These mutations range from a single point mutation to an entire gene deletion and can result in various AR dysfunction, including impaired androgen binding, impaired DNA binding, impaired cofactor interaction, blocked formation of a functional receptor (deletions, nonsense mutations, splice-junction alterations), decreased AR expression, and an unstable androgen-receptor complex. Depending on the type of dysfunction, various degrees of functional impairment of the AR occur, resulting in a wide spectrum of symptoms, ranging from a total female phenotype to

normal male phenotype with decreased secondary sexual hair, infertility, or gynecomastia. Although various individual mutations have been characterized, no correlation between the severity of the syndrome and a particular gene defect has been observed.

A variety of genetic defects impair the normal functioning of the AR. Generally, mutations that delete the entire AR gene or interrupt the AR open reading frame, blocking the formation of a functional receptor resulting from premature termination, aberrant splicing, or deletion of partial or complete exon segments, are associated with a phenotype of complete androgen insensitivity. This is due to the fact that AR DNA- and hormone-binding domains, critical for AR function, are located at the carboxyl terminus of the AR protein. As a consequence, defects that truncate the receptor protein at any point during its synthesis result in the removal of a portion of one or both of these important functional domains and will lead to a complete loss of AR function.

Mutations in the ligand-binding domain represent approximately 67% of AR mutations related to androgen insensitivity. The majority of these mutations are single nucleotide substitutions and cause defects in



**Figure 3** Pedigree of a kindred with complete androgen insensitivity indicates the maternal transmission of this disease. ○, normal females; □, normal males; ●, androgen insensitive subjects.

androgen binding that have been categorized as receptor-negative (the absence of detectable ligand binding), receptor-positive (quantitatively normal but qualitatively abnormal ligand binding), and receptor-reduced or receptor-deficient (reduced capacity or reduced affinity of ligand binding) defects, as demonstrated in patient genital skin fibroblasts. Substitution mutations in the AR ligand-binding domain have been identified in patients with the entire spectrum of androgen-insensitive phenotypes.

Approximately 15% of AR mutations are located in the DNA-binding domain, resulting in either complete or partial androgen insensitivity. Studies of these mutations have shown that the mutant receptors bind androgens with normal or near-normal affinity but fail to bind normally to the target DNA sequences by *in vitro* DNA-binding assays, resulting in a defect in androgen-AR function as demonstrated by *in vitro* transfection analyses.

It has been reported that some AR mutations result in decreased AR protein levels and decreased androgen-binding capacity in the genital skin fibroblasts of affected individuals. These mutant ARs may have subtle differences in function on selected target genes when analyzed by *in vitro* transfection. However, the levels of AR in these individuals are significantly decreased due to the mutations that alter AR gene transcription, translation, or posttranslational processes.

AR mutations that alter the stability of the androgen-receptor complex can also cause androgen insensitivity. It has been shown that replacement of arginine 774 of AR with cysteine residue leads to androgen resistance and undetectable androgen binding in genital skin fibroblasts. However, substitution of

the same residue by histidine leads to normal levels of androgen binding in fibroblasts that display a marked thermal lability in *in vitro* assays, suggesting that the stability of the androgen-receptor complex is important for normal AR activity.

A new type of defect has been reported at the postreceptor level in an individual with CAIS as diagnosed based on clinical and biochemical features. The AR gene in this patient is normal. However, the transmission of the activation signal from the AF-1 domain of the AR is disrupted, probably due to a defect in an AR coregulator or coactivator, indicating that defects in AR signal transmission to target gene expression can also result in androgen insensitivity despite the absence of AR mutation.

## Diagnosis and Management

Since the phenotypic appearance of CAIS is totally feminine, the condition is usually diagnosed following breast development at puberty when patients seek medical advice for primary amenorrhea. Complete androgen insensitivity is the most likely diagnosis in a phenotypic female who presents with primary amenorrhea, breast development, scanty or absent pubic and axillary hair, a short vagina, and absent cervix and uterus. These subjects also have a clear, smooth, and acne-free complexion.

Occasionally, individuals with CAIS are diagnosed before or soon after birth. This diagnostic evaluation results from the discrepancy between the findings of a 46,XY karyotype on amniocentesis and the presence of a female phenotype on prenatal ultrasound examination or at birth. Some individuals with CAIS are

diagnosed during infancy when they develop inguinal herniae, often bilaterally. Occasionally, a 46,XY individual with 17 $\alpha$ -hydroxylase deficiency can also have the same phenotypic presentation at puberty.

5 $\alpha$ -Reductase-deficient individuals are most frequently born with severe ambiguity of the external genitalia (pseudovaginal and perineoscrotal hypospadias) and can occasionally present with female genitalia. However, puberty is significantly different from that in androgen-insensitive patients because distinct virilization without gynecomastia occurs. Urinary 5 $\beta$ /5 $\alpha$  C21 steroid metabolite ratios are markedly elevated in 5 $\alpha$ -reductase-2-deficient patients, whereas these are only slightly changed or normal in patients with complete androgen insensitivity.

In summary, the diagnosis of CAIS can usually be made based on the clinical presentation and the biochemical features. Genetic analysis with identification of an AR mutation and functional demonstration of the mutation can assist and confirm the diagnosis.

Once diagnosed, the management of patients with CAIS depends on whether the individuals are prepubertal or postpubertal. CAIS subjects are phenotypic and psychosexual females. It has been reported that up to 25% of patients with CAIS develop testicular neoplasms, although seldom before 25 years of age. If a diagnosis is made before puberty, the patient is usually followed until puberty is completed and breasts have developed naturally before being referred to a surgeon for gonadectomy. However, the optimal time for gonadectomy with regard to psychological adjustment is still a matter of debate. It should be realized that waiting for the postpubertal period to perform the gonadectomy means that it will often be done during the tumultuous years of adolescence or soon thereafter. Gonadectomy performed at this time, however, has the advantage of involving adolescents or young adults in decision making.

It can be argued that gonadectomy should be performed in childhood, thus avoiding the difficulties of yet another psychologic adjustment during adolescence. However, some adult patients with this condition, who were gonadectomized in childhood, believe that gonadectomy should have been delayed until early adulthood, when they could have been more actively involved in decision making; thus making the psychologic adjustment easier. If a child has an inguinal hernia and is brought to a physician for this reason and the diagnosis is made, consideration should be given to perform gonadectomy at the time of hernia repair, thereby avoiding a second operation.

It is important that the patient learn about the condition from his or her doctor in a careful and

sensitive way. This will avoid severe psychologic problems. Therefore, it is critical that the doctor develop a good relationship with the patient and the parents. The timing of the information from initiation of the topic to a detailed discussion must be individualized. Psychologic counseling is needed for the patient and the family.

The postpubescent patient should be gonadectomized due to the probability of testicular neoplasms. Cyclic estrogen replacement therapy should be prescribed at the appropriate time to prevent osteoporosis and to maintain breast development, and patients should be followed regularly. For most patients, the vagina is of sufficient size to allow normal coital function. In a patient whose vagina is too small, vaginal dilation is the usual corrective measure, and vaginal surgery is rarely required.

## PARTIAL (INCOMPLETE) ANDROGEN INSENSITIVITY SYNDROME

In partial or incomplete forms of androgen insensitivity, a spectrum of clinical phenotypes is present and includes gynecomastia and severely ambiguous genitalia, mild hypospadias, gynecomastia only, normal male genitalia with infertility, and decreased body hair in adulthood. A range of phenotypic abnormalities have been reported in 11 members of one family, indicating that a single mutant gene, variably expressed, may be a factor in the variant phenotypic forms of partial androgen insensitivity syndrome (PAIS).

The endocrine profile, analogous to that in patients with CAIS, shows elevated LH and testosterone levels. Total 17 $\beta$ -estradiol produced and secreted by the testes has been reported to be greater than in patients with CAIS. However, the degree of feminization during puberty is not as prominent. This may indicate a less severe androgen and estrogen imbalance at the cellular level due to some androgen responsiveness.

Genetic analysis has revealed a variety of AR mutations responsible for PAIS. These mutations occur throughout the entire encoding region and result in various AR dysfunctions. There is no correlation between a specific AR gene mutation and phenotypic expression, although numerous mutations have been identified and characterized.

Treatment of patients with PAIS is dependent on the degree of virilization. In most cases, patients with profound genital ambiguity should be raised as females since deficient masculinization and gynecomastia will undoubtedly occur at puberty. Surgical correction of external genitalia and gonadectomy are indicated and female sex hormone therapy should be prescribed at

puberty. Surgical correction of mild hypospadias is necessary in patients raised as males, and surgical correction of gynecomastia may be necessary.

Supplemental high-dose androgen therapy to increase virilization is controversial due to its possible deleterious effect on the cardiovascular system. However, due to the fact that some AR mutations display conditional androgen-binding defects, the use or development of specific androgen analogs that can overcome the androgen-binding defects manifest in natural androgens is a therapeutic strategy in the management of PAIS that remains to be investigated.

## SUMMARY

Androgen insensitivity syndrome due to AR mutations is a natural model for elucidating androgen actions in male sexual development, physiology, and pathophysiology. The identification and characterization of various AR mutations provide important information for understanding AR structure and function. The elucidation of AR–coregulator interaction in androgen action and the identification of an AR–coregulator defect in androgen insensitivity further indicate the complexity of androgen–AR action and open the door to a new therapeutic strategy.

## Acknowledgments

This study was supported in part by National Institutes of Health Grant M01-RR-00047 (General Clinical Research Center) and a Merck Foundation fellowship.

## See Also the Following Articles

Adrenal Androgens • Androgen Biosynthesis and Gene Defects • Androgen Biosynthesis Inhibitors and Androgen

Receptor Antagonists • Androgens, Gender and Brain Differentiation • Anti-Müllerian Hormone • Endocrine Disrupters and Male Sexual Differentiation • Gender Assignment and Psychosocial Management • Genes and Gene Defects Affecting Gonadal Development and Sex Determination • Gynecomastia • Testes, Embryology of

## Further Reading

- Adachi, M., Takayanagi, R., Tomura, A., Imasaki, K., Kato, S., Goto, K., Yanase, T., Ikuyama, S., and Nawata, H. (2000). Androgen-insensitivity syndrome as a possible coactivator disease. *N. Engl. J. Med.* **343**(12), 856–862.
- Imperato-McGinley, J., and Canovatchel, W. J. (1992). Complete androgen insensitivity—Pathophysiology, diagnosis and management. *Trends Endocrinol. Metab.* **3**(3), 75–81.
- Imperato-McGinley, J., and Zhu, Y. S. (2002). Gender and behavior in subjects with genetic defects in male sexual differentiation. In “Hormones, Brain and Behavior” (D. W. Pfaff *et al.*, eds.), Vol. 5, pp. 303–345. Academic Press, San Diego.
- Imperato-McGinley, J., Peterson, R. E., Gautier, T., Cooper, G., Danner, R., Arthur, A., Morris, P. L., Sweeney, W. J., and Shackleton, C. H. L. (1982). Hormonal evaluation of a large kindred with complete androgen insensitivity: Evidence for secondary 5 $\alpha$ -reductase deficiency. *J. Clin. Endocrinol. Metab.* **54**(5), 931–941.
- Patterson, M. N., McPhaul, M. J., and Hughes, I. A. (1994). Androgen insensitivity syndrome. *Baillieres Clin. Endocrinol. Metab.* **8**(2), 379–404.
- Quigley, C. A., DeBellis, A., Marschke, K. B., El-Awady, M. K., Wilson, E. M., and French, F. S. (1995). Androgen receptor defects: Historical, clinical, and molecular perspectives. *Endocrine Rev.* **16**(3), 271–321.
- Sultan, C., Lumbroso, S., Paris, F., Jeandel, C., Terouanne, B., Belon, C., Audran, F., Poujol, N., Georget, V., Gobinet, J., Jalaguier, S., Auzou, G., and Nicolas, J. C. (2002). Disorders of androgen action. *Semin. Reprod. Med.* **20**(3), 217–228.
- Zhu, Y. S., Cai, L. Q., Cordero, J. J., Canovatchel, W. J., Katz, M. D., and Imperato-McGinley, J. (1999). A novel mutation in the CAG triplet region of exon 1 of androgen receptor gene causes complete androgen insensitivity syndrome in a large kindred. *J. Clin. Endocrinol. Metab.* **84**, 1590–1594.



# Androgens, Gender and Brain Differentiation

Julianne Imperato-McGinley and Yuan-Shan Zhu  
Weill Medical College of Cornell University, New York, New York, United States

## Glossary

**androgen receptor** An intracellular protein that binds to androgens and mediates androgen action.

**dihydrotestosterone** A derivative of testosterone formed by the action of 5 $\alpha$ -reductase isozymes that is involved in male sexual differentiation and development.

**5 $\alpha$ -reductase isozymes** Microsomal nicotinamide adenine dinucleotide phosphate (NADPH)-dependent proteins that reduce the double bond at the 4–5 position of a variety of C<sub>19</sub> and C<sub>21</sub> steroids such as testosterone.

**gender identity** The sense of being male or female; the self-awareness of knowing one's sex.

**gender role** The expression of one's gender identity to the public; manifested by one's behavior or actions.

**male pseudohemaphrodite** A 46,XY individual who has the testicular tissue but whose external genitalia and other sexual characteristics are so ambiguously developed that the sex of the individual is uncertain.

**testosterone** A male hormone that is produced by the testes and that induces and maintains male secondary sexual characteristics.

Genetic and hormonal influences on phenotypic sex determination in humans have been shown to be similar to those in other mammals. Many animal studies have also demonstrated that hormones are essential for sexual differentiation of the brain during development and for the maintenance of sexually dimorphic behavior throughout life. However, the effect of hormonal influences on sexual dimorphic differences in the human nervous system and on gender identity and sex differences in human behavior is still an emerging field. This article discusses the importance of androgens in (1) determination of male gender identity and (2) cognitive function by reviewing specific inherited genetic defects in androgen biosynthesis and action. The complex interaction of nature versus nurture is also addressed.

## ROLE OF ANDROGENS IN BRAIN MORPHOLOGY

Studies of brain morphology of various animals demonstrate the presence of sexual dimorphism. Androgens administered perinatally to female rats result in structural changes in the sexually dimorphic nucleus of the preoptic area of the brain, making its similar to that of male rats. In canaries (*Serinus canarius*) and zebra finches (*Poephila guttata*), areas in the brain that control the vocal cords are noticeably larger in males than in females. The area X of the lobus parolfactorius is well developed in males of both species but is hardly identifiable in females. These size differences correlate with differences in singing ability; males produce a complex song, whereas females do not normally sing at all (zebra finches) or sing an infrequent simple song (canaries). The influence of testosterone at a critical period in development induces enhancement of these song areas of the brain in female canaries and zebra finches. The action of testosterone on the song areas of the brain may be mediated by conversion to estradiol given that it has been shown that masculine patterns of song area development in genetic females can be induced with estrogen. However, feminine patterns cannot be reproduced in males with antiestrogens or inhibitors of estrogen synthesis. An unresolved issue involves studies of genetic females that have functional testicular tissue and virtually no ovarian tissue but that still have feminine song circuitry.

Morphological sex differences that can be induced by androgens are also present in the auditory system of the tree frog (*Eleuthero dactylus coqui*), the spinal cord of the cat, the brain of the juvenile macaque monkey, and the neurons innervating the bulbocavernosus and levator ani muscles as well as the neurons innervating the ischiocavernosus and levator ani muscles (dorsolateral motor nucleus [DLN]) of the rat.

Studies of the human brain have found that the volume of the nuclei in the preoptic area is  $2.5 \pm 0.6$

times larger in men, containing approximately twice as many cells. One area (INH-3) is larger by a factor of three in the male compared with the female. Another cell group is twice as large in the male brain, varies in women according to hormone levels, and is 3.7-fold larger in women of childbearing age. However, it is still not clear whether these nuclei are subdivisions of the paraventricular nucleus or are separate anatomical entities.

Sex differences in areas such as the left cortical language regions, as well as in the shape and surface area of the human callosum, have been reported in some studies but not in others. Because the callosum consists of myelinated connecting fibers, larger callosal volumes in women are interpreted as providing better interhemispheric communication, resulting in less functional specialization of the two hemispheres.

## ROLE OF ANDROGENS IN MALE BEHAVIOR

In animal studies, androgens administered early in females have been shown to organize neural systems in such a way as to stimulate or induce male sexual response during adulthood and to inhibit female sexual response. The critical period for this central nervous system effect of androgens might not coincide with the critical period for external genital differentiation. In some animals, pre- or perinatal androgen stimulation also affects nonsexual masculine patterns of behavior that are independent of hormonal exposure during adulthood. Comparable effects have been seen in animal models such as the guinea pig, rabbit, hamster, rat, dog, and rhesus monkey. The critical period of exposure (pre- and/or perinatal) differs among the various species. In offspring of pregnant rhesus monkeys, treatment with androgens during a specific period of gestation increases mounting and "rough and tumble" play in females. Thus, sexually dimorphic behavior in animals is secondary to sex steroid-induced differentiation and activation of the brain at critical periods.

## GENDER IDENTITY DEVELOPMENT

The relative influence of hormonal versus environmental factors in the determination of gender identity has been an ongoing debate for many years.

### Environmental Influence

In 1955, Money proposed that human sexuality was undifferentiated at birth and becomes differentiated as

a consequence of various experiences. This theory was revised later to acknowledge that male sexually dimorphic behavior is expressed at birth but that it can be incorporated into either a male or female gender identity pattern. This theory was tested by matching individuals with ambiguous genitalia that were "chromosomally, gonadally, and otherwise diagnostically the same." The "matched pairs" were reported as differing only in their sex assignment and sex of rearing. The results demonstrated that gender identity correlated with sex of rearing, and not with chromosomal or gonadal sex, leading to the conclusion that sex of rearing was predominant in establishing gender identity. However, at the time of the studies, adequate hormone evaluation was unavailable and individuals were not assessed for their hormonal environment. Therefore, the degree of androgen exposure was assumed to be similar for the phenotypically matched pairs but was not objectively known to be so. Consequently, the issue of nature (androgen imprinting) versus nurture (sex of rearing) was not resolved due to incomplete hormonal characterization of the individuals.

Individuals with inadequate testosterone production or action are not suitable models to assess the relative importance of environment (sex of rearing) versus nature (androgen imprinting) in determining male gender identity because deficient androgen exposure would result in a decrease in the androgen effect on the brain. In addition, when castration and sex hormone therapy are initiated in an individual born with ambiguous genitalia to coincide with the determined sex of rearing, the natural hormonal sequence of events in the evolution of gender is interrupted. As a consequence, no valid conclusions can be drawn about the relative importance of these factors.

## Hormonal Influence

### **Gender Identity in Research Subjects with 5 $\alpha$ -Reductase-2 Deficiency**

In research subjects with 5 $\alpha$ -reductase-2 deficiency, the unique biochemical defect affects male external genital development without altering testosterone response. In these individuals, the biosynthesis of testosterone and its peripheral action are normal; thus, prenatal and neonatal testosterone exposure of the brain proceeds as in the normal male. However, deficient 5 $\alpha$ -reductase-2 enzyme activity impairs conversion of testosterone to dihydrotestosterone (DHT), and the deficiency *in utero* results in genital ambiguity. As a result, many affected individuals with

this condition are believed to be female at birth and are raised as girls. However, at puberty, significant virilization occurs under the influence of normal to elevated plasma testosterone levels and a gender change occurs. Thus, 5 $\alpha$ -reductase-2-deficient individuals provide a unique opportunity to evaluate the relative influences of nature (testosterone) versus environment (sex of rearing) in the determination of gender identity in humans.

In the affected subjects from the Dominican kindred, those who were raised as females began to realize that they were different from other girls in the village because they did not develop breasts and/or they felt masses in either the inguinal canal or the "labia." They passed through a number of stages that included no longer feeling like girls, feeling like men, and finally identifying as men. When they became convinced of their maleness, a change in gender role occurred either during puberty or during the postpubertal period. In some individuals, the gender role change to male was delayed until they were confident of their ability to defend themselves. The average age of the gender role change was 16 years.

With one exception, those who were raised as girls changed to a male gender role and perform male work in a society where there is a definite division of labor; women perform household activities, whereas men work as farmers, miners, or woodsmen. Females are the affected males' choice for sexual intimacy. The adequacy of sexual intercourse depends on the severity of the chordee and the size of the phallus. With one exception, all of the subjects who changed to a male gender role either live or have lived with women in common law marriages. Some chose women with children from previous unions. One man lives alone in the hills working as a farmer since adopting a male gender role at 20 years of age.

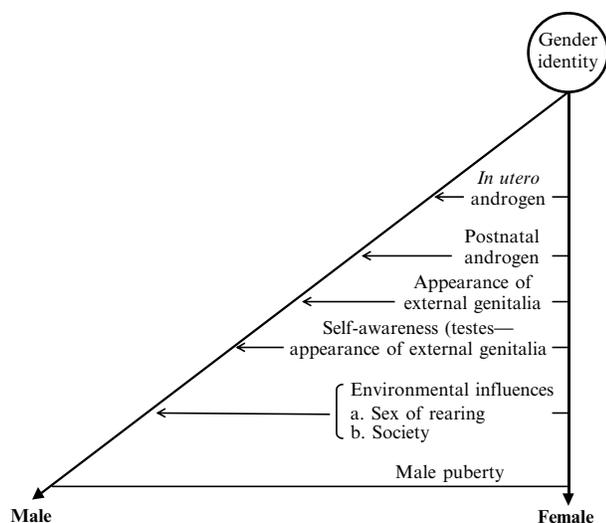
The social and psychosexual development of New Guinean research subjects of the Sambian tribe with 5 $\alpha$ -reductase-2 deficiency has been recorded in field studies of Carlton Gajdusek over decades with similar observations. The Sambian tribal culture of the eastern highlands is rigidly gender segregated. Women are caretakers, whereas men are hunters and warriors. Male initiation rites during puberty include ritualized fellatio and other rituals for men of premarital age. Gender segregation during pubertal initiation was allegedly strictly enforced and included killing any female who accidentally witnessed these events. In the past, New Guinean 5 $\alpha$ -reductase-2-deficient individuals who were raised as girls until puberty, when they made the transition to boys, were reported. But today, as in the Dominican kindred, the condition is usually

recognized at birth by experienced midwives or is recognized during childhood. The female-to-male gender change produced the term "turnim man" from the Melanesian pidgin and was incorporated into the Sambian lexicon, connoting that these individuals are innately and biologically driven to change gender.

Some believe that New Guinean research subjects with 5 $\alpha$ -reductase-2 deficiency, as well as others with this condition, are regarded as a "third sex" and change to a male gender role to adapt to their "male-admiring" society. It has been our experience, as well as Gajdusek's experience, that these individuals clearly regard themselves as male. Three affected New Guinean 5 $\alpha$ -reductase-2 patients specifically requested and obtained genital correction during adulthood so that they could be, in their words, "complete men." In addition, gender change in 5 $\alpha$ -reductase-2 deficiency has been noted in affected individuals from many other countries such as Brazil, Turkey, Mexico, Cyprus, Algeria, Italy, Lebanon, Pakistan, Saudi Arabia, and Sweden as well as in Dominican individuals not part of the large kindred described previously. When puberty is allowed to occur without medical intervention, the majority of individuals identify as male.

It can be theorized that in normal males, the sex of rearing and androgen imprinting of the brain act in concert to determine male gender expression. Studies of 5 $\alpha$ -reductase-2-deficient individuals have shown that when the sex of rearing (female) is discordant with the testosterone-mediated biological sex, the biological sex will prevail when puberty occurs in a non-intervening environment. Thus, under the influence of testosterone-mediated puberty, "masculinization" of the brain theoretically occurs and a male gender identity develops, overriding the female sex of rearing. Therefore, androgens appear to act as inducers and activators in the evolution of male gender identity in humans. It is unclear whether it is testosterone or DHT that directly mediates this process given that both type 1 and type 2 5 $\alpha$ -reductase isozymes are present in the brain, and that the type 1 isozyme is normal in individuals with 5 $\alpha$ -reductase-2 deficiency.

Gender identity has been postulated to become fixed at around the time of language development, between 18 months and 4 years of age. Studies of individuals with 5 $\alpha$ -reductase-2 deficiency appear to demonstrate that the development of gender identity is not fixed during early life but rather evolves throughout childhood and becomes fixed during or after puberty.



**Figure 1** Schematic illustration of the critical factors involved in the evolution of a male gender identity in humans.

In summary, environmental and/or sociocultural factors are not the sole factors responsible for the formation of male gender identity; androgens make a strong and definite contribution (Fig. 1). Analogous to the induction of a male phenotype from an inherent female phenotype, the formation of male gender identity in humans appears to be at least partially induced by androgens from an undifferentiated and/or inherently female nervous system.

### **Gender Identity in Research Subjects with 17 $\beta$ -Hydroxysteroid Dehydrogenase-3 Deficiency**

Genetic male research subjects with 17 $\beta$ -hydroxysteroid dehydrogenase-3 (17 $\beta$ -HSD-3) deficiency are born with severe ambiguity of the genitalia due to deficiency in conversion of androstenedione to testosterone. At puberty, they develop male secondary sexual characteristics (phallic enlargement and abundant body and facial hair), consequent to the peripheral conversion of androstenedione to testosterone by other 17 $\beta$ -HSD isozymes. Gender change from female to male has been reported in subjects with this condition, including those from a large Arab kindred in the Gaza Strip extending over eight generations. From published cases with 17 $\beta$ -HSD-3 deficiency, at least half of the research subjects raised as female change their gender from female to male spontaneously or in consultation with a physician and/or psychiatrist at various ages.

In the Arab kindred, some individuals raised as girls exhibited aggressive behavior, leading to their dismissal from school. Those who changed gender

did so on their own initiative, and some did so without parental consent or psychiatric help. The affected individuals are capable of having erections with ejaculations. Three individuals were castrated after diagnosis at the decision of the physicians, and they were raised as women. It is interesting that no individuals from this kindred living as females are known to have married. Some have stated that they would rather have been raised as males, revealing a questionable female identity. One elderly individual was found to have a female gender role but apparently identifies as male. This individual was a farm laborer and was proud to be stronger and more productive than male colleagues.

Conversion of androstenedione to testosterone is possible in the human brain. Except for the 17 $\beta$ -HSD-3 isozyme that is deficient in these research subjects, other 17 $\beta$ -HSD isozymes are normally expressed in the brains of both humans and animals. Thus, alternate pathways for testosterone and/or DHT formation via other 17 $\beta$ -HSD isozymes are in the brain as well as in extragonadal tissues of patients with 17 $\beta$ -HSD-3 deficiency and can theoretically cause androgen imprinting of the brain.

These studies suggest that hormones play a significant role in gender identity formation and that both environmental and hormonal factors influence male gender identity formation in humans.

### **SEX DIMORPHISM IN COGNITIVE FUNCTION AND LATERALITY**

Meta-analyses of sex differences in spatial abilities have shown that spatial abilities of males are consistently greater than those of females on certain tasks. A number of tests showed significant male advantages that were stable across age and have not decreased during recent years. Tests included mental rotation tasks, the primary mental abilities (PMA) spatial relations subtest, and the rod and frame test. Male superiority in mathematics performance on the SAT has remained constant, and on a computerized version of a task of spatial ability, males outperform females across a variety of measures. On functional magnetic resonance imaging (fMRI) using a three-dimensional maze, sexually dimorphic differences in areas of brain activation have also been noted.

Women, on average, have slightly better verbal skills than do men. When all language measures are combined, women excel in tests of speech production and verbal fluency. Using echo-planer fMRI, a sex difference in brain organization was found during letter recognition, rhyme, and lexical-semantic tasks.

There is a trend toward greater right ear superiority in men. Auditory testing of laterality produces the most robust effect, with dichotic consonant–vowel (CV) syllable pairs tasks being the most reliable approach to testing of this type. From the past emerging data, it seems reasonable to suggest that the sex differences described previously are linked to differences in functional organization of the brain.

## COGNITIVE ABILITIES IN ANDROGEN-INSENSITIVE INDIVIDUALS

Individuals with complete androgen insensitivity are phenotypic and psychosexual females due to the lack of androgen response consequent to androgen receptor gene mutations. Individuals with complete androgen insensitivity were studied using the Spanish version of the Wechsler Intelligence Scale (EIWA) to evaluate the relationship between androgen unresponsiveness and cognitive abilities, with particular attention to subtests of visual–spatial ability. In this study, affected individuals, as well as control males and females, all are members of a large kindred with this condition. Subjects with androgen insensitivity demonstrated significantly lower scores on subtests of spatial ability than did either control males or females from the kindred.

Because androgen-insensitive subjects are raised as females having a totally female psychosexual orientation, it should be considered that their cognitive performance may reflect the influence of their sex roles as a factor of socialization. However, this consideration does not explain their significantly lower overall performance on the perceptual organization factor and subtests of spatial ability when compared with control females from the same kindred. This exaggerated female pattern of performance suggests an effect of androgen unresponsiveness.

Studies of individuals with complete androgen insensitivity by other investigators demonstrate a modest but consistent and significant tendency toward superiority of verbal abilities over space–form abilities using the Wechsler Adult Intelligence Scale. Performance–perceptual scores are poorer in affected individuals than in both male and female controls.

## CONCLUSION

Over the past two decades, studies of natural human genetic models with deficiency in androgen production or action have demonstrated the importance of

androgens in male sexual differentiation, development of male gender identity/role, and male-patterned behavior and cognitive function. These studies have also revealed that androgen, mainly testosterone, plays a vital “imprinting” role during a critical period of development. These actions appear to be mediated via androgen interacting with the androgen receptor, not via conversion of androgen to estrogen interacting with estrogen receptors as demonstrated in animal studies. This is further supported by the fact that 46,XY individuals with an estrogen receptor mutation or aromatase mutation, resulting in deficiency in estrogen production and action, develop a male gender identity and role. The fact that the majority of research subjects with 5 $\alpha$ -reductase-2 deficiency or 17 $\beta$ -HSD-3 deficiency, who were raised as females, changed their gender identity and role to male during or after puberty suggests that the development of gender identity evolves throughout childhood and is fluid until the events of puberty.

It is important to bear in mind that androgen acts together with social or environmental factors within the endocrine milieu in the development of the male gender identity/role (Fig. 1). The critical period of androgen exposure of the brain for this development remains uncertain.

## See Also the Following Articles

Adrenal Androgens • Androgen Biosynthesis and Gene Defects • Androgen Biosynthesis Inhibitors and Androgen Receptor Antagonists • Androgen Insensitivity Syndrome • Endocrine Disrupters and Male Sexual Differentiation • Estrogen and the Male • Germ Cell Differentiation Signaling Events, Male • Hyperandrogenism, Gestational • Müllerian Inhibiting Substance: New Insights • Pseudohermaphroditism, Male, Due to 5 $\alpha$ -Reductase-2 Deficiency • Sexual Function and Androgens

## Further Reading

- Arnold, A. P. (1997). Sexual differentiation of the zebra finch song system: Positive evidence, negative evidence, null hypotheses, and a paradigm shift. *J. Neurobiol.* **33**, 572–584.
- Goy, R. W., and McEwen, B. S. (1980). “Sexual Differentiation of the Brain.” MIT Press, Cambridge, MA.
- Imperato-McGinley, J., Peterson, R. E., Gautier, T., and Sturla, E. (1979). Androgens and the evolution of male-gender identity among male pseudohermaphrodites with 5 $\alpha$ -reductase deficiency. *N. Engl. J. Med.* **300**, 1233–1237.
- Imperato-McGinley, J., Peterson, R. E., Stoller, R., and Goodwin, W. E. (1979). Male pseudohermaphroditism secondary to 17 $\beta$ -hydroxysteroid dehydrogenase deficiency: Gender role change with puberty. *J. Clin. Endocrinol. Metab.* **49**, 391–395.
- Imperato-McGinley, J., Pichardo, M., Gautier, T., Voyer, D., and Bryden, M. P. (1991). Cognitive abilities in androgen-insensitive

- subjects: Comparison with control males and females from the same kindred. *Clin. Endocrinol. (Oxf.)* **34**, 341–347.
- Imperato-McGinley, J., and Zhu, Y. S. (2002). Gender and behavior in subjects with genetic defects in male sexual differentiation. In “Hormones, Brain, and Behavior” (D. W. Pfaff, A. P. Arnold, A. M. Etgen, S. E. Fahrbach, and R. T. Rubin, eds.), vol. 5, pp. 303–345. Academic Press, Orlando, FL.
- Poletti, A., Coscarella, A., Negri-Cesi, P., Colciago, A., Celotti, F., and Martini, L. (1998).  $5\alpha$ -reductase isozymes in the central nervous system. *Steroids* **63**, 246–251.
- Rosler, A., and Kohn, G. (1983). Male pseudohermaphroditism due to  $17\beta$ -hydroxysteroid dehydrogenase deficiency: Studies on the natural history of the defect and effect of androgens on gender role. *J. Steroid Biochem. Mol. Biol.* **19**, 663–674.
- Shaywitz, B. A., Shaywitz, S. E., Pugh, K. R., et al. (1995). Sex differences in the functional organization of the brain for language. *Nature* **373**, 607–609.
- Voyer, D., Voyer, S., and Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychol. Bull.* **117**, 250–270.



ANF

see Atrial Natriuretic Factor



# Angiogenesis

Andreas Bikfalvi

University of Bordeaux I, Bordeaux, France

## Glossary

**angiogenesis** In a strict sense, the formation of vessels from preexisting vessels. Generally, the development of new vessels.

**angiopoietins** Factors implicated in the recruitment of accessory cells to the vasculature and vessel remodeling.

**arteriogenesis** Formation of larger blood vessels including arterioles.

**endothelial cells** Major cell type that is targeted by angiogenesis factors or inhibitors. Endothelial cells are in close contact with the blood and organize into tubular structures.

**fibroblast growth factors** Pleiotropic growth factors that also induce angiogenesis.

**hypoxia-inducible transcription factors (HIFs)** There are two HIFs, HIF-1 $\alpha$  and HIF-2 $\alpha$ , that regulate genes involved in angiogenesis.

**integrins** Heterodimeric cell surface receptors for matrix molecules.

**lymphangiogenesis** Formation of lymphatic vessels.

**mural cells** Cells lining the vessel wall in close contact with the subendothelial matrix and endothelial basement membranes. Pericytes are among the principal mural cells and are found in capillaries.

**platelet-derived growth factors** Growth factors implicated in the recruitment of accessory cells (mainly pericytes) to the vasculature.

**proteolytic enzymes** Enzymes that degrade proteins such as extracellular matrix molecules. The major proteolytic systems for the degradation of matrix molecules are the plasmin/plasminogen activator–inhibitor system and matrix metalloproteinases.

**pruning** Process that yields a mature remodeled vascular network.

**RIP-Tag mouse model** Transgenic mouse model with targeted expression of large T antigen in pancreatic  $\beta$  cells. A model for multistage carcinogenesis.

**vascular endothelial growth factors (VEGFs)** The major regulators of angiogenesis.

**vasculogenesis** Formation of blood vessels from progenitor cells.

The formation of vascular channels, angiogenesis, is a fundamental process that takes place during the embryonic life and also plays a crucial role in the adult organism.

## INTRODUCTION

Generally, angiogenesis refers to two different mechanisms of vasoformation and types of vessels. Angiogenesis, in a strict sense, describes the formation of new vessels from preexisting functional vessels. On the other hand, vasculogenesis involves the differentiation of endothelial progenitor cells that are incorporated subsequently into vessels. Until recently, it was believed that angiogenesis occurs in the embryo and the adult organism, whereas vasculogenesis occurs

only during embryonic development. This view has been challenged by the identification of circulating endothelial progenitor cells in the adult that are thought to arise in the bone marrow. These circulating progenitors are able to actively participate in angiogenesis processes, such as tissue ischemia, tumor angiogenesis, or ocular neovascularization.

In the past, angiogenesis researchers mainly studied the mechanisms of formation of blood vessels. However, lymphatic vessels have become a focus of intense research. The study of arteriogenesis has also broadened the angiogenesis field. Arteriogenesis describes the formation of larger functional blood vessels—a process that is particularly important in the formation of collaterals that limit tissue ischemia.

This article provides a brief overview of the mechanisms of angiogenesis and the impact of angiogenesis research on the pathophysiology and therapy of disease.

## BASIC MECHANISMS OF VASOFORMATION

Vasoformation is dependent on molecular regulations in both healthy and pathological tissue. This process is dependent on paracrine angiogenesis signals that induce proliferation and migration of vascular cells and their assembly into functional vessels. Vessel stabilization and remodeling are very important events that occur at a later stage. In microvessels, two major cell types, endothelial cells and pericytes, participate in these processes. In larger vessels, smooth muscle cells are involved instead of pericytes. A number of soluble factors, receptors, and extracellular matrix molecules play a role in vascular morphogenesis. Expression of these factors is under the control of a molecular switch inside the cells.

Hypoxia is a driving force for angiogenesis in tumors or ischemic tissue (Fig. 1). Hypoxia regulates angiogenesis via an increase in hypoxia-inducible transcription factor-1 $\alpha$  (HIF-1 $\alpha$ ) that initiates a program of survival and adaptive gene expression. In angiogenesis, the major factor regulated through the HIF-1 $\alpha$  system is vascular endothelial growth factor (VEGF). In the presence of oxygen, the enzyme prolyl 4-hydroxylase (PHD) binds molecular oxygen and hydroxylates proline residues in HIF-1 $\alpha$ . Hydroxylated HIF-1 $\alpha$  associates with the von Hippel-Lindau (VHL) gene product, passes to the proteasome,

and is rapidly degraded. Under hypoxic conditions, hydroxylation is inhibited and HIF-1 $\alpha$  levels are stabilized. There are three sequence-related PHDs. Probably only one of these forms, PHD2, which resides in the cytoplasm, is involved in HIF-1 $\alpha$  regulation during angiogenesis. PHD2 is also transcriptionally induced by HIF-1 $\alpha$  in a low-oxygen environment. This provides an autoregulatory feedback loop.

In vascular cells, another form of HIF, HIF-2 $\alpha$ , is responsible for regulating gene expression. For example, an important receptor that binds VEGF, VEGFR2, is up-regulated through a HIF-2 $\alpha$ -dependent mechanism.

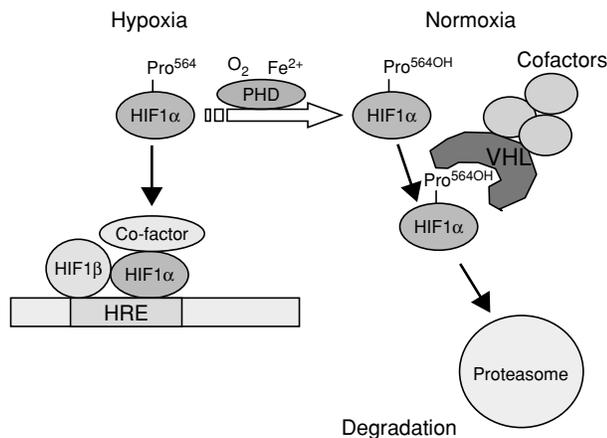
The molecular angiogenesis switch is dependent on hypoxia, oncogenic transformation, which occurs in tumor cells, or on autocrine growth factor loops. For example, activation of the ras gene product induces the expression of VEGF, a stimulator of angiogenesis, and down-regulates inhibitors such as thrombospondins in tumor cells. Similar effects are also observed when autocrine growth factor loops are present in tumor cells.

## KEY FACTORS REGULATING VASCULAR DEVELOPMENT

Among the most important regulators of angiogenesis are the VEGFs. This family is composed of VEGF-A, -B, -C, and -D and the related placental growth factors (PLGFs). VEGFs are essential in embryonic and postnatal vascular development. They also play an important role in ischemia-driven or tumor angiogenesis. In fact, one of the VEGF prototypes, VEGF-A, is a permissive factor for multistage carcinogenesis, as evidenced in the RIP-Tag mouse model.

VEGFs bind three types of tyrosine-kinase receptors: VEGFR1 (flt1), VEGFR2 (flk1 or KDR), and VEGFR3 (flt3). VEGF-A binds VEGFR2 and VEGFR1. In contrast, VEGF-B binds only VEGFR1. VEGF-C and VEGF-D both preferentially bind VEGFR3 but also interact with VEGFR2. Finally, PLGFs bind only VEGFR1.

VEGFR2 is the critical receptor for angiogenesis in blood vessels and for vascular permeability. VEGFR1 may synergize with VEGFR2 in postnatal and pathological angiogenesis. VEGFR1 is also found on hematopoietic cells that accumulate at angiogenic sites. However, only VEGFR2 seems to be necessary for embryonic vascular development. Nevertheless, both receptors are needed for repair-associated, tumor, or retinal neoangiogenesis in the adult.



**Figure 1** The oxygen-sensing pathway. HIF1 $\alpha$ , hypoxia-inducible factor-1 $\alpha$ ; HIF1 $\beta$ , hypoxia-inducible factor-1 $\beta$ ; HRE, hypoxia response element; VHL, von Hippel-Lindau.

VEGFR3/flt4 is required for the growth and the maintenance of the lymphatic vessels and is also mutated in primary human lymphedema (Milroy's disease), which illustrates the importance of this receptor in the lymphatic tissue. Targeted inactivation of VEGFR3 in mice results in abnormal lymphatic vessel growth. VEGF prototypes also use coreceptors for binding to target cells. Among the coreceptors, neuropillins seem to be the most important. For instance, neuropillin-1 is a coreceptor of VEGFR2. This interaction is critical for VEGF-A binding to VEGFR2.

VEGF-A is the main VEGF form that regulates angiogenesis in blood vessels, whereas lymphangiogenesis is primarily dependent on VEGF-C. Several splice variants of VEGF-A (VEGF-165, VEGF-121, and VEGF-189) have been identified that exhibit variable affinities for heparan sulfates. This interaction is important for VEGF's bioavailability. VEGF expression is regulated by a number of factors, including hypoxia, oxidative stress, reactive oxygen species, tumor suppressor genes or growth factors, or cytokines that activate the MAP kinase pathway.

VEGF alone appears to be unable to direct blood vessel organization or maturation. A number of other factors act not only on vascular endothelial cells but also on other vascular cells, such as pericytes. They may also contribute to vessel stabilization or pruning. Among these factors are fibroblast growth factors (FGFs), platelet-derived growth factors (PDGFs), angiopoietins, developmental gene products such as Notch/Delta, or molecules classically involved in axonal guidance. Particularly noteworthy among the latter are ephrins and neuropillins.

The major system involved in vessel stabilization is the angiopoietin system. Angiopoietin-1 (Ang-1) is thought to stabilize vessels by rendering them less sensitive to VEGF. The stabilizing effect of Ang-1 is disrupted by angiopoietin-2 (Ang-2) binding to the Ang-1 receptor Tie-2. Ang-1 stimulates endothelial cell migration by activating tie-2 and inducing the recruitment of Dok-R, Nck, and Pak. Ang-1, but not Ang-2, has been shown to be in the extracellular matrix. Ang-1 induces platelet-derived factors (PDGF), which is released and stimulates migration of pericytes, smooth muscle cells, and other accessory cells and, thus, favors mural cell coverage. Ang-1 is able to trigger vessel remodeling, as demonstrated in the retina.

Angiogenesis is also under the control of proteolytic enzymes and inhibitors, including the plasminogen activator system and matrix metalloproteinase and their inhibitors. The basic principle of the activity of

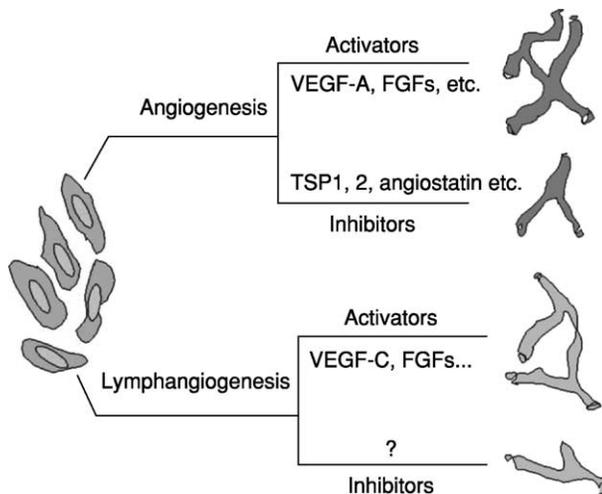
these enzymes is that they must be present at a critical concentration at the cell surface to promote invasion of vascular tubes. Furthermore, inhibitors such as tissue inhibitor of metalloproteinase-2 for matrix metalloproteinase (MMPs) may be required to localize the proteolytic activity at the cell surface, thus promoting activation of MMPs. This may account for the paradoxical stimulatory effects observed for these inhibitors in some circumstances.

Cell adhesion molecule receptors such as integrin  $\alpha_v\beta_3$  (and, to a lesser extent,  $\alpha_v\beta_5$ ) also play a critical role in angiogenesis. For example, the integrin  $\alpha_v\beta_3$  is highly expressed in proliferating endothelial cells, and both a monoclonal antibody to  $\alpha_v\beta_3$  and a low-molecular-weight antagonist have been shown to inhibit angiogenesis in *in vivo* models. This indicates that integrin  $\alpha_v\beta_3$  has a promoting role in angiogenesis and may constitute a potential interesting therapeutic target. However, recent observations of  $\alpha_v\beta_3$  knockout mice have challenged this view. In particular, it has been reported that mice lacking  $\beta_3$  integrin show enhanced pathological angiogenesis. The reason for these apparent differences is not clear. Cheresh and collaborators attempted to explain the differences using the concept of ligated and unligated integrins.

Unligated integrins provide dead signals to endothelial cells that are blocked by ligation. Knocking out integrins affects both ligated and unligated integrins and may artificially increase the invasiveness of endothelial cells by increasing endothelial cell survival, assuming that more unligated than ligated integrins are present.

## ANGIOGENESIS AND PATHOLOGY

Angiogenesis is a driving force for a number of pathologies, such as cancer, ocular neovascular disease, ischemic disease, and chronic inflammatory diseases. Cancer cells express a complex molecular repertoire that critically impacts the surrounding vascular stroma (Fig. 2). It is clearly recognized that tumor cells produce both negative and positive regulators of vasof ormation, and that the net effect on the vasculature is the outcome of the balance between these two types of regulators. That the growth of tumors is dependent on the vasculature was recognized by Algire in 1945 and later formulated as a paradigm by Judah Folkman. In general, tumors less than 2 or 3 mm are avascular and grow slowly. Tumor cells are then activated intracellularly by a mechanism called angiogenic switch and then start to favor the positive



**Figure 2** Regulation of tumor angiogenesis. Tumors produce both positive (activators) and negative (inhibitors) regulators and regulate both angiogenesis and lymphangiogenesis. Inhibitors of lymphangiogenesis have not been identified. VEGF, vascular endothelial growth factor; FGFs, fibroblast growth factors; TSP, thrombospondins.

regulators over the negative regulators. First insight into the molecular mechanisms of the angiogenic switch was gained through the analysis of tumor progression in the RIP-Tag mouse model. Overexpression of the T antigen under the control of the insulin promoter induces an insulinoma that progresses to highly malignant and vascularized lesions. Tumor nodules from the early stage do not induce angiogenesis when cocultured with endothelial cells *in vitro*. At the later stage, however, angiogenesis is potently induced. A number of molecules are involved in the angiogenic switch. VEGF is required since mice with conditional inactivation of VEGF do not undergo the angiogenic switch. MMP3 and MMP9 are also required. These proteases allow the translocation of VEGF from the tumor matrix to the blood vessels. FGF prototypes also seem to be needed in this model.

Tumors are not identical and may be more or less dependent on one or the other angiogenesis factor or inhibitor. For example, prostate carcinoma and melanoma are more dependent on FGF and VEGF, whereas other tumors, such as ovarian carcinoma, are primarily dependent on VEGF. Furthermore, tumors may modify their angiogenesis repertoire during their lifetime and switch from one to the other growth factor/growth factor receptor system. This may constitute an escape mechanism for tumors due to environmental constraints.

Tumors may also express angiogenesis inhibitors, such as thrombospondin-1 (TSP-1) and TSP-2. Involvement of TSPs in tumor angiogenesis has been particularly well studied in mouse models of skin carcinogenesis. TSP-1 and TSP-2 both seem to inhibit tumor angiogenesis and tumor growth in models of skin carcinoma in mice. On the other hand, TSP-2, but not TSP-1, increases with tumor progression, indicating that TSP-2, but probably not TSP-1, is part of a host antitumor defense mechanism.

In addition to blood vessels, lymphatic vessels are also required for dissemination of tumor cells. Factors critically involved in the development of lymphatics, such as VEGF-C and FGFs, may also play a role in lymphangiogenesis in tumors. Blocking lymphangiogenesis in a highly metastatic human lung cancer cell line by inhibiting VEGF-C suppresses lymph node metastasis. Furthermore, crossing RIP-Tag mice with mice expressing VEGF-C under the control of the insulin promoter yields bigenic mice that develop pancreatic tumors with metastatic spread through the lymphatic system. Most important, Dadrás *et al.* reported that intratumoral lymphangiogenesis in melanoma patients is associated with poor survival.

Ocular neovascular disease is another pathological condition in which abnormal angiogenesis plays a leading role. Diseases include diabetic retinopathy and age-related macular dystrophy (ARMD). VEGF seems to be one of the principal factors in the pathophysiology of these diseases. In the case of ARMD, the fas/fas ligand system may also have a critical role. Indeed, fas/fas ligands are expressed in the retinal pigmented epithelium and the choroid vessels. Knockout mice for fas/fas ligand exhibit aberrant ocular neovascularization resembling ARMD.

Angiogenesis is also implicated in ischemic disease in two ways. First, capillary growth within the walls of large arteries may contribute to the establishment of a proliferative lesion and invasion into the intima. It has been reported that inhibition of plaque neovascularization reduces the accumulation of macrophages and progression to advanced atherosclerotic lesions. Second, the growth of the collateral circulation that limits the extent of the ischemic lesion is also controlled by angiogenesis factors such as VEGF or FGFs. This has offered novel therapeutic opportunities to salvage the ischemic area and to improve morbidity and mortality in patients with coronary or peripheral artery disease.

Finally, chronic inflammatory disorders such as chronic polyarthritis are also angiogenesis dependent. In the early phase, this disease is characterized by a

proliferative lesion of synoviocytes in the synovia. Within the inflamed synovia, the number and quality of microvessels are also altered. VEGF and integrin  $\alpha_v\beta_3$  seem to play an essential role since blocking their activity in animal models results in disease improvement.

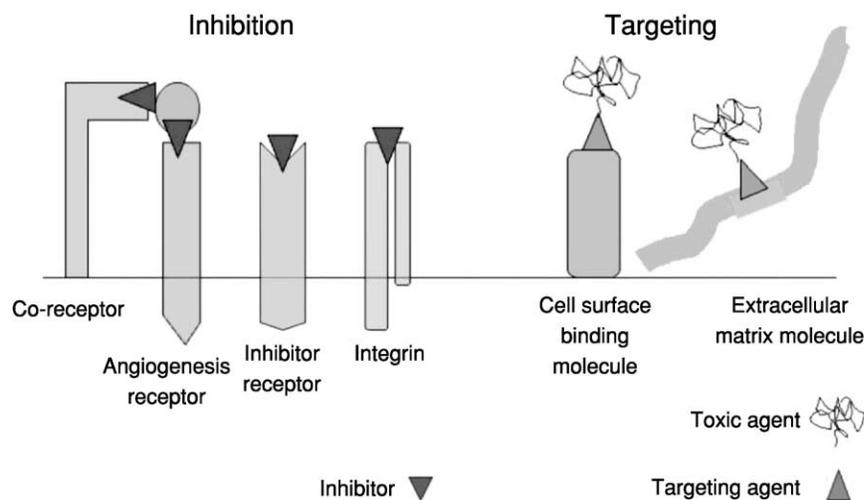
## ANGIOGENESIS AND THERAPY

A number of molecules are under preclinical or clinical evaluation for angiogenic or antiangiogenic therapy. Clinical trials using VEGF or FGFs to stimulate the collateral circulation in patients with coronary disease or peripheral arterial ischemic disease are under way. On the other hand, inhibition of angiogenesis is therapeutically relevant in cancer, retinal neovascular disease, or chronic inflammatory disease. The major disease for which antiangiogenesis strategies have been investigated is cancer. Targeting the vasculature for angiogenesis inhibition may be achieved through extracellular or intracellular mechanisms (Fig. 3). Extracellularly, inhibitors may block binding of angiogenesis factors to receptors, interfere with coreceptor/ligand/receptor interactions, activate inhibitor receptors, or modulate integrin–extracellular matrix interactions or protease activity.

Intracellularly, inhibitors may directly inhibit tyrosine kinase receptors or signaling modules. Among inhibitors, molecules interfering with VEGF activity are a major focus of research. VEGF activity may be

blocked with anti-VEGF or VEGF receptor antibodies or specific VEGF receptor kinase inhibitors. Ferrara and collaborators were the first to show that antibodies to VEGF slowed tumor growth. Human forms of the antibody (Avastin) are now in phase 3 clinical trial for the treatment of solid tumors. Recently, remarkable results were announced in a phase 3 clinical trial with Avastin in combination with chemotherapy in colon carcinoma. Low-molecular-weight inhibitors of the tyrosine kinase activity of the VEGF receptor (VEGFR-KIs) are another venue for antiangiogenesis tumor therapies. Encouraging results have been obtained with VEGFR-KIs in kidney and colon carcinoma. Another target for development of angiogenesis inhibitors is molecules encrypted within larger regulatory proteins, including endostatin, angiostatin, thrombospondin, platelet factor-4 (PF-4), and endorepellin. For example, a peptidomimetic of thrombospondin has been synthesized that exhibits high antiangiogenesis activity and is currently in clinical trials. A C-terminal fragment of PF-4 has also been developed that exhibits increased antiangiogenic and antitumor activity in glioblastoma.

Furthermore, antiangiogenesis properties of molecules already known to exhibit inhibitory activities on other cell types have recently been discovered. Particularly noteworthy is the fact that chemotherapy, when given at low dose and metronomically, has mainly antiangiogenic effects and greatly inhibits tumor development in mice.



**Figure 3** Angiogenesis inhibition versus vascular targeting. For angiogenesis inhibition, inhibitors may block the interaction of angiogenesis factors with their receptor by disrupting specific binding to receptors or coreceptors. Furthermore, inhibitors may activate inhibitor receptors such as CD36 for thrombospondin-1. Inhibitors may also disrupt the interaction between integrins and matrix molecules. Targeting agents bind surface molecules that are specifically present on the surface of angiogenic endothelial cells or in the extracellular matrix around blood vessels.

Combinatory approaches involving several molecules or treatment regimens have been developed to increase therapeutic efficacy. Antiangiogenic molecules may be combined with chemotherapy or radiotherapy. These approaches have been validated in the mouse, and phase 3 clinical trials are under way involving, for example, the association of chemotherapy with Avastin in colon or kidney carcinoma. Another approach is to combine several antiangiogenesis drugs. For example, the combination of endostatin, angiostatin, and TNP470 has been shown to be more effective than any single agent alone. Bergers and colleagues associated a VEGFR-KIs with Gleevec, a drug used for the treatment of chronic myelogenous leukemia. In addition to inhibiting the BCR-abl tyrosine kinase activity, Gleevec also inhibits c-kit and PDGFR tyrosine kinase. Since PDGFs are critically involved in pericyte recruitment and vessel stabilization, inhibition of this receptor tyrosine kinase may be important to resensitize vessels to VEGF inhibition. This is indeed the case, as demonstrated by Bergers and colleagues using the RIP-Tag mouse model. Treatment of established tumors leads to regression when the two inhibitors are combined.

Vascular targeting is different from the approaches described previously (Fig. 3). It involves selective targeting of tumor blood vessels and their subsequent destruction. This approach is based on the assumption that the tumor vasculature is different from that in normal tissues. Much evidence supports this claim. Tumor endothelial cells are often hypoxic, lack nutrients, and exposed to a number of cytokines and stress factors. This modifies the gene expression profile, making them angiogenic. The proof of concept that targeting is a valuable approach for cancer treatment was provided in 1993 when Burrows and Thorpe used an anti-MHC class II antibody coupled to ricin toxin to destroy tumor vasculature in experimental animal models. Other compounds that target the vasculature include the combretastatins, which are currently in phase 1 trials. They are selectively toxic to tumor vasculature by disrupting the tubulin cytoskeleton.

Recently, impressive results were reported by Halin and collaborators using a vascular-specific antibody (L19) that was isolated from an antibody phage library and that recognized the ED-B domain of fibronectin. This antibody is selectively targeted to the tumor endothelium. Chimeras of the antibody with interleukin-12 or tissue factor caused regression of established tumors when injected in experimental mouse tumor models.

This oxygen-sensing pathway also offers opportunities for therapeutic intervention. PHD inhibitors should be proangiogenic, whereas molecules that

abrogate HIF-1 $\alpha$  binding to coactivators and/or the hypoxia response element reduce angiogenesis and tumor growth. Modulation of the iron content of the cell may also affect the activity of the PHDs because these enzymes are sensitive to the local Fe<sup>2+</sup> concentration.

## CONCLUSION

Angiogenesis is a fundamental mechanism in embryonic and postnatal development that also plays a major role in pathologies such as cancer, ocular neovascular disease, ischemic disease, and chronic inflammatory disease. Many molecules, receptors, and intracellular signaling modules have been implicated in vascular morphogenesis. These discoveries have offered novel opportunities for therapeutic intervention. There is an increasing repertoire of drugs with which to manipulate angiogenesis and new endothelial-specific genes with which to target the vasculature. Thus, angiogenesis research is currently at an exciting stage.

## See Also the Following Articles

Fibroblast Growth Factor (FGF) • Platelet-Derived Growth Factor (PDGF) • Prostate Cancer

## Further Reading

- Algire, G. H., Chalkely, H. W., Legallais, F. Y., and Park, H. (1945). Vascular reactions of normal and malignant tumors in vivo: I. Vascular reactions of mice to wounds and to normal and neoplastic transplants. *J. Natl. Cancer Inst.* **6**, 73–85.
- Alitalo, K., and Carmeliet, P. (2002). Molecular mechanisms of lymphangiogenesis in health and disease. *Cancer Cell* **1**, 219–227.
- Bergers, G., Song, S., Meyer-Morse, N., Bergsland, E., and Hanahan, D. (2003). Benefits of targeting both pericytes and endothelial cells in the tumor vasculature with kinase inhibitors. *J. Clin. Invest.* **111**, 1287–1295.
- Bikfalvi, A., and Bicknell, R. (2002). Recent advances in angiogenesis, anti-angiogenesis and vascular targeting. *Trends Pharmacol. Sci.* **23**, 576–582.
- Carmeliet, P. (2000). Mechanisms of angiogenesis and arteriogenesis. *Nature Med.* **6**, 389–395.
- Cheresh, D. A., and Stupack, D. G. (2002). Integrin-mediated death: An explanation of the integrin knockout phenotype? *Nature Med.* **8**, 193–194.
- Dadras, S. S., Paul, T., Bertocini, J., Brown, L. F., Muzikansky, A., Jackson, D. G., Ellwanger, U., Garbe, C., Mihm, M. C., and Detmar, M. (2003). Tumor lymphangiogenesis: A novel prognostic indicator for cutaneous melanoma metastasis and survival. *Am. J. Pathol.* **162**, 1951–1960.
- Ferrara, N. (2003). Role of vascular endothelial growth factor in physiologic and pathologic angiogenesis: Therapeutic implications. *Semin. Oncol.* **29**(6 Suppl. 16), 10–14.
- Folkman, J. (1971). Tumor angiogenesis: Therapeutic implications. *N. Engl. J. Med.* **285**, 1182–1186.

- Hagedorn, M., and Bikfalvi, A. (2000). Target molecules for anti-angiogenic therapy: From basic research to clinical trials. *Crit. Rev. Oncol. Hematol.* **34**, 89–110.
- Halin, C., Rondini, S., Nilsson, F., Berndt, A., Kosmehl, H., Zardi, L., and Neri, D. (2002). Enhancement of the antitumor activity of interleukin-12 by targeted delivery to neovasculature. *Nat. Biotechnol.* **20**, 264–269.
- Javerzat, S., Auguste, P., and Bikfalvi, A. (2002). The role of fibroblast growth factors in vascular development. *Trends Mol. Med.* **8**, 483–489.
- Kerbel, R., and Folkman, J. (2002). Clinical translation of angiogenesis inhibitors. *Nature Cancer Rev.* **2**, 727–739.
- Oliver, G., and Detmar, M. (2002). The rediscovery of the lymphatic system: Old and new insights into the development and biological function of the lymphatic vasculature. *Genes Dev.* **6**, 773–783.
- Rak, J., Yu, J. L., Kerbel, R. S., and Coomber, B. L. (2002). What do oncogenic mutations have to do with angiogenesis/vascular dependence of tumors? *Cancer Res.* **62**, 1931–1934.
- Reyes, M., et al. (2002). Origin of endothelial progenitors in human postnatal bone marrow. *J. Clin. Invest.* **109**, 337–346.
- Ruoslahti, E. (2002). Specialization of tumour vasculature. *Nature Cancer Rev.* **2**, 83–90.
- Semenza, G. (2000). HIF-1: Mediator of physiological and pathological responses to hypoxia. *J. Appl. Physiol.* **88**, 1474–1480.
- Vale, P. R., Isner, J. M., and Rosenfield, K. (2001). Therapeutic angiogenesis in critical limb and myocardial ischemia. *J. Interv. Cardiol.* **14**, 511–528.



# Angiotensin, Evolution of

Werner Kloas

Leibniz Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany

## Glossary

**angiotensin receptors** All angiotensin receptors identified in several classes of vertebrates share a high degree of homology and belong to a superfamily of seven-transmembrane G protein-coupled receptors.

**angiotensins** Bioactive peptide hormones of varying length between 6 and 10 amino acids. The main functions of angiotensins are associated with osmomineral and water balance as well as maintenance of blood pressure in vertebrates.

**evolution of the renin–angiotensin system (RAS)** The functioning systems to produce angiotensin are present in all classes of vertebrates and some invertebrate groups, indicating an early appearance of endocrine function of angiotensin during evolution.

Angiotensin (Ang) is a classical peptide hormone important in osmomineral regulation and blood pressure maintenance of vertebrates. The Ang-producing renin–angiotensin system (RAS) and corresponding receptors for Ang are present in all classes of vertebrates. Comparative studies indicate that homologous Ang and RAS are present in several groups of invertebrates, implicating an important biological role of Ang during early evolution.

## INTRODUCTION

The presence of the peptide hormone Ang produced by the renin–angiotensin system (RAS) and its corresponding receptors is reported in all classes of vertebrates. Ang plays an important role in the control of blood pressure and osmomineral regulation directly by vasoconstrictory actions in the circulatory system or indirectly by regulating osmominerals, affecting hormones such as corticosteroids, catecholamines, or nonapeptides, and by influencing drinking behavior as a dipsogenic agent. However, studies indicate that the evolution of Ang occurred at an earlier evolutionary period in several groups of invertebrates.

## COMPONENTS OF THE RENIN–ANGIOTENSIN SYSTEM

In vertebrates, the major pathway of systemic Ang is the generation of the decapeptide Ang I (angiotensins 1–10) by the enzyme renin from liver-produced angiotensinogen. Ang I is subsequently converted to the octapeptide Ang II (angiotensins 1–8) by Ang-converting enzyme (ACE) and then metabolized to smaller peptides exhibiting biological activities, such as Ang III (angiotensins 2–8), Ang IV (angiotensins 3–8), and Ang V (1–7). In addition, the tissue-specific existence of Ang-producing RAS has been demonstrated within the brain, pituitary, gonads, and adrenal, where Ang is processed not only via renin and ACE but also by various other enzymes.

## Angiotensins

Amino acid sequences of Angs are highly preserved throughout evolution and differ only in positions 1, 3, 5, and 9, where exchanges of amino acids may occur (Table I). However, species-specific differences in the structure of Angs can markedly modify their biological activities depending on the corresponding Ang receptors of the respective species. It is noteworthy that the structure of Ang I in humans and leeches is identical.

## Renins and Angiotensin-Converting Enzymes

Renins and renin-like enzymes cleaving Ang I by proteolytic cleavage from angiotensinogen are present in all classes of vertebrates and also found in leeches and insects. The presence of ACEs generating the biologically active octapeptide Ang II from Angs I is established for all vertebrates, and functioning ACEs have been demonstrated in invertebrates such as insects, leeches, and mollusks.

Table I Angiotensin I Amino Acid Sequences from Vertebrates and Annelids<sup>a</sup>

	Amino acid sequence									
	1	2	3	4	5	6	7	8	9	10
Common structure	—	Arg	—	Tyr	—	His	Pro	Phe	—	Leu
Species variation										
Mammals (human)	Asp		Val		Ile				His	
Birds (chicken)	Asp		Val		Val				Ser	
Reptiles (alligator)	Asp		Val		Val				Ala	
Amphibians (bullfrog)	Asp		Val		Val				Asn	
Bony fishes (eel)	Asn		Val		Val				Gly	
Cartilaginous fishes (dogfish)	Asn		Pro		Ile				Gln	
Cyclostomes (hagfish)	Asn		Val		Val				Thr	
Annelids (leech)	Asp		Val		Ile				His	

<sup>a</sup>Native angiotensin I from all vertebrate classes and annelids. Among vertebrate classes additional species-specific differences in amino acid sequences in positions 1, 3, 5, or 9 are present or may be found.

## ANGIOTENSIN RECEPTORS AND CELL SIGNALING PATHWAYS

All Ang receptors identified and cloned in vertebrates belong to the superfamily of seven-transmembrane G protein-coupled receptors, which have several second messenger systems, including phosphatidylinositol-4,5-bisphosphate, intracellular Ca<sup>2+</sup>, endothelium-derived relaxing factor, nitric oxide, and cGMP. In invertebrates, evidence of functioning Ang receptors in leeches was obtained by binding studies and physiological experiments.

## BIOLOGICAL ACTIONS

The classical functions of Ang in vertebrates are the maintenance of water and osmomineral contents and blood pressure. In higher vertebrates, Ang decreases the glomerular filtration rate in the kidney and stimulates the adrenal to secrete mineralocorticoid aldosterone by the cortex and to release the catecholamines adrenaline and noradrenaline by the medulla. The effects on the central nervous system include dipsogenic behavior and increased hypophyseal secretion of nonapeptides (vasopressin or vasotocin), which act positively on water conservation and blood pressure. In addition, Ang acts on memory processes, reproduction, and immune response modulation of mammals.

The endocrine counterpart on nearly all target organs for Ang and its biological actions is the family of natriuretic peptides, which counteract the

maintenance of water and osmomineral homeostasis in all classes of vertebrates.

However, main target organs for Ang may differ in vertebrates during early evolution. In amphibians and fish, stimulatory effects of Ang on the adrenal homologues producing catecholamines or corticosteroids are found in some species but absent in others. Additional Ang targets in fish are osmoregulatory organs, such as gill, intestine, and chondrichthyan rectal gland, all of which are very important for osmoregulation. Angiotensin-induced effects on dipsogenic behavior and renal functions also vary widely among lower vertebrates, whereas vasopressor activities are always prominent.

In invertebrates, biological actions of Ang are reported for leeches, in which Ang has a diuretic effect by influencing Cl transport, demonstrating its physiological significance for maintenance of osmoregulation after a blood meal in these annelids.

Angiotensin is one of the classical vertebrate hormones involved in osmomineral and blood pressure regulation. The development of more complex endocrine effects of Ang on several target organs during vertebrate evolution is associated with changes from aquatic to semiterrestrial and terrestrial lifestyles, for which salt and water conservation as well as maintenance of blood pressure are essential. However, comparative studies indicate that Ang originated much earlier during evolution in annelids, in which it is also associated with water and osmomineral regulation.

## See Also the Following Articles

Hypertension, Renin and • Renin

## Further Reading

- Anderson, W. G., Cerra, M. C., Wells, A., Tierney, M. L., Tota, B., Takei, Y., and Hazon, N. (2001). Angiotensin and angiotensin receptors in cartilaginous fishes. *Comp. Biochem. Physiol. A* **128**, 31–40.
- Kobayashi, H., and Takei, Y. (1996). The renin–angiotensin system. Comparative aspects. *Zoophysiology* **35**, 77–92.
- Nishimura, H. (2001). Angiotensin receptors—Evolutionary overview and perspectives. *Comp. Biochem. Physiol. A* **128**, 11–30.
- Russell, M. J., Klemmer, A. M., and Olson, K. R. (2001). Angiotensin signaling and receptor types in teleost fish. *Comp. Biochem. Physiol. A* **128**, 41–51.
- Salzet, M., Deloffre, L., Breton, C., Vieau, D., and Schoofs, L. (2001). The angiotensin system in elements in invertebrates. *Brain Res. Rev.* **36**, 35–45.
- Sandberg, K., and Ji, H. (2001). Comparative analysis of amphibian and mammalian angiotensin receptors. *Comp. Biochem. Physiol. A* **128**, 53–75.



Anorchia

see Agonadism



# Anorexia Nervosa

Jesús Argente

University Autónoma and Hospital Universitario Infantil Niño Jesús, Madrid, Spain

María Teresa Muñoz

Hospital Infantil Universitario Niño Jesús, Madrid, Spain

## Glossary

**amenorrhea** The absence of menstruation during a minimum period of 3–6 months in women who have previously experienced menstruation (secondary amenorrhea) or the lack of menarche at 16 years of age (primary amenorrhea).

**bone mass** Total quantity of bone tissue, including the total volume of bone tissue and the total quantity of mineralized extracellular matrix.

**malnutrition** A pathological condition occurring when energy and nutrient requirements are not met by dietary intake. There is a wide spectrum of clinical forms, depending on the severity and duration of the deficit, the age of the subject, and the cause of the condition.

**osteopenia** Significant diminution of bone mass per unit volume in relation to what is considered normal for the age, pubertal stage, and sex of the subject.

**osteoporosis** A disease characterized by low bone mass and microarchitectural deterioration of bone tissue, leading to enhanced bone fragility and a consequent increase in fracture risk.

Anorexia nervosa (AN) is a childhood psychiatric disorder characterized by patient-induced and maintained weight loss that leads to progressive malnutrition and specific pathophysiological signs (disturbance of body image and fear of obesity). Diagnostic criteria were proposed for the first time in 1972 by Feighner *et al.* and were later modified by the American Academy of Psychiatry in the third edition of the *Diagnostic and Statistical Manual of Mental Disorders*. These criteria allowed uniform classification of patients with these characteristics and also suggested the existence of less severe forms, called nonspecified eating disorders. These forms can also have a dangerous evolution and therefore require therapy equal to that of the clinical forms. The diagnostic criteria for AN were updated in 1994 in the fourth edition of the *DSM*.

## PREVALENCE

The prevalence of anorexia nervosa (AN) among adolescents and young adults is approximately 0.5–1%, with a bimodal age distribution including peak ages of 14 and 18 years. However, a significant number of cases in girls during the initial stages of puberty, as well as before the onset of puberty, have been observed. The female/male ratio ranges from 5:1 to 20:1. A study of a Spanish female population between 12 and 21 years of age showed a prevalence of 0.3% for AN, 0.8% for bulimia, and 3.1% for nonspecified eating disorders; thus, a total of 4.1% of the population suffers from some type of eating disorder. However, 50–67% of adolescent females are dissatisfied with their weight and body shape, and most adolescents have been on a diet. Many of these teens use unhealthy weight control methods, including fasting, diet pills, and vomiting.

Adolescents with chronic diseases, especially females, have a higher incidence of eating disorders and special attention should be given to these patients to ensure that they do not develop clinically significant cases of anorexia or bulimia. Indeed, a cross-sectional study indicates that the prevalence is approximately twice that seen in age-matched controls.

## ETIOPATHOGENESIS

The etiology of AN is multifactorial, including genetic, biological, psychological, and cultural factors. The coexistence of various risk factors increases the possibility of developing this disease (Table I).

### Individual Factors

Patients with AN frequently have personality disorders, including low self-esteem and high anxiety. They are introverted, obsessive, and perfectionists, and feel as if they are little effective, although the

**Table 1** Diagnostic Criteria for Anorexia Nervosa<sup>a</sup>


---

Refusal to maintain body weight at or above a minimally normal weight for age and height (e.g., weight loss leading to maintenance of body weight less than 85% of that expected; or failure to make expected weight gain during period of growth, leading to body weight less than 85% of that expected)
An intense fear of gaining weight or becoming fat, even though underweight
A disturbance in the way in which one's body weight or shape is experienced, undue influence of body weight or shape on self-evaluation, or denial of the seriousness of the current low body weight
In postmenarcheal females, amenorrhea (the absence of at least three consecutive menstrual cycles)
Two subtypes are described: restricting subtype and bingeing/purging subtype

---

<sup>a</sup>From the American Psychiatric Association (1994).

reverse is usually the case. In addition, those patients that use purgative methods have a tendency to steal and often have problems with alcohol and drugs.

It is significant that in these patients the loss of self-esteem and autocontrol culminates in them going on a diet. To this culturally accepted activity, they often add other useful activities, such as studying many hours and intense physical exercise, which often result in a brief improvement in their condition. The malnutrition provokes a series of alterations, both physical and mental, and the capacity to relate socially is altered, producing a new decrease in self-esteem and autocontrol. The response is a more strict diet, which can then perpetuate a dangerous spiral of events.

Some high-level athletes develop AN, but others present incomplete or subclinical forms with a more difficult diagnosis. A form called athletic anorexia has been described that includes high-level female athletes who have an intense fear of gaining weight or becoming fat, even though they are below their ideal weight, with low-caloric intake and often excessive exercise.

## Genetic Factors

No specific marker has been identified that suggests an increased possibility of AN. Investigation has concentrated on analysis of polymorphisms of genes related to weight control and the serotonergic and dopaminergic pathways. One of the most frequently studied genes is that of the serotonin receptor. Serotonin [5-hydroxytryptamine (5-HT)] is involved in a broad range of biological, physiological, and behavioral functions and has been implicated in the development of eating disorders. Polymorphisms within different 5-HT receptor genes and the tryptophan hydroxylase gene have also been analyzed. Some studies have described an association between the-1438 A allele of the -1438 G/A polymorphism within the promoter region of the 5-HT<sub>2A</sub> receptor

gene and AN. These studies indicate that patients with AN have enhanced 5-HT<sub>2A</sub> receptor binding and provide further evidence for a serotonergic dysfunction in eating disorders, although further studies are needed.

The hypothalamus is also an important organ in the control of energy metabolism because it is responsible for the sensation of hunger and satiety and hence energy intake. In addition, via modulation of the sympathetic nervous system, it is involved in thermogenesis and, as a result, energy expenditure. Various neuropeptides control these functions. Regulation of acute eating behavior incorporates a system of satiety signals originating in the food. Cholecystokinin, bombesin, gastrin-releasing peptide, and others are involved in this signaling system and reach the brain via peripheral innervation or the circulation to activate their receptors.

Long-term energy balance is regulated via a system composed of different hormones secreted in proportion to corporal adiposity, such as leptin and insulin, that act at the level of the central nervous system. These respond to changes in body fat by activating anabolic or catabolic pathways, the first through production of neuropeptide Y (NPY), which stimulates food intake, and the second via the hypothalamic melanocortin system, which reduces food intake and stimulates weight loss (Fig. 1).

Leptin, a hormone synthesized by adipose tissue, plays an important role in the regulation of food intake and energy expenditure. Its mechanisms of action are unknown, although its primary target seems to be the hypothalamus. Leptin acts at the level of the hypothalamus to decrease appetite, resulting in weight loss. Plasma leptin levels and secretory pattern vary during the night and day and are influenced by food intake. Molecular analysis of the coding region and part of the promoter region of the leptin gene in patients with AN has yielded negative results. Hence, involvement of the leptin gene in the etiology of AN seems unlikely.

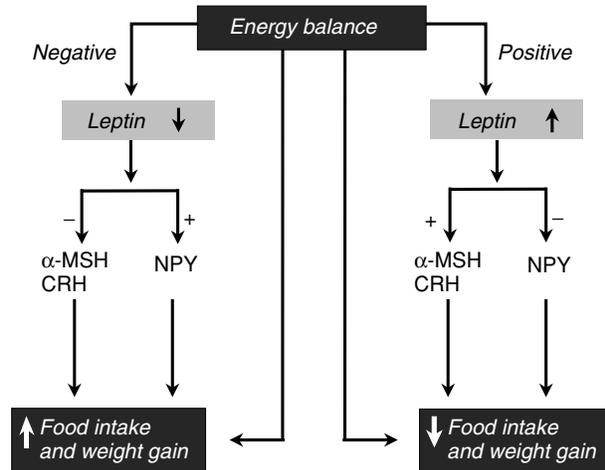


Figure 1 Hormonal response to malnutrition.

It has been demonstrated that a gastrointestinal peptide hormone, ghrelin, stimulates growth hormone secretion in rats and humans. Ghrelin also plays an important role in the regulation of energy balance. Plasma ghrelin levels are regulated by acute and chronic changes in energy balance (e.g., fasting increases, whereas feeding decreases, circulating ghrelin concentrations).

The actions of NPY, incrementing ingestion and decreasing thermogenesis, are opposite those of leptin. Neuropeptide YY5 and YY1 receptors in rats and humans are assumed to play a major role in NPY-induced food intake. The neuropeptide YY5 receptor gene (NPYY5R) is expressed in brain regions involved in the central regulation of feeding behavior, including the lateral hypothalamus, the paraventricular nucleus, and the arcuate nucleus. Systematic mutation screening within the coding region of the NPYY5R revealed a rare Glu-4-Ala variant in a single patient with AN. This allele was transmitted from the mother, who had no antecedents of any eating disorder. Association and transmission disequilibrium studies pertaining to variations and polymorphisms within NPYY1R and NPYY5R and AN were negative.

Glucocorticoids are also implicated in energy regulation. Via their effects on NPY, they act as endogenous antagonists of leptin and insulin. Other neuropeptides that stimulate food intake and energy storage are melatonin-concentrating hormone and orexin A and B, which increase in response to fasting and stimulate appetite. The melanocortins (MCs) are peptides derived from the precursor proopiomelanocortin (POMC) and act on specific receptors. The endogenous MC most implicated in food intake and body weight is  $\alpha$ -melanocyte-stimulating hormone

( $\alpha$ -MSH), which has a high affinity for the MC receptors, especially MC3 and MC4. Mutation screening of the coding region of MC4-R in patients with AN and bulimia nervosa revealed two common polymorphisms in both groups. Allele and genotype frequencies did not differ between these groups and probands of different weight extremes.

### Familial Factors

The families of patients with AN have certain characteristics in common. They are often overprotective, strict, and have a relative incapacity to solve their own conflicts. The mother figure is described as the boss of the family and the father figure as distant. In many families, the patient is often recognized as an individual only after the onset of the illness; consequently, the patient continues with the illness in order to remain the center of the family's attention.

### Sociocultural Factors

These adolescents are often very vulnerable, receiving a large quantity of information that they cannot assimilate, which creates tension regarding problems normal for their age, including sexuality, competitiveness, individuality, and independence within the family.

It is well-known that every historic period determines the prototypes for fashion and beauty. During the past two decades, thinness for women and strong physic for men have been fashionable. These stereotypes have been perpetuated mainly via publicity. The adolescent is constantly bombarded with information regarding the ideal weight and figure, how to have the perfect body, what type of exercise one must practice to achieve this perfect body, and miracle diets.

### EVALUATION

Multiple organ system complications are seen, including those involving the cardiovascular and peripheral vascular systems and gastrointestinal, hematological, renal, skeletal, endocrine, and metabolic disorders. These alterations are related not only to the state of malnutrition but also to the conduct of these patients to control their weight. A number of endocrine and metabolic disturbances described in patients with AN indicate hypothalamic dysfunction, including amenorrhea-oligomenorrhea, delayed puberty, hypothyroidism, hypercortisolism, interferon growth

factor-1 (IGF-1) deficiency, electrolyte abnormalities, hypoglycemia, and hypophosphatemia.

### Medical Complications

The clinical manifestations of AN are broad, affecting all systems of the organism, and depend largely on whether the form is restrictive or purgative. Some anorexic patients (10–20%) have bulimic tendencies, which mainly include provocation of vomiting, the use of laxatives, and a compulsive increase in physical activity.

Cardiovascular problems occur in up to 80% of patients, including bradycardia and hypotension, due to autonomic nervous system imbalances. Electrocardiographic abnormalities may include atrial and ventricular arrhythmias and QTc abnormalities. In addition, changes in myocardial function have been reported with decreases in myocardial tissue mass, mitral valve prolapse, and pericardial effusions.

Gastrointestinal complications are also common. AN can cause a decrease in gastrointestinal motility, resulting in chronic constipation. Laxative abuse can lead to cathartic colon syndrome and chronic constipation that is sometimes refractory to treatment. Cases of acute gastric dilatation have been described during the phase of realimentation of extremely affected anorexia patients since gastric emptying of solids is retarded, and that of liquids is retarded in some patients. Esophageal problems include severe esophagitis and even ruptures of the esophagus associated with induced vomiting.

Neurological consequences result from severe malnutrition (Fig. 2). Computed tomography and magnetic resonance imaging have demonstrated cortical atrophy and ventricular dilatation. Malnourished patients have greater cerebrospinal fluid volumes and reduced white and gray matter volumes. Abnormalities on computed tomography scans are reversible with refeeding and nutritional recovery.

### Biochemical Abnormalities

Hematological findings include anemia, leucopenia (relative neutropenia and lymphocytosis), thrombocytopenia, low erythrocyte sedimentation rate, and decreased fibrinogen levels in plasma. The anemia and occasional pancytopenia appear to be due to hypoplasia of the bone marrow, which is filled with a gelatinous mucopolysaccharide.

Vomiting results in the loss of sodium, hydrogen, and potassium, causing metabolic alkalosis. The use of



Figure 2 Patient with anorexia nervosa and severe malnutrition.

laxatives provokes the loss of potassium and bicarbonate, which can result in metabolic acidosis. Finally, the use of diuretics can cause increased loss of sodium, potassium, and calcium in the urine, depending on the dose and drug used.

Renal complications are present in 7% of these patients and can include a decrease in glomerular filtration, an increase in plasma urea and creatinine levels, electrolyte alterations, edema, and hypokalemic nephropathy. Renal concentration ability is impaired and polyuria may occur. An abnormal response of arginine vasopressin to an osmotic stimulus may be seen.

Cellular immune functions are also altered as a consequence of poor nutrition. These include modifications in some immunoglobulin fractions (IgG and IgA and complement factors C3 and C4), low response in the cutaneous delayed hypersensitivity test, and alterations in the lymphocyte subpopulations CD3, CD4, and CD57. However, infections are infrequent in these patients.

Plasma protein levels are usually normal, although in some cases hypoalbuminemia is present. Elevated amylase has been observed in the absence of clinical signals of pancreatitis. Predisposition factors are duodenal and jejunal dilation, and these are more frequent in patients with bulimia. There is an increase in

$\beta$ -carotenes and vitamin A, although the reason for this is unclear. In contrast, blood levels of copper and zinc are reduced and iron and ceruloplasmin levels are normal.

Mild hypercholesterolemia is frequent in AN, with the elevation occurring in the low-density lipoprotein fraction, whereas both high-density lipoprotein and very low-density lipoprotein levels are normal. Plasma triglyceride levels are normal despite low values for hepatic lipase and lipoprotein lipase activities. The cause of the hypercholesterolemia is not known.

## DIFFERENTIAL DIAGNOSIS

A differential diagnosis must be made in situations characterized by loss of weight in young persons, such as brain tumors and lymphomas, and in cases with gastrointestinal symptoms, such as chronic inflammatory disease. In addition, different endocrine pathologies, such as Addison's disease, hyperthyroidism, and diabetes mellitus, must be taken into consideration. Depending on the predominant psychopathological symptoms, depression and obsessive-compulsive alterations, social phobia, and schizophrenia must be ruled out.

## COMPLICATIONS

### Endocrine and Neurotransmitter Disturbances

#### **Hypothalamic–Pituitary–Ovarian Axis**

Patients with AN exhibit isolated hypogonadotropic–hypogonadism of hypothalamic origin. The etiology is uncertain, although multiple factors may play a role, including hypothalamic dysfunction, reduction of weight, sex steroids and neurotransmitters alterations, as well as physical exercise (Table II).

Adolescents with AN exhibit low basal levels of both luteinizing hormone (LH) and follicle-stimulating hormone (FSH) as well as low estradiol levels, indicating the abnormal function of the hypothalamic–pituitary–gonadal axis. In addition, spontaneous secretion of LH during a 24-h period is decreased in both frequency and amplitude of the secretory bursts. With weight gain, serum levels of both LH and FSH are increased, suggesting that malnutrition may play a role in the regulation of gonadotropin secretion. Disturbances in neurotransmitters have been described, including the dopaminergic system and endogenous opioids, both peptides related to GnRH regulation. Whether these alterations in

neurotransmitters are primary or secondary to malnutrition remains to be elucidated.

Malnutrition may be responsible for the delayed puberty and reduction in growth seen in these patients. This phenomenon has been interpreted as a mechanism of adaptation to the reduction in nutrients. Delayed puberty is present when symptoms appear in prepubertal patients; in contrast, if the disease begins after development has begun, puberty is detained and the growth spurt delayed and smaller. Finally, if symptoms appear after puberty, secondary amenorrhea is present. One of the indications that the process of adaptation to malnutrition has begun is hypoinsulinemia, present as a consequence of low glucose and amino acid levels. Furthermore, growth hormone (GH) abnormalities and low IGF-1 levels contribute to poor growth in prepubertal patients,

**Table II Endocrine Changes in Anorexia Nervosa<sup>a</sup>**

<b>Hypothalamic–pituitary axis</b>
LH ↓
FSH ↓
GH ↓↑
IGF-1 ↓
IGFBP-1 ↑
IGFBP-3 ↓
GHBP ↓
Thyrotropin (delayed response to TRH)
Corticotropin (↓ response to CRF)
Prolactin ⇔ ↓
ADH: abnormal regulation
<b>Thyroid gland</b>
T <sub>4</sub> ↓
T <sub>3</sub> ↓
Reverse T <sub>3</sub> ↑
<b>Adrenal gland</b>
Cortisol ⇔ ↑
Urinary free cortisol ↑
Abnormal dexamethasone suppression
<b>Ovary</b>
Estradiol ↓
Estrone ↓
Progesterone ↓
<b>Testis</b>
Testosterone ↓

<sup>a</sup>Abbreviations used: normal, ⇔; decreased, ↓; increased, ↑; LH, luteinizing hormone; FSH, follicle-stimulating hormone; LHRH, luteinizing hormone-releasing hormone; IGF-1, insulin-like growth factor-1; IGFBP-1, insulin-like growth factor-binding protein type 1; IGFBP-3, insulin-like growth factor-binding protein type 3; GHBP, growth hormone-binding protein; TRH, thyrotropin-releasing hormone.

leading to a reduction in their final height. Reliable data are not available to establish the percentage of growth lost in relation to the target height. However, one study estimated that some patients were 3 cm shorter than their target height.

Whether leptin is a permissive factor or plays a central role in the initiation of puberty is unknown. In subjects with reduced fat stores, problems with reproductive system functioning are frequent, including a reduction in serum sex steroid levels. A similar phenomenon, the shutdown of the hypothalamic–pituitary–gonadal axis, occurs in patients with anorexia nervosa after the loss of fat stores. In both cases, it is speculated that the problems with gonadal function could be related to the decreased serum leptin levels as a result of the loss of fat tissue. We have observed that in patients with AN during partial weight recovery, there is no significant increase in leptin levels or in the recovery of gonadal function. Of course, it is possible that a longer weight recovery period is necessary for this to occur, which would result in loss of the linear correlation with the body mass index (BMI). Studies suggest that BMI is the most important control factor for the secretion of leptin in situations of modified food intake and that there is a loss in circadian rhythm in patients with AN. Of great interest is whether leptin is necessary for the recovery of menstruation in these patients. However, more studies are necessary to answer this question.

There is evidence of an association between melatonin levels and gonadal function in humans, with woman with hypothalamic amenorrhea having elevated nocturnal levels of melatonin. Patients with AN are also reported to have elevated nocturnal levels of 6-sulfatoxymelatonin (the principal metabolite excreted in the urine) both at diagnosis and after weight recovery if amenorrhea exists.

The percentage of total body fat can be evaluated by using dual-energy X-ray absorptiometry. In patients with AN and moderate malnutrition, the percentage of total body fat is a better indicator of the nutritional state than BMI. There is a significant correlation between leptin levels and the percentage of total body fat that is not found between BMI and leptin.

It has been reported that if patients with AN recuperate weight to obtain at least 90% of the weight adequate for their height, their menstrual cycles will return within 6 months. It therefore follows that one of the decisive factors for the normalization of gonadal function is the recuperation of the nutritional state.

### **Hypothalamic–Pituitary–GH Axis**

Most studies indicate that a large percentage of patients with AN have elevated basal and GH-releasing hormone (GHRH)-stimulated GH levels. There are alterations in the GH response to different stimuli (decreased response after hypoglycemia, clonidine, and hexarelin) as well as paradoxical hormone responses [elevated GH after thyrotropin-releasing hormone (TRH) thyrotropin or after intravenous glucose], although these responses are heterogeneous. Few studies have analyzed spontaneous GH secretion (SGHS) in AN patients. We studied SGHS in a group of anorexic patients at diagnosis and at two different times during weight recovery and found that at diagnosis SGHS is heterogeneous. In 40% of subjects, mean 24-h GH secretion was >3 ng/ml (lower limit of normality), and the remaining 60% had levels below the normal range. The difference between these groups and the controls was due to modification in the amplitude of the GH peaks and not to pulse frequency. In both groups, recovery of at least 10% of initial weight resulted in the normalization of the parameters of SGHS. These observations suggest that alterations in GH secretion in these patients are due to modifications in its neuroendocrine control, with an increase in GHRH release and decreased somatostatin tone.

The GH pattern in conjunction with the negative correlation between basal and pulsatile GH secretion and BMI suggests that the observed alterations in GH secretion are directly related to malnutrition. One possible mechanism may involve the reduced IGF-1 levels caused by malnutrition. This would effect a reduction in the negative feedback action that IGF-1 exerts on GH secretion at the level of the hypothalamus and pituitary. Another variable that may be involved is the hypoestrogenism that accompanies amenorrhea. It has been suggested that malnutrition underlies the increase in the amount of GH secreted in each pulse and that hypoestrogenism is responsible for the increased pulse frequency. Definite conclusions, however, can not be drawn from these studies.

In patients with AN, plasma ghrelin levels are significantly increased and rapidly normalize after partial weight recovery. Whereas decreased leptin levels in AN patients might simply reflect reduced body fat mass, increased gastric ghrelin secretion in AN might reflect a physiological effort to compensate for the lack of nutritional intake and stored energy.

Serum GH-binding protein (GHBP) levels in patients with AN are dramatically reduced and tend to normalize with weight recovery. The reduction in GH

receptors is most likely one of the principal causes of GH resistance.

In malnutrition, the low GHBP levels may be related to hypoinsulinemia, alterations in thyroid function, or hypoestrogenism. On the other hand, many studies have demonstrated a correlation between serum GHBP levels and BMI or the percentage of body fat or, more specifically, visceral fat. Given that it has not been demonstrated that circulating GHBP is uniquely or even preferentially derived from liver GH receptors, it is possible that other tissues, such as adipose tissue, may contribute to plasma GHBP levels. If this is the case, the extreme reduction in adipose tissue in patients with AN may cause the observed decrease in serum GHBP levels.

Patients with AN have extremely reduced serum IGF-1 levels that tend to normalize with weight recovery; however, as observed for other forms of malnutrition, the time necessary for this to occur may be prolonged. Circulating IGF-1 is largely dependent on GH, but it is also very sensitive to nutritional changes. In AN patients, serum IGF-1 levels do not correlate with GH secretion, which suggests that the decrease in IGF-1 is independent of GH and probably directly due to the state of malnutrition. The coexistence of reduced IGF-1 levels and normal or elevated GH secretion suggests that these patients exhibit resistance to GH action.

Data on serum levels of free IGF-1 are limited and contradictory. Some authors have found normal levels, whereas others report a decrease; in both cases, weight recovery tended to increase free IGF-1 levels. Likewise, some authors have found serum IGF-2 levels to be normal at the time of diagnosis and to increase with weight recovery. In contrast, others report IGF-2 levels to be decreased, although not significantly, and to normalize with weight recovery. Serum levels of free IGF-2 are reported to be decreased in patients with AN and to increase with weight recuperation.

Patients with AN have elevated serum IGFBP-1 and IGFBP-2 levels that tend to normalize with weight recovery. Both are reported to be GH independent and very sensitive to nutritional regulation. The increase in IGFBP-1 in these patients is most likely related to hypoinsulinism, although other metabolic or hormonal factors, such as increased glucagon and glucocorticoids levels, as well as decreased intracellular glucose or other specific substrates may be involved. In AN, as in other forms of malnutrition, the increase in IGFBP-2 most likely depends on the combined influence of caloric-protein restriction, hypoinsulinism, and GH resistance.

Serum IGFBP-3 levels are decreased in AN patients as a consequence of GH resistance and tend to normalize after weight recovery. Indeed, all components of the trimolecular complex formed by the union between IGFBP-3, IGF, and the acid-labile subunit (ALS) are decreased. Given that these proteins are all GH dependent and regulated by the nutritional state, this is not unexpected. IGFBP-3 decreases significantly with caloric restriction, but in adults it decreases only with protein restriction. In contrast to other catabolic situations, in AN increased proteolysis of IGFBP-3 has not been observed.

Although present in serum in low concentrations, IGFBP-4 and IGFBP-5 are very important in bone formation: At the cellular level, they regulate the actions of the IGFs. In AN, serum levels of both of these IGFBPs are dramatically reduced and do not normalize with partial weight recovery.

The biological significance of the changes in IGFBPs that occur in AN or that are due to malnutrition is difficult to explain, in part because the physiological roles of the binding proteins are not totally understood. The decrease in serum IGFBP-3, and therefore the trimolecular complex IGFBP-3-IGF-ALS, impedes the retention of IGF in the vascular space, favoring the decrease in plasma levels. On the other hand, the increase in IGFBP-1 and IGFBP-2, two proteins with low molecular weights that can cross the vascular barrier, favors even more the movement of the IGFs to the tissues on which they can act. Therefore, modification of the serum and tissue levels of the IGFBPs could be one of the mechanisms by which malnutrition regulates the concentration and actions of the IGFs and the somatotrope axis.

#### **Hypothalamic-Pituitary-Thyroid Axis**

Thyroid function is affected by malnutrition in AN. Clinically, patients appear to be in a relative hypothyroid state, sometimes called euthyroid sick syndrome. Clinical manifestations include hair loss, dry skin, hypothermia, and bradycardia. All these findings are reversible with appropriate refeeding and successful treatment.

Laboratory findings include low-normal levels of thyroxine ( $T_4$ ) and thyroid-stimulating hormone (TSH), below normal levels of triiodothyronine ( $T_3$ ), and elevated levels of reverse  $T_3$ . All these findings are the result of malnutrition and weight loss. In fact, the low  $T_3$  level correlates with the amount of weight loss. The extremely reduced  $T_3$  levels in these patients are due to altered peripheral deiodination that preferentially transforms  $T_4$  into the inactive metabolite, reverse  $T_3$ . Finally, blunting of the TSH response to

TRH is indicative of hypothalamic–pituitary dysfunction. Ultrasonographic methods in patients with AN demonstrate that the thyroid is markedly reduced in comparison to that of age- and sex-matched controls. This glandular atrophy is not due to low TSH levels because TSH levels are usually normal in AN. However, thyroid size is influenced by IGF-1, and the low IGF-1 levels in these patients may contribute to thyroid atrophy. These alterations normalize with weight recovery and thyroid hormone replacement is usually not indicated.

### **Hypothalamic–Pituitary–Adrenal Axis**

Adrenocorticotropin (ACTH) and the endogenous opioids are derived from the same precursor, POMC. Increased ACTH secretion is preceded by activation of the POMC system. The opioid system has both direct and indirect influences over food intake and the level of physical activity. In laboratory animals, opioid administration stimulates appetite via receptors in the paraventricular nucleus, and opioid antagonists such as naloxone reduce appetite. Corticotropin-releasing hormone (CRH), synthesized in neurons of the paraventricular nucleus, is regulated, at least in part, by leptin and insulin. Its administration intracerebroventricularly induces a reduction in food intake and weight loss.

In AN, plasma cortisol levels are often elevated while circadian rhythm is conserved. Dexamethasone can partially suppress this hypercortisolemia, which is similar to what is observed in patients with depression and Cushing's disease. In acute situations of AN, the dexamethasone test has no medical significance; however, in patients who are gaining weight, it may have prognostic value. Refeeding studies of anorexic patients have shown that, irrespective of the initial weight, weight gains as low as 10% are associated with the normalization of cortisol secretion. The mean plasma half-life of cortisol is prolonged and ACTH levels are within the normal range, but the ACTH response to CRH is inferior to that of control patients. This hypercortisolemia may reflect a defect at the level of the hypothalamus, or even higher, that results in hypersecretion of CRF. Taken together, these observations suggest CRH hypersecretion more than cortisol resistance.

## **Osteoporosis or Osteopenia**

### **Bone Mineral Density**

Osteopenia is a frequent and often persistent complication in patients with AN, leading to clinical

fractures and increased fracture risk through life. According to the international literature, more than 50% of patients with AN present osteopenia at diagnosis.

The pathogenesis of osteopenia and osteoporosis is not completely known; however, numerous studies have highlighted a number of factors, including severe malnutrition, poor calcium intake, excessive exercise patterns despite malnutrition, hypoestrogenism, increased serum cortisol levels, and hormonal imbalances, such as decreased progesterone levels and decreased IGF-1 levels.

During infancy and puberty, obtaining adequate bone mineralization requires normal nutrition and metabolic and endocrine functions and the absence of chronic pathologies. Hormones and growth factors that play a major role in regulating bone metabolism include T<sub>3</sub>, sex steroids, vitamin D, parathormone, and GH, together with the IGF–IGFBP system and an intricate local system of growth and transcriptional factors.

The degree of osteopenia possibly depends on the age at which the amenorrhea began as well as its duration. Indeed, patients with primary amenorrhea have a more severe osteopenia than those that present with secondary amenorrhea. There are significant concerns about the lasting impact and irreversibility of osteoporosis. Evaluation of bone density is recommended in patients who have been amenorrheic for 6 months to 1 year.

Investigation of the changes in bone mineral density (BMD) showed that nonrecovered AN patients with the binge eating/purging type have a significantly reduced BMD compared to patients with the restricting type. These results suggest that patients with the binge eating/purging type are at high risk for osteoporosis and may need additional therapy to prevent bone loss.

Karllson *et al.* reported that a substantial proportion of bone mass deficit in anorexic patients was due to smaller bone size. Recovery from illness was associated with near normal bone size and volumetric BMD. However, incomplete recovery of lean and fat mass may account for part of the remaining deficit in bone size but not volumetric BMD.

Administration of estrogens and gestagens to adolescents with reduced bone mass and amenorrhea for at least 1 year indicated that osteopenia cannot be reversed. In contrast, in patients who recovered menstruation spontaneously, a 20% increase in bone mass compared to that at diagnosis was seen.

The effects of estrogens on bone metabolism have been described as inhibitory for the resorption

process, although direct effects on osteoblastic activity have been described. It has been reported that AN occurring during adolescence impairs both mineral accrual and bone size. Although reduced volumetric BMD may be related to estrogen deficiency, there was no reduced bone size after adjusting for fat and lean mass. Weight, but not estrogen use, is a significant predictor of BMD in anorexic women at all skeletal sites. The reason why estrogens are incapable of increasing bone mass in adolescents with AN and amenorrhea is unknown. It may be due to the failure to administer estrogen therapy at diagnosis, poor compliance, or perhaps the short duration of recovery. In addition, decreases in other nutritional and hormonal factors are also involved in the pathogenesis of bone mass loss; hence, estrogen replacement alone may not be sufficient for BMD recovery.

### **IGF-1 and Leptin**

IGF-1 is one of the most important regulators of bone metabolism. Circulating serum levels of IGF-1 correlate with BMD in the normal population. IGF-1 exerts a double effect on bone metabolism by stimulating osteoblastic activity and the resorption process.

Deficiency in growth factors, especially IGF-1, most likely due to the state of malnutrition, as well as the slow recuperation of their plasma levels with weight gain, is known to occur in these patients. However, it is not known whether these patients would benefit from the administration of biosynthetic GH or recombinant IGF-1. Several trials have analyzed the effect of recombinant human IGF-1 (rhIGF-1) on bone formation in AN patients. The administration of rhIGF-1 at a dose of 100  $\mu\text{g}/\text{kg}$  subcutaneously (sc) twice a day for 6 days increased metabolic markers of bone formation as well as resorption, whereas the injection of rhIGF-1 at a dose of 30  $\mu\text{g}/\text{kg}$  sc per day stimulated only bone production formation marker.

Improvement in nutritional status in AN patients via intravenous hyperalimentation therapy results in a rapid increase in serum IGF-1 levels, followed by a progressive increase in osteocalcin. This indicates that bone formation begins immediately. Nevertheless, increased bone resorption appears to continue for at least 5 weeks.

The effect of leptin administration compared to estrogen therapy in ovariectomy-induced bone loss in rats has also been reported. Leptin was effective in reducing trabecular bone loss, trabecular architectural changes, and periosteal bone formation. These findings suggest that leptin may regulate bone remodeling, and this effect may be, at least in part, mediated

by the osteoprotegerin/RANK (receptor for activation of nuclear factor kappa B) ligand pathway. RANK and RANK ligand (RANKL) are members of the tumor necrosis factor (TNF) and TNF receptor superfamilies, which are essential for osteoclast differentiation. In the bone microenvironment, the stimulatory effects of RANKL are neutralized by the secreted decoy receptor, osteoprotegerin (OPG). It follows that the balance between OPG and RANKL secretion by stroma cells is critical for the regulation of osteoclast formation.

The dramatic decline in leptin levels observed in AN may be one of the major hormonal factors involved in the pathogenesis of the associated bone fragility through diminishing cortical bone formation rates and skeletal growth. Leptin may play an important protective role in bone metabolism by inhibitory bone resorption.

### **Bone Markers**

The available data are few and contradictory regarding markers of bone formation. The bone isoenzyme of alkaline phosphatase (bAP) and the amino-terminal pro-peptide of procollagen I (PNIP) have the greatest diagnostic sensitivity in detecting anomalies in bone remodeling, at least in osteoporetic women. Among the markers of bone resorption, the telopeptide carboxy terminal of the 1 chain of type 1 collagen (CTX) has demonstrated great specificity and sensitivity in the investigation of bone metabolism.

Osteoporosis in adolescents with AN is related to decreased bone formation and actual bone resorption. It is clear that markers of bone formation are decreased and markers of bone mineral resorption are increased. Trabecular bone seems to be more vulnerable than cortical bone. A significant positive correlation of BMI, IGF-1, and IGFBP-3 with osteocalcin as a bone formation marker was demonstrated, whereas a negative correlation for the bone resorption marker CrossLaps was found only with BMI.

Patients with AN exhibit a high degree of osteopenia that correlates with their bAP levels. Furthermore, the urinary fragments of CTX in patients with AN are derived primarily from old bone ( $\beta$ -CTX), whereas in young adolescents they are primarily from new bone ( $\alpha$ -CTX). Therefore,  $\alpha$ -CTX is more adequate for measuring bone resorption in controls, whereas  $\beta$ -CTX is more adequate in anorexic patients.

One study demonstrated that during undernutrition and amenorrhea, with low IGF-1 and extremely low circulating estradiol, biochemical markers of bone metabolism indicated a shift toward

bone resorption because bone formation markers (osteocalcin, bAP, and PNIP) were normal and bone resorption markers (CTX) elevated. However, in fully recovered patients, bone metabolism markers indicate accelerated bone turnover with an increase in both formation and resorption. Bone formation markers correlated positively with IGF-1 and bone resorption markers negatively with estradiol, indicating that IGF-1 is a major bone formation stimulator, whereas estradiol action predominately inhibits bone resorption.

In patients with AN, the mechanism of bone loss does not appear to be due to an increase in absorption over formation. It is possible that the observed increase in bone remodeling is a mechanism developed in an attempt to restore bone mass. However, the large deficit of calcium in these patients (the loss of exogenous sources due to the deficit in alimentation produces the liberation of bone calcium to maintain the homeostasis of the extracellular fluid) and the deficit in amino acids as a result of fasting make it very difficult to restore bone mass. The best predictors for osteopenia are BMI and the duration of amenorrhea, followed by the duration of regular menses before amenorrhea.

Treatment protocols emphasize the importance of nutritional rehabilitation and recovery. The addition of 1500 mg/day of calcium with vitamin D is recommended and lifestyle counseling with an emphasis on smoking cessation is relevant. The issue of exercise is controversial; for most women, including teens, weight-bearing exercise is an important factor contributing to bone mineral accretion and is more important than calcium intake in adolescent bone health. Hormone replacement therapy with oral contraceptives has been studied as a treatment for osteopenia/osteoporosis; results have been mixed. Some preliminary studies have demonstrated a minor protective effect on bone density of the lumbar spine, whereas others have not.

## TREATMENT

An integrated treatment program should be instituted and carried out by a multidisciplinary team, including a pediatrician, endocrinologist, psychiatrist, psychologist, nurse, and possibly others. It is very important that the correct diagnosis is made and that the patient and family are made aware of the importance of the disease and the various aspects of treatment. They should be aware that treatment will necessarily last for a period not less than 5 years.

The role of the physician in the control of this process and the establishment of an adequate relationship with the patient and family is fundamental for successful treatment. Treatment should be on an outpatient basis if the diagnosis was made early, the degree of undernutrition is not too severe, and the mental disturbance is not incapacitating. If this is not the case, the patient must be hospitalized.

Treatment objectives should have a strict priority: prevent the death of the patient, prevent the disease from becoming chronic, and start physical and mental recuperation.

## Nutritional Treatment

### **Refeeding**

To successfully begin refeeding, it is fundamental that a therapeutic alliance with the patient be established in which the patient understands and accepts that he or she has a disease. This can be accomplished by asking the patient if he or she has the various signs and symptoms of AN so that, little by little, the patient can identify with this condition.

The physician should use height and weight graphs to explain what percentile the patient is currently in and what the patient's weight should be for his or her age and height. A target weight, acceptable to both the patient and the physician, should be agreed upon. The patient's understanding of caloric requirements to maintain a normal weight should be explored because very few possess knowledge of the appropriate requirements for their height, age, and sex. Patients must understand that their growth and physical activity depend on the adequate intake of calories, including proteins, fats, carbohydrates, and vitamins and minerals.

### **Obtaining and Maintaining an Adequate Weight**

After the target weight has been achieved, a maintenance diet must be prescribed, and intake of foods not on the diet and eating between scheduled meals must be prohibited. This helps the patient overcome the fear of losing control or gaining weight.

## Psychological Treatment

Psychotherapeutic help can consist of individual, familial, or group treatment. In AN, the initial phases of psychotherapy consist of helping the patient begin refeeding by alleviating the feelings of guilt of the patient or family. The next step is to treat the overall

psychological well-being of the patient such that eventually the medical problems become secondary. It is fundamental to work with the family using familial therapy or counseling.

### Pharmacological Treatment

Psychotropic medication should not be used as the only or principal treatment but, rather, as a prevention of relapse in patients who have gained weight or for treatment of symptoms associated with AN, such as depression or obsessive-compulsive behavior. A variety of substances have been used to treat AN because it is basically unresponsive to pharmacological treatment.

### PROGNOSIS AND EVOLUTION

The normal evolution of AN consists of cycles of recuperation and relapse even in the best controlled cases, and the evolution rarely lasts less than 2 or 3 years. Approximately 50% of AN patients achieve total recuperation, whereas 20% have residual problems and 30% become chronic. The mortality rate is between 0.5 and 1% per year of observation. The most frequent causes of death are severe undernutrition, gastrointestinal complications, infections, and suicide. With prolonged follow-up, the mortality rate increases slightly (up to 20% for patients older than 20 years of age).

### CONCLUSION

The Academy Pediatric Association recommends a multidisciplinary treatment. The team should include a medical physician, dietician, and psychiatrist/psychologist. Evidence suggests that there is a hypothalamic dysfunction in patients with AN, and in general this normalizes with weight recovery. Disturbances in neurotransmitter, neuropeptide, and neuroendocrine systems have been reported in acutely ill patients and in patients during follow-up.

Due to scarce macro- and micronutrients, intact nonvital processes that increase energy output and are not necessary for survival, such as growth, pubertal development, and reproduction, are shut down until the nutritional situation improves. Early implementation of appropriate psychological and nutritional therapy and the best treatment for osteopenia/osteoporosis in these patients is necessary. The subtypes of AN and BMI at follow-up are the best

predictors of bone mineral density, and leptin may play an important protective role in bone metabolism.

Analysis of the genetic mechanisms underlying weight regulation is progressing rapidly. The genetic analysis of AN may help to define new drug targets and therefore lead to new treatment strategies.

The prognosis for adolescents with eating disorders is favorable. With early and aggressive treatment, most adolescents with AN recover, as demonstrated by a number of long-term follow-up studies at adolescent medicine centers.

### See Also the Following Articles

Body Proportions • Constitutional Delay of Growth and Puberty (CDGP) • Eating Disorders and the Reproductive Axis • Obesity, Childhood and Adolescence • Obesity Regulation • Osteoporosis, Overview • Puberty: Physical Activity and Growth

### Further Reading

- American Psychiatric Association. (1994) "Diagnostic and Statistical Manual of Mental Disorders." American Psychiatric Association Washington, DC.
- Argente, J., Barrios, V., Chowen, J. A., Sinha, M. K., and Considine, R. V. (1997). Leptin plasma levels in healthy Spanish children and adolescents, children with obesity and adolescents with anorexia nervosa and bulimia nervosa. *J. Pediatr.* **131**, 833–838.
- Argente, J., Caballo, N., Barrios, V., Muñoz, M. T., Pozo, J., Chowen, J. A., Morandé, G., and Hernández, M. (1997). Multiple endocrine abnormalities of the growth hormone and insulin-like growth factor axis in patients with anorexia nervosa: Effect of short- and long-term weight recuperation. *J. Clin. Endocrinol. Metab.* **82**, 2084–2092.
- Audí, L., Vargas, D. M., Gussinyé, M., Teste, D., Martí, G., and Carrascosa, A. (2002). Clinical and biochemical determinants of bone metabolism and bone mass in adolescent female patients with anorexia nervosa. *Pediatr. Res.* **51**, 497–504.
- De la Piedra, C., Calero, J. A., Traba, M. L., Asensio, M. D., Argente, J., and Muñoz, M. T. (1999). Urinary and C-telopeptides of collagen I: Clinical implications in bone remodeling in patients with anorexia nervosa. *Osteop. Int.* **10**, 480–486.
- Ducy, P., Ambling, M., Takeda, S., Priemel, M., Schilling, A. F., Beil, F. T., *et al.* (2000). Leptin inhibits bone formation through a hypothalamic relay: A central control of bone mass. *Cell* **100**, 197–207.
- Gorwood, P., Ades, J., Bellodi, L., Cellini, E., Collier, D. A., DiBella, D., *et al.* (2002). The 5-HT (2A)-1438 G/A polymorphism in anorexia nervosa: A combined analysis of 316 trios from six European centres. *Mol. Psychiatr.* **7**, 90–94.
- Karlsson, M. K., Weigall, S. J., Duan, Y., and Seeman, E. (2000). Bone size and volumetric density in women with anorexia nervosa receiving estrogen replacement therapy and in women recovered from anorexia nervosa. *J. Clin. Endocrinol. Metab.* **85**, 3177–3182.

- Muñoz, M. T., and Argente, J. (2002). Anorexia nervosa: Hypogonadotropic hypogonadism and bone mineral density. *Hormone Res.* **57**(Suppl.), 57–62.
- Muñoz, M. T., Morandé, G., Garcís, F., Pozo, J., and Argente, J. (2002). The effects of estrogen administration on bone mineral density in adolescents with anorexia nervosa. *Eur. J. Endocrinol.* **46**, 45–50.
- Otto, B., Cuntz, U., Fruehauf, E., Wawarta, R., Folwaczny, C., Riepl, R. L., *et al.* (2001). Weight gain decreases elevated plasma ghrelin concentrations of patients with anorexia nervosa. *Eur. J. Endocrinol.* **145**, R5–R9.
- Pozo, J., and Argente, J. (2002). Delayed puberty in chronic illness. *Best Pract. Res. Clin. Endocrinol. Metab.* **16**(1), 73–90.
- Støving, R. K., Flyvbjerg, A., Frystyc, J., Fisker, S., Hangaard, J., Hansen-nord, M., and Hagen, C. (1999). Low serum levels of free and total insulin-like growth factor I (IGF-I) in patients with anorexia nervosa are not associated with increased IGF-binding protein-3 proteolysis. *J. Clin. Endocrinol. Metab.* **84**, 1346–1350.



# Anti-Müllerian Hormone

Nathalie Josso and Nathalie di Clemente  
Unité INSERM 693, Clamart, France

## Glossary

**activin receptor-like kinases (ALKs)** ALKs act as signaling (type I) receptors for the transforming growth factor- $\beta$  (TGF- $\beta$ ) family. They may be involved in the signaling pathway of several ligands [e.g., bone morphogenetic proteins (BMPs) and anti-Müllerian hormone (AMH)].

**bone morphogenetic proteins (BMPs)** Members of the TGF- $\beta$  family; they are involved in many morphogenetic processes, namely cartilage and gonadal differentiation.

**GATA-4** Members of the GATA family of proteins (GATA-1–6) recognize a consensus sequence on DNA due to conserved zinc fingers in their DNA-binding domain. They play important roles in erythroid cell differentiation. GATA-4 is abundantly expressed in the Sertoli cells of the differentiating testis and binds to the AMH promoter through a conserved proximal element.

**high-mobility group (HMG)** A 79-amino acid DNA-binding domain shared by SOX proteins and SRY.

**Müllerian ducts** Precursors of the uterus, Fallopian tubes, and upper part of the vagina; the main target organ of AMH.

**Smad proteins** Smad proteins are a family of intracellular signal mediators for members of the TGF- $\beta$  family. They are classified into three broad categories. Receptor-specific Smads bind specifically either to type I receptors of TGF- $\beta$  or activin (Smads 2 and 3) or to members of the BMP family (Smads 1, 5, and 8). After activation by their specific type I receptor, they bind to a co-Smad (Smad-4) and enter the nucleus to modulate gene transcription. Inhibitory Smads, Smads 6 and 7, prevent the activation of receptor-specific Smads.

**SOX-9** Members of the SOX gene family are critical transcription factors for many developmental processes. SOX-9 is a key regulator of cartilage and testis development; mutations in humans lead to a malformative syndrome known as campomelic dysplasia and to gonadal aplasia in XY subjects. SOX-9 is a powerful activator of AMH transcription.

**steroidogenic factor-1 (SF-1)** Also known as Ad4BP, SF-1 is an orphan nuclear receptor that activates AMH transcription. It also activates P450 steroid hydroxylases and aromatase. In its absence, gonads, adrenals, the pituitary, and the hypothalamus do not develop normally.

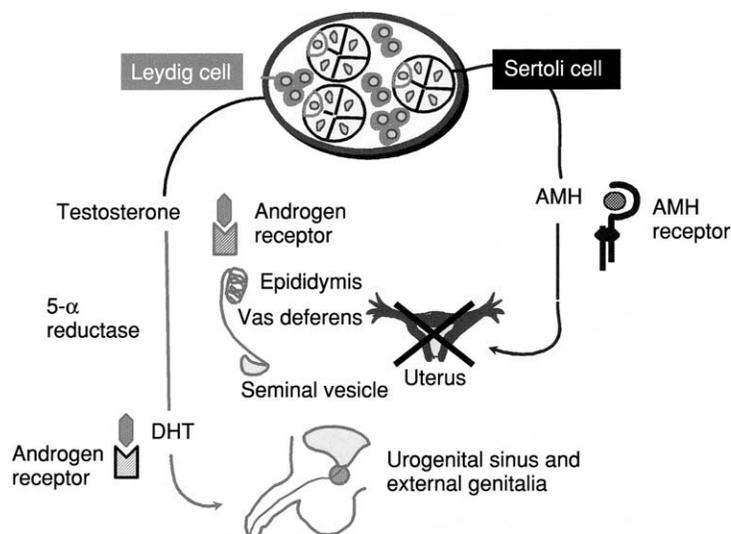
Anti-Müllerian hormone, also called Müllerian inhibiting substance, is a member of the transforming growth factor- $\beta$  family, whose main effect is the initiation of regression of the Müllerian ducts in the male fetus.

## INTRODUCTION

Anti-Müllerian hormone (AMH), also known as Müllerian inhibiting substance, is a “sexually specialized” member of the transforming growth factor- $\beta$  (TGF- $\beta$ ) family in the sense that it is expressed exclusively by gonadal somatic cells and that its effects are targeted to the gonads and reproductive tract. Its main biological effect, the inhibition of Müllerian duct development in male fetuses, represents the first

step of phenotypic male sex differentiation. Müllerian ducts are sensitive to AMH only during a short developmental window, before 8 weeks in the human fetus.

The existence of a specific factor dedicated to Müllerian regression was predicted in 1953 by Alfred Jost, who proposed that the fetal testis has a dual hormonal secretion: testosterone, responsible for the virilization of Wolffian ducts and external genitalia, and a second hormone, responsible for Müllerian regression (Fig. 1), that was identified and purified only many years later. AMH inhibits meiotic prophase maturation and the growth of the fetal ovary, in which it sometimes induces the differentiation of structures resembling seminiferous tubules. AMH also has effects postnatally. In the male, it inhibits the differentiation and function of Leydig cells; in



**Figure 1** (Right) The role of AMH in fetal sex differentiation. AMH, secreted by fetal Sertoli cells, binds to an AMH receptor located on the membrane of target organs and induces the regression of Müllerian ducts, which would otherwise give rise to the uterus. (Left) The effect of testosterone secreted by Leydig cells. After binding to the androgen receptor, testosterone induces the differentiation of male excretory ducts and masculinizes the penis and urogenital sinus after reduction to DHT.

the female, it inhibits follicular maturation and breast cancer cell growth. However, the postnatal effects of AMH are not crucial because testicular or ovarian functions are not disrupted by mutations of AMH or its receptors.

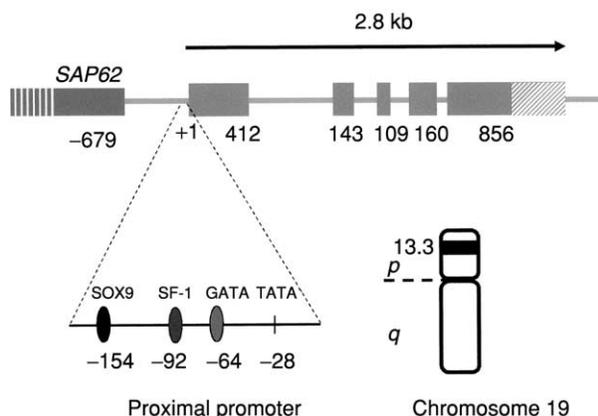
## BIOCHEMISTRY AND MOLECULAR BIOLOGY OF AMH AND ITS RECEPTORS

AMH is a 560-amino acid glycoprotein formed by two 70-kDa monomers linked by disulfide bonds. The hormone is cleaved at a proteolytic site 109 amino acids upstream of the C terminus, yielding a short bioactive C-terminal domain with homology to other members of the TGF- $\beta$  family and a long N-terminal prodomain with no bioactivity of its own but which enhances the bioactivity of the C terminus.

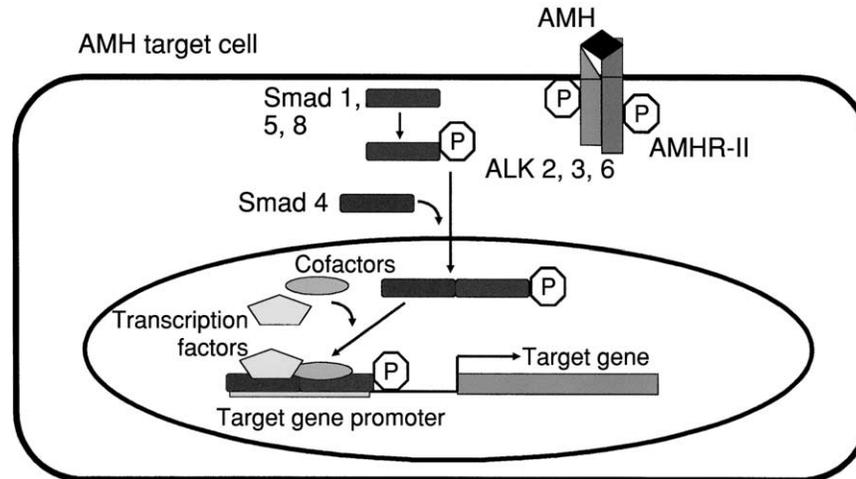
The human AMH gene was cloned in 1986. It is located on the tip of the short arm of chromosome 19, band p13.3. Only 2.75 kilobase pairs (kbp) in length, it contains five exons; the 3'-end of the fifth one is particularly GC rich and codes for the bioactive C-terminal domain. The 200-bp minimal promoter is flanked by a housekeeping gene, *SAP-62*, coding for a spliceosome, and contains binding sites for various transcription factors (Fig. 2).

In keeping with its status as a TGF- $\beta$  family member, AMH signals through two membrane bound receptors, both with serine/threonine kinase activity. Paradoxically, receptor type II is the primary

receptor; it binds to AMH and then recruits the signaling, so-called type I, receptor. The gene for the type II receptor, AMHR-II, cloned in 1994, is AMH specific; it is divided into 11 exons, has a molecular mass of 82 kDa, and maps to chromosome 12q13. Based on evidence obtained from different tissues and cell lines, it is believed that three type I receptors—ALK2,-3, and-6—involved in the signaling pathway of the bone morphogenetic protein (BMP) family also function as AMH type I receptors. As expected, all three signal through the BMP receptor-specific Smad molecules 1, 5, and 8 (Fig. 3).



**Figure 2** The AMH gene. The 3'-end of exon 5 (hatched) codes the 109-amino acid biologically active C terminus. The 200-bp promoter contains binding sites for transcription factors and is adjacent to a housekeeping gene, *SAP62*. AMH maps to chromosome 19 p13.3.



**Figure 3** AMH signal transduction. AMH binds to autophosphorylating receptor type II (AMHR-II), which then recruits a type I receptor, ALK 2, 3, or 6. The complex transiently binds and phosphorylates a receptor-specific Smad (1, 5, or 8), which then binds to a common Smad (co-Smad), Smad 4. The receptor-specific and co-Smad then enter the nucleus and bind to the promoter of target genes with the help of cofactors and transcription factors.

## EXPRESSION AND REGULATION

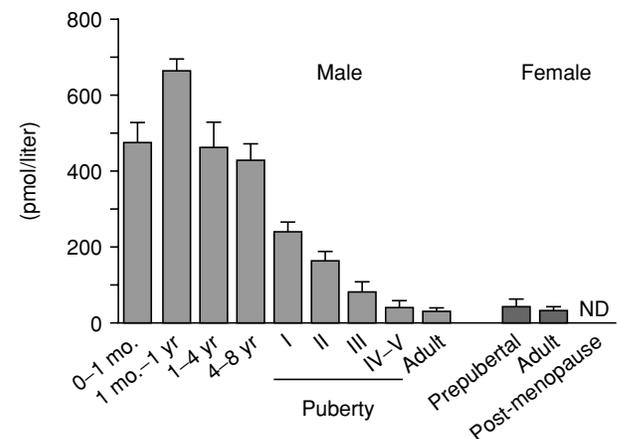
In the male, AMH is produced in large amounts by Sertoli cells just prior to the differentiation of fetal seminiferous tubules until puberty (Fig. 4). In sexually mature individuals, Sertoli cells continue to synthesize small amounts of AMH; after puberty, however, the hormone is preferentially secreted toward the lumen of the seminiferous tubule. Because of the formation of the blood–testis barrier, AMH concentration is higher in the seminal plasma than in serum. AMHR-II is expressed in the mesenchymal cells surrounding the fetal Müllerian duct, in Sertoli cells, and, to a lesser degree, in Leydig cells.

In the female, low amounts of AMH and AMHR-II are coexpressed by granulosa cells of preantral and small antral follicles. Expression decreases slightly as follicle maturation progresses. When follicles disappear from the ovary at the end of reproductive life, AMH is no longer detectable in serum.

The chronological expression of AMH is tightly regulated. In the male, it is essential that the hormone be produced during the short window of Müllerian responsiveness. Several transcription factors—*SOX-9* (a member of the high-mobility group gene family), steroidogenic factor-1, and *GATA-4*, all of which are expressed in the developing testis but not the ovary—bind to response elements on the AMH promoter (Fig. 2) and cooperate to trigger the early initiation of AMH transcription in the testis. After birth, AMH is regulated positively by follicle-stimulating hormone (FSH) and negatively by androgens. During puberty, when Leydig cells begin to produce testosterone, the

expression of AMH by Sertoli cells is repressed, provided normal androgen receptors are present. In their absence, the stimulatory influence of FSH becomes apparent.

In the female, initiation of AMH expression begins as soon as follicle maturation reaches the preantral stage, either after birth or in the last weeks of pregnancy. Little is known, however, about the mechanism of regulation.



**Figure 4** Serum AMH in postnatal life measured by ELISA. Serum AMH can be measured by ELISA in both sexes. In the human male, serum AMH is high until puberty, except in the first month of life. It decreases slowly until puberty, when its serum concentration is inversely correlated to pubertal maturation. In the adult male, serum AMH is very low or undetectable. In the female, serum AMH is very low and becomes undetectable after menopause.

## AMH IS A MARKER OF GONADAL FUNCTION

AMH is produced postnatally by the gonad of both sexes; ascertainment of its serum level is useful to assess gonadal function.

### Testicular Function

#### **Seminiferous Tubule Function in Prepubertal Subjects**

AMH is expressed constitutively by prepubertal Sertoli cells. It follows that its circulating level reflects the presence and function of seminiferous tubules, with no need for prior stimulation by chorionic gonadotropin (Fig. 4). Pediatric endocrinologists assay serum AMH in children with impalpable testes to discriminate between anorchia and bilateral cryptorchidism and in patients with ambiguous genitalia to distinguish testicular dysgenesis from disorders affecting testosterone synthesis or sensitivity.

#### **Androgen Action**

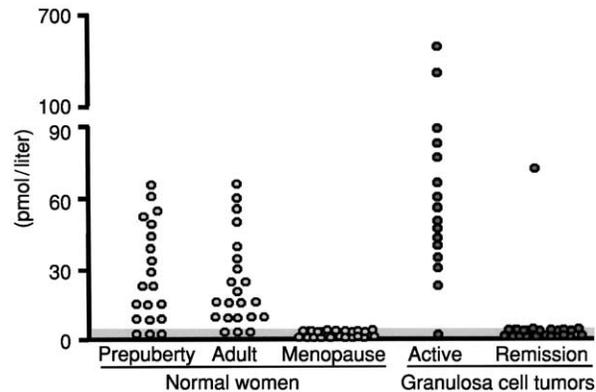
Normally, AMH is down-regulated by testosterone produced by Leydig cells during puberty. A significant reduction of serum AMH is observed when testosterone concentration reaches 7 nmol/ng/liter. If testosterone is not synthesized or if the androgen receptor is abnormal, not only does serum AMH fail to decrease but also it increases higher than normal childhood levels because FSH is now produced and stimulates AMH transcription in Sertoli cells. Serum AMH may also be used as a sensitive marker for the effectiveness of medication inactivating the androgen receptor, such as cyproterone acetate.

#### **Azoospermia**

In oligospermic adult males, the level of AMH in seminal fluid is correlated with sperm concentration. In patients with nonobstructive azoospermia, it is higher if spermatozoa are detectable in a testicular biopsy. Thus, it may aid in predicting the success of an intracytoplasmic sperm injection procedure.

### Adult Ovarian Function

In women, serum AMH reflects the function and number of granulosa cells. Between birth and menopause, serum AMH levels are less than 75 pmol/liter or may be undetectable. This is always the case after natural menopause or ovariectomy (Fig. 5).



**Figure 5** Serum AMH in women with or without granulosa cell tumors. In normal women, serum AMH is low (<75 pmol/liter) or undetectable. It is always undetectable after menopause. In women with an active recurrence of a granulosa cell tumor, serum AMH is very high except during remission. Serum AMH is a sensitive marker of granulosa cell tumor recurrence, preceding the emergence of clinical symptoms.

#### **Granulosa Cell Tumors**

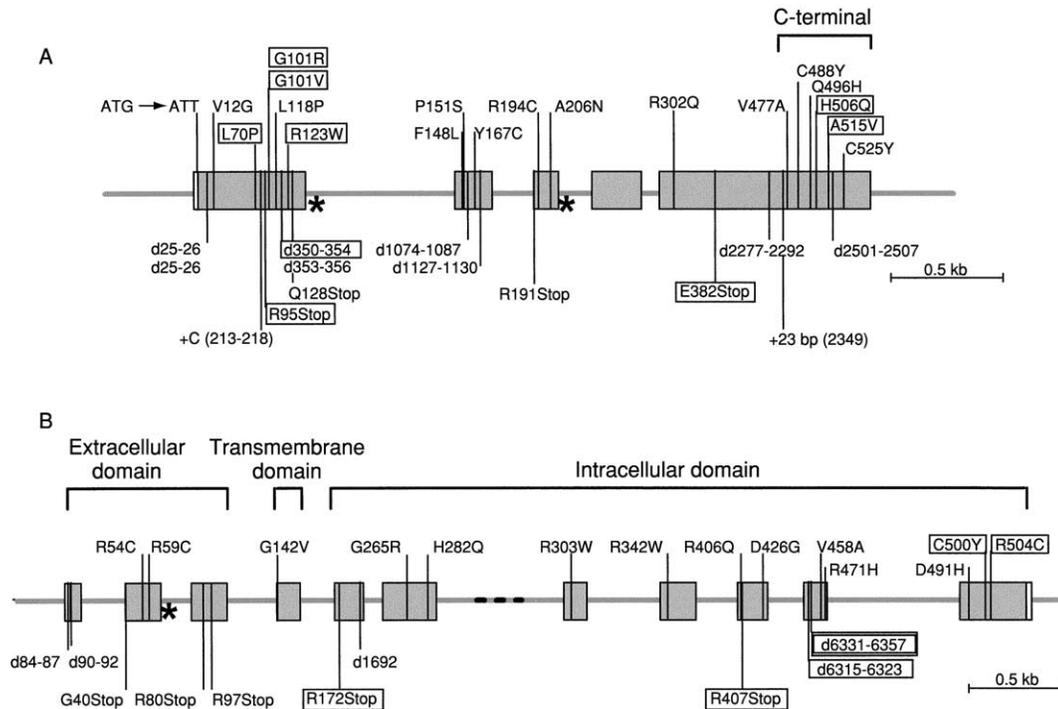
Granulosa cell tumors are relatively rare, but they tend to recur after treatment. Serum AMH is a sensitive and early marker of evolutive granulosa cell tumors (Fig. 5) and allows the detection of recurrence before clinical signs become evident.

#### **Assessment of Ovarian Reserve: Application to *In Vitro* Fertilization**

Studies of the human ovary suggest a correlation between the stock of resting primordial follicles (ovarian reserve) and the number of small growing follicles. No markers of the former exist, but because AMH is produced exclusively by small preantral follicles, it has been used as a marker of the depletion in follicle stock associated with ovarian aging. Serum AMH has also been shown to be a predictive factor for the number of oocytes retrieved during an *in vitro* fertilization procedure. In contrast, serum AMH is elevated in women with polycystic ovary syndrome.

#### **Persistent Müllerian Duct Syndrome: Mutations of AMH and the AMH Receptor Gene**

Persistent Müllerian duct syndrome is a rare intersex condition that is transmitted in an autosomal recessive mode and characterized by the persistence of uterus and Fallopian tubes in otherwise normally virilized genetic males. The condition is discovered at surgery performed for cryptorchidism and/or inguinal hernia. Genetic analysis (Fig. 6) usually reveals mutations of either the AMH gene, associated with a very low or



**Figure 6** AMH and AMHR-II mutations in the persistent Müllerian duct syndrome. Missense mutations are shown above the gene; insertions or deletions, usually leading to a stop codon, are shown below. Splice mutations are indicated by an asterisk. Boxed mutations are those found in more than one family. (A) AMH gene mutations. Mutations affect all exons, especially the 3'-end of exon 5, which codes the bioactive C-terminal fragment, and the 3' half of exon 1. Most mutations are restricted to a given family. (B) AMHR-II mutations. Again, the mutations are spread more or less evenly along the gene. Many mutations are specific for a given family, with the exception of the deletion between bases d6331 and d6357, which is found in the homozygous or heterozygous state in approximately half of patients with an AMHR-II mutation.

undetectable serum AMH concentration, informative only before puberty, or the type II AMH receptor (AMHR-II), associated with normal serum AMH levels. The proportion of AMH and AMHR-II mutations is approximately equal. The goal of treatment should be the preservation of fertility. This involves replacement of the testes in the scrotum and requires careful partial resection of the Müllerian ducts. The uterus should never be removed *in toto* because the male excretory ducts are usually contained in the uterine wall. Women with homozygous mutations of either the AMH or the AMHR-II gene are normal and fertile.

## CONCLUSION

Initially associated exclusively with male sex differentiation, AMH has gained recognition as a sensitive marker of gonadal function, useful in pediatric endocrinology, gynecology, and andrology. The availability of a commercial enzyme-linked immunosorbent

assay kit places AMH assay within the reach of the average clinician.

## See Also the Following Articles

Genes and Gene Defects Affecting Gonadal Development and Sex Determination • In Vitro Fertilization (IVF) • Ovarian-Follicular Apparatus • Sexual Function and Androgens • Sexual Maturation, Female • Sexual Maturation, Male

## Further Reading

- Al-Attar, L., Noël, K., Dutertre, M., Belville, C., Forest, M. G., Burgoyne, P. S., Josso, N., and Rey, R. (1997). Hormonal and cellular regulation of Sertoli cell anti-Müllerian hormone production in the postnatal mouse. *J. Clin. Invest.* **100**, 1335–1343.
- Arango, N. A., Lovell-Badge, R., and Behringer, R. R. (1999). Targeted mutagenesis of the endogenous mouse *Mis* gene promoter: In vivo definition of genetic pathways of vertebrate sexual development. *Cell* **99**, 409–419.
- Baarends, W. M., Uilenbroek, J. T. J., Kramer, P., Hoogerbrugge, J. W., Vanleeuwen, E. C. M., Themmen, A. P. N., and Grootegoed, J. A. (1995). Anti-Müllerian hormone and anti-Müllerian

- hormone type II receptor messenger ribonucleic acid expression in rat ovaries during postnatal development, the estrous cycle, and gonadotropin-induced follicle growth. *Endocrinology* **136**, 4951–4962.
- Cate, R. L., Mattaliano, R. J., Hession, C., Tizard, R., Farber, N. M., Cheung, A., Ninfa, E. G., Frey, A. Z., Gash, D. J., Chow, E. P., Fisher, R. A., Bertoni, J. M., Torres, G., Wallner, B. P., Ramachandran, K. L., Ragin, R. C., Manganaro, T. F., MacLaughlin, D. T., and Donahoe, P. K. (1986). Isolation of the bovine and human genes for Müllerian inhibiting substance and expression of the human gene in animal cells. *Cell* **45**, 685–698.
- de Vet, A., Laven, J. S. E., de Jong, F. H., Themmen, A. P. N., and Fauser, B. C. J. M. (2002). Antimüllerian hormone serum levels: A putative marker for ovarian aging. *Fertil. Steril.* **77**, 357–362.
- Fujisawa, M., Yamasaki, T., Okada, H., and Kamidono, S. (2002). The significance of anti-Müllerian hormone concentration in seminal plasma for spermatogenesis. *Hum. Reprod.* **17**, 968–970.
- Jost, A. (1953). Problems of fetal endocrinology: The gonadal and hypophyseal hormones. *Rec. Prog. Horm. Res.* **8**, 379–418.
- Long, W. Q., Ranchin, V., Pautier, P., Belville, C., Denizot, P., Cailla, H., Lhommé, C., Picard, J. Y., Bidart, J. M., and Rey, R. (2000). Detection of minimal levels of serum anti-Müllerian hormone during follow-up of patients with ovarian granulosa cell tumor by means of a highly sensitive ELISA. *J. Clin. Endocrinol. Metab.* **85**, 540–544.
- Rajpert-de Meyts, E., Jorgensen, N., Graem, N., Müller, J., Cate, R. L., and Skakkebaek, N. E. (1999). Expression of anti-Müllerian hormone during normal and pathological gonadal development: Association with differentiation of Sertoli and granulosa cells. *J. Clin. Endocrinol. Metab.* **84**, 3836–3844.
- Rey, R. A., Belville, C., Nihoul-Fékété, C., et al. (1999). Evaluation of gonadal function in 107 intersex patients by means of serum anti-Müllerian hormone measurement. *J. Clin. Endocrinol. Metab.* **84**, 627–631.



# Antiadrenergic Agents

P. A. van Zwieten

University of Amsterdam, Amsterdam, The Netherlands

## Glossary

**adrenergic nervous system** The sympathetic part of the autonomous nervous system, where the neurotransmitters (norepinephrine and epinephrine) are adrenergic.

**cholinergic nervous system** The parasympathetic part of the autonomic nervous system, where the neurotransmitter (acetylcholine) is cholinergic.

**pre/postganglionic** Anatomical position of the sympathetic neuron with respect to the ganglion.

**pre/postsynaptic** Anatomical position of receptors with respect to the synapse/nerve ending.

**synapse** Anatomic/functional structure allowing the transition of nerve activity from the nerve ending (postganglionic) through the synaptic cleft and toward the target organ (blood vessel, heart, etc.).

**A**ntiadrenergic agents are drugs that suppress the activity of the sympathetic (adrenergic) nervous system, targeting the cardiovascular system and other organs.

## INTRODUCTION

The autonomic nervous system is responsible for the homeostasis/regulation of several internal organs, tissues, and metabolic systems. Well-known examples of such targets of the autonomic nerves are the cardiovascular system (heart and blood vessels), the gastrointestinal system (stomach, intestine, biliary duct, etc.), the respiratory tract (lungs and airways), and glucose/lipid metabolism.

The autonomic nervous system is subdivided into the sympathetic and parasympathetic systems. The sympathetic or adrenergic system is activated with the purpose of adapting the individual organism for “flight, fright, or fight”—that is, the short-term availability and release of energy and alertness. In contrast, the parasympathetic nervous system facilitates the

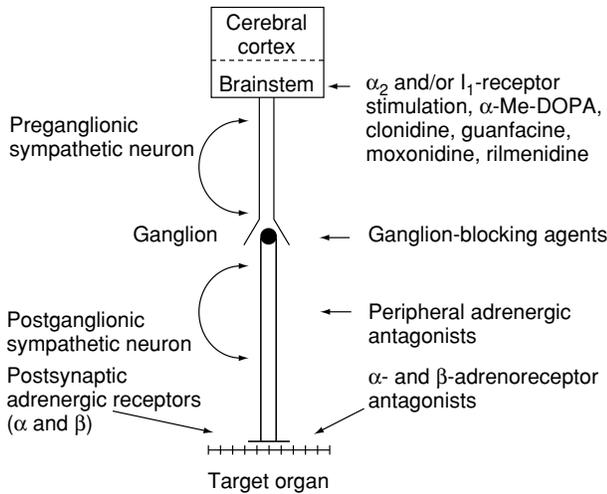
storage of energy and metabolic components to allow the release of energy in a situation of flight, fright, or fight. As such, the sympathetic and parasympathetic systems counteract each other’s activities in a sort of “yin-yang” balance. Activation of the sympathetic nervous system implies the stimulation of the cardiovascular system, thus increasing heart rate, cardiac output, cardiac contractility, and blood pressure. During sympathetic stimulation, the activity of the gastrointestinal system is attenuated. Stimulation of the parasympathetic system implies the activation of the digestive system, particularly the gastrointestinal tract, with the aim of facilitating the storage of energy. During parasympathetic activation, the cardiovascular system’s activity is reduced.

Both the sympathetic and parasympathetic systems are subject to modulation by the central nervous system, particularly the brainstem. However, higher (cortical) brain centers can also modulate both subgroups of the peripheral autonomic nervous system.

Several diseases are associated with dysregulation of either subgroup of the autonomic nervous system. Relevant examples of this pathophysiological involvement particularly concern the sympathetic nervous system, which is known to be overactive in important cardiovascular derangements or diseases, such as hypertension, congestive heart failure, or cardiac tachyarrhythmia. For this reason, a vast body of research has accumulated with the aim of finding drugs that can attenuate the hyperactivity of the sympathetic nervous system. Drugs with such activity are called antiadrenergic agents.

## SYMPATHETIC (ADRENERGIC) NERVOUS SYSTEM AND ITS RECEPTORS

The brainstem, consisting predominantly of the pons/medullary region, controls the peripheral sympathetic nervous system (Fig. 1). The peripheral sympathetic nervous system contains a kind of relay station called



**Figure 1** Schematic representation of the sympathetic (adrenergic) nervous system. Adrenergic structures and receptors in the brainstem modulate the activity of the peripheral sympathetic nervous system. The targets of all relevant antiadrenergic drugs are shown. Virtually every structure/receptor of the system can be manipulated selectively by drugs.

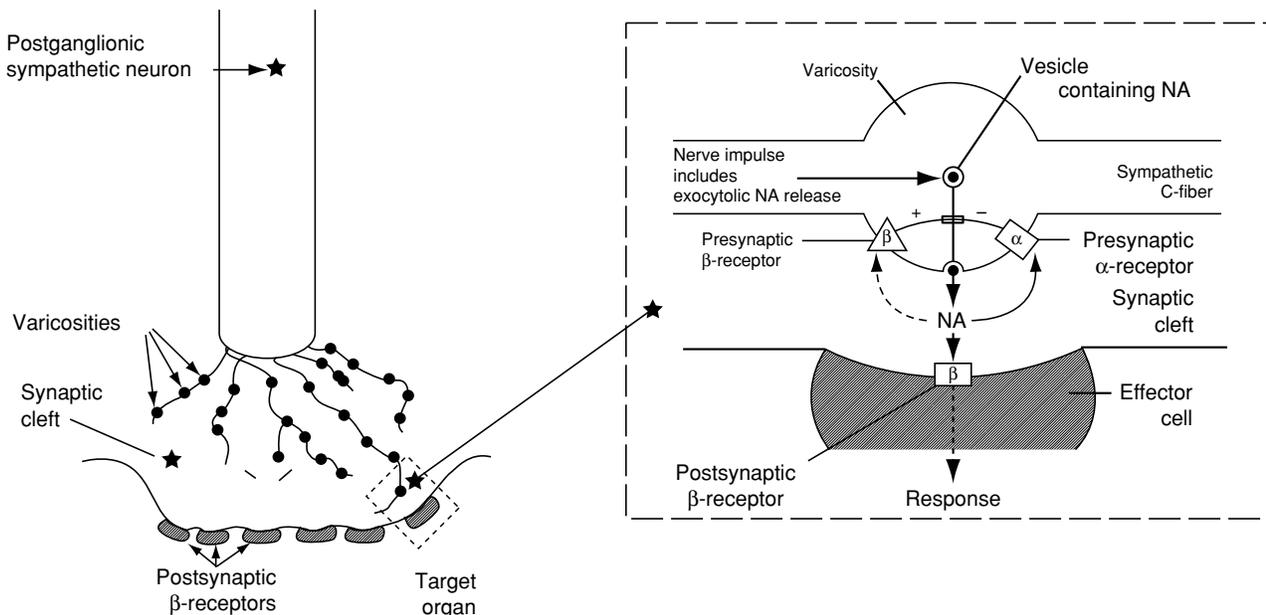
the sympathetic ganglion. Transmission in the ganglion is performed by the neurotransmitter acetylcholine (ACh), which targets the nicotine-like ACh<sub>x</sub>

receptors. Peripheral sympathetic neurons are subdivided into pre- and postganglionic types according to their anatomical position with respect to the ganglion.

The sympathetic activity is transferred from the postganglionic neuron to the target organ (e.g., the heart or a blood vessel). The anatomical transition structure required for this process is called the synapse (Fig. 2). The process of humoral neurotransmission within the synapse occurs as follows: The neurotransmitter norepinephrine (NE; also called noradrenaline) is released from the postsynaptic nerve ending, diffuses through the synaptic cleft, and then binds to the receptors in the target organ. The receptor thus activated mediates a physiological/functional process (e.g., constriction of a blood vessel) or an increase in cardiac frequency or contractility.

The receptor in the target organ is activated by the endogenous neurotransmitter NE but also by synthetic drugs, such as methoxamine, phenylephrine, and isoprenaline, that imitate the effects of NE. These activator drugs are called sympathomimetic agents. They are agonists with respect to the receptor.

Antagonists are chemical compounds that block the stimulating effect of NE (the neurotransmitter) and that of sympathomimetic agents. The antagonist or blocker binds to the receptor, thus preventing the stimulating effect of the endogenous neurotransmitter



**Figure 2** Schematic representation of a sympathetic (adrenergic) synapse. Nerve activity releases the endogenous neurotransmitter from the varicosities at the nerve endings. Stimulation of postsynaptic  $\alpha$ -adrenoceptors causes a physiological/pharmacological effect (e.g., vasoconstriction). Blockade of the  $\alpha$ -adrenoceptor (postsynaptic) counteracts the effect of norepinephrine. Stimulation of presynaptic  $\alpha$ -adrenoceptors inhibits the release of the neurotransmitter. Conversely, presynaptic  $\alpha$ -receptor blockade stimulates the release of the neurotransmitter.

Table I Functional Effects of the Stimulation/Blockade of  $\alpha$ -Adrenoceptors by an Agonist and an Antagonist

Receptor	Stimulation by an agonist causes	Blockade by an antagonist (blocker) causes
$\alpha_{1A}$	Vasoconstriction	Vasodilation
$\alpha_{1A}$ (prostate)	Contraction	Smooth muscle relaxation
$\alpha_2$ (postsynaptic) vascular	Vasoconstriction	Vasodilatation
$\alpha_2$ (presynaptic)	NE release↓	NE release↑
$\alpha_2$ (brainstem)	Sympathoinhibition Blood pressure↓	

NE. In the sympathetic nervous system, the receptor antagonists are also called antiadrenergic agents.

The sympathetic or adrenergic receptors are subdivided into a variety of subtypes.  $\alpha$ -Adrenoceptors are predominantly located in blood vessels but also in the brainstem.  $\beta$ -Adrenoceptors are mainly present in various structures of the heart but also in the bronchi. In a later stage, there is a further subdivision into  $\alpha_1/\alpha_2$  and  $\beta_1/\beta_2/\beta_3$ . The functional roles of the most important  $\alpha$ - and  $\beta$ -adrenoceptors are summarized in Tables I and II. The tables also note the effects of stimulation by an agonist and the effects of blockade by an antagonist.

The  $\alpha$ -adrenoceptors in the brainstem deserve special attention. These receptors in the pons/medulla region are involved in the central nervous regulation of the cardiovascular system, particularly that of blood pressure. The receptors belong to the  $\alpha_2$  subpopulation. Their stimulation by an agonist causes sympathoinhibition in the periphery, resulting in a reduced release of NE from the nerve endings and a decrease in blood pressure. These central  $\alpha_2$ -adrenoceptors are also the target of centrally acting antihypertensives, such as clonidine and  $\alpha$ -methyl-DOPA (through its active metabolite  $\alpha$ -methylnorepinephrine).

## ANTIADRENERGIC AGENTS

As illustrated in Fig. 1, virtually any structure or receptor of the sympathetic nervous system can be

modulated more or less selectively by antiadrenergic drugs. Several of these drugs play a role (or have played a role) as therapeutic agents in the treatment of cardiovascular or respiratory diseases.

### $\alpha$ -Adrenoceptor Antagonists ( $\alpha$ -Blockers)

Competitive inhibition of  $\alpha$ -adrenoceptors by appropriate antagonists predominantly causes vasodilatation and a reduction of blood pressure, particularly in hypertensive patients. Doxazosin and prazosin are the prototypes of  $\alpha$ -blockers. They are selective for the postsynaptic  $\alpha_1$ -adrenoceptor subtype: This means that presynaptic  $\alpha_2$ -adrenoceptors are not blocked, thus preventing the enhanced release of NE from the sympathetic nerve endings (Fig. 2). Nonselective, older  $\alpha$ -blockers, such as phentolamine or phenoxybenzamine, are no longer used because they are poorly tolerated.

Doxazosin and prazosin are moderately effective antihypertensive drugs. Orthostatic hypotension, headache, and reflex tachycardia are well-known adverse reactions, which are caused by vasodilation (also in the venous blood vessels).

A few newer  $\alpha$ -adrenoceptor antagonists display a certain degree of selectivity for the  $\alpha_{1A}$ -receptor in the smooth muscle of the prostate. These agents cause relaxation of this smooth muscle tissue and hence facilitate urinary flow in patients with benign prostate

Table II Functional Effects of the Stimulation/Blockade of  $\beta$ -Adrenoceptors by an Agonist and an Antagonist

Receptor	Stimulation by an agonist causes	Blockade by an antagonist causes
$\beta_1$ (cardiac)	Heart rate↑ Contractility↑ A-V conduction↑	Heart rate↓ Contractility↓ A-V conduction↓
$\beta_2$ (vascular)	Vasodilatation	Vasoconstriction
$\beta_2$ (bronchi)	Bronchodilatation	Bronchoconstriction
$\beta_2$ (metabolic)	Blood glucose↑	Blood glucose↓

hyperplasia (BPH). Alfuzosine, terazosin, and tamsulosin are examples of such agents.

### $\beta$ -Adrenoceptor Antagonists ( $\beta$ -Blockers)

$\beta$ -Adrenoceptor antagonists are competitive inhibitors of the effects of endogenous catecholamines (norepinephrine and epinephrine) at the level of  $\beta$ -adrenoceptors in various tissues. In clinical medicine, they are widely used in the treatment of hypertension, angina pectoris, supraventricular tachyarrhythmias, congestive heart failure, and as secondary prevention following myocardial infarction. Their main therapeutic activity is to counteract the detrimental effects of catecholamines released by the hyperactive sympathetic nervous system.

Well-known adverse reactions are bronchoconstriction, cold hands and feet (caused by peripheral vasoconstriction), and sleep disturbances. The therapeutic actions are all mediated by the blockade of  $\beta_1$ -adrenoceptors, whereas blockade of  $\beta_2$ -adrenoceptors is the cause of most of the side effects. For this reason,  $\beta_1$ -selective blockers are preferable. Atenolol, bisoprolol, and metoprolol are examples of  $\beta_1$ -selective blockers. Newer  $\beta$ -blockers, such as carvedilol, celiprolol, and nebivolol, contain an additional vasodilator component that may offer a hemodynamic advantage.

### Postganglionic Neuron Blockers

The electrical membrane activity in postganglionic neurons can be blocked (via the inhibitory influx of  $\text{Na}^+$  ions) by drugs such as guanethidine, cyclazanine, and bretylium. Although they are effective blood pressure-lowering drugs, these agents have been abandoned because of their poor tolerability.

### Ganglion Blocking Drugs (Ganglioplegic Agents)

These agents are competitive blockers of the  $\text{ACh}_n$  receptors in the sympathetic and also the parasympathetic ganglia. Ganglionic blockade in the sympathetic ganglia accounts for their effective lowering of blood pressure, particularly in hypertensive patients. Pentamethonium, hexamethonium, and trimetaphan are examples of ganglioplegic agents. These agents have been abandoned as therapeutics because of their low tolerability, which is partly due to the blockade of the parasympathetic ganglia.

### Rauwolfia Alkaloids

For the sake of completeness, I briefly mention these older drugs, which have lost their importance as therapeutics because of their poor tolerability. Reserpine and related alkaloids (from *Rauwolfia serpentina*) destroy the storage sites (granules) of norepinephrine in the sympathetic nerve endings and hence impair the function of the sympathetic system. Although reasonably effective blood pressure-lowering agents, their adverse reactions preclude their use in the treatment of hypertension.

### Centrally Acting Antihypertensives

$\alpha$ -Adrenoceptors present in the pontomedullary region of the brainstem (e.g., the nucleus tractus solitarius and the vasomotor center) will cause peripheral sympathoinhibition when stimulated with endogenous norepinephrine and also centrally acting antihypertensives.

Clonidine, guanfacine, and  $\alpha$ -methyl-DOPA (through its active metabolite  $\alpha$ -methylnoradrenaline) are  $\alpha_2$ -adrenoceptor agonists and, accordingly, cause peripheral sympathoinhibition and the reduction of elevated blood pressure in hypertensive patients. The centrally acting drugs are effective antihypertensives, and their mode of action appears to be attractive on hemodynamic grounds. Their adverse reactions are unpleasant (sedation, dry mouth, and male sexual impotence). These side effects are also mediated by  $\alpha_2$ -adrenoceptors (although not located in the brainstem) and thus it is virtually impossible to develop drugs with the same mode of action as clonidine, etc., but with a better profile of adverse reactions.

Imidazoline ( $\text{I}_1$ ) receptors in the brain (in the ventrolateral medulla) have been identified as targets for centrally acting drugs. Stimulation of central  $\text{I}_1$  receptors also causes peripheral sympathoinhibition and lowering of blood pressure. Newer drugs causing this effect are moxonidine and rilmenidine, which display lower affinity for  $\alpha_2$ -adrenoceptors than clonidine or  $\alpha$ -methyl-DOPA. It is hoped that these newer agents display a better tolerability.

## THERAPEUTIC USE OF ANTIADRENERGIC DRUGS

### Essential Hypertension

A subtle association between essential hypertension and hyperactivity of the sympathetic nervous system has been repeatedly demonstrated. On the other hand,

the treatment of hypertension was initially based on the use of various antiadrenergic drugs, which seems a rational approach. Other types of drugs have been introduced, such as diuretics, calcium antagonists, and suppressants of the renin-angiotensin system (ACE inhibitors and AT<sub>1</sub> blockers).

β-Blockers continue to be widely used in antihypertensive treatment, and this approach is supported by numerous clinical trials and a meta-analysis. α-Blockers, particularly doxazosin, have obtained a moderate position in antihypertensive treatment, which is not clearly supported by evidence on an epidemiological scale.

Centrally acting agents such as clonidine and α-methyl-DOPA have been largely abandoned because of their poor tolerability. This approach may be somewhat improved by the introduction of the imidazoline agonists (moxonidine and rilmenidine), but the clinical follow-up for the use of these agents is insufficient. Rauwolfia alkaloids, postganglionic neuron blockers, and ganglioplegic agents have been abandoned because of their poor tolerability.

### Angina Pectoris

Stable angina pectoris is routinely treated with a variety of anti-ischemic drugs in addition to invasive interventions for revascularization, such as balloon dilatation of a stenosis, or cardiac surgery. Nitroglycerin and other nitrates are used for the treatment of acute symptoms. β-Blockers are first choice for long-term treatment, unless contraindications preclude the use of these drugs. The reduction in heart rate is the major basis of the anti-ischemic action of β-blockers, thus reducing myocardial oxygen consumption. Calcium antagonists, long-acting nitrates, low-dose aspirin, and a statin are usually included in the treatment schedule.

### Secondary Prevention after Myocardial Infarction

Recurrent infarction and/or sudden death after myocardial infarction (MI) are major risks in post-MI patients. β-Blockers have been demonstrated to be significantly protective against these risks, probably as a result of their antiarrhythmic activity, which protects against the toxic effects of endogenous catecholamines released by the sympathetic system.

### Supraventricular Tachyarrhythmia

The antisympathetic effects of β-blockers explain their antiarrhythmic potency in the treatment of supraventricular tachyarrhythmia. Sotalol, frequently used as an antiarrhythmic agent, is a β-blocker with additional antiarrhythmic potency because of its class III (Vaughan-Williams' classification) activity.

### Congestive Heart Failure

β-Blockers are recognized as beneficial therapeutics in the management of congestive heart failure (CHF). CHF is usually accompanied by sympathetic overactivity, which is detrimental for the decompensated heart. β-Blockers counteract this process. Inotropic activity of the β-blocker that is too negative can be avoided by uptitrating the dose over a period of 3 weeks. Beneficial activity in CHF has been demonstrated for carvedilol, bisoprolol, and metoprolol. β-Blockers are applied when treatment with ACE inhibitors and diuretics is not successful.

### Glaucoma Simplex

In patients with glaucoma simplex (open-angle glaucoma), intraocular pressure can be lowered by the topical application of β-blockers such as timolol. The therapeutic activity is based on reduction of the production of aqueous humor.

### Benign Prostate Hyperplasia

The moderately beneficial effect of α<sub>1A</sub>-blockers in the symptomatic treatment of BPH was mentioned previously. Drugs used for this purpose are alfuzosine, terazosin, and tamsulosin. Hypotension is a side effect of these agents.

### See Also the Following Articles

Adrenergic Mechanisms • Adrenergic Receptors • Stress and Endocrine Physiology

### Further Reading

Bylund, D. B., Eikenberg, D. C., Hieble, J. P., *et al.* (1994). International Union of Pharmacology. Nomenclature of adrenoceptor. *Pharmacol. Rev.* **46**, 121–136.

Cohn, J. N. (1998). Betablockers in heart failure. *Eur. Heart J.* **19**(Suppl. F), F52–F55.  
“Goodman and Gilman’s The Pharmacological Base of Therapeutics,” 10th ed. (2003). McGraw-Hill, New York.  
Messerli, F. (1996). “Cardiovascular Drug Therapy,” 2nd ed. Saunders, Philadelphia.

van Zwieten, P. A. (1997). Central imidazoline receptors as targets of centrally acting antihypertensives: Moxonidine and rilmenidine. *J. Hypertension* **15**, 117–122.  
World Health Organisation/International Society of Hypertension. (1999). Guidelines for the management of hypertension. *J. Hypertension* **17**, 151–183.



## Antiestrogens

see SERMS (Selective Estrogen Receptor Modulators)



# Antithyroid Drugs

Anthony D. Toft

Royal Infirmary, Edinburgh, United Kingdom

## Glossary

**Graves' disease** The most common cause of hyperthyroidism, accounting for approximately 75% of cases and resulting from the inappropriate production of thyrotropin-receptor antibodies, which stimulate the thyroid gland to secrete excessive quantities of thyroxine and triiodothyronine.

**hyperthyroidism (thyrotoxicosis)** The clinical condition arising from sustained raised concentrations of thyroxine and/or triiodothyronine in the circulation. With rare exceptions (e.g., thyrotropin-secreting pituitary tumor), serum thyrotropin concentrations are low.

**thionamides** A group of antithyroid drugs comprising carbimazole, its active metabolite, methimazole, and propylthiouracil. Their main action is to inhibit the iodination of tyrosyl residues in thyroglobulin, one of the early stages of thyroid hormone synthesis.

The antithyroid drugs are principally used in the management of hyperthyroidism of Graves' disease. The antithyroid drugs carbimazole, methimazole, and propylthiouracil are heterocyclic compounds known as thionamides that contain a thiourylene group.

## INTRODUCTION

Carbimazole (CBZ) is immediately metabolized in serum to methimazole (MMI), and it has been calculated that 10 mg of CBZ yields 6 mg of MMI. All the antithyroid drugs are rapidly and almost completely absorbed from the gastrointestinal tract, with peak serum concentrations at 1 or 2 h. Antithyroid drugs are actively transported into the thyroid, where they inhibit the synthesis of triiodothyronine ( $T_3$ ) and thyroxine ( $T_4$ ), principally by interfering with the iodination of tyrosine by serving as preferential substrate for the iodinating intermediate of thyroid peroxidase. Oxidized iodine is thus diverted from the tyrosyl iodination sites in thyroglobulin. The iodinated antithyroid

drugs are desulfurated and further oxidized to inactive metabolites. In addition, propylthiouracil (PTU) inhibits  $T_4$  to  $T_3$  conversion, but this is of little clinical significance other than perhaps in the management of thyrotoxic crisis, when it is important to lower the raised serum  $T_3$  concentration as quickly as possible. There is also evidence that the thionamides have an immunosuppressive action, but any effect is short-lived because patients with Graves' hyperthyroidism frequently relapse after drug withdrawal.

Historically, CBZ has been the drug of choice in the United Kingdom, but MMI has been the drug of choice in all other areas of the world. PTU is also widely employed in the Americas but elsewhere tends to be restricted to use during pregnancy and breastfeeding and in patients who have reacted adversely to CBZ or MMI.

## PHILOSOPHY OF TREATMENT OF GRAVES' DISEASE

Each of the treatments for Graves' disease is effective but none is perfect, and the conclusions of Hershman and colleagues in the 1960s remain valid today, namely that the experience with antithyroid drugs, thyroidectomy, and radioiodine have not provided clear-cut criteria that can be used to select the best treatment for individual patients. A course of antithyroid drugs is appropriate for the minority of patients in whom a single episode of hyperthyroidism is followed by prolonged remission (30%). The majority, however, have a relapsing and remitting course over many years, and efforts to predict the natural history of the hyperthyroidism in most patients at diagnosis, using biochemical and immunological markers, have proved disappointing. On a group basis, a small goiter, a low serum concentration of thyrotropin-receptor antibodies (TRAbs), and older age favor remission after a course of antithyroid drugs, whereas the risk of relapse in a young male with severe hyperthyroidism and a large vascular goiter is so high that most would advocate surgery as the primary treatment. Similarly, a

patient with a high concentration of TRAbs, significant ophthalmopathy, and pretibial myxoedema at the end of a course of antithyroid drugs is unlikely to remain in remission for very long.

Because of the difficulty in determining the future pattern of the hyperthyroidism, treatment remains empirical and is inevitably influenced by the prejudices of the physician and the patient and by the expertise locally available, such as provided by nuclear medicine facilities and an experienced thyroid surgeon.

Management varies from center to center and between countries such that the preferred treatment of a 43-year-old female presenting with hyperthyroidism of moderate severity who did not plan further pregnancy was antithyroid drugs (77%) by European physicians but iodine-131 (69%) by their North American counterparts. There was an even greater contrast in choice of therapy when the index case was that of a 19-year-old female. One-third of physicians in the United States regarded iodine-131 as most appropriate, whereas only 4% of physicians in Europe thought it was most appropriate.

Frustration caused by the failure to target antithyroid therapy more specifically and by the gradual acceptance that iodine-131 is neither carcinogenic nor teratogenic has led to a move during the 10 years since the previously discussed survey was performed to advocate the more liberal use of radioactive iodine, irrespective of age. However, there remain anxieties among the public about irradiation in general.

Antithyroid drugs are not normally indicated in the treatment of toxic nodular goiter, unless they are used to render the patient euthyroid prior to surgery since recurrence of hyperthyroidism is invariable after drug withdrawal. There is no role for antithyroid drugs in subacute or postpartum thyroiditis, in which the thyrotoxicosis is due to the release of preformed thyroid hormones.

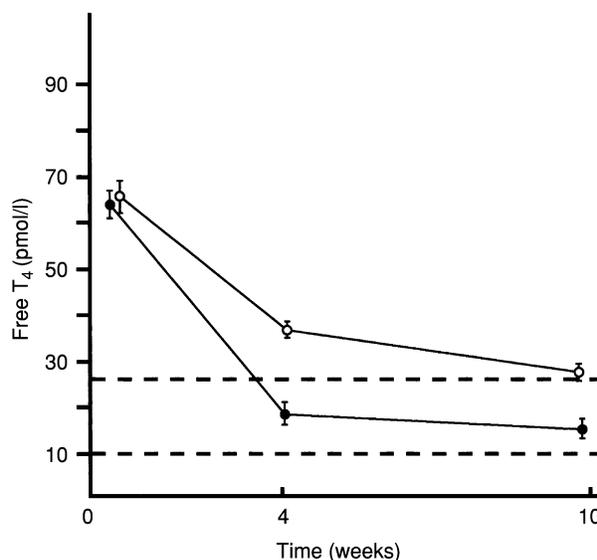
## DURATION OF TREATMENT

The conventional period of antithyroid drug therapy is best viewed as a method by which those destined to have a single short-lived episode of hyperthyroidism are identified and destructive therapy with iodine-131 or surgery avoided. Remission rates vary slightly, depending on the iodine status of the population, but approximately 50% remain in remission after 2 years of follow-up. The percentage is significantly less after only 6 months of treatment and is not improved by prolonging therapy for up to 3 years. Relapse can occur at any stage but is most common in the first 2 years.

Long-term treatment with antithyroid drugs is appropriate, however, in patients with underlying autonomous thyroid function (Graves' disease or nodular goiter) in whom hyperthyroidism has been precipitated by amiodarone and in whom chronic treatment with the antidysrhythmic drug is planned because iodine-131 is likely to be ineffective.

## DOSAGE

Methimazole is available as 5- and 10-mg tablets and CBZ as 5- and 20-mg tablets. The initial dose is 20–30 and 30–40 mg daily, respectively, depending on the severity of the hyperthyroidism (Fig. 1). Once-daily dosage is appropriate in all but the most severely thyrotoxic patients, who benefit from a twice-daily regimen. After 3 or 4 weeks, the dose of MMI can be reduced to 10–20 mg daily and that of CBZ to 20–30 mg daily. Further adjustment should be made on the basis of measurements of serum concentrations of  $T_3$ ,  $T_4$ , and thyroid-stimulating hormone (TSH) until a maintenance dose of 5 and 5–15 mg of MMI or CBZ, respectively, is achieved. Some patients can be maintained on as little as 5 mg CBZ on alternate days and, although this may a homeopathic dose, will promptly relapse if the drug is stopped. Patients begin to feel an improvement after 10–14 days. Initial changes in drug



**Figure 1** Mean  $\pm$  SE serum-free  $T_4$  concentrations during the first 10 weeks of treatment of hyperthyroidism of Graves' disease with either 20 mg (open circles) or 40 mg (closed circles) of CBZ daily. Although patients in both groups were clinically euthyroid by week 10, there seems little point in delaying the reversal of thyrotoxicosis by using the lower dose in the absence of evidence that adverse effects of the thionamides are dose related. Adapted from Page *et al.* (1996).

dosage should be based on thyroid hormone concentrations because delayed recovery of thyrotropes, previously exposed to high levels of  $T_3$  and  $T_4$  in the serum, may result in an inappropriately low serum TSH concentration. After 10–12 weeks of treatment, serum TSH is the best guide for adjustment of the dose, with high and low concentrations indicating excessive and inadequate therapy, respectively.

PTU is available as 50-mg tablets and the starting dose is 100–150 mg three times daily, reflecting its shorter half-life. The maintenance dose is 50–150 mg daily.

### Block-and-Replace Therapy

In this regime, the antithyroid drug is continued at the high initial dose after the patient is euthyroid, and hypothyroidism is avoided in the long-term by adding thyroxine at a dose of 100–150  $\mu\text{g}$  daily. The dose of thyroxine, but not that of the antithyroid drug, is adjusted to maintain serum TSH at the lower reference range. This combination has long been thought to be beneficial for patients with significant ophthalmopathy, presumably by avoiding hypothyroidism and possibly also due to the immunosuppressive action of the high-dose thionamide. It is also of value for those with so-called “brittle thyrotoxicosis,” often attributed to poor compliance but now known in some patients to be due to fluctuating concentrations and activities of TRAbs. Remission rates are not improved by standard block-and-replace therapy.

Claims that a negligible relapse rate can be achieved by following block-and-replace therapy for 18 months with thyroxine alone for up to an additional 3 years, thereby “putting the thyroid to rest,” reducing antigen release and TRAb concentrations, have not been substantiated.

### ADVERSE REACTIONS

The adverse effects of antithyroid drugs can occur at any time but almost always do so within 3–6 weeks of starting treatment. There is some cross-sensitivity between CBZ or MMI and PTU. Although it is common practice to change to the alternative antithyroid drug in the event of a minor adverse reaction, such as a skin rash, opinion is divided over whether the development of agranulocytosis is an absolute contraindication to further drug therapy.

#### Life-Threatening Reaction

The most serious adverse reaction is agranulocytosis, which develops in 0.2–0.5% of patients.

Agranulocytosis is characterized by fever, systemic upset, oropharyngeal bacterial infection, and a granulocyte count of less than  $0.25 \times 10^9/\text{liter}$ . The onset is sudden, and the consensus is that routine monitoring of the white blood cell count serves no purpose. Patients should simply be instructed to contact their medical practitioner immediately in the event of developing a sore throat or mouth ulceration. After stopping antithyroid drug therapy, the white blood cell count returns to normal within 1–3 weeks, during which time the affected patient should be isolated and treated with broad-spectrum antibiotics. Recovery of the white blood cell count may be hastened by the use of granulocyte colony-stimulating factor, but its value in patients with the most profound reduction in granulocyte count ( $< 0.1 \times 10^9/\text{liter}$ ) is unclear. Mild leukopenia with a relative lymphocytosis is common in Graves' disease and is not a contraindication to the use of antithyroid drugs.

### Other Reactions

The most common reactions are nausea, loss of taste, headache, and hair loss, which may be self-limiting and do not necessarily require drug withdrawal. The most troublesome is a skin rash, which is usually urticarial and affects 1 to 2% of patients. A migratory polyarthritis may occur alone or in association with the rash and resolves within 4 weeks of stopping treatment. Much rarer adverse effects include cholestatic jaundice, vasculitis (almost exclusively in patients taking PTU, which may be associated with anti-neutrophil cytoplasmic antibody), a lupus-like syndrome, and the nephrotic syndrome.

### ADJUNCT TO TREATMENT WITH IODINE-131

Iodine-131 requires approximately 6–8 weeks to be effective, and during this latent period hyperthyroidism may be exacerbated, with an increase in morbidity and even mortality in those with severe thyrotoxicosis and associated cardiovascular disease. For this reason, it is not uncommon not only to render the patient euthyroid before radioiodine treatment but also to continue the antithyroid drug for 6 weeks. The thionamides confer on the thyroid a degree of radioresistance. Unless PTU or MMI (CBZ) is stopped for at least 15 and 2 days, respectively, before and for a similar period after iodine-131 therapy, the cure rate, defined as abolishing hyperthyroidism, is reduced. If it is not clinically possible to discontinue PTU for such a prolonged period in relation to iodine-131 therapy, the dose of isotope should be increased by at least 25%.

## ANTITHYROID DRUGS IN PREGNANCY

Maternal hyperthyroidism in pregnancy is usually due to Graves' disease. TRAb crosses the placenta and if the mother is thyrotoxic, it must be assumed that the fetus is similarly affected. Before effective treatment was available, the fetal death rate was as high as 50%. Fortunately, antithyroid drugs also cross the placenta and by careful monitoring of maternal thyroid function, normal fetal development can be achieved, even though cord blood may show evidence of overtreatment. Like other organ-specific autoimmune diseases, Graves' hyperthyroidism tends to improve or even remit during pregnancy. A small dose of antithyroid drug, such as 5 mg MMI daily, will maintain free T<sub>4</sub> and TSH concentrations in the reference ranges. It is good clinical practice to examine the mother every 4 weeks during pregnancy and to stop the antithyroid drug 4 weeks before the expected date of delivery to avoid any possibility of fetal hypothyroidism when brain development is at a maximum. There is evidence that maternal hypothyroidism may cause subtly impaired neuropsychological development in the subsequent offspring. Apart from meticulous control of maternal thyroid function during pregnancy, it is wise for thyrotoxic patients not to become pregnant until stable normal thyroid function has been achieved.

Measurement of the thyrotropin-receptor antibody concentration in maternal serum during the last examination before delivery may be helpful because a high level is a predictor of neonatal thyrotoxicosis. Since thyroid hormones cross the placenta relatively poorly, the block-and-replace regime is not recommended during pregnancy.

### Methimazole (CBZ) or Propylthiouracil?

Aplasia cutis congenita (ACC) is a rare disorder of the skin, usually affecting the scalp and less than 3 cm in diameter, that has been reported in a small number of

neonates whose mothers received MMI during pregnancy. ACC has not been reported in association with PTU, which is widely used in North America, and some believe that PTU is the drug of choice during pregnancy or for those planning pregnancy. The consensus, however, is that there is insufficient evidence to establish a direct causal relationship between ACC and MMI. Because MMI (CBZ) and PTU are equally effective in controlling Graves' hyperthyroidism during pregnancy, it makes sense to use the preparation with which one has most experience.

If hyperthyroidism recurs after delivery, is due to Graves' disease and not postpartum thyroiditis, and the mother wishes to breast-feed, PTU is the drug of choice because it is transferred to the milk one-tenth as much as the other thionamides.

### See Also the Following Articles

Graves' Disease • Graves' Disease, Hyperthyroidism in • Hyperthyroidism, Subclinical • Iodine, Radioactive • Thyroid Disease, Epidemiology of • Thyroid Function Tests • TSH Receptor (Thyrotropin Receptor)

### Further Reading

- Hancock, L. D., Tuttle, R. M., LeMar, H., Bauman, J., and Patience, T. (1997). The effect of propylthiouracil on subsequent radioactive iodine therapy in Graves' disease. *Clin. Endocrinol.* **47**, 425-430.
- Hershman, J. M., Givens, J. R., Cassidy, C. E., and Astwood, E. B. (1966). Long-term outcome of hyperthyroidism treated with antithyroid drugs. *J. Clin. Endocrinol. Metab.* **26**, 803-807.
- McIver, B., Rae, P., Beckett, G., Wilkinson, E., Gold, A., and Toft, A. (1996). Lack of effect of thyroxine in patients with Graves' hyperthyroidism who are treated with an antithyroid drug. *N. Engl. J. Med.* **334**, 220-224.
- Mestman, J. H. (1998). Hyperthyroidism in pregnancy. *Endocrinol. Metab. Clin. North Am.* **27**, 127-149.
- Page, S. R., Sheard, C. E., Herbert, M., Hopton, M., and Jeffcoate, W. J. (1996). A comparison of 20 or 40 mg per day of carbimazole in the initial treatment of hyperthyroidism. *Clin. Endocrinol.* **45**, 511-515.



Appetite

see Hunter and Satiation



# Asherman's Syndrome

Theodore A. Baramki and Carolyn J. Alexander  
*Johns Hopkins University School of Medicine, Baltimore, Maryland, United States*

## Glossary

**amenorrhea** The complete absence of menses.

**curettage** Scraping of the lining of the uterus.

**hypomenorrhea** Uterine bleeding less than normal in amount.

**hysteroqram** Radiologic image of the uterine cavity using radio-opaque medium.

**hysteroscopy** A procedure using an endoscope to visualize the uterine cavity.

**missed abortion** Nonviable pregnancy that has not been expelled.

**placenta accreta** Abnormal adherence of part or all of the placenta to the uterine wall.

**sonography** Imaging using ultrasonic waves.

**synechiae** Adhesions.

**A**sherman's syndrome denotes the presence of adhesions or synechiae in the endometrial cavity resulting in infertility or recurrent spontaneous abortions with or without amenorrhea.

## INTRODUCTION

In 1948, Asherman described a clinical condition presenting with amenorrhea due to endometrial adhesions that he termed amenorrhea traumatica (atretica). The adhesions may partially or completely obliterate the endometrial cavity, the internal os, or a combination of these areas. The patient may present with amenorrhea, hypomenorrhea, or recurrent spontaneous abortion. The adhesions may be filmy or dense, arising in the center of the endometrial cavity or arising from a side wall and encroaching on the cavity. Histologically, these adhesions are formed of endometrial, myofibrous, or connective tissue.

## ETIOLOGY

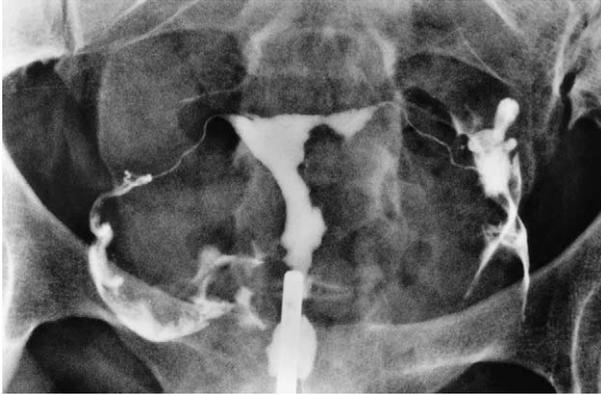
The most frequent cause of Asherman's syndrome is endometrial curettage for retained products of conception. In addition to curettage, mechanical evacuation methods for elective, incomplete, or missed abortion may lead to synechiae formation. Postpartum instrumentation of an infected uterus represents a significant risk factor for the development of endometrial scarring. Also, endometritis caused by intrauterine devices or pelvic inflammatory disease may lead to synechiae. Endometrial tuberculosis can destroy the endometrium. This can be confirmed by culture of the menstrual discharge or tissue obtained by endometrial biopsy.

Uterine synechiae may also result from myomectomy, cesarean section, or metroplasty. Synechiae are rare after diagnostic curettage.

## DIAGNOSIS

The possibility of intrauterine synechiae (Asherman's syndrome) must be considered in individuals who develop amenorrhea following pregnancy-related curettage or endometritis. Despite the amenorrhea, patients will continue to have cyclic changes in the breasts and have a biphasic basal body temperature chart if they are ovulatory.

Hysteroqram with the aid of fluoroscopy can diagnose the extent and location of the synechiae (Fig. 1). The filling defects do not change appearance with extra distension of the uterine cavity or with changes of position of the patient. In case of severe Asherman's syndrome, intravasation of the dye into the venous system is not unusual (Fig. 2). However, careful placement of the tip of the cannula or catheter and avoidance of excessive force when injecting the contrast medium will give a clear picture of the uterine cavity (Fig. 3). With complete obliteration of the uterine cavity, only the cervical canal is visualized (Fig. 4).

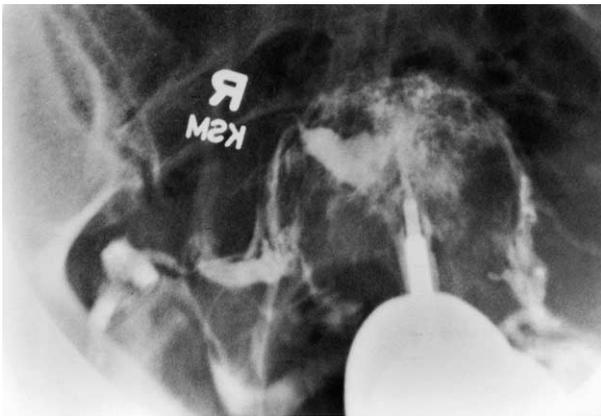


**Figure 1** Hysterosalpingogram showing partial obliteration of the uterine cavity due to synechiae.

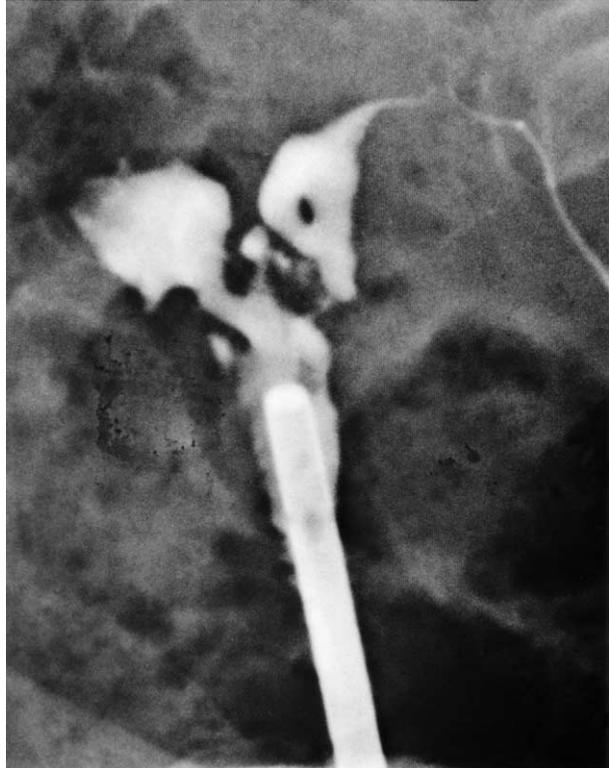
Sonography, especially three-dimensional ultrasound, improves recognition of the uterine anatomy. Sonohysteroscopy may also be helpful in the diagnosis of synechiae. Hysteroscopy gives direct visualization of the texture and extent of the synechiae (Fig. 5) and is a valuable therapeutic tool.

## TREATMENT

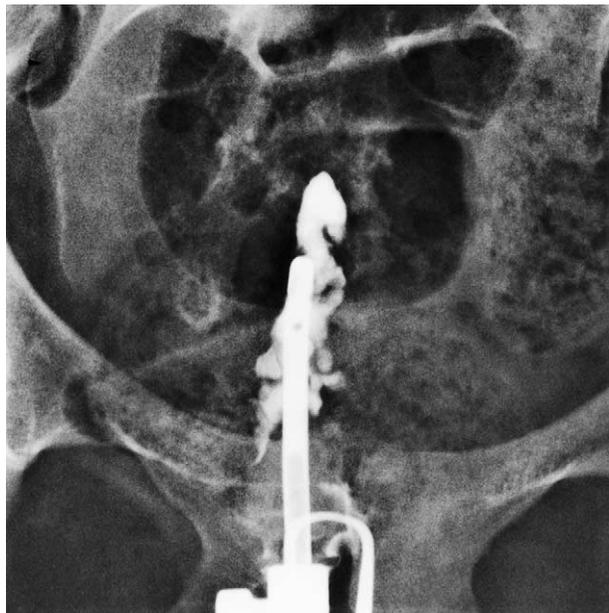
In the past, uterine adhesions were treated with dilatation and curettage, and postoperative management included an intrauterine device to separate the walls of the uterus while estrogen and antibiotics were being given. Occasionally, a post-treatment hysterosalpingogram was performed to ensure normalization of the cavity. Hysteroscopy now plays a major role in the approach to the diagnosis and treatment of uterine synechiae. The fibrous nature of the lesion and its precise location can be determined. Filmy adhesions



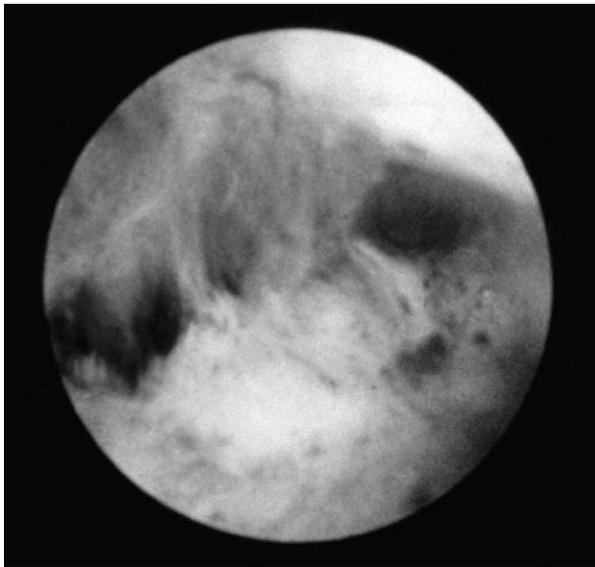
**Figure 2** Hysterosalpingogram showing intravasation of dye into the venous system of a uterus involved with marked synechiae. The tip of the cannula does not seem to be free in the uterine cavity.



**Figure 3** Hysterosalpingogram of the same patient in Fig. 2 performed at a later date, with proper placement of the cannula, showing the extensive adhesions. Only the left tube is visualized.



**Figure 4** Hysterosalpingogram of a patient who developed amenorrhea following postpartum infection. The tenaculum is applied to the cervix at the level of the external os. The cervical canal is visualized with total uterine cavity occlusion (Asherman's syndrome).



**Figure 5** Hysteroscopic view of central synechiae showing the lateral recesses.

frequently rupture under the pressure of the distending medium or may be dislodged by the uterine sound, cervical dilator, or the tip of the hysteroscope.

In cases of severe Asherman's syndrome, one can use fluoroscopy to improve the hysteroscopic resection of adhesions. Radio-opaque medium is injected through a 16-Touhy needle through bands of adhesions into pockets of endometrium previously identified by sonography. This technique provides an intraoperative fluoroscopic view of pockets of endometrium behind an otherwise blind-ending endocervical canal in women with severe Asherman's syndrome, allowing guided hysteroscopic division of adhesions and reducing the likelihood of perforation and formation of a false passage. Hysteroscopy with simultaneous laparoscopy can also help prevent, or help make a prompt diagnosis of, uterine perforation. In severe cases, repeated hysteroscopic attempts at adhesiolysis may be needed.

In order to prevent recurrence of adhesions, a silastic triangular balloon, a No. 14 pediatric Foley catheter, or an intrauterine device may be used to separate the walls of the uterine cavity. The balloon or catheter may be kept in place for 7 days, during which time the patient is given broad-spectrum antibiotics. Estrogen in the form of conjugated estrogens given in a dose of 2.5 mg daily for 1 or 2 months will help the regeneration of the endometrium. During the final 10 days of this course of estrogen, 10 mg daily of medroxyprogesterone acetate is given to slough the endometrium. If an intrauterine device is

used, it can be removed at the time of withdrawal bleeding.

In patients with intrauterine adhesions severe enough to produce amenorrhea, biologically active endometrium can undergo malignant change. One case of endometrial carcinoma arising within intrauterine synechiae has been reported.

## IMPLICATIONS FOR PREGNANCY

Hysteroscopic treatment of Asherman's syndrome can be effective in reconstructing a functional uterine cavity. In a series of 187 patients who were treated with hysteroscopic lysis of adhesions during a 10-year period, Valle and Sciarra found 43 patients with mild or filmy adhesions, 97 with moderate or fibromuscular adhesions, and 47 with severe or connective tissue adhesions. Normal menstruation was restored in 88% of patients who had menstrual abnormalities. Among the 187 patients, 143 (76%) became pregnant; of these, 114 (80%) achieved a term pregnancy, 26 (18%) had a spontaneous abortion, and 3 (2%) had ectopic pregnancies. The reproductive outcome parallels the severity of the disease. In patients with mild disease, the term pregnancy rate was 81%, in those with moderate disease it was 60%, and in patients with severe disease it was 32%.

There is a distinct risk of placenta accreta after lysis of intrauterine adhesions. Of three such women, two required treatment by cesarean hysterectomy. Deaton *et al.* reported uterine rupture after aggressive treatment of Asherman's syndrome that seemed to occur not from uterine contractions but rather from the uterine expansion of pregnancy. Their patient underwent two hysteroscopic resections of adhesions, both complicated by uterine perforation. She subsequently conceived and required hospitalization at 25 weeks of gestation for vaginal bleeding without labor. After 40 days in the hospital, still without evidence of contractions, she had an acute uterine rupture that required an emergency cesarean hysterectomy. The fundus was described as "paper-thin from cornu to cornu."

Uterine synechiae may be diagnosed during pregnancy using ultrasonography. A retrospective case-control study was performed by Buchmeier and Longnecker between March 1988 and March 1995. Cases and controls were part of a population of 29,543 pregnant patients who underwent ultrasonographic examination during the second trimester. Synechiae were diagnosed by the sonographic finding of a shelf-like protrusion into the amniotic cavity. The overall prevalence of uterine synechiae was 0.47% (140 of

29,543) in the scanned population. No significant difference was found between cases and controls with respect to maternal age, reproductive losses, and medical problems. The mean gestational age at the time of diagnosis was  $18.3 \pm 4.2$  weeks. No differences in outcome existed between cases and controls except for mean birth weight. The researchers concluded that the presence of uterine synechiae does not appear to confer an increased risk for poor pregnancy outcome.

### See Also the Following Article

Endometriosis

### Further Reading

- Asherman, J. G. (1948). Amenorrhoea traumatica (atretica). *J. Obstet. Gynecol. Br. Empire* **55**, 23.
- Broome, J. D., and Vancaillie, J. G. (1999). Fluoroscopically guided hysteroscopic division of adhesions in severe Asherman's syndrome. *Obstet. Gynecol.* **93**, 1041–1043.
- Buchmeier, S. E., and Longnecker, M. (1997). Clinical significance of sonographically detected uterine synechiae in pregnant patients. *J. Ultrasound Med.* **16**, 465–469.
- Bukulmez, O., Yarali, H., and Gurgan, T. (1999). Total corporal synechiae due to tuberculosis carry a very poor prognosis following hysteroscopic synechialysis. *Hum. Reprod.* **14**, 1960–1961.
- Capella-Allouc, S. (1999). Hysteroscopic treatment of severe Asherman's syndrome and subsequent fertility. *Hum. Reprod.* **14**, 1230–1233.
- Deaton, J. L., Maier, D., and Androch, J. (1989). Spontaneous uterine rupture during pregnancy after treatment of Asherman's syndrome. *Am. J. Obstet. Gynecol.* **160**, 1053–1054.
- Georgakopoulos, P. (1974). Placenta accreta following lysis of uterine synechiae (Asherman's syndrome). *J. Obstet. Gynecol. Br. Commonwealth* **81**, 730–733.
- Valle, R. F., and Sciarra, J. J. (1988). Intrauterine adhesions: Hysteroscopic diagnosis, classification, treatment, and reproductive outcome. *Obstet. Gynecol.* **158**, 1459–1470.



# Assisted Reproductive Technology (ART)

Shahar Kol

Rambam Medical Center, Haifa, Israel

## Glossary

**anovulation** Absence of ovulation even though menstruation may continue.

**AID** Artificial insemination with donor's sperm.

**AIH** Artificial insemination with husband's sperm.

**assisted hatching** The process of helping the blastocyst hatch from its shell (zona pellucida) by making a hole in the zona pellucida; may increase the chance of embryo implantation in some patients.

**asthenospermia** Poor motility of sperm in the seminal fluid.

**azoospermia** An absence of sperm in the seminal fluid.

**blastocyst** The stage at which the embryo has reached 5 to 6 days after fertilization, a fluid-filled cyst has formed, and the cells have begun to differentiate into the inner cell mass (which will form the fetus) and the trophoblast (which will form the placenta and fetal membranes).

**chromosome/chromosomal abnormalities** Thread-like structures within the nucleus that contain the hereditary material DNA, where mistakes can occur either in chromosome numbers (46 in humans) or in structure.

**cleavage** The mitotic division of the fertilized egg (zygote) into two or more cells; usually starts at about 26 h after fertilization.

**corpus luteum** A small endocrine organ that develops within the ruptured follicle after the oocyte has been released from the ovary; mainly produces the hormone progesterone.

**cryobiology** The use of low-temperature techniques on cells, tissues, and organs, usually by freezing and storage.

**cryopreservation** Freezing tissues and storing them in liquid nitrogen at  $-196^{\circ}\text{C}$  (e.g., sperm, oocytes, ovarian and testicular tissue, embryos).

**egg donation** When a woman who has no fertility problems donates her eggs (oocytes) to another woman to enable her to carry pregnancy and have a child.

**egg recovery (collection, pickup)** When eggs are aspirated from the ovary; performed by vaginal ultrasound-directed method or by laparoscopy.

**embryo** The product of fertilization; at early stage of development, usually in the uterus, but also *in vitro* in the laboratory.

**embryo biopsy** Taking one or more cells (blastomeres) from an embryo (often at the 8-cell stage but can be at the blastocyst stage) for genetic or other analysis.

**embryo transfer/replacement** Stage in *in vitro* fertilization when the embryo is transferred to the uterus.

**endometrium** The lining of the uterus that develops at the beginning of each cycle so that it is ready to receive an embryo; endometrium breaks down, leading to menstruation, if there is no embryo.

**endometriosis** The presence of endometrium outside the uterus.

**fallopian tubes** A pair of small, fine, delicate tubes where fertilization usually takes place; transport ova and sperm to the fertilization site and transport the developing embryo to the uterus.

**fertilization** The penetration of the oocyte by the sperm, a process that results in the formation of an embryo.

**follicle-stimulating hormone (FSH)** A hormone produced by the pituitary gland in the brain; in women, stimulates ovulation and the production of estrogen; in men, stimulates the production of sperm.

**gamete intra-fallopian transfer (GIFT)** When eggs retrieved (usually under laparoscopy) are mixed with a prepared semen sample and introduced into the fallopian tube.

**gametes** Male and female reproductive cells (sperm and oocyte).

**gonadotropin-releasing hormone (GnRH)** Responsible for initiating the production of FSH and luteinizing hormone from the pituitary.

**GnRH agonist** Drugs that have a similar effect to GnRH and that activate the receptor sites in the gonadotropin secretory cells of the pituitary (e.g., buserelin, triptorelin, leuprolide) and finally suppress (down-regulate) the pituitary.

**GnRH antagonist** Drugs that compete with natural GnRH on its receptor but do not activate it, thereby blocking its effect (e.g., ganirelix, cetrorelix).

**human chorionic gonadotropin (hCG)** The hormone produced by placental cells in pregnancy; detection of this hormone in blood or urine is the basis of pregnancy testing; this hormone mimics the activity of luteinizing hormone and so is used for triggering ovulation.

**hysterosalpingogram (HSG)** An X-ray study in which a radio-opaque fluid is injected into the uterus so that the uterus and the fallopian tubes are visualized.

**hysteroscopy** The use of a fine optical instrument to view the inside of the uterus via the cervix.

**implantation** The embedding of an embryo in the endometrium of the uterus.

**infertility** Arbitrarily defined as the inability of a couple to produce a pregnancy after 1 year of intercourse with no contraception; one in six couples affected; can be described as primary or secondary infertility.

**intrauterine insemination (IUI)** Insemination of a prepared sperm sample into the uterus through the cervix.

**intra-cytoplasmic sperm injection (ICSI)** Injecting the sperm directly into the egg; major advance in male factor infertility.

**in vitro fertilization (IVF)** Fertilizing the egg in the laboratory for subsequent embryo transfer; requires egg retrieval/pickup; can be done in a natural or simulated cycle (when the ovaries are made to produce more eggs); after recovery, eggs placed in a plastic dish with a specially prepared semen sample, where fertilization may take place; pre-embryo then transferred to the uterus, where it may implant and develop further into a fetus and eventually a baby.

**laparoscopy** Procedure allowing visualization of abdominal organs using an optical instrument introduced through the abdominal wall.

**luteinizing hormone (LH)** A pituitary gonadotropic hormone that stimulates maturation and rupture of the follicle in women.

**micromanipulation** Procedures used (as an extension of IVF) to manipulate the embryo by blastomere biopsy, zona drilling, and the like.

**microsurgical epididymal sperm aspiration (MESA)** When sperm cells are aspirated from the epididymis and used in the IVF procedure.

**ovarian hyperstimulation syndrome (OHSS)** A complication of gonadotropin therapy in which there is excessive stimulation of the ovaries; women having treatment undergo frequent scans and hormone testing to prevent OHSS.

**oligospermia** A condition in which there are fewer sperm cells in the ejaculate than are considered "normal."

**oocyte** Female gamete or egg.

**polycystic ovarian disease (PCOD)** A condition where multiple small follicles appear in the ovary; abnormal imbalance can arise, causing problems in ovulation and infertility.

**progesterone** A hormone produced by the corpus luteum after ovulation has occurred; also produced by the placenta in pregnancy.

**teratospermia** A condition in which less than 50% of the sperm are morphologically normal.

**testicular sperm aspiration (TESA)** Obtaining sperm directly from the testis.

**ultrasound scanning** A procedure in which high-frequency sound waves are emitted into the pelvis and, as they bounce back, are used to build up a picture; used in many areas of medicine; can be performed abdominally or via the vagina in reproductive medicine; many organs (e.g., kidneys, liver, heart, uterus, ovaries) can be visualized.

**Y chromosome** One of the chromosomes (X or Y) that determine the sex of an individual; XX determines that the embryo will be female; presence of the Y chromosome determines that it will be male.

**zygote intra-fallopian transfer (ZIFT)** Similar to GIFT except that the embryo is transferred into the fallopian tube rather than sperm and eggs, where the advantage is that fertilization is established before transfer.

**zona drilling** The process of making a hole (by acid tyrode solution or laser) in the zona pellucida (the egg shell) to assist in the hatching process.

On July 25, 1978, Louise Joy Brown (Fig. 1), the world's first *in vitro* fertilization baby, was born in Great Britain. Infertility affects millions of couple worldwide. Before the birth of Louise Brown, those women who were found to have blocked or ruined fallopian tubes had no hope of becoming pregnant if surgical procedures failed to correct the situation.

## INTRODUCTION

Patrick Steptoe and Robert Edwards (Fig. 2), two physicians from the United Kingdom, had been actively working on finding an alternative solution for conception since 1966. Although Steptoe and



**Figure 1** Louise Joy Brown, the world's first *in vitro* fertilization baby.

Edwards had successfully found a way in which to fertilize an oocyte outside a woman's body, they were still troubled by problems after replacing the fertilized egg back into the woman's uterus. By 1977, all of the pregnancies resulting from their procedure (approximately 80) had lasted only a few short weeks.

Lesley Brown, Louise Brown's mother, changed that when she successfully passed the first few weeks of pregnancy. This amazing achievement opened a new era in reproductive medicine, an era of rapid developments and achievements.

## INDICATIONS

Pregnancy is always a matter of chance. A normal couple attempting pregnancy in their 20s has about a 20% chance of getting pregnant in any one month,

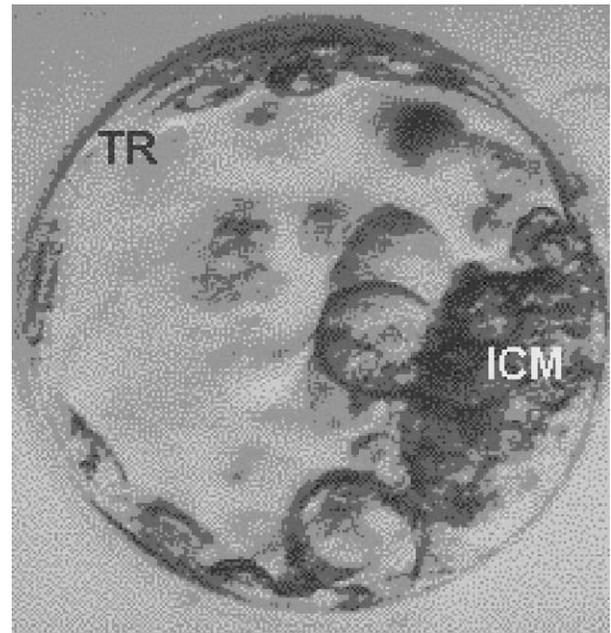


**Figure 2** (A) Dr. Patrick Steptoe and (B) Dr. Robert Edwards.

and couples in their 30s have about a 10% chance in any one month. Sooner or later, a couple will achieve pregnancy if the partners have (1) a reasonably high monthly chance of fertility (more than 5% per month) and (2) a reasonable time in which to keep trying (say, 2 years). If pregnancy has not happened within 1 or 2 years, then it is likely that the monthly chance of pregnancy will be less than 5%. Tests are then done.

For normal fertility, sperm in the vagina must swim up through the cervix, uterus, and fallopian tubes to meet an ovulated egg that has been carried from the surface of the ovary to the middle part of the fallopian tube. The embryo that results (strictly speaking, it is still a "pre-embryo") develops for 3 days in the fallopian tube. It then travels to the uterus, where it floats and develops for another 3 or 4 days before attaching to and implanting within the endometrium (lining of the uterus), thereby establishing pregnancy. A few days before the period is missed, a blood pregnancy test will be positive. It is at about this time of implantation, a week after fertilization, that the first few cells in the center of the "embryo" actually differentiate into what will be the fetus. All other cells (the majority of cells at this stage) go to form the placenta (Fig. 3).

In general, a couple may be relatively infertile (with a reduced monthly chance of conception) or completely infertile (with no chance of conception, sometimes called sterility). The following are the leading causes:



**Figure 3** Embryo at the blastocyst stage. Inner cell mass (ICM) will produce the fetus itself. Trophoblast (TR) will produce the placenta and membranes.

- Problems with ovulation (the release of the egg from the ovary). There will be complete infertility if ovulation does not take place and periods are absent (amenorrhoea), although many such patients are treatable with hormones or drugs (patients with polycystic ovary syndrome). If there are no responsive eggs in the ovaries (ovarian failure), egg donation is required.
- Problems with sperm production (often not treatable except by assisted conception). There will be complete infertility if there are no sperm cells in the ejaculate (azoospermia).
- A blockage between the uterus and the ovary, preventing fertilization (Fig. 4). The most common site of blockage is the fallopian tubes, sometimes treatable with microsurgery or assisted conception.
- Endometriosis, a common condition in which the endometrium grows outside the uterus, disturbing a number of events essential to conception and implantation of the embryo in the uterus. Treatment can be medical, surgical, or with assisted conception. Infertility is usually relative rather than complete.

Fertility tests involve the following:

- Progesterone level in the second half of the menstrual cycle (luteal phase) to establish ovulation. Other hormonal tests (e.g., prolactin, luteinizing hormone [LH], follicle-stimulating hormone [FSH], thyroid-stimulating hormone [TSH], testosterone, estradiol) are also often done to assess endocrine integrity and ovarian reserve.
- Sperm count (or a postcoital test looking for sperm in the mucus of the cervix).
- X ray of the uterus and fallopian tubes (a hysterosalpingogram) or a laparoscopy to look at the pelvic organs.



**Figure 4** Mechanical infertility. Both fallopian tubes are greatly dilated and blocked.

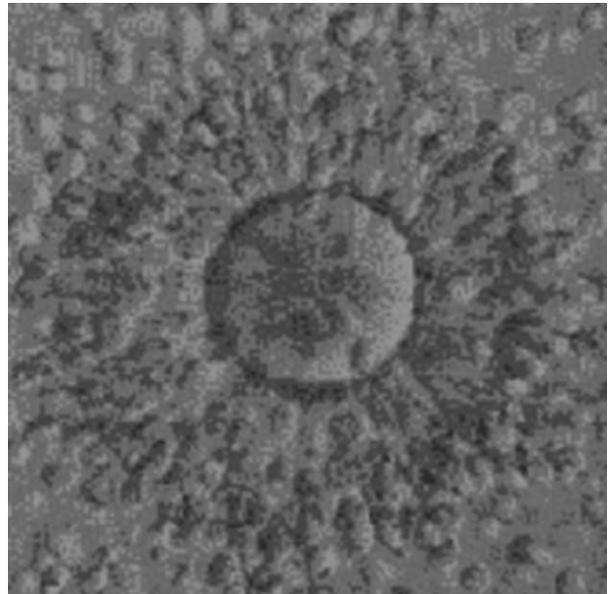
These tests may show the following:

- Complete infertility (“sterility”): (a) ovarian failure, with no chance of inducing ovulation; (b) complete absence of sperm (“azoospermia”); or (c) complete obstruction of the fallopian tubes.
- Relative infertility: (a) infrequent ovulation or absent ovulation, resolved partly by treatment; (b) a decrease in the sperm count; (c) partial blockage of the tubes or the presence of scar tissue around the tubes or ovary; (d) endometriosis of any degree; (e) an abnormality of the uterus such as fibroids, polyps, or scarring of the lining; (f) an abnormality of the cervix such as a previous cone biopsy or inflammation (“cervicitis”); and/or (g) an immune reaction against sperm cells (“anti-sperm antibodies”) in either partner.
- Unexplained infertility: sometimes no abnormality is obvious.

Some causes of infertility can be overcome with drugs or surgical intervention. Otherwise, assisted reproductive technology (ART) is needed, with *in vitro* fertilization (IVF) being the cornerstone of this treatment.

## SUPEROVULATION FOR IVF

The first step of the IVF procedure involves stimulation of egg growth. Spontaneously, a cycling woman will ovulate only one egg each month. This single egg (Fig. 5) may be used for IVF. In fact, the first IVF baby was produced in this way. However, the procedure is



**Figure 5** Mature oocyte surrounded by cumulus cells.

cumbersome and laborious because the natural cycle must be followed very carefully to time egg retrieval. To increase the chance of obtaining pregnancy, an effort is made to recruit 10 to 12 eggs, with an eye on patient safety to prevent ovarian hyperstimulation syndrome. This goal is achieved by stimulating the ovaries with drugs (e.g., Pergonal, Puregon, Humegon, Metrodin, Gonal-F, Menogon). These drugs contain gonadotropins produced from urine of postmenopausal women (e.g., Pergonal, Humegon, Menogon, Metrodin) or recombinant human FSH (e.g., Puregon, Gonal-F). Because spontaneous ovulation might occur, leading to cycle cancellation, other drugs are used to prevent the endogenous production of LH, which triggers ovulation.

Gonadotropin-releasing hormone (GnRH) agonists are a group of drugs (e.g., Decaptyl, Leuprolide, Buserelin, Nafarelin) that activate the pituitary GnRH receptors. Following an initial burst of gonadotropins, prolonged exposure of these receptors to the GnRH agonist leads to pituitary down-regulation, with prevention of endogenous LH surge. Newer compounds act as GnRH antagonists (e.g., Orgalutran, Antagon, Cetrotide). They bind to the receptors but do not activate them (competitive inhibition).

The growth and development of the ovarian follicles are closely monitored by repeated ultrasound studies and blood tests for hormone levels (estradiol and progesterone). Based on the information obtained from these tests, the optimal timing for ovulation is determined. Ovulation itself is triggered by an injection of human chorionic gonadotropin (hCG) (e.g., Chorigon, Profasi, Pregnyl), after which the egg retrieval is scheduled (usually after 34–36 h).

## THE LABORATORY

### Egg Retrieval

The procedure itself is performed under general anesthesia or sedation at the hospital by a transvaginal route. A needle, guided by ultrasound imaging, is inserted through the vaginal wall into the ovaries, where the follicles containing the eggs are punctured and aspirated. The released eggs are transferred to the lab, where their developmental stage is assessed under the microscope.

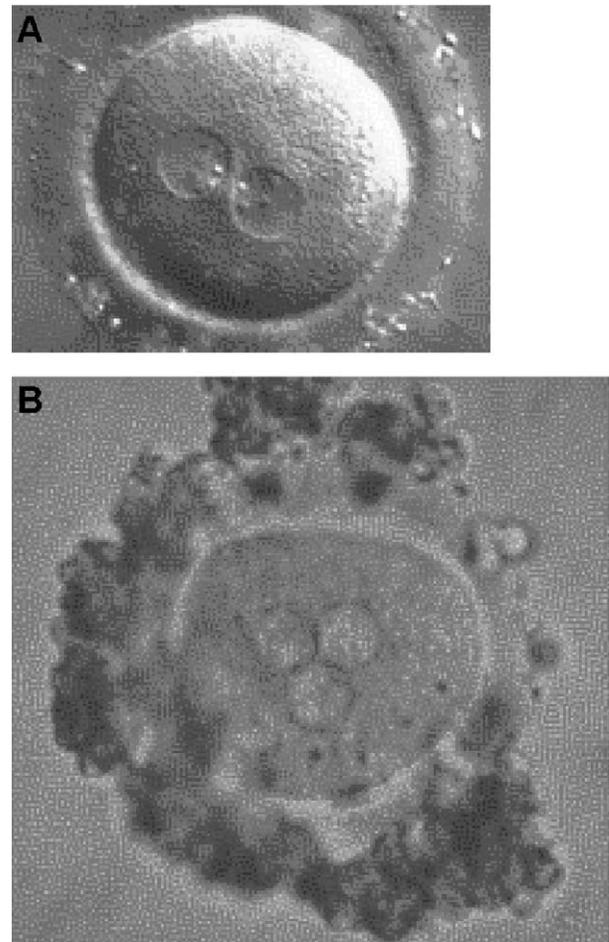
### Sperm Preparation

Sperm is obtained by masturbation. It then undergoes a series of lab procedures to prepare it for interaction

with the egg. During these procedures, the sperm is washed and resuspended in special medium.

### Fertilization

Eggs are kept in dishes, to which sperm is added at the proper concentration. The dishes are kept in an incubator where the environment (temperature, humidity, and gas composition) is carefully monitored. In given time intervals, the eggs are assessed for fertilization and subsequent cleavage (divisions). This assessment is based on morphology; hence, the detailed chromosomal composition of the embryos cannot be addressed without further tests. Approximately 18 h after fertilization, the eggs are examined for the appearance of 2 pronuclei (maternal and paternal) (Fig. 6A). If the number of pronuclei is greater than 2, the eggs are discarded (Fig. 6B). On day 2 postfertilization, the eggs are examined for cleavage. The number

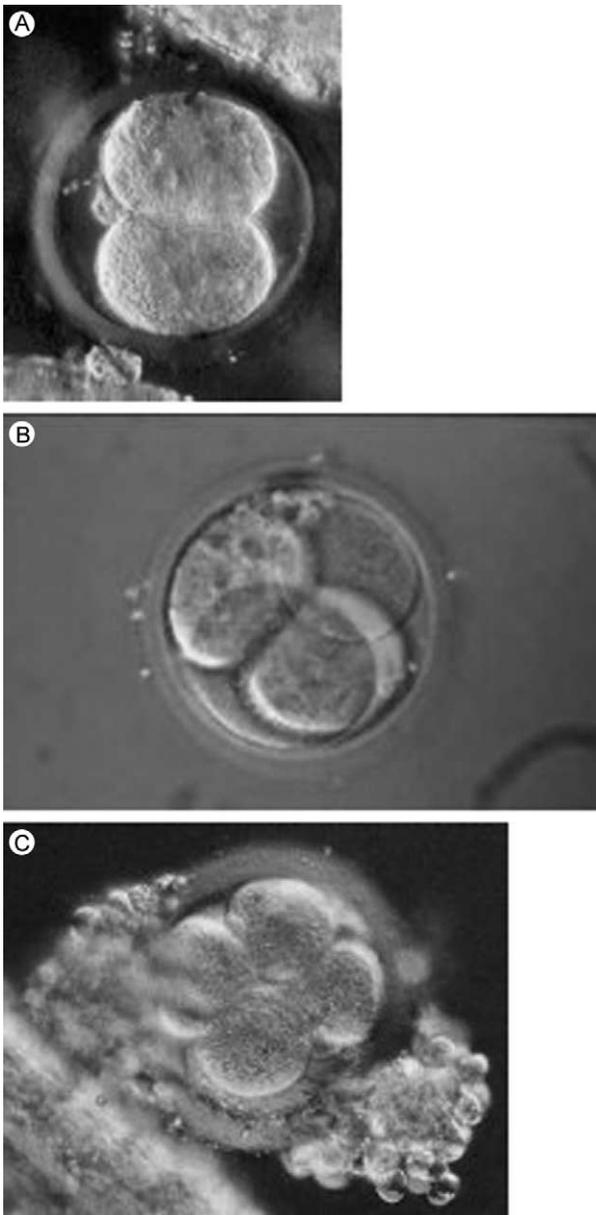


**Figure 6** (A) Fertilized oocyte with 2 pronuclei clearly visible, indicating normal fertilization. (B) Fertilized oocyte with 3 pronuclei visible, indicating abnormal fertilization.

of cells and their morphological appearance are recorded, and based on these parameters, each embryo is graded for quality.

### Embryo Transfer

Embryos that have satisfactorily divided are transferred to the uterus 2 to 3 days after fertilization (Fig. 7). This simple and painless procedure does not



**Figure 7** (A) A 2-cell embryo, day 2 postfertilization. (B) A 4-cell embryo, day 2 postfertilization. (C) An 8-cell embryo, day 3 postfertilization.

require sedation or anesthesia. The embryos are laden on a small plastic catheter, which is gently introduced through the cervix into the uterus. Once in the uterine cavity, the embryos are gently released and the catheter is withdrawn. Given the nature of cycle stimulation, medications are prescribed to support the young embryos hormonally. With an eye to preventing multiple pregnancy, the number of embryos transferred to the uterus should be considered very carefully. In good prognosis cases (young patients in first IVF treatment), 1 or 2 high-quality embryos are usually transferred. Patients with a history of repeated failures, or who are more than 40 years of age, may receive more embryos.

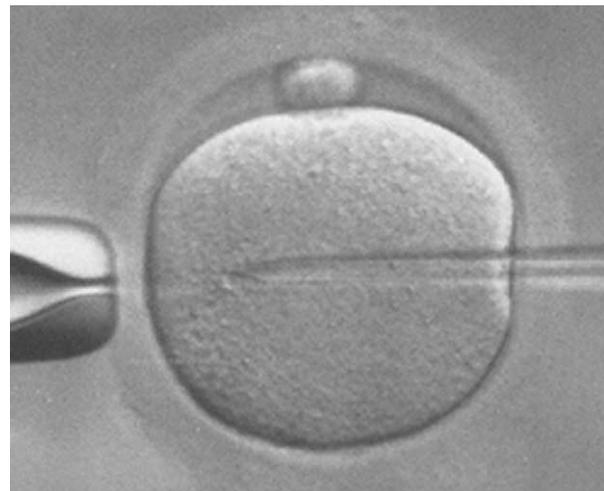
### Other Technologies

#### **Cryopreservation**

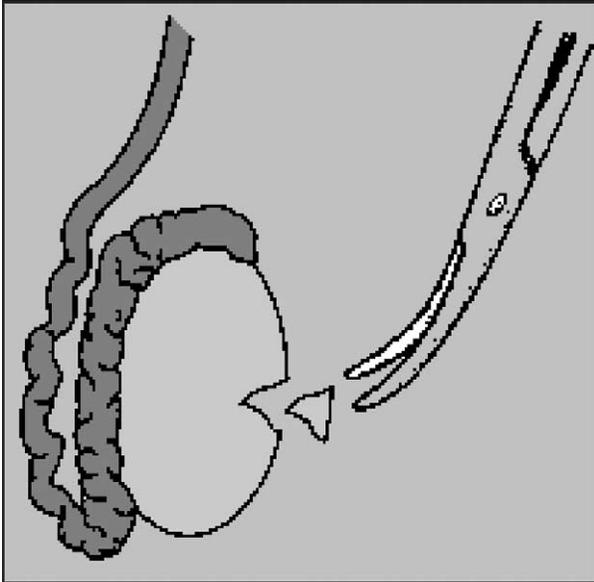
The average number of retrieved oocytes is 10 to 12, of which about 8 are expected to fertilize and cleave. Because only 1 or 2 “fresh” embryos are transferred, the rest are cryopreserved to be used later. Frozen embryos that are kept in  $-196^{\circ}\text{C}$  (liquid nitrogen) can be used successfully many years after they were created.

#### **Intracytoplasmic Sperm Injection**

The intracytoplasmic sperm injection (ICSI) technique (Fig. 8), first introduced in 1993, revolutionized the treatment of male infertility: each mature oocyte is injected with a single sperm cell. ICSI is a very successful solution in cases of extreme oligospermia, even when only a few viable sperm cells can be found in the ejaculate.



**Figure 8** ICSI: A mature oocyte is injected with a single sperm.



**Figure 9** Testicular sperm extraction: A piece of testicular tissue is surgically removed, and sperm cells will be obtained from this tissue to be used in ICSI.

#### **Microsurgical Epididymal Sperm Aspiration**

Microsurgical epididymal sperm aspiration (MESA) involves obtaining immature sperm cells from the epididymis (which joins the testicle to the vas) in cases where obstruction in the genital tract leads to absence of sperm in the ejaculate. The recovered sperm can be used for ICSI, either immediately or after cryostorage.

#### **Testicular Sperm Aspiration**

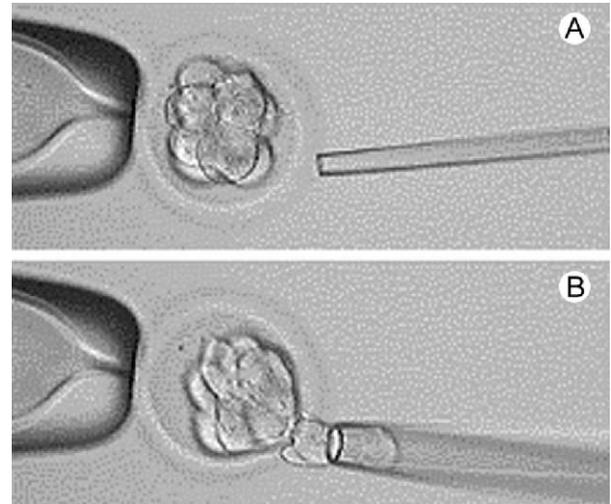
Testicular sperm aspiration (TESA or TESE) (Fig. 9) is a surgical procedure to obtain sperm from within the testicular tissue in azoospermic patients.

#### **Preimplantation Genetic Diagnosis**

In preimplantation genetic diagnosis (PGD), a single blastomere is obtained by embryos biopsy (Fig. 10) at the 6- to 8-cell stage (3 days postfertilization). The chromosomal composition of this cell is assessed by DNA hybridization technology, or a potential abnormal gene is targeted by polymerase chain reaction (PCR). If found to be normal, the embryo is transferred to the uterus. Abnormal embryos are discarded.

#### **Blastocyst Culture**

The first stage, when embryo differentiation is observed, is the blastocyst stage. Embryos reach that stage 5 to 6 days after egg retrieval (Fig. 11). In “natural pregnancy,” the embryo reaches the uterus at this stage. In some cases (e.g., patients after repeated IVF failure),

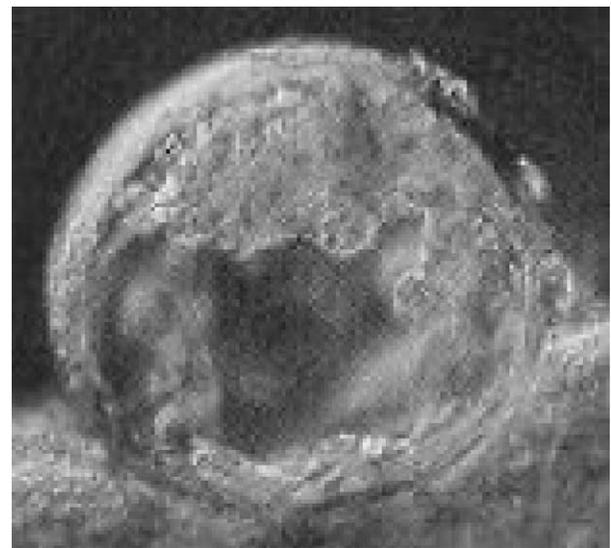


**Figure 10** (A) Embryo biopsy: An 8-cell embryo is held, while a biopsy capillary is introduced through a hole in the zona pellucida. (B) Embryo biopsy: A single cell is gently removed from the embryo.

embryo culture up to the blastocyst stage should be considered. In addition, this technology may help to minimize the multiple pregnancy rate because the implantation potential of embryos at the blastocyst stage is higher than that of embryos at the cleavage stage (days 2–3); therefore, 1 or 2 transferred blastocysts have a good chance of achieving pregnancy.

## **OUTCOME**

In general, it is recommended that the patient should refrain from physical exertion following embryo



**Figure 11** An embryo at the blastocyst stage.

transfer; however, complete bed rest does not seem to be necessary. The presence of pregnancy is established by using a sensitive test for hCG, which is the hormone secreted by the placenta. The test is done 14 days after egg retrieval. If pregnancy is established, monitoring is continued until ultrasound imaging allows a direct visualization of the developing fetus. Success depends on many variables. The most important predictor of success is the age of the female patient. Current IVF technology can deliver approximately a 20% implantation rate (the number of gestational sacs that develop in the uterus divided by the number of transferred embryos). This is close to the natural implantation rate.

## Complications

### **Multiple Pregnancy**

Usually, 1 to 3 embryos are transferred to the uterus in a given cycle, with the hope that 1 or at most 2 will implant and develop. However, occasionally the procedure “over-succeeds” and it is found that a patient has a triplet pregnancy. The outcome of these pregnancies is not optimal given that premature delivery is the rule, with potential severe impact on the newborns. In these cases, selective fetal reduction should be considered.

### **Ovarian Hyperstimulation Syndrome**

Fertility drugs override the natural process of egg development for the purpose of obtaining many eggs. Occasionally, a patient may develop a large number of eggs, a process that can give rise to a clinical syndrome known as ovarian hyperstimulation syndrome (OHSS). The syndrome is characterized by lower abdominal pain, ovarian cysts, and (in its severe form) accumulation of fluid within the abdomen. Occasionally, hospitalization is required, and in extreme cases, termination of pregnancy is necessary.

### **Increased Risk of Cancer?**

About 1 in 10 women will develop cancer of the breast at some stage of their lives. About 1 in 90 women who live to their 70s will develop cancer of the ovary. Because IVF has become a common procedure, it is understandable that quite a number of cancers will develop among women once they are treated in this way. However, like pregnancy itself, stimulating the ovaries during IVF causes the ovarian hormones to reach high levels, and this may accelerate the development of a breast cancer (or an ovarian cancer) that is already present but that has not been detected. It is not known whether the long-term risk of breast

cancer is increased after repeated IVF treatments, although studies are currently under way. There is a strong presumption that repeated IVF cycles, especially if there has been no pregnancy, are likely to increase the risk of cancer of the ovary during later life. There are two reasons for this. First, some studies have implied that the use of fertility drugs is associated with an increase in risk. Second, it is known that anything that stops ovulation and menstruation, such as having been on birth control pills for a number of years or having been pregnant and having breast-fed, is rather protective against later cancer of the ovary. Ovulating 10 eggs in 1 month in an IVF program may have the same effect as 10 months of normal ovulation in terms of risk.

## THE FUTURE

Current efforts focus on technologies that will let us know which embryo in the lab has the potential to implant and become a healthy baby. Routine use of techniques such as PGD and assessing biochemical markers will improve implantation rates.

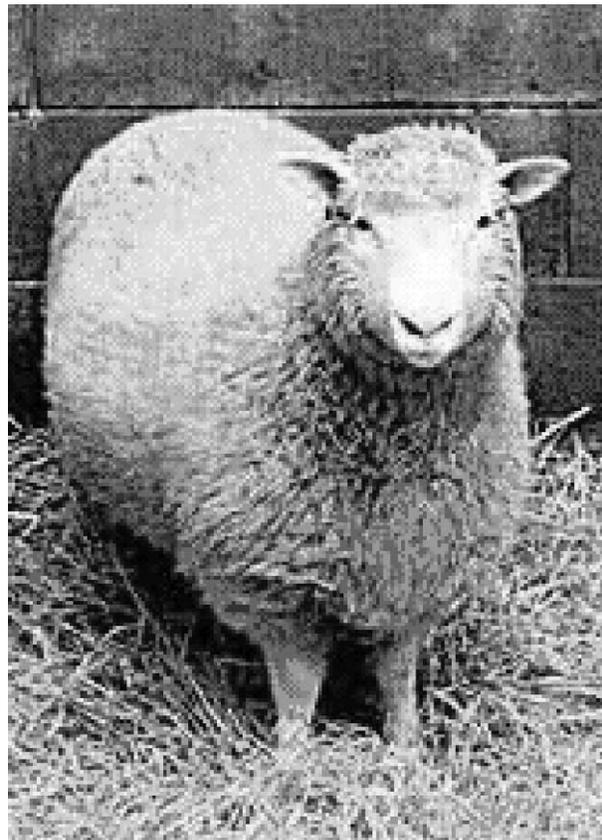


Figure 12 Dolly, the cloned sheep.

The ability to produce embryos *in vitro* has introduced us to cloning and stem cell technology. In short, cloning is based on enucleation of a mature unfertilized oocyte and injection of a nucleus taken from an adult diploid cell. The oocyte cytoplasm has the capacity to reprogram and de-differentiate the donor nucleus to be followed by cleavage and embryo development. Healthy newborns have been produced in domestic animals (with Dolly the cloned sheep being the first [Fig. 12]), although with low efficiency and high rates of abortion, congenital malformations, and neonatal deaths.

The inner cell mass detected at the blastocyst stage is composed of cells that have the potential to differentiate to any tissue. These cells can be kept undifferentiated in culture for a long time, and can serve as a source for any needed tissue, once the signal for differentiation is known. Theoretically, it will be possible to produce any type of tissue or organ to suit the individual patient's needs. A major advantage of "therapeutic cloning" is that immunity-based rejection will be avoided because the genetic material of the implanted cells/organs will be identical to that of the patient.

### See Also the Following Articles

Endometriosis • Fertilization • Gonadotropin-Induced Ovulation • In Vitro Fertilization (IVF) • Infertility, Overview • Ovarian Failure Treatment Strategies: Egg Donation • Pregnancy Endocrinology • Premature Ovarian Failure • Superovulation and Intrauterine Insemination

### Further Reading

- Bolton, V. N., Wren, M. E., and Parsons, J. H. (1990). Preimplantation genetic diagnosis. *Brit. Med. J.* **301**, 1277.
- Bolton, V. N., Wren, M. E., and Parsons, J. H. (1991). Pregnancies after *in vitro* fertilization and transfer of human blastocysts. *Fertil. Steril.* **55**, 830–832.
- Brinsden, P. R. (ed.) (1999). "A Textbook of *in Vitro* Fertilization and Assisted Reproduction: The Bourn Hall Guide to Clinical and Laboratory Practice," 2nd ed. Parthenon, Pearl River, NY.
- Kol, S., and Itskovitz-Eldor, J. (2000). Severe OHSS: Yes, there is a strategy to prevent it! *Hum. Reprod.* **15**, 2266–2267.
- Nygren, K. G., and Andersen, A. N. (2001). Assisted reproductive technology in Europe, 1997: Results generated from European registers by ESHRE, European IVF-Monitoring Programme (EIM), for the European Society of Human Reproduction and Embryology (ESHRE). *Hum. Reprod.* **16**, 384–391.
- Out, H. J., Mannaerts, B. M. J. L., Driessen, S. G. A. J., et al. (1995). A prospective, randomized, assessor-blind, multicentre study comparing recombinant and urinary follicle stimulating hormone (Puregon versus Metrodin) in *in vitro* fertilization. *Hum. Reprod.* **10**, 2534–2540.
- Palermo, G., Jpris, H., Devroey, P., et al. (1992). Pregnancies after intracytoplasmic sperm injection of a single spermatozoon into an oocyte. *Lancet* **340**, 17–18.
- Porter, R. N., Smith, W., Craft, I. L., et al. (1984). Induction of ovulation for *in-vitro* fertilization using buserelin and gonadotropins. *Lancet* **2**, 1284–1285.
- Silber, S. J., Van Steirteghem, A. C., Liu, J., et al. (1995). High fertilization and pregnancy rate after intracytoplasmic sperm injection with spermatozoa obtained from testicle biopsy. *Hum. Reprod.* **10**, 148–152.
- Society for Assisted Reproductive Technology and American Society for Reproductive Medicine. (2000). Assisted reproductive technology in the United States: 1997 results generated from the American Society for Reproductive Medicine/Society for Assisted Reproductive Technology Registry. *Fertil Steril.* **74**, 641–654.
- Stephens, P. C., and Edwards, R. G. (1978). Birth after the re-implantation of a human embryo. *Lancet* **2**, 366.
- Trounson, A., and Mohr, L. (1983). Human pregnancy following cryopreservation, thawing, and transfer of an eight cell stage human embryo. *Nature* **305**, 707–709.
- Wilmut, I., Schnieke, A. E., McWhir, J., et al. (1997). Viable offspring derived from fetal and adult mammalian cells. *Nature* **385**, 810–813.



# Atherogenesis

Lina Badimon and Vicenta Llorente-Cortés

Cardiovascular Research Center, ICCV/CSIC-Hospital Santa Creu i Sant Pau, Barcelona, Spain

## Glossary

**atherogenesis** Early changes within the artery wall that are associated with atherosclerosis.

**atherosclerotic plaque** Thickening of the arterial wall through the accumulation of lipids, macrophages, T lymphocytes, smooth muscle cells, extracellular matrix, calcium, and necrotic debris.

**endothelium** Continuous monolayer composed of endothelial cells that covers the inner surface of the entire vascular system.

**inflammation** Response of the arterial wall to endothelial injury that leads to elevated microbial antigens and elevated circulating levels of immune cytokines.

**lipid accumulation** One of the main events that leads to atherosclerotic plaque progression and that is caused by extracellular and intracellular abnormal lipid clearance.

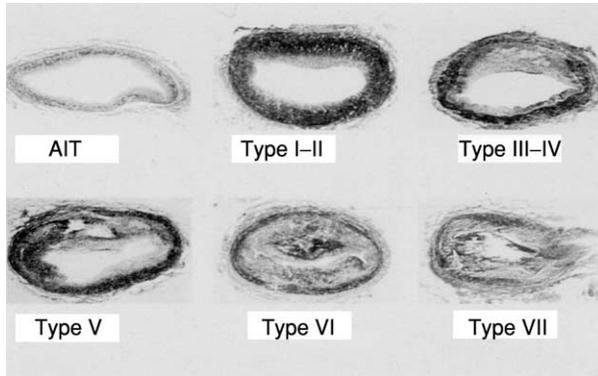
Atherosclerosis has been defined as a thickening of the arterial wall through the accumulation of lipids, macrophages, T lymphocytes, smooth muscle cells, extracellular matrix, calcium, and necrotic debris. Atherosclerosis is the main cause of cardiovascular disease, affecting large and medium-size arteries and coronary arteries. Atherosclerotic plaques may remain silent, without any clinical symptoms, but when plaques suffer fissuring, rupture, or disruption, the ensuing thrombotic complication produces the clinical manifestation of the disease. The pathogenesis of atherosclerosis is complex and multifactorial. Among the identified risk factors are age, insulin resistance syndrome and diabetes, obesity, lack of exercise, smoking, hypertension, and hypercholesterolemia. Various epidemiological studies suggest a crucial role of hyperlipidemia in early atherogenesis. However, high cholesterol levels are not the only causative factor.

## INTRODUCTION

The American Heart Association classified the various stages of lesion progression into six phases and lesion

types (Fig. 1). The first stage has been called arterial intimal thickening (AIT) (type I) and starts with the migration of vascular smooth muscle cells (VSMCs) to the intima and the small accumulation of low-density lipoprotein (LDL) and macrophages in the subendothelial-intimal space. Modified LDL is a chemottractant for monocytes and VSMCs; these cells, by taking up modified LDL in a noncholesterol down-regulated manner, become foam cells initiating the fatty streak formation (type II). The extracellular lipid accumulation coming from death cells besides the foam cells is characteristic of lesion type III and lesion type IV typical of the third decade of life. In more advanced lesions (types V, VI, and VII), pathological changes occur in VSMCs of the media layer that contain a significant number of foam cells. In type V lesions, several collagen layers cover the lipid core. These lesions may have a cap of VSMCs and extracellular matrix (ECM) covering the lipid core that modulates their susceptibility to rupture and thrombus formation. The successive accumulation of collagen fibers leads to a progressive lumen narrowing. During the fourth decade of life, lesions can rupture and trigger thrombotic depositions (type VI lesions). During this stage, the formation of an occlusive thrombus can have dramatic effects. During the fifth decade of life, most of the lesions show an advanced stage and are composed mainly of calcifications (type VII lesions) or fibrous tissue (type VIII lesions).

Still not much is known about the disease initiation and progression, and the major risk factors can explain only approximately 50% of the coronary events found in humans. A major mystery in human atherogenesis is the well-known variation in lesion progression among individuals with similar plasma lipid profiles or risk factors. Furthermore, other factors affecting the molecular and cellular mechanisms in the arterial wall may be critical for the pathogenesis of the disease. Various findings suggest a diversity of agents that can induce endothelial injury, leading to a vascular response in a harmful chronic state. Besides LDL,



**Figure 1** Classification of the atherosclerotic lesions according to the American Heart Association. AIT, arterial intimal thickening; type I-II; type III-IV; type V; type VI; type VII. (Original preparations.)

cytokines and growth factors are important regulators of the inflammatory response, cellular growth, and lipid deposition. All of these processes are responsible for atherosclerotic plaque progression. During the final stages of the disease, arterial plaques become complex aggregations of lipids, ECM, and necrotic and apoptotic cells covered by a fibrous cap consisting primarily of smooth muscle cells surrounded by collagens and proteoglycans (PGs). The vulnerable plaques contain a lipid-rich core and a soft fibrous cap.

## ROLE OF INFLAMMATION

It has been postulated that inflammation is involved in atherogenesis due to the observation that elevated levels of cytokines can be measured during acute coronary syndrome (ACS) presentation. In fact, one of the hypotheses for the initiation of atherosclerosis is the so-called “response to injury hypothesis” involving the response of the innate immune system to the accumulation and modifications of lipoproteins in the arterial intima. Various constituents of the modified lipoproteins trigger the production of mediators of innate immunity. There are also nonlipid mediators involved in inflammation such as homocysteine, angiotensin II, and microbial products that can induce the elaboration of cytokines from atheroma-associated cells.

In normal circumstances, the endothelial monolayer in contact with flowing blood is inert to the adhesion of leukocytes. The situation changes in dysfunctional endothelium. One of the endothelial-leukocyte adhesion molecules involved mainly in the early adhesion of mononuclear leukocytes to arterial endothelium is the vascular cell adhesion molecule-1 (VCAM-1), which binds particularly

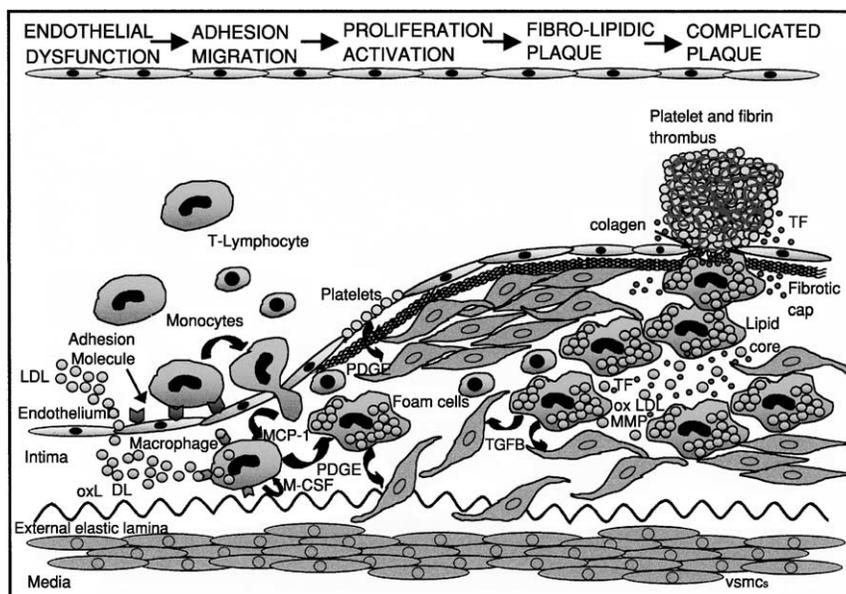
those classes of leukocytes found in nascent atheroma: the monocyte and the T lymphocyte. In addition to VCAM-1, P- and E-selectin and intercellular adhesion molecule-1 (ICAM-1) also seem to contribute to leukocyte recruitment. As shown in Fig. 2, once the leukocyte adheres to the endothelium, it enters into the intima by diapedesis, a process that is facilitated by various chemokines such as monocyte chemoattractant protein-1 (MCP-1), interleukin-8 (IL-8), and a trio of CXC chemokines induced by interferon- $\gamma$ . Once they are in the intima, monocytes acquire characteristics of macrophage by expressing certain scavenger receptors (SRs), such as scavenger receptor A (SRA-I and SRA-II) and CD36, that internalize modified lipoproteins, leading to foam cell formation.

One of the most important clinical markers of inflammation is the C-reactive protein (CRP), a marker of inflammation that has been shown in multiple epidemiological studies to predict incidence of myocardial infarction, stroke, peripheral arterial disease, and sudden cardiac death. In terms of clinical application, some data indicate that CRP seems to be a strong predictor of cardiovascular events and adds prognostic information at all levels of calculated Framingham risk and at all levels of metabolic syndrome. The feature that distinguishes CRP from LDL cholesterol is the fact that inflammation (but not elevated cholesterol LDL) plays a major role in nearly all processes associated with metabolic syndrome.

## ROLE OF LIPOPROTEINS

Cholesterol is transported into the vessel wall as a component of the lipoproteins. LDLs are considered the most atherogenic lipoproteins because they accumulate in the intima and carry large amounts of plasmatic cholesterol (up to 70%). The coronary risk lipid profile that uses the total cholesterol/high-density lipoprotein (HDL) ratio (or the LDL/HDL ratio) predicts risk of vascular disease and is used as a tool for therapeutic management of patients at risk for vascular disease. Although these methods are predictive of coronary artery disease (CAD) in general, it is also well known that the extent of occlusive disease and CAD varies greatly among individuals with similar cholesterol and HDL lipid profiles. For this reason, the National Cholesterol Education Program Expert Panel revised these guidelines and recommends monitoring LDL and HDL in the context of coronary heart disease (CHD) risk factors and “risk equivalents.”

High levels of circulating LDL have shown to be predictive markers of high risk of cardiovascular disease (CVD) in individual persons. It appears that



**Figure 2** Illustration showing the molecular mechanisms that lead to initiation and progression of the atherosclerotic lesions. PDGF, platelet-derived growth factor; M-CSF, macrophage colony-stimulating factor; MCP-1, monocyte chemoattractant protein-1; oxLDL, oxidized LDL; TF, tissue factor.

discrete LDL subclasses carry various levels of atherogenicity. The “atherogenic lipoprotein phenotype” describes a combination of moderate hypertriglyceridemia, low HDL cholesterol, and a predominance of small dense LDL particles. This dyslipemia is prevalent in patients with metabolic syndrome, in those with type 2 diabetes, and in postmenopausal women.

### Oils, Lipoproteins, and Coronary Risk

Dietary fats and oils differ in the chain lengths of their constituent fatty acids and the number and geometry of their double bands. These differences markedly affect concentrations of lipids in plasma, and differences in the amount and type of fat in the diet can induce differences of 30 to 40% in serum LDL concentrations. [Table I](#) shows the main saturated fatty acids (SFAs), monounsaturated fatty acids (MUFAs), and polyunsaturated fatty acids (PUFAs). When SFAs are replaced by unsaturated fats, total plasma cholesterol is lowered. A review of metabolic studies, prospective cohort studies, and clinical trials indicates that there are multiple mechanisms by which diet potentially influences risk of CHD, and there are dietary strategies that are effective in preventing CHD ([Table II](#)). As such, the substitution of nonhydrogenated unsaturated fats for saturated and *trans* fats; the increase in the consumption of omega-3 fatty acids from fish, fish oil supplements, and plant sources; and the consumption

of a diet high in fruits, vegetables, nuts, and whole grains and low in refined grain products are good strategies to prevent CHD.

### Modified Forms of LDL

Extracellular accumulation of lipids in the arterial intima occurs very early in response to increased plasma lipoprotein levels in animals. PGs and protein-bound lipoprotein particles, perhaps in a micro-environment shielded from plasma antioxidants, can undergo modifications. As shown in [Table III](#), such modifications include oxidation, aggregation, and enzymatic and nonenzymatic modifications of LDL. Modified forms of LDL are associated with increased atherogenicity because the physicochemical

**Table I** Main Dietary Fatty Acids

Saturated fatty acids	Monounsaturated fatty acids	Polyunsaturated fatty acids
Lauric acid (12:0)	Oleic acid (18:1n-9)	Linoleic acid (18:2n-6)
Myristic acid (14:0)	<i>Trans</i> (16:1 + 18:1)	$\alpha$ -Linolenic acid (18:3n-3)
Palmitic acid (16:0)		Eicosapentenoic acid (20:5n-3)
Stearic acid (18:0)		Docosahexenoic acid (22:6n-3)

**Table II Mechanisms by Which Diet Potentially Influences Risk of CHD**

- 
- Lipid levels
    - LDL cholesterol
    - HDL cholesterol
    - Triglycerides
    - Lipoprotein
  - Blood pressure
  - Thrombotic tendency
  - Cardiac rhythm
  - Endothelial function
  - Systemic inflammation
  - Insulin sensitivity
  - Oxidative stress
  - Homocysteine level
- 

*Source.* Modified from Hu, F. B., and Willett, W. C. (2002). Optimal diets for prevention of coronary heart disease. *J. Am. Med. Assoc.* **288**, 2569–2578.

properties of the lipoprotein become altered. This results in changes in the biological properties of the LDL and also in an increase in the susceptibility of the LDL to other types of modifications.

## Cellular Effects of Modified LDL

### **Endothelial Dysfunction**

Hypercholesterolemia induces an increase of leukocyte recruitment by an increased expression of adhesion molecules per se and those induced by cytokines, a decrease in endothelium-dependent vasodilatation,

**Table III Modified Forms of LDL**

LDL modification	Taken up by	Reference(s)
Interaction with PGs	MØ, VSMCs	Camejo <i>et al.</i> (1993)
Glycosylation	MØ, VSMCs	Witztum <i>et al.</i> (1989) Parlinski <i>et al.</i> (1995)
(4-HNE)-LDL	MØ	Jürgens <i>et al.</i> (1987) Hoff <i>et al.</i> (1989)
MDA-LDL	MØ	Fogelman <i>et al.</i> (1980) Haberland <i>et al.</i> (1988)
Self-aggregation	MØ, VSMCs	Khoo <i>et al.</i> (1988)
Carbamylation and maledylation	MØ	Weisgraber <i>et al.</i> (1978)
Desialylation	MØ, VSMCs	Orekhov <i>et al.</i> (1989) Tertov <i>et al.</i> (1992)
Oxidation	MØ	Witztum <i>et al.</i> (1994)

*Note.* MØ, macrophage; VSMCs, vascular smooth muscle cells; (4-HNE)-LDL, 4-hydroxynonenal-modified LDL, MDA-LDL, malondialdehyde-modified LDL.

and alterations in the thrombosis/fibrinolysis balance. Our group has demonstrated that systemic hypercholesterolemia can induce endothelial dysfunction by altering the expression of genes that are regulated through the down-regulation of sterol regulatory element-binding proteins (SREBPs) in endothelial cells.

### **Synthesis and Degradation of Extracellular Matrix**

Modified LDL can modulate the synthesis of PGs in various cell types. The increase of PG synthesis induced by LDL might have important consequences for the intimal LDL retention. In addition, the exposure of endothelial cells to apolipoprotein E (apoE)-containing HDL has been shown to stimulate the production of heparan sulfate proteoglycans (HS-PGs) that have increased sulfation.

Lipoproteins also modulate the expression of metalloproteinases (MMPs), that is, enzymes that are able to digest various connective tissue components. An increase in metalloproteinase expression and activity is associated with the disruption of the fibrous cap of lesions and plaque rupture. In addition, the breakdown products of the ECM may be biologically active and might increase processes that are fundamental for the pathogenesis of the atherosclerosis.

### **Foam Cell Formation**

Modified lipoproteins are taken up through mechanisms not regulated by cholesterol, leading to high intracellular cholesteryl ester accumulation and foam cell formation. The accumulation of lipid-laden foam cells is one of the earliest steps in the progression of the atherosclerotic plaque. In macrophages, the SRs are mainly responsible for modified LDL uptake. Our group has described a main role for LDL receptor-related protein (LRP) in the uptake of aggregated LDL (agLDL) by VSMCs. SRs and the LRP have been detected in human atherosclerotic lesions and, therefore, could play an important role in foam cell formation in the arterial intima. Several modified lipoproteins have demonstrated to up-regulate their own receptors; for example, oxidized LDL (oxLDL) increases the expression of several SRs such as CD36, SR-A, and lysil oxidase-1 (LOX-1) in various cell types. In addition, agLDL up-regulates the expression of their receptor LRP in VSMCs. Most of the modified lipoproteins could lead to a progressive cholesteryl ester accumulation, not only by being taken up through non-down-regulated receptors but also by up-regulating their own receptors.

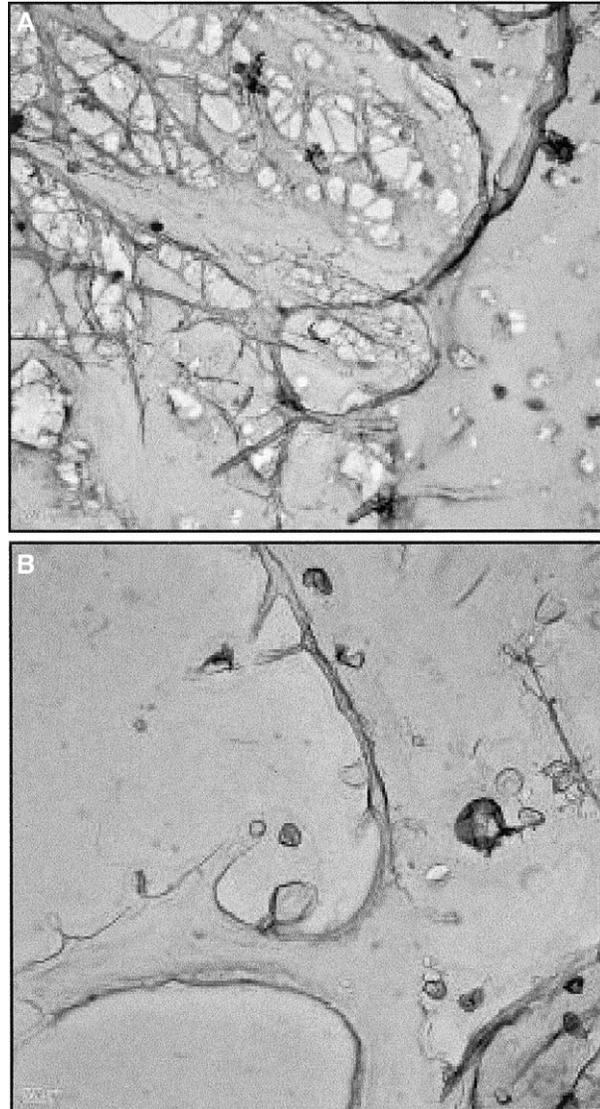
## Lipoproteins and Metabolic Syndrome

Metabolic syndrome was defined by the National Institutes of Health in 2001 as a cluster of disorders that include abdominal obesity, insulin resistance, diabetes, endothelial dysfunction, elevated blood pressure, and impaired fibrinolysis. The risk factors that constitute metabolic syndrome consist of atherogenic dyslipidemia, elevated blood pressure, elevated plasma glucose, and a prothrombotic state. Metabolic syndrome is closely linked to the metabolic derangement called insulin resistance. This condition is characterized by a generalized defect in the insulin-signaling pathway. Because insulin induces a myriad of metabolic responses, a defect in insulin signaling results in several metabolic changes. The presence of insulin resistance predisposes to the development of type 2 diabetes. There are four major causes of insulin resistance: genetics, obesity, lack of exercise, and diet composition. Insulin resistance and metabolic syndrome are related to the atherogenic dyslipidemia also called “atherogenic lipoprotein phenotype,” which is characterized by increased triglyceride-rich lipoproteins, increased small LDL particles, and reduced HDL levels. These changes in lipoprotein and fatty acid profile observed in insulin resistance may influence various proatherosclerotic mechanisms such as membrane lipid composition, metabolism, and signal–transduction pathways.

## ROLE OF THE VARIOUS COMPONENTS OF THE VASCULAR WALL

### Extracellular Matrix

The ECM of the arterial intima is a relatively large compartment made of collagen, elastin, complex PGs, hyaluronate, and multidomain proteins such as fibronectin, laminin, and tenascin (Fig. 3). The ECM occupies 60% of the arterial intima and regulates numerous cellular functions. The main PGs structuring the ECM are chondroitin sulfate (CS)-PGs, such as versican and biglycan, which have the longest negatively charged glycosaminoglycan (GAG) chains and are synthesized mainly by VSMCs. HS-PGs, such as perlecan, are constituents of the basement membrane and are synthesized mainly by endothelial cells and VSMCs. In addition, other HS-PGs, such as syndecan and glypican, are found in the cell membranes of the vascular cells. Whereas CS-PGs play a major role in LDL retention in the arterial intima, cell surface HS-PGs are dynamic molecules that mediate



**Figure 3** Extracellular matrix ultrastructure of VSMCs. Cells were washed, fixed, and processed for freeze-drying and replication in electron microscopy. (A) Nonenzymatically treated human VSMCs. (B) Heparinase-treated human VSMCs. (Modified from Llorente-Cortés *et al.* (2002). *Arterioscler. Thromb. Vasc. Biol.* 22, 1905–1911.)

ligand catabolism. Collagens play a central role not only in maintaining the integrity and stability of the wall but also in many cellular functions.

### Heparan Sulfate Proteoglycans

HS-PGs consist of a protein core and at least one heparan sulfate side chain. HS-PGs can be secreted (perlecan) or shed from the cell surface (perlecan, syndecan, and glypican). Syndecan and perlecan can function as receptors for several growth factors and

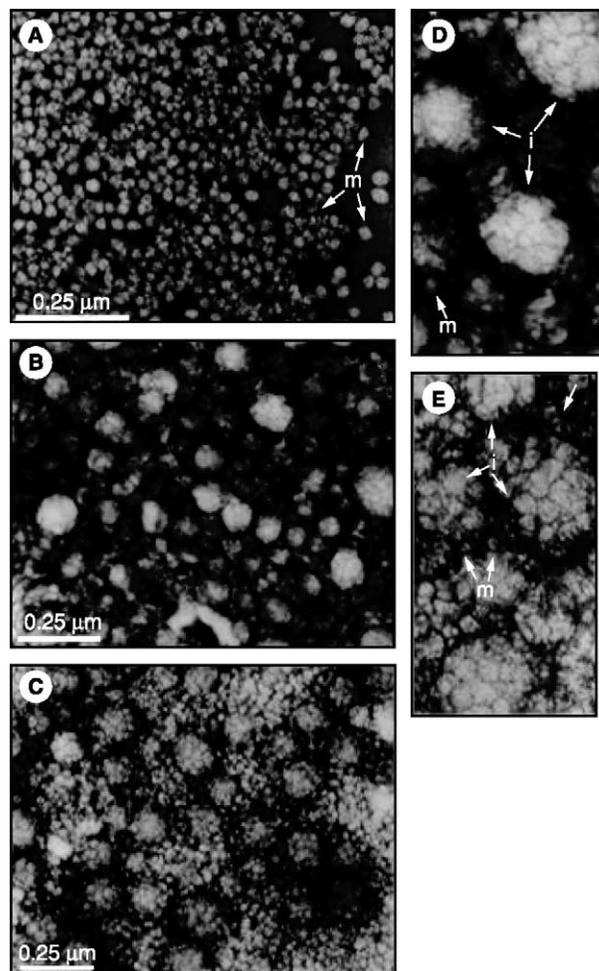
lipoproteins. HS-PGs act as a receptors for basic fibroblast growth factor (bFGF), which stimulates VSMC proliferation. They may act as potential receptors for atherogenic lipoproteins, such as apoE-triglyceride-rich lipoprotein particles and lipoprotein lipase, or may facilitate the uptake of ligands by a process called ligand transfer to lipoprotein receptors, such as the LRP.

### Chondroitin Sulfate Proteoglycans

CS-PGs are considered to be atherogenic due to their ability to trap and aggregate cholesterol-rich lipoproteins. Because one of the key initiation events in early atherogenesis is the subendothelial retention of atherogenic lipoproteins, CS-PGs play a main role in the initiation of atherogenesis (“response to retention” hypothesis). Apoproteins have been found to colocalize with specific PGs in human atherosclerotic plaques, providing support to the “response to retention” hypothesis. Several studies demonstrate that CS-PGs act as sites for apoB-100 lipoprotein retention by the interaction between positively charged heparin-binding domains on apoB-100 or apoE and negatively charged GAG chains of the PGs. The GAGs of versican induce alterations in the LDL particles that facilitate oxidative enzymatic modification, fusion, and aggregation of LDL particles. **Figure 4** shows an electron microscopy image of versican-modified LDL in comparison with nLDL and agLDL. Versican clearly induces the aggregation of LDL particles. Aggregation substantially increases its binding to arterial PGs due to the creation of a multimeric ligand with better exposure of key positive residues on apolipoprotein B. Modified LDL is avidly taken up by macrophages and VSMCs, leading to foam cell formation. In addition, some CS-PGs such as decorin, a small DS/CS PG, regulate cell growth and growth factor activity.

### Collagen

Collagen is the major component of the ECM of the vessel wall and has a critical impact in atherogenesis. Collagen is essential for the maintenance of vessel wall integrity and elasticity. Collagen is involved in the process of cell differentiation, adhesion, migration, proliferation, and apoptosis. It has been reported that the bulk of vascular collagen is produced by smooth muscle cells. However, collagens can also be produced by endothelial cells, adventitial fibroblasts, and macrophages. Among the 19 collagens described, 13 are found in the vessel wall or are expressed by cells of the vessel wall *in vitro*. The most abundant are type I and type III collagen, which form an interconnected



**Figure 4** Electron microscopy of nLDL, agLDL, and versican-modified LDL. (A) nLDL. (B) Precipitable fraction from vortexed LDL. (C) Precipitable fraction from versican LDL. (D) Doubly magnified section from panel B. (E) Doubly magnified section from panel C. Samples were negatively stained and observed by electron microscopy. m, monomeric; f, fused; a, aggregated. (Modified from Llorente-Cortés *et al.* (2002). *Arterioscler. Thromb. Vasc. Biol.* 22, 387–393.)

network of cross-banded fibers along with smaller amounts of type VI and type VIII collagen, associated with the fibers and the elastic lamellae in the blood vessel wall. Type I and type III collagen seems to play an important role in arterial wall remodeling in response to hypertension. It has also been reported that type I collagen expression was localized mainly in the adventicia, outer media, and intima and that type III collagen expression was uniformly localized across the arterial wall in response to an elevation of blood pressure. Type III collagen accumulates in atherosclerotic lesions, and in various *in vitro* studies in cultured endothelial cells. The involvement of type

VIII collagen in processes of endothelial differentiation and organization and *in vitro* angiogenesis has been demonstrated.

## Cellular Components

### Endothelium

In health, the vascular endothelium forms a multi-functional interface between circulating blood and the various tissues and organs in the body. It constitutes a selectively permeable barrier for macromolecules as well as a nonthrombogenic surface that actively maintains the fluidity of blood. The endothelium is a complex and dynamic organ that responds to environmental stimuli and activates vasoactive substances, including vasoconstrictors such as angiotensin II and vasodilators such as nitric oxide (Table IV). These vasoactive substances mediate vascular tone, structure, and function influencing VSMC growth, apoptosis, platelet aggregation, monocyte and leukocyte adhesion, and thrombosis. The homeostasis of vasoactive substances is disrupted by endothelial dysfunction, leading to changes in vascular structure and function. Hypertension and other risk factors for CVD are associated with endothelial dysfunction and vascular remodeling. Elevated angiotensin II activity,

which is strongly correlated with hypertension, is a major trigger of endothelial dysfunction in hypertensive patients. Angiotensin II stimulates nicotinamide adenine dinucleotide phosphate (NADPH)/nicotinamide adenine dinucleotide (NADH) oxidase in the endothelium, VSMCs, and adventitia of blood vessels to generate reactive oxygen species, leading to endothelial dysfunction, cell growth, and inflammation. These changes result in up-regulation of endothelin-1, adhesion molecules, nuclear factor- $\kappa$ B (NF- $\kappa$ B), and other inflammatory mediators as well as increased breakdown of nitric oxide and uncoupling of nitric oxide synthase. Thus, the balance of vasoconstriction and vasodilation is disrupted, with resultant vascular remodeling and injury. Through these processes, endothelial dysfunction aggravates existing hypertension, accelerating the progression of atherosclerosis. Endothelial dysfunction is now recognized as a key early risk factor for cardiovascular morbidity and mortality. Certain consequences of endothelial dysfunction are directly related to the pathogenesis of atherosclerosis and its complications. Pathophysiological stimuli of arterial endothelial dysfunction that are especially relevant to atherogenesis include activation by cytokines and bacterial endotoxin, infection by viruses, advanced glycosylation end products that are generated in diabetes and with aging, hyperhomocysteinemia, and hypercholesterolemia and oxidized LDL. In addition to these humoral stimuli, it is clear that biochemical forces generated by flowing blood can also influence endothelial cell structure and function, modulating the expression of pathophysiologically relevant genes. The possibility that hemodynamic forces can act as pathophysiological stimuli for endothelial dysfunction provides a conceptual rationale for the long-standing observation that the earliest lesions of atherosclerosis characteristically develop in a nonrandom pattern, the geometry of which correlates with branch points and other regions of altered blood flow. There have been documented a variety of changes in the metabolic and synthetic activities of endothelial cells in response to defined biomechanical forces. These include the production of arachidonate metabolites, growth factors, coagulation and fibrinolytic components, ECM components, and vasoactive mediators. Certain of these more acute shear-induced changes appear to involve regulation at the level of rate-limiting enzymes and/or substrate availability. However, especially in the case of delayed responses in which *de novo* synthesis occurs, up-regulation of gene expression appears to occur as a direct consequence of exposure to fluid mechanical forces. There are genes, such as platelet-derived

Table IV Factors Synthesized by the Endothelium

Factors	Process	Contribution
PDGF	Growth	Stimulator
TGF- $\beta$		Inhibitor
Heparin-like GAG		Inhibitor
NO	Vasodilatation	Stimulator
Endothelin		Inhibitor
Angiotensin II		Inhibitor
TF	Thrombosis	Stimulator
Trombomodulin		Stimulator
vWF		Inhibitor
t-PA	Fibrinolysis	Stimulator
PAI-1		Inhibitor
Prostacyclin	Others	
E-selectin, P-selectin, ICAM-1, VCAM-1.		
IL-1, IL-6, IL-8, MCP-1, GM-CSF		

Note. PDGF, platelet-derived growth factor; TGF- $\beta$ -transforming growth factor; NO, nitric oxide; TF, tissue factor; vWF, von Willebrand factor; tPA, tissue plasminogen activator; PAI-1, tissue plasminogen activator inhibitor; ICAM-1, intercellular cell adhesion molecule; VCAM-1, vascular cell adhesion molecule; IL, interleukin; MCP-1, monocyte chemoattractant protein; GM-CSF, granulocyte macrophage colony-stimulating factor.

Table V Cell Surface Receptors That Mediate the Uptake of Modified LDL

Receptors	Ligands	Cellular type	Atherosclerotic plaque	References
LDL receptor	Unmodified LDL	EC, VSMCs and MØ	±	Goldstein (1979)
	LDL interacting with PGs			Kodama (1990)
SRs-AI, SR-AII	OxLDL			Acton <i>et al.</i> (1994)
CD36	OxLDL,HDL	MØ	+	Rigotti <i>et al.</i> (1997)
				Endeman <i>et al.</i> (1993)
CD68	OxLDL	MØ	++	Ramprasad <i>et al.</i>
FcR2/2-B2	OxLDL aggregates	MØ	++	Stanton (1992)
				Sawamura <i>et al.</i> (1997)
LOX-1	OxLDL	EC, VSMCs, and MØ	++	
SRs-PSOX	OxLDL	MØ	++	
LRP	AgLDL	VSMCs and MØ	+++	Llorente <i>et al.</i> (2000)
	Versican-fused LDL			Luoma (1994)
	ApoE-lipid complexes			

*Note.* Expression levels: ±, low or absent; +, moderate; ++, high; + + +, very high. EC, endothelial cells; VSMCs, vascular smooth muscle cells, MØ, macrophages; SRs-AI, SR-AII, scavenger receptors class A; SRs-BI, scavenger receptors class B; CD68, scavenger receptors class D; FcR2/2-B2, receptor for the Fc domain of immunoglobulins; LOX-1, scavenger receptors class E; SRs-PSOX, scavenger receptors for phosphatidylserine and oxidized lipoproteins; LRP, low-density lipoprotein receptor-related protein.

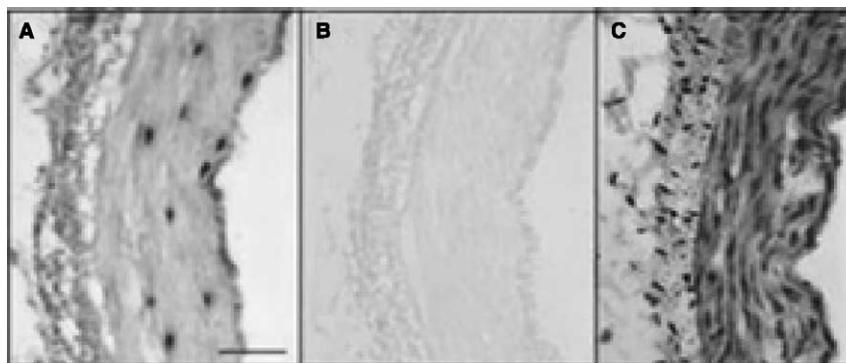
growth factor-B (PDGF-B), MCP-1, VCAM-1, and endothelin-1, that have a “shear stress response element” in the promoter.

### Monocytes/Macrophages

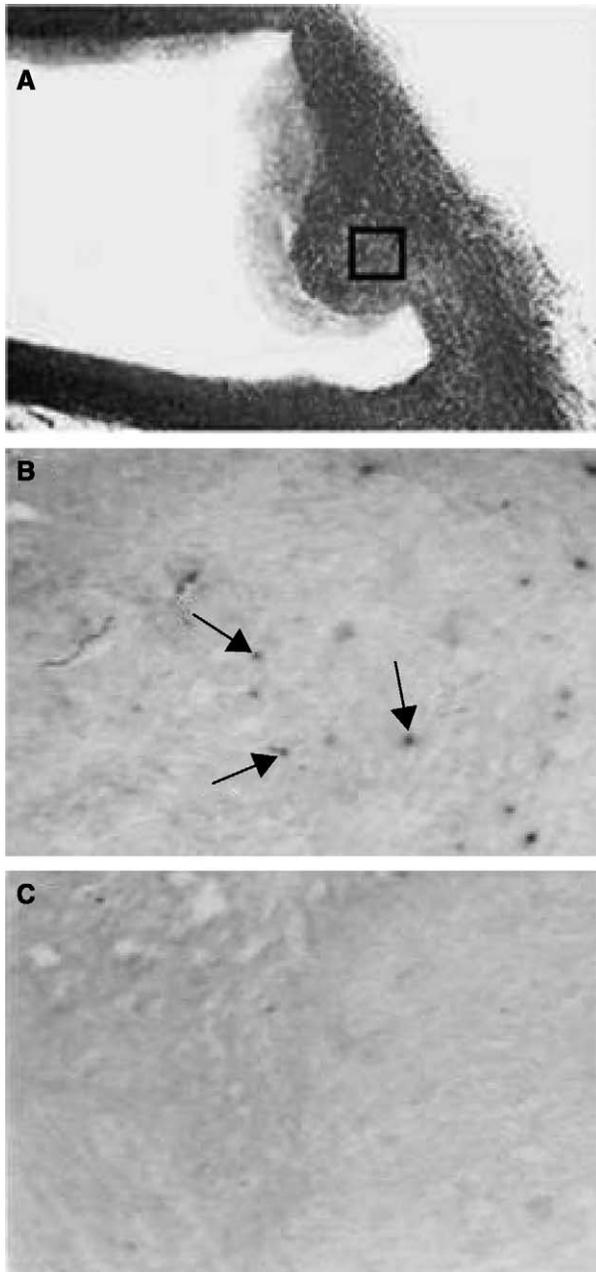
The state and function of the macrophages in the atherosclerotic lesions may be critical for the development of atherosclerosis. Macrophages play an important role in innate immune responses, cellular adhesion, phagocytosis of apoptotic cells, and lipid uptake. Most of these macrophage functions are mediated by SRs. Since the cloning of the first two macrophage SRs (now called SRA-I and SRA-II), the broad SR family has

grown considerably (Table V). On the basis of functional studies and expression in the arterial intima, only some of the SRs are good candidates to contribute to atherosclerotic foam cell formation.

Besides their role in lipid accumulation, macrophages may contribute to atherosclerosis through secretory inflammatory products. In fact, atherosclerosis can be considered a chronic inflammatory process. Activated lymphocytes and macrophages with a wide expression of class II histocompatibility antigen have been found at every stage of atherosclerotic lesions, indicating that macrophages may participate in local immune responses.



**Figure 5** NOR-1 overexpression in angioplasty. *In situ* reverse transcription polymerase chain reaction (RT-PCR) shows the induction of NOR-1 in medial VSMCs after PTCA, with serial sections from a dilated artery. (A) Vessel showing NOR-1 expression detected by antisense oligonucleotide anti-NOR-1. (B) Negative RT-PCR controls. (C) Hematoxylin- and eosin-stained vessel. (Modified from Martínez-González *et al.* (2003). *Circ. Res.* **92**, 96–103.)



**Figure 6** *In situ* hybridization analysis of LRP expression in a hypercholesterolemic pig. (A) Masson trichromic stain of samples from abdominal aorta. (B) *In situ* hybridization with LRP antisense riboprobe. (C) Controls with LRP sense riboprobe. Arrows indicate LRP mRNA expression. Magnification  $\times 100$ . (Modified from Llorente-Cortés *et al.* (2002). *Circulation* **106**, 3104–3110.)

### Vascular Smooth Muscle Cells

VSMCs represent an average of 50% of the cellular component in advanced atherosclerotic plaque and may reach 90 to 95% in early lesions. In response to multiple stimuli, VSMCs from the arterial tunica media are activated and migrate to the intima, where

they proliferate. These seem to be early steps in the onset of the atherosclerotic process. The proliferation and migration of VSMCs from the media to the intima is one of the key events in early atherosclerosis. The understanding of the molecular mechanisms involved in VSMC activation and differentiation requires an accurate mapping of the cascade of transcription factors induced by atherogenic stimuli. Recently, various nuclear receptors, including peroxisome proliferator-activated receptors (PPARs), retinoid receptors, both retinoid X receptors (RXRs) and retinoic acid receptors (RARs), and retinoid-related orphan receptors (RORs), have been identified in VSMC activation/proliferation and, consequently, have been implicated in atherogenic processes. Our group has identified neuron-derived organ receptor-1 (NOR-1) as a new early-response gene in VSMCs. NOR-1 is strongly increased by growth factors and thrombin. NOR-1 is overexpressed in atherosclerotic lesions from patients with CAD (Fig. 5), and balloon angioplasty transiently induces NOR-1 in porcine coronary arteries. These results suggest that NOR-1 may play a role in the molecular mechanisms underlying both spontaneous and accelerated atherosclerosis.

VSMCs also contribute to the lesion by synthesizing ECM. In fact, proliferative VSMCs have a high capacity to synthesize sulfated PGs, and it is well established that PGs in the arterial wall are involved in the focal deposition of cholesterol-rich particles during the early phases of atherogenesis.

Finally, VSMCs also have great importance in foam cell formation. Our group demonstrated that agLDL can cause high intracellular cholesteryl ester accumulation in VSMCs. We demonstrated, for the first time, that in VSMCs, LRP mediates the binding and internalization of agLDL. We also demonstrated that in the absence of LRP function, VSMCs are unable to accumulate cholesterol. In addition, this receptor is up-regulated by agLDL uptake, leading to intracellular lipid accumulation. Figure 6 shows LRP up-regulation in VSMCs of hypercholesterolemic animals.

### See Also the Following Articles

ABCA1 Defects • Atherosclerosis • Collagen Metabolism Disorders • Dysbetalipoproteinemia and Type III Hyperlipidemia

### Further Reading

Barnes, M. J., and Farndale, R. W. (1999). Collagens and atherosclerosis. *Exp. Gerontol.* **34**, 513–525.

- DeFronzo, R. A., and Ferrannini, E. (1991). Insulin resistance: A multifactorial syndrome responsible for NIDDM, obesity, hypertension, dyslipidemia, and atherosclerotic cardiovascular disease. *Diabetes Care* **14**, 173–194.
- Falk, E., Prediman, K. S., and Fuster, V. (1995). Coronary plaque disruption. *Circulation* **92**, 657–671.
- Fuster, V., Badimon, L., Badimon, J. J., *et al.* (1992). The pathogenesis of coronary artery disease and the acute coronary syndromes (part I). *N. Engl. J. Med.* **326**, 242–250.
- Fuster, V., Badimon, L., Badimon, J. J., *et al.* (1992). The pathogenesis of coronary artery disease and the acute coronary syndromes (part II). *N. Engl. J. Med.* **326**, 310–318.
- Grundty, S. M. (1998). Hypertriglyceridemia, atherogenic dyslipidemia, and the metabolic syndrome. *Am. J. Cardiol.* **81**, B18–B25.
- Grundty, S. M., Abate, N., and Chandalia, M. (2002). Diet composition and the metabolic syndrome: What is the optimal fat intake? *Am. J. Med.* **113**, S25–S29.
- Hu, F. B., and Willett, W. C. (2002). Optimal diets for prevention of coronary heart disease. *J. Am. Med. Assoc.* **288**, 2569–2578.
- Libby, P. (2002). Inflammation in atherosclerosis. *Nature* **420**, 868–874.
- Llorente-Cortés, V., Martínez-González, J., and Badimon, L. (2000). LDL receptor-related protein mediates uptake of aggregated LDL in human vascular smooth muscle cells. *Arterioscler. Thromb. Vasc. Biol.* **20**, 1572–1579.
- Llorente-Cortés, V., Otero-Viñas, M., Sanchez, S., Rodríguez, C., and Badimon, L. (2002). Low density lipoprotein upregulates low-density lipoprotein receptor-related protein expression in vascular smooth muscle cells. *Circulation* **106**, 3104–3110.
- Martínez-González, J., Rius, J., Castelló, A., Cases-Langhoff, C., and Badimon, L. (2003). Neuron-derived orphan receptor-1 (NOR-1) modulates vascular smooth muscle cell proliferation. *Circ. Res.* **92**, 96–103.
- Marx, N., Schonbeck, V., Lazar, M. A., Libby, P., and Plutzky, J. (1998). Peroxisome proliferator-activated receptor  $\gamma$  activators inhibit gene expression and migration in human vascular smooth muscle cells. *Circ. Res.* **83**, 1097–1103.
- Rekhter, M. D. (1999). Collagen synthesis in atherosclerosis: Too much and not enough. *Cardiovasc. Res.* **41**, 376–384.
- Ridker, P. M. (2003). Clinical application of C-reactive protein for cardiovascular disease detection and prevention. *Circulation* **107**, 363–369.
- Rodríguez, C., Martínez-González, J., Sánchez-Gómez, S., and Badimon, L. (2001). LDL downregulates CYP51 in porcine vascular endothelial cells and in the arterial wall through a sterol regulatory element binding protein-2-dependent mechanism. *Circ. Res.* **88**, 268–274.
- Ross, R. (1995). Cell biology and atherosclerosis. *Annu. Rev. Physiol.* **57**, 791–804.
- Ross, R. (1999). Atherosclerosis: An inflammatory disease. *N. Engl. J. Med.* **340**, 115–126.
- Ross, R., and Harker, L. (1976). Hyperlipidemia and atherosclerosis. *Science* **193**, 1094–1100.
- Schiffirin, E. L. (2002). Beyond blood pressure: The endothelium and atherosclerosis progression. *Am. J. Hypertension* **15**, S115–S122.
- Stary, H. C., Chandler, A. B., Glagov, S., *et al.* (1994). A definition of initial, fatty streak, and intermediate lesions of atherosclerosis: A report from the Committee on Vascular Lesions of the Council on Arteriosclerosis, American Heart Association. *Circulation* **89**, 2462–2478.
- Wessby, B. (2003). Dietary fat, fatty acid composition in plasma, and the metabolic syndrome. *Curr. Opin. Lipidol.* **14**, 15–19.
- Williams, K. J. (2001). Arterial wall chondroitin sulfate proteoglycans: Diverse molecules with distinct roles in lipoprotein retention and atherogenesis. *Curr. Opin. Lipidol.* **12**, 477–487.



# Atherosclerosis

Jesus A. Araujo and Aldons J. Lusis  
*University of California, Los Angeles, California, United States*

## Glossary

**atherosclerotic plaque** Deposit of lipids and cellular and fibrous elements in the artery wall.

**foam cells** Lipid-engorged cells, mostly macrophages.

**infarction** Tissue necrosis caused by hypoxia or anoxia resulting from insufficient or absent blood perfusion.

**ischemia** Insufficient blood flow to satisfy the tissue metabolic needs.

**vascular thrombosis** Thrombus formation frequently occurring at the level of an unstable atherosclerotic plaque.

**A**therosclerosis is a chronic inflammatory disease characterized by the progressive accumulation of lipids and cellular and fibrous elements in large and medium-sized elastic and muscular arteries.

## INTRODUCTION

Atherosclerosis is a vascular syndrome that is the primary cause of ischemic heart disease and stroke. In Western societies, it is the underlying cause of approximately 50% of total mortality. For many decades, it was considered a degenerative disease with no major potential for intervention, but it is better understood as a very dynamic, chronic inflammatory condition that is converted to an acute clinical event by the induction of plaque rupture leading to thrombosis, with a great potential for intervention. Epidemiological studies have revealed several important environmental and genetic risk factors. Work with genetically modified mice has made it feasible to experimentally examine mechanistic hypotheses, resulting in a clearer understanding of the cellular and molecular processes involved. Important tools for the management of the various clinical entities have been developed for diagnosis, risk stratification, and

therapeutics. Risk factors can be classified as those with a strong genetic component and those that are mostly environmental (Table I). However, approximately 50% of the casualties related to ischemic vascular disease are not explained by known pathogenic factors, posing a challenge to achieve a more complete understanding.

## PATHOGENESIS

### Progression of the Atherosclerotic Lesions

Stary described the microscopical evolution of atherosclerosis. The earliest stage (Stary I lesion), which is not apparent macroscopically, consists of subendothelial accumulations of lipid-loaded macrophages, called foam cells, and is found in 45% of infants up to 8 months of age. By puberty, it is common to see the classical “fatty streak” (Stary II lesions), visualized macroscopically with Sudan IV stain as a flat or slightly raised streak, consisting of lipid-loaded macrophages and smooth muscle cells containing lipid droplets and minimal, scattered, extracellular lipid. More advanced lesions include those with multiple extracellular lipid cores (Stary III lesion) that appear as a raised fatty streak or an atheroma (Stary IV lesion), characterized by a single confluent extracellular lipid core. By the third decade of life, fibrous lesions (Stary V lesions) appear when Stary III or Stary IV lesions are surrounded and/or capped by smooth muscle cells and collagen. These plaques can harbor calcification and areas of microthrombi in different stages of fibrotic organization. Different components of an advanced atherosclerotic lesion are shown in Fig. 1. The majority of acute coronary ischemic events occur as a result of plaque fissuring/rupture, with exposure of the subendothelial matrix and the subsequent development of an occluding thrombus on the surface of the plaque.

Table 1 Genetic and Environmental Factors Associated with Atherosclerosis and CAD<sup>a</sup>

Factor	Evidence
<b>Strong genetic component</b>	
Elevated levels of LDL/VLDL	Epidemiologic studies, studies of genetic disorders, and animal models; Small, dense LDL particles are particularly strongly associated with CAD. Clinical trials have shown benefits of cholesterol reduction.
Reduced levels of HDL	Epidemiological studies, studies of genetic diseases, and studies of animal models
Elevated levels of Lp(a)	Epidemiological studies: Animal studies have yielded both positive and negative results.
Elevated blood pressure	Epidemiological studies: Clinical trials have demonstrated benefits of blood pressure reduction, with a particularly strong effect on stroke.
Elevated levels of homocysteine	Epidemiological studies: Animal studies have been inconclusive.
Family history	Independent factor in nearly all studies.
Diabetes and obesity	Epidemiological and animal studies
Elevated levels of hemostatic factors	Associations have been observed with elevated levels of fibrinogen, plasminogen activator inhibitor type 1, and platelet reactivity.
Depression and other behavioral traits	Associations have been observed in several studies. The results are complicated by associated traits.
Gender (male)	Men younger than 60 years of age develop CAD at more than twice the rate as women. Recent trials has raised concerns about the benefits of postmenopausal estrogen replacement.
Systemic inflammation	Elevated levels of inflammatory molecules, such as sPLA <sub>2</sub> , or inflammation markers, such as CRP, are associated with CAD.
The metabolic syndrome	This cluster of metabolic disturbances, with insulin resistance as a central feature, is strongly linked to CAD.
<b>Environmental factors</b>	
High-fat diet	Population migration studies have revealed very strong associations with lifestyle, and diet appears to be the most significant factor. High-fat, high-cholesterol diets are usually required for development of atherosclerosis in experimental animals.
Smoking	Epidemiological studies: It has been estimated that approximately 30% of CAD deaths are due to smoking. Clinical trials have demonstrated benefits of cessation of smoking.
Low antioxidant levels	Antioxidants protect against atherosclerosis in experimental animals. However, clinical trials with antioxidants are mixed and are not conclusive.
Lack of exercise	A significant independent relationship has been observed.
Infectious agents	Animal studies: Epidemiological studies have shown associations with various agents, such as <i>Chlamydia pneumoniae</i> . Clinical trials have been inconclusive.

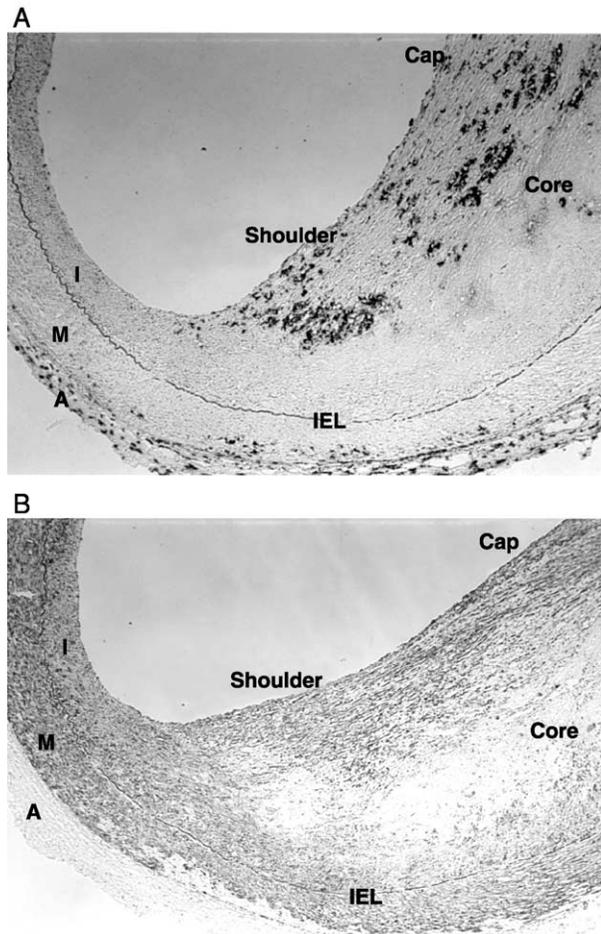
<sup>a</sup>Abbreviations used: CAD, coronary artery disease; HDL, high-density lipoprotein; LDL, low-density lipoprotein; VLDL, very low-density lipoprotein.

### Initial Events: Foam Cell Formation and Generation of the Fatty Streak

The events of atherosclerosis have been greatly clarified by studies in animal models, including rabbits, pigs, nonhuman primates, and rodents. Mice deficient in apolipoprotein E or the low-density lipoprotein (LDL) receptor develop advanced lesions and are the models most used in genetic and physiological studies. [Figure 2](#) depicts a model of the main events that occur in atherosclerotic plaque formation. The first observable change in the artery wall following the feeding of a high-fat, high-cholesterol diet is the accumulation of lipoprotein particles and their aggregates in the intima at sites of lesion predilection. LDL diffuses passively through endothelial cell (EC) junctions, and it is

retained in the vessel wall, which seems to involve interactions between the LDL constituent apolipoprotein B (apoB) and matrix proteoglycans. In addition to LDL, other apoB-containing lipoproteins, namely lipoprotein(a) and remnants, can accumulate in the intima and promote atherosclerosis.

The buildup of LDL in the subendothelial space is determined by the rate of retention as well as the rate of influx. Trapped LDL is modified by oxidation, lipolysis, proteolysis, and aggregation. Lipid oxidation as a result of exposure to oxidative waste of vascular cells is the most significant modification. Such modifications initially give rise to “minimally oxidized” LDL species, which can stimulate the overlying ECs to produce a number of proinflammatory molecules, including adhesion molecules and growth



**Figure 1** Human coronary artery lesion. (A) All three main layers of the arterial wall are seen: intima (I), media (M), and adventitia (A). The internal elastica lamina (IEL) clearly separates the intima from the media. The fibrous plaque core is rich in extracellular lipids and foam cells. Staining with mAb to macrophages (HAM-56) allows visualization as dark infiltrates in the atherosclerotic plaque, prominent at the shoulder of the lesion. Notice the remarkable difference in the intima thickness between the neighbor normal arterial wall and the affected wall by the atherosclerotic plaque. (B) Staining with mAb to muscle actin (HHF35) allows visualization of the smooth muscle cells on top of the lesion core.

factors such as macrophage colony-stimulating factor (M-CSF). The retention of LDL and the biological effects of the trapped LDL depend on the extent that they are oxidatively modified, which is related to the balance of pro-oxidant and antioxidant elements. Two important factors in the antioxidant repertoire are heme oxygenase-1 (HO-1) expression and high-density lipoprotein (HDL). HO-1 seems to be both an anti-inflammatory and an antioxidant gene. HDL is strongly protective against atherosclerosis, presumably because of its capacity to remove cholesterol from the vascular tissue and also to inhibit lipoprotein

oxidation. The antioxidant properties of HDL derive in part from serum paraoxonase, an esterase carried on HDL that can degrade certain biologically active oxidized phospholipids.

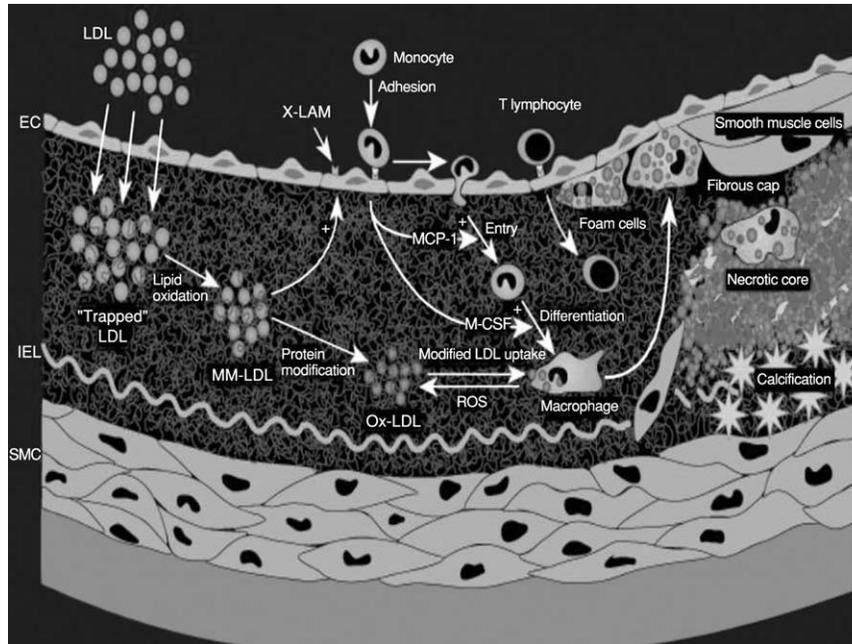
Within days or weeks of diet initiation, monocytes can be observed adhering to the surface of the endothelium. The first step in adhesion, the “rolling” of leukocytes along the endothelial surface, is mediated by selectins that bind to carbohydrate ligands on leukocytes. The firm adhesion of monocytes to endothelium can be mediated by the integrin VLA-4 on these cells, which interacts with both VCAM-1 on the endothelium and the CS-1 splice variant of fibronectin. Monocyte chemoattractant protein-1 and its receptor, CCR2, are also important in monocyte recruitment. Once present in the subendothelial space, monocytes are stimulated to proliferate and differentiate into macrophages, a process partly mediated by M-CSF.

LDL is extensively modified to generate the highly oxidized LDL by the action of reactive oxygen species and by several enzymes, such as lipoxygenases, myeloperoxidase, sphingomyelinase, and secretory phospholipase produced by the vascular cells. Oxidized LDL is then rapidly taken up by macrophages through their scavenger receptors, especially SR-A and CD36, to form foam cells. Macrophage scavenger receptor expression is regulated by peroxisome proliferator-activated receptor- $\gamma$ , an important transcription factor whose ligands include oxidized fatty acids, and by cytokines such as tumor necrosis factor- $\alpha$ , interferon- $\gamma$  (IFN- $\gamma$ ), and M-CSF. Oxidized LDL can also inhibit the production of nitric oxide, a chemical mediator with multiple antiatherogenic properties, including vasorelaxation. When sufficient foam cells accumulate in addition to extracellular lipid, a fatty streak lesion can be seen macroscopically by staining with Sudan IV.

### Advanced Lesions: Progression to Fibrous Plaques and Complicated Plaques

The increasing number of macrophages and T lymphocytes secrete cytokines and growth factors (e.g., platelet-derived growth factor) that are responsible for smooth muscle cell (SMC) migration, proliferation, and extracellular matrix production. In addition to the growing mass of extracellular lipid, this leads to the formation of fibrous plaques.

Various factors contribute to the development of advanced lesions. The engagement of CD40-CD40L between all major vascular cell types results in the production of inflammatory cytokines, matrix-degrading



**Figure 2** Atherosclerosis model. The artery wall architecture has three main segments: the intima, from the endothelial cell (EC) monolayer to internal elastic lamina membrane (IEL); the media, consisting of smooth muscle cells (SMC); and the adventitia, which is the outer layer. Low-density lipoproteins (LDL) or other atherogenic lipoproteins enter and accumulate in the subendothelial space. Trapped LDL particles become oxidized probably as a result of interactions with reactive oxygen species (ROS) produced by vascular cells. Such minimally modified LDLs (MM-LDL) induce endothelial cells to express adhesion molecules for monocytes (X-LAM), monocyte chemoattractant protein (MCP-1), and macrophage colony-stimulating factor (M-CSF), resulting in more recruitment of blood monocytes and subsequent differentiation into macrophages. With time, LDL particles become highly oxidized (Ox-LDL) and are then recognized by scavenger receptors on macrophages that endocytose them, giving rise to cholesterol-engorged foam cells. Such foam cells, the hallmark of fatty streaks, contribute to the development of advanced lesions by the production of various cytokines and growth factors that stimulate smooth muscle proliferation and fibrous cap formation. Calcification, hemorrhage, and thrombosis complicate plaques at advanced stages.

proteases, and adhesion molecules. Elevated homocysteine levels appear to injure ECs and to stimulate proliferation of vascular SMCs. Activation of the renin-angiotensin system also stimulates SMC growth and the production of extracellular matrix. A vicious cycle of inflammation ensues as inflammatory conditions promote the retention and modification of LDL and increasingly more modified LDL stimulates inflammation.

As the atherosclerotic plaque grows, active intimal calcification, development of new vessels from the media, and rupture of vessels with resulting intramural hemorrhage or thrombosis may occur. The progressive and insidious growth of plaque burden into the lumen of the vessel leads to narrowing of the arterial lumen. When this reaches a critical level, ischemia occurs, causing various symptoms depending on the affected tissue.

However, it is the unstable plaque that ruptures and exposes its subendothelial components with subsequent thrombus formation that has the most significance, at least in the coronary arteries, and is

responsible for the majority of myocardial infarctions. Plaque vulnerability is related to the fibrous cap thickness, which reflects the dynamics of matrix production and degradation. Vulnerable plaques generally have thin fibrous caps, increased numbers of inflammatory cells, and frequently rupture at the lesion edges where macrophages produce proteases that degrade extracellular matrix, including collagenases, gelatinases, and stromelysin. At the same time, matrix production by SMCs may be inhibited by IFN- $\gamma$  produced by T lymphocytes.

## CLINICAL MANIFESTATIONS

Atherosclerotic plaques may compromise the lumen of the vessels and obstruct blood flow to the respective tissue. Ischemia and/or infarction occur depending on the magnitude of the obstruction (partial, subtotal, or total), the duration of the obstruction, the time course (acute vs chronic), the collateral circulation, the degree of reversibility, and/or spontaneous reperfusion of

Table II Clinical Manifestations due to Atherosclerosis

Organ/system	Clinical entity
Heart	Coronary artery disease (stable and unstable angina and myocardial infarction) Congestive heart failure
Central nervous system	Cerebrovascular disease (transient ischemic attacks and strokes)
Peripheral vasculature	Peripheral vascular disease (intermittent claudication, vascular insufficiency, and gangrene)

the affected vascular territory. The most common clinical entities are shown in [Table II](#).

Atherosclerosis is a long and silent disease. By the time the patient presents with chest pain due to coronary artery disease or intermittent claudication due to arterial insufficiency, years and decades of lipid deposition and plaque progression have occurred. Atherosclerosis is a debilitating disease. Stroke and congestive heart failure due to ischemic cardiomyopathy incapacitate a large number of individuals, imposing an incredible burden on the patients, their families, and society in general. It is of extreme concern that the first manifestation is frequently sudden death, eliminating opportunities for intervention. Symptomatic patients may be fortunate in the sense that there is an opportunity for diagnosis, risk factor modification, and therapeutic intervention. It is clear that we cannot rely on the clinical manifestations for the diagnosis of these entities since it may be too late by then.

## ASSESSMENT

In animals, assessment has been performed mostly by postmortem evaluation of atherosclerotic lesions that originate at different levels in the arterial vasculature (aorta, carotid, coronary or femoral arteries, etc.) under the defined experimental variables employed. The evaluation can be quantitative or qualitative and either at the macroscopic or at the microscopic level.

In humans, the situation is much more complex given the necessity of achieving a phenotypic characterization *in vivo*. Postmortem correlations can also be established and are useful, but they provide only a narrow view of the very dynamic and multifactorial process in question. When present, the specific clinical manifestations in the corresponding clinical scenario help in deciding which diagnostic modality to use. It is even more difficult to approach the asymptomatic individual. The same strategies used with clinical atherosclerosis have been used here, but often these miss the earlier stages of the disease, when preventive actions are desirable. Diagnostic modalities can be classified as invasive and noninvasive

techniques. Invasive techniques, such as angiography (the gold standard) and intravascular ultrasound, offer the greatest degree of accuracy available, limited by cost and potential complications. Angioscopy allows direct visualization of the arterial wall, revealing plaque surface traits that are not representative of the internal heterogeneity of the plaque.

Noninvasive techniques can be classified as anatomic methods aimed at detecting the atherosclerotic plaque per se and physiologic methods aimed at detecting a physiologic abnormality created by impairment of blood flow in an arterial bed or impairment of vasodilation. Among the former, transesophageal echocardiography, electron-beam computed tomography, magnetic resonance angiography, and magnetic resonance imaging stand out. Among the latter, stress testing, coronary positron emission tomography, ultrasonic brachial vasodilatation, and the ankle–arm systolic blood pressure index techniques are commonly used. There is no one single and optimal method for phenotypic assessment, and the choice must be based on the specific indication.

## THERAPEUTICS

[Table III](#) lists various treatment strategies for atherosclerosis-related disorders, mainly coronary artery disease. These include drugs that improve symptoms or modify risk factors, invasive medical procedures, and surgical procedures. Some agents improve both symptoms and survival; for example, beta-blockers decrease cardiac work and myocardial oxygen demand by decreasing heart rate and cardiac contractility. Some agents improve symptoms without affecting mortality; thus, nitrates decrease cardiac work and oxygen demand by increasing venous capacitance, decreasing venous return, and decreasing cardiac volume preload. Also, some agents improve symptoms but may have harmful effects (e.g., short-acting calcium channel blockers).

The development of interventional cardiology in recent years has revolutionized the treatment of coronary disease. Angioplasty with or without stent

Table III Strategies for Treatment of Symptoms or Improvement of Life Expectancy for Atherosclerosis-Related Disorders

Therapy	Intervention	Mechanism
Lipid-modifying drugs	Statins	Decrease LDL cholesterol, and/or increase HDL cholesterol, and/or decrease triglycerides (varies with the specific drug)
	Gemfibrozil	
	Niacin	
Antiplatelet drugs	ASA	Decrease platelet reactivity
	Clopidogrel, ticlopidine	Decrease thrombus formation
	Glycoprotein IIB–IIIa inhibitors	
Antithrombotic drugs	Thrombolytic agents	Lysis of thrombus
	Unfractionated heparin	Decrease thrombus formation
	Low-molecular-weight heparin	
Antiangina drugs	Beta blockers	Improvement of myocardial oxygen offer–demand balance
	Nitrates	
	Calcium channel blockers	
Antihypertensive drugs	Diuretics	Blood pressure control
	ACE inhibitors	
	Beta blockers	
	Calcium channel blockers	
	Alpha blockers	
	Direct vasodilators	
Antioxidant drugs	Vitamin E	Increase antioxidant reserve
Interventional techniques	Angioplasty	Mechanical resolution of coronary and peripheral artery blockages
	Stents	
	Catheter-based endarterectomy	
Surgical techniques	Coronary artery bypass graft surgery	Surgical bypass or repair of coronary, carotid, or peripheral artery blockages
	Peripheral artery bypass surgery	
	Endarterectomy for carotid and peripheral arteries	
	Aortic aneurysm repair	

placement is the customary choice of one- and two-vessel coronary disease. It is particularly useful after medical therapy has been optimized without resolution of symptoms. With the better understanding of this disease and the development of better agents, surgical cases are decreasing in frequency.

However, cardiovascular disease is still the most common cause of death in the Western world. The known risk factors explain only about 50% of susceptibility to the disease. A better understanding of how to optimize therapy for the individual patient is required.

## GENETIC DISSECTION

Atherosclerosis is a multifactorial and complex disease. Risk factors can be classified as those with an important genetic component and those that are largely environmental (Table I). Common forms

of coronary artery disease result from the combination of genetic susceptibility and an unhealthy environment.

Rare mendelian forms, such as familial hypercholesterolemia and Tangier disease, have provided important insights into the disease. Studies of candidate genes associated with predisposing conditions, such as hyperlipidemia, low HDL levels, diabetes, hypertension, and pro-coagulant disorders, have revealed a number of genes with significant or suggestive association or linkage with traits relevant to atherosclerosis.

As a result of the genome project and large-scale sequencing, thousands of single-nucleotide polymorphisms are being identified and a catalogue of all common variations in humans will be generated during the next few years. It can be anticipated that profiles of genetic factors, perhaps in the form of risk haplotypes, can be generated and that individuals in the future may be screened for these and therapies

may be tailored based on the particular risks for each individual.

## NEW VENUES

Active and intense research is being conducted at all levels, from the basic aspects of genetics and molecular and cellular biology to the development of better diagnostic resources and risk-stratifying tools and the formulation of better therapeutic strategies. In the middle of the great genomic revolution, as our understanding of the genetics of the disease grows, genetic diagnosis will become increasingly important. It is hoped that this will allow genetic screening for diagnostic and risk-stratification purposes, which will enable better assignment of management resources and adequate and tailored pharmacological therapy.

## CONCLUSION

Atherosclerosis is a chronic inflammatory disease characterized by progressive accumulation of lipids and cellular and fibrous elements in large and medium-sized elastic and muscular arteries. Elevated cholesterol levels, especially in the LDL fraction, are a principal risk factor. Lipoprotein retention and oxidation in the vascular wall leads to endothelium activation, monocyte infiltration, foam cell formation, inflammation, and proliferation of smooth muscle cells. Subsequently, a necrotic core of lipid and cell debris develops and vulnerable plaques rupture, leading to thrombosis. Clinical manifestations result from ischemia and vascular thrombosis in the respective vessels, leading to myocardial infarction, stroke, vascular insufficiency, and gangrene of the extremities.

Various resources are available for diagnosis and risk stratification, but all are limited and far from ideal. Despite an array of therapeutic options tailored for the different clinical manifestations, better strategies are needed to stop or reverse the disease and to be able to adjust pharmacological agents to each individual. It is hoped that the genetic dissection of this syndrome will allow genetic diagnosis to support screening and tailored therapy.

## Acknowledgments

This work was supported by grants from the National Institutes of Health (HL30568, HL28481, and Training Grant T32 HL07895) and the Laubisch Fund, UCLA. We thank Dr. Tom Drake for providing micrographs of atherosclerotic lesions and Eric Lusic for artwork.

## See Also the Following Articles

Atherogenesis • Cardiovascular Disease in Diabetes

## Further Reading

- Fuster, V., Badimon, L., Badimon, J. J., and Chesebro, J. H. (1992). The pathogenesis of coronary artery disease and the acute coronary syndromes. *N. Engl. J. Med.* **326**, 242.
- Glass, C. K., and Witztum, J. L. (2001). Atherosclerosis: The road ahead. *Cell* **104**, 503–516.
- Libby, P. (2002). Atherosclerosis: The new view. *Sci. Am.* **286**, 46–55.
- Lusic, A. (2000). Atherosclerosis. *Nature* **407**, 233–241.
- Sary, H. C. (1989). Evolution and progression of atherosclerotic lesions in coronary arteries of children and young adults. *Arteriosclerosis* **99**(Suppl. 1), 19–32.
- Steinberg, D., Parthasarathy, S., Carew, T. E., Khoo, J. C., and Witztum, J. L. (1989). Beyond cholesterol: Modifications of low density lipoprotein that increase its atherogenicity. *N. Engl. J. Med.* **320**, 915–924.



# Atrial Natriuretic Factor and Family of Natriuretic Peptides

Kenneth K.-Y. Ma and Adolfo J. de Bold  
University of Ottawa, Ottawa, Ontario, Canada

## Glossary

**cis-acting elements** Short DNA consensus sequences located upstream (5'-flanking region) of the transcription start point. They can be either enhancers, which increase transcription, or silencers, which decrease transcription.

**diuresis** The increased excretion of urine, especially in greater than normal amounts.

**exon** Segment of a gene that corresponds to a sequence in the final processed RNA transcript of that gene.

**heart failure** Defective cardiac filling and/or impaired contraction and emptying, resulting in a heart's inability to pump a sufficient amount of blood to meet the needs of the body tissues or to be able to do so only with an elevated filling pressure.

**hypertrophy** An increase in the size of a tissue or an organ due to an increase in the size of its component cells. Contrast with hyperplasia.

**immediate-early gene (IEG)** Class of genes whose expression is low in quiescent cells, but whose transcription is activated within minutes of stimulation. Many IEGs encode *trans*-acting factors.

**intron** A noncoding sequence of DNA within a gene that is transcribed into nuclear RNA, but is subsequently removed by RNA splicing in the nucleus and rapidly degraded.

**natriuresis** The excretion of sodium in the urine, especially in greater than normal amounts.

**TATA-box** A consensus segment, 5'-TATAAA-3', that is 19–27 bp upstream from the start point of eukaryotic structural genes to which RNA polymerase II binds. Also known as a Goldberg-Hogness box.

**trans-acting factors** DNA-binding proteins that recognize *cis*-acting elements on the 5'-flanking region of the regulated gene.

(BNP), and C-type natriuretic peptide (CNP). These polypeptide hormones play central roles in maintaining fluid and electrolyte balance and circulatory homeostasis. In mammals, the bulk of ANF and BNP is produced and secreted by the cardiac muscle cell (cardiocyte), whereas CNP is produced mainly by the vascular endothelium and brain.

## PHYSIOLOGY, STRUCTURE, BIOSYNTHESIS, AND REGULATION OF ANF

### Physiology, Structure, and Biosynthesis of ANF

#### **Physiological Actions of ANF**

In 1981 it was demonstrated that injections of heart atrial extracts into rats gave rise to pronounced diuresis, natriuresis, and lowered blood pressure. Within a short time, a peptide with the same biological properties as the crude atrial extracts was isolated and its amino acid sequence was identified. This peptide, now known as atrial natriuretic factor (also known as atrial natriuretic peptide), firmly established the heart as an endocrine organ and heralded a new era of research on the control and maintenance of blood pressure, blood volume, and vascular tone.

ANF is a potent natriuretic and diuretic agent, owing these properties to both its renal hemodynamic actions and its direct tubular actions. By simultaneous dilation of afferent arterioles and constriction of efferent arterioles, ANF increases the glomerular filtration rate and filtration fraction. ANF will directly inhibit water reabsorption by the renal cortical collecting duct and contributes to inhibition of Na<sup>+</sup> reabsorption by the renal inner medullary collecting duct.

ANF also has profound effects on the cardiovascular system. Acute administration of ANF causes a fall in arterial pressure due to a reduction in cardiac output mediated by a decreased preload and vascular resistance. The decreased preload is believed to be a consequence of venodilation and reduction of the intravascular volume. Furthermore, ANF also reduces

The atrial natriuretic factor family of natriuretic peptides (NPs) consists of three members: atrial natriuretic factor (ANF), brain natriuretic peptide or B-type natriuretic peptide

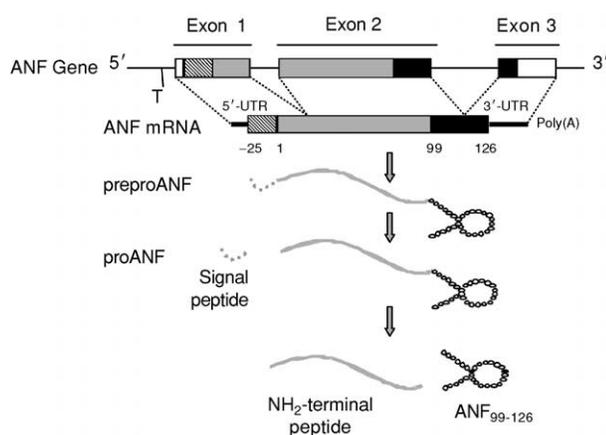
sympathetic tone by inhibiting arterial baroreceptor response and suppression of catecholamine release from autonomic nerve endings.

The renin–angiotensin–aldosterone system (RAAS) is also a target for ANF. ANF will directly reduce renin secretion, which has the cascade effect of lowering circulating levels of angiotensin II (Ang II), a potent vasoconstrictor and stimulator of aldosterone. ANF can also directly inhibit aldosterone synthesis and secretion from the glomerulosa cells of the adrenal cortex.

The central actions of ANF include inhibition of secretion of vasopressin and of salt and water intake. ANF also possesses important anti-mitotic actions on vascular endothelial cells, smooth muscle cells, and cardiac fibroblasts.

### Structure and Biosynthesis of ANF

The sequencing of ANF permitted the generation of oligonucleotide probes that were used to isolate cDNA clones for ANF-coding sequences. The mRNA is approximately 1 kb long and encodes a preproANF that contains between 149 and 153 amino acids depending on the species (Fig. 1). The human amino acid sequence shares strong homology with peptides from other species including rat, mouse, and pig. In the human atria, ANF is stored mainly within organelles referred to as specific atrial granules as a prohormone called proANF<sub>1–126</sub>. Subsequent processing releases into the bloodstream the biologically active hormone, ANF<sub>99–126</sub>. Normal human plasma concentrations of ANF are approximately



**Figure 1** Structure of the gene and biosynthetic pathway of human ANF. Solid black sections are those that code for the mature ANF<sub>99–126</sub>. Gray sections code for NH<sub>2</sub>-terminal fragments and the striped section codes for the signal peptide. Numbers indicate the amino acid position relative to the sequence encoded in proANF. T indicates the location of the TATAA-box. UTR, untranslated region.

6 fmol/ml. The biological half-life of ANF<sub>99–126</sub> is very short (0.5–2 min).

All members of the ANF family of natriuretic peptides [ANF, B-type natriuretic peptide (BNP), and C-type natriuretic peptide (CNP)] share a common central ring structure formed by a disulfide bridge (positions 105 and 121 for ANF). Disruption of this 17-member ring leads to a loss of biological activity (Fig. 2).

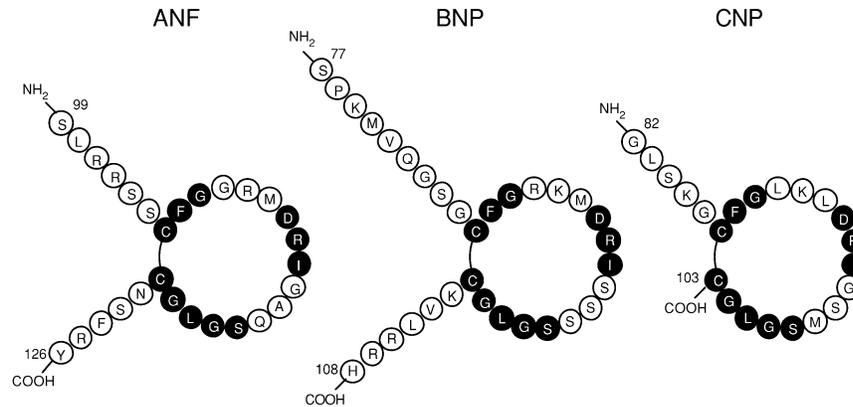
### ANF Gene Regulation

The ANF gene is mapped to chromosome 1 in humans, chromosome 4 in mouse, and chromosome 5 in rats. Structurally, the ANF gene is similar in all mammals and consists of three exons and two introns (Fig. 1).

### Tissue-Specific and Developmental Regulation of ANF Gene Expression

Atrial expression of ANF rises continuously during fetal and postnatal development to comparatively high levels (1–3% of total atrial mRNA), whereas ventricular expression of ANF, although quite high during fetal development, falls quickly after birth (1% of atrial levels). Extracardiac (aortic arch, hypothalamus, pituitary, and kidney) sites of synthesis have also been documented, but the major site of ANF production remains the atria. In adulthood, ANF expression by the heart can be markedly increased by mechanical stretch of the cardiocyte and/or by neuroendocrine hormones.

*Cis*-acting elements responsible for tissue- or stage-specific expression of ANF have been identified on the 5'-flanking promoter of the ANF gene. The characteristic TATAA-box motif for binding the ubiquitous transcription initiation factor IID (TFIID) transcription factor is 30 bp upstream of the transcription initiation site (CAP site). Within the –700 bp ANF promoter, certain regions that confer chamber and temporal-specific activity have been identified. A CARG motif at –405 bp plays a critical role in the ventricular and stage-specific activity of ANF expression. Furthermore, located between –380 and –300 is cardiac regulatory element (CARE). CARE and the *trans*-acting factor Catf-1 are active in the embryonic atria and ventricle but not in the postnatal ventricular cardiocyte. Finally, there are a few elements that are important in maintaining basal expression of ANF. These are located in the proximal promoter from –135 to –1 and include GATA-, Nkx2.5 response element (NKE), and T-box 5 (Tbx5)-binding elements.



**Figure 2** The biologically active forms of the natriuretic peptides. Amino acids that are shared by all members of the natriuretic peptide family are identified by filled circles. A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine.

### Mechanical and Neuroendocrine Regulation of ANF Gene Expression

#### Acute Mechanical Myocardial Stretch

ANF is continuously released from the atria of the heart, but appropriate mechanical stretch can increase the rate of ANF secretion. In acute experiments with isolated atrial preparations, increases in atrial wall stretch are met with corresponding increases in ANF secretion with no change in ANF gene expression. Despite continued mechanical stimulation, the acute increase in ANF release eventually decays to baseline values within minutes. The precise mechanism underlying mechanotransduction of stretch-induced ANF release remains to be elucidated.

#### Neuroendocrine Hormones or Agents

Numerous neuroendocrine agents can also increase ANF gene expression and release.  $\alpha_1$ -Adrenergic agonists (e.g., phenylephrine) and vasoactive peptides, such as endothelin-1 (ET-1) and Ang II, are potent stimulators of ANF gene expression and release. The vasoconstrictor ET-1 is synthesized and released by the endothelial and mesothelial cells. It is believed that ET-1 can act in a paracrine fashion *in vivo* to modulate ANF gene expression and release. Ang II, an effector molecule of RAAS, is a potent vasoconstrictor as well as a growth factor for cardiocytes.

Ang II, ET-1, and catecholamines bind to the heterotrimeric G protein heptahelical transmembrane family of receptors. In general, activation of the receptor leads to the recruitment of secondary effector molecules that activate cytosolic and/or nuclear substrates. In particular, Ang II, ET-1, and phenylephrine will rapidly stimulate protein kinase C and the mitogen-activated protein kinase, p44/42 (ERK), which in

turn will activate *trans*-acting factors responsible for the increase in ANF expression. These include the immediate-early genes (IEGs), such as the *c-fos* and *c-jun* heterodimer, that can bind to activator protein-1 (AP-1) consensus elements on the 5'-flanking region of the ANF promoter.

Hormonal ligands for nuclear receptors, such as glucocorticoids, have also been shown to up-regulate ANF gene expression and release. A cytokine produced by the cardiocyte, called cardiotrophin I, has been shown to up-regulate ANF gene expression.

#### Chronic Hemodynamic Overload and Changes in ANF Expression

Under chronic hemodynamic overload, mature terminally differentiated cardiocytes will respond with cellular hypertrophy to normalize wall stress. Importantly, ANF gene expression and release are markedly increased not only from the atrial cardiocytes, but also from the ventricular cardiocytes by the hypertrophic stimuli. This can translate into a 100-fold increase in ANF plasma concentrations in certain pathophysiological conditions such as chronic congestive heart failure. Hypertrophic stimuli, such as prolonged myocardial stretch or humoral growth factors, will activate the IEGs, such as *c-myc*, *c-fos*, and/or *c-jun*. This is followed by activation of the characteristic embryonic gene program seen during cardiac development. Molecularly, this gene program accounts for the re-expression of ANF, BNP,  $\beta$ -myosin heavy chain, and skeletal  $\alpha$ -actin in the hypertrophic adult ventricle.

The precise mechanisms responsible for the chronically increased ventricular ANF gene expression and secretion have not been fully elucidated. *In vivo*

experiments have demonstrated that there may be two components responsible for these changes in natriuretic peptide expression: one that is dependent on hemodynamic load (i.e., mechanical stretch) and one that is load independent (i.e., direct effects of humoral growth factors, such as ET-1 and Ang II, on cardiocytes).

## OTHER MEMBERS OF THE ANF FAMILY OF NATRIURETIC PEPTIDES

### Physiology, Structure, Biosynthesis, and Regulation of BNP

#### Physiological Actions of BNP

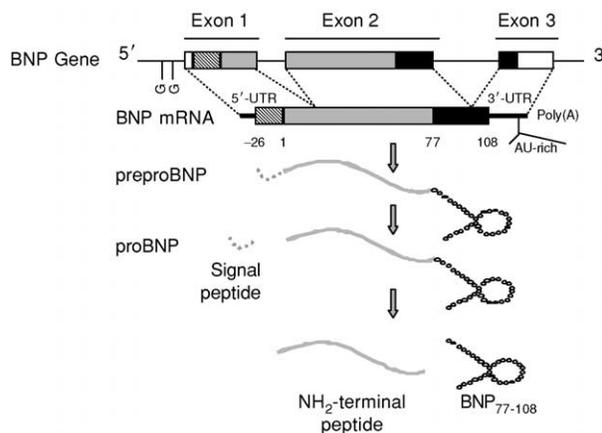
Brain natriuretic peptide was originally isolated from porcine brain. However, it was later discovered that the highest concentrations were found in the heart. BNP, like ANF, is a potent natriuretic, diuretic, vasorelaxant, anti-mitotic factor as well as an antagonist of the renin-angiotensin-aldosterone axis. Of interest, however, BNP may have functions not associated with ANF. BNP has been shown to have anti-fibrotic properties. This effect may be particularly important in cardiovascular diseases in which cardiac fibrosis contributes to the progression to heart failure.

#### Structure and Biosynthesis of BNP

BNP is synthesized as a 121- to 134-amino-acid preprohormone (Fig. 3). In humans, subsequent processing releases a mature biologically active 32-amino-acid carboxy-terminal fragment. Unlike ANF or CNP, the amino acid sequence of BNP is not as highly conserved and may differ by as much as 50% between species. The plasma half-life of BNP is approximately 22 min, which is approximately six times longer than that of ANF. The normal circulating BNP level in humans is approximately 0.9 fmol/ml.

#### Regulation of BNP Gene Expression

The BNP gene is mapped to chromosome 1 in human and to chromosome 4 in mouse. Structurally the BNP gene is similar to the ANF gene in that it is composed of three exons and two introns. However, there are important differences between the regulation of BNP and ANF at the transcriptional and posttranscriptional levels. Hemodynamic overload increases both atrial and ventricular expression of BNP (and ANF) dramatically. In pathological states such as chronic congestive heart failure, the plasma BNP level can increase 1000-fold.



**Figure 3** Structure of the gene and biosynthetic pathway of human BNP. Solid black sections are those that code for the mature BNP<sub>77-108</sub>. Gray section codes for the NH<sub>2</sub>-terminal fragment and the striped section codes for the signal peptide. Numbers describe the amino acid position relative to the sequence encoded in proBNP. G's indicate the location of the GATAAA sequences (TATAAA-box). UTR, untranslated region.

The 5'-flanking region of the BNP gene resembles more of an erythroid- than of a muscle-specific promoter. *Cis*-acting elements within 114 bp upstream of the CAP site appear to be responsible for conferring not only tissue-specific expression but also hemodynamic stress responsiveness. This region of the promoter contains conserved AP-1 and GATA recognition sites. GATA transcription factors are believed to be important in both basal and stimulated expression of BNP.

In response to certain hypertrophic stimuli, BNP mRNA is rapidly up-regulated in a protein synthesis-independent manner with quick turnover. These features are characteristic of an IEG. In support of this, the BNP 3'-untranslated region contains many AU-rich elements that may confer instability to the BNP transcript (mRNA  $t_{1/2}$  ~60 min) not found in ANF mRNA, which is quite stable. Interestingly, BNP mRNA stability can be enhanced after treatment with phenylephrine or ET-1. In addition to neurohumors such as ET-1 and phenylephrine, BNP gene expression and secretion can also be up-regulated by pro-inflammatory cytokines, such as interleukin-1 $\beta$  and tumor necrosis factor  $\alpha$  (TNF $\alpha$ ).

In summary, the significant increase in BNP expression during overload may be due a combination of specific transcriptional and posttranscriptional mechanisms.

## Physiology, Structure, Biosynthesis, and Regulation of CNP

### Physiological Actions of CNP

CNP is found in the cerebellum, hypothalamus, anterior pituitary, kidney, and the vascular endothelial cells but not in the heart. The central actions of CNP include antagonizing angiotensin II-mediated increases in vasopressin, which in turn decreases salt and water intake.

The natriuretic activity of CNP is only approximately 1% of that of ANF. CNP has potent anti-growth properties on vascular smooth muscle cells, suggesting that CNP may be an important regulator of blood vessels.

### Structure and Biosynthesis of CNP

CNP is the most conserved of all natriuretic peptides, with 90% homology observed among human, mouse, pig, and rat. Produced from a 126-amino-acid preprohormone, the bioactive hormone, unlike ANF or BNP, is only 22 amino acids long, because CNP has no carboxy-terminal tail (Fig. 2).

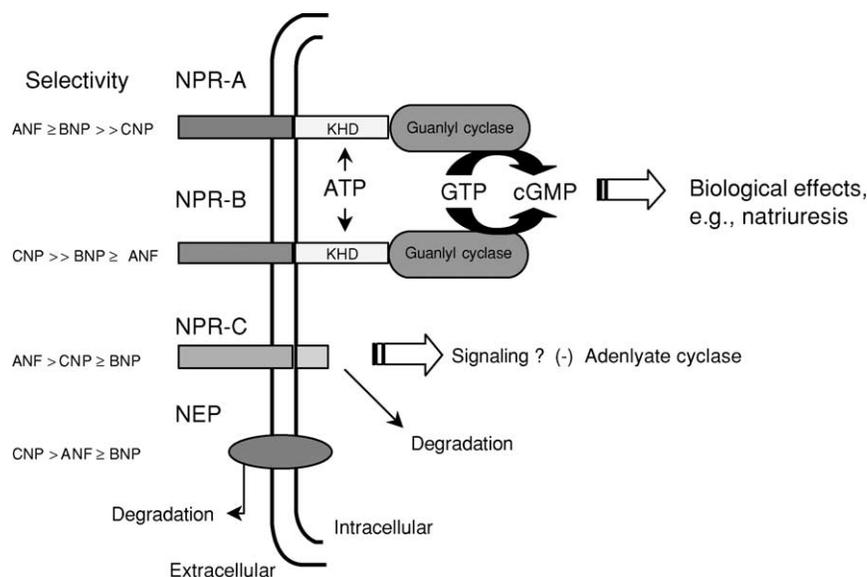
### Regulation of CNP Gene Expression

The gene for CNP is located on chromosome 2 in humans, in contrast to the ANF and BNP genes, which are located on chromosome 1 in humans. The

mouse CNP gene consists of least two exons and one intron. The 5'-flanking region of CNP contains numerous *cis*-acting regulatory elements including dinucleotide CA repeats, cyclic AMP-response element-like, nuclear factor  $\kappa$ B, and shear stress recognition sites. Cytokines, such as transforming growth factor- $\beta$  and TNF $\alpha$ , can up-regulate CNP.

## NATRIURETIC PEPTIDE RECEPTORS

Many of the biological actions of the natriuretic peptides are mediated through association with specific high-affinity receptors on the surface of target cells and the generation of cyclic guanosine monophosphate (cGMP) (Fig. 4). There are three natriuretic peptide receptors (NPR-A, NPR-B, and NPR-C). NPR-A and NPR-B are linked to guanylate cyclase and, on activation of the receptor, cGMP is formed. cGMP targets may include cGMP-dependent protein kinases and cGMP-gated ion channels. NPR-A binds both ANF and BNP, with preference for ANF. NPR-B binds CNP with far less preference for either ANF or BNP. NPR-C is the clearance receptor and binds CNP with slightly greater affinity than ANF or BNP. Circulating natriuretic peptides are also inactivated by neutral endopeptidases present within renal tubular and vascular cells.



**Figure 4** The natriuretic peptide receptors. Binding of ANF and BNP to NPR-A and CNP to NPR-B stimulates intrinsic guanylyl cyclase activity of the receptor. The transduction of ligand binding to cGMP generation is ATP dependent and requires participation from the KHD portion of the receptor. NPR-A, natriuretic peptide receptor-A; NPR-B, natriuretic peptide receptor-B; NPR-C, natriuretic peptide receptor-C; KHD, kinase-like homology domain; NEP, neutral endopeptidase 24.11.

## NATRIURETIC PEPTIDES AND PATHOPHYSIOLOGY

A hallmark of chronic congestive heart failure (CHF) is enhanced atrial and ventricular synthesis and release of ANF and BNP secondary to increased hemodynamic load and humoral stimulation. The severity of heart failure correlates well with plasma levels of ANF and BNP such that these peptides are reliable diagnostic and prognostic markers. In severe cases, the plasma level of BNP may surpass that of ANF. It is very likely that the increases in ANF and BNP serve to limit or delay the progression of CHF, in part by its renal, vascular, and endocrine actions.

## SUMMARY

The atrial natriuretic factor family of natriuretic peptides plays an important role in regulating blood pressure, extracellular fluid homeostasis, and cardiovascular growth under normal physiological conditions and in diseased states. Natriuretic peptides defend against excess salt and water retention, promote vascular relaxation, inhibit RAAS, inhibit sympathetic outflow, and have potent anti-mitotic properties. For many pathophysiological cardiovascular conditions, ANF and particularly BNP have become important diagnostic and prognostic tools. At present, many clinical treatment strategies for

heart failure are based on the application of the knowledge base developed about the natriuretic peptides. These include systemic administration of natriuretic peptides and pharmacological approaches, such as the dual angiotensin-converting enzyme and neutral endopeptidase inhibitors. The beneficial outcomes of these agents only highlight the importance of further study of the endocrine heart and the ANF family of natriuretic peptides.

## See Also the Following Articles

Aldosterone in Congestive Heart Failure • Natriuretic Peptides • Renin

## Further Reading

- de Bold, A. J., and Bruneau, B. G. (2000). Natriuretic peptides. *In* "Handbook of Physiology, Section 7, The Endocrine System, Volume III" (J. C. S. Fray, ed.), pp. 377–409. Oxford University Press, London.
- de Bold, A. J., Bruneau, B. G., and Kuroski de Bold, M. L. (1996). Mechanical and neuroendocrine regulation of the endocrine heart. *Cardiovasc. Res.* **31**, 7–18.
- Durocher, D., Grépin, C., and Nemer, M. (1998). Regulation of gene expression in the endocrine heart. *Rec. Prog. Horm. Res.* **53**, 7–23.
- Gardner, D. G., Wu, J., and Kovacic-Milivojevic, B. (1998). Cellular and molecular aspects of the A-type natriuretic peptide. *In* "Natriuretic Peptides in Health and Disease" (W. K. Samson and E. R. Levin, eds.), pp. 71–84. Humana Press, Totowa, NJ.



# Autoimmune Polyglandular Syndrome

Andrew Muir

Medical College of Georgia, Augusta, Georgia, United States

## Glossary

**adrenocortical failure** Insufficient production of steroid hormones from the cortex of the adrenal gland, including clinically important deficiencies of cortisol and aldosterone.

**autoimmune polyendocrinopathy–candidiasis–ectodermal dystrophy (APECED)** A commonly used alternate name for type 1 autoimmune polyglandular syndrome.

**autoimmune thyroiditis** Inflammation of the thyroid gland caused by abnormal self-reactive immune responses that can cause either low thyroid function (Hashimoto's thyroiditis) or high thyroid function (Graves' disease).

**autoimmunity** Abnormal responsiveness of cells of the immune system against the body's own (self) organs.

**hypoparathyroidism** Insufficient parathyroid hormone production, resulting in serum calcium levels below normal.

**mucocutaneous candidiasis** Infection of mucosal surfaces or skin by the yeast *Candida albicans*.

**nonobese diabetic (NOD) mouse** An inbred strain of mice that spontaneously develop immune-mediated destruction of pancreatic beta cells, causing diabetes.

Type 1 and type 2 autoimmune polyglandular syndromes (APS-1 and APS-2) include clusters of immune-mediated diseases that overlap in their clinical presentations but appear to have distinct underlying pathogeneses. The components of APS-1 are mucocutaneous candidiasis, hypoparathyroidism, and adrenocortical failure; the diagnosis requires that two of these three conditions be present. APS-2 is defined by the presence of adrenocortical failure and either autoimmune thyroiditis or type 1 diabetes mellitus. Individuals with either form of APS frequently develop autoimmune disease in target organs that are not included in the syndrome definitions (e.g., gonadal failure, atrophic gastritis, vitiligo, alopecia), and some individuals with autoimmune polyglandular failure do not fulfill the strict criteria for either form of APS but have close relatives who do.

## INTRODUCTION

The explanation for the immune system's faulty targeting of specific groups of organs in patients with either type 1 or type 2 autoimmune polyglandular syndrome (APS-1 or APS-2, respectively) has been elusive, but genetic, immunological, and environmental influences are believed to be important. A normal immune system eliminates foreign antigens from its host while remaining tolerant to self-antigens. Multiple mechanisms exist to establish and maintain tolerance in lymphocytes. Through an interaction with stromal bone marrow cells, self-reactive B-lymphocyte precursors usually die. Immature B lymphocytes that have not completed B-cell receptor (BCR) gene rearrangement may "edit" their light chain gene in an attempt to produce a specificity that will not be subject to negative selection. B lymphocytes do not survive more than a few days in the circulation unless their BCR is bound in a manner that activates the cell.

Intrathymic tolerance is the process whereby T-lymphocyte precursors from the bone marrow are first positively selected if their antigen-specific T-cell receptor (TCR) is likely to be recognized by the host's immune system and, therefore, useful for fending off foreign antigen. Cells that are not positively selected will die. Interaction between young T lymphocytes and epithelial cells in the thymic cortex mediate this process. The T lymphocytes that survive positive selection are then subjected to negative selection, a process whereby bone marrow-derived antigen-presenting cells (and, to a lesser extent, epithelial cells) trigger apoptosis (programmed death) of cells whose TCR specificities bind self-antigens too avidly. Once they are released from the thymus, naive T lymphocytes must receive ongoing "survival signals" as they circulate through blood, lymphoid organs, and lymphatics. Those cells that bind antigen to their TCR may either be activated, become anergic (unresponsive), or die (activation-induced cell death). Control of this fate is dependent on multiple factors, including the avidity of the TCR for the human

leukocyte antigen (HLA) molecular complex containing the antigen on the surface of antigen-presenting cells, accompanying signals generated by accessory molecule interactions, the preexisting activation state of the lymphocyte, and the presence of soluble mediators of immunity (e.g., cytokines).

Immune function experiments and genome mapping have delineated both similarities and differences between people with isolated target organ destruction and those with polyglandular autoimmune disease. In addition, important distinctions between APS-1 and APS-2 have been demonstrated. Therefore, the two syndromes are addressed as separate entities in this article.

## TYPE I AUTOIMMUNE POLYGLANDULAR SYNDROME

### Epidemiology

APS-1 is a rare condition, even among the three ethnic groups in which it is seen most frequently. Prevalence estimates place the rate among Iranian Jews at 1 in 9000, among Sardinians at 1 in 14,400, and among Finns at 1 in 25,000. Even a Norwegian prevalence estimate of 1 in 80,000 is still probably below the rate seen in most of North America and Western Europe. A slight female predominance among APS-1 patients has been reported by some groups, but this bias was not obvious in a group of 68 Finnish patients or in 23 affected Iranian Jews. Immune system dysfunction typically becomes clinically apparent at an early age, often during the first 5 years of life. However, initial signs may be delayed until adulthood.

Mucocutaneous candidiasis is uncommon among Iranian Jews, but in most populations, infections of the nails, skin, and gastrointestinal tract affect 73 to 100% of patients. Indeed, it is the most frequent first sign of APS-1. Hypoparathyroidism, diagnosed in 76 to 96% of patients, usually appears during the first decade of life, and the diagnosis of adrenocortical insufficiency, seen in 72 to 100% of patients, is made an average of approximately 5 years later. About one-third of patients worldwide develop all three of the defining diseases, although variations of the degrees of severity exist among different ethnic groups.

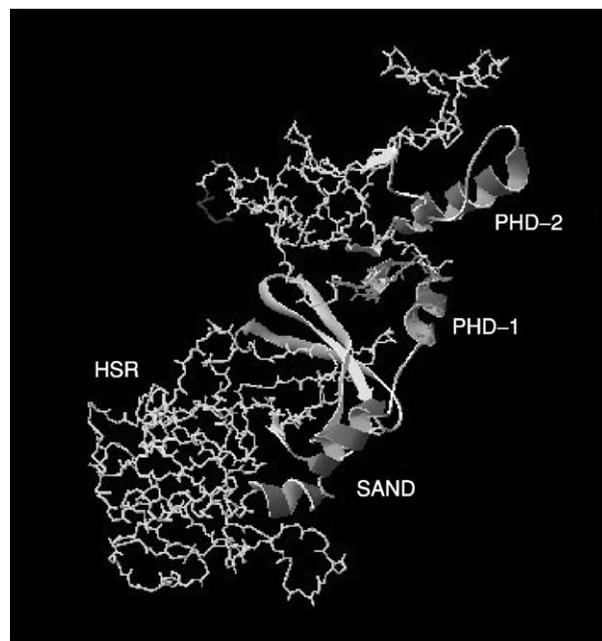
### Etiology

#### Genetics

Single-gene, autosomal recessive inheritance with complete penetrance marks APS-1 as unusual among autoimmune diseases. Its relatively simple genetics

provides hope that an understanding of the molecular basis for deranged immune tolerance in APS-1 may be achieved more easily than that for more common autoimmune conditions that are polygenic. The gene conferring the condition, designated the autoimmune regulator (*AIRE*) gene, was first mapped to chromosome 21q22.3 and then positionally cloned. The gene consists of 14 exons that transcribe a 2445-bp mRNA transcript. The translation product contains 545 amino acids with a predicted molecular weight of 57.5 kDa.

A proposed molecular model for *AIRE* is presented in Fig. 1. The protein contains at least three DNA-binding domains (two PHD or zinc finger domains and one SAND domain). The presence of these domains and the localization of *AIRE* to the nucleus suggested that the protein was a transcription factor. This postulate has received strong support from subsequent functional studies. Inclusion of an *AIRE* construct in a eukaryotic transactivation system, in which the protein was fused to a heterologous DNA-binding domain, increased expression of reporter genes by one to two orders of magnitude. Cotransfection of an *AIRE* construct and an interferon- $\beta$  (IFN- $\beta$ ) minimal promoter upstream of luciferase into monkey kidney cells resulted in a 5.5-fold



**Figure 1** A model for the *AIRE* protein. Reprinted with permission from Kumar, P. G., Laloraya, M., Wang, C-Y., Ruan, Q-G., Davoodi-Semiromi, A., Kao, K-J., and She, J-X. (2001). The autoimmune regulator (*AIRE*) is a DNA-binding protein. *J. Biol. Chem.* 276, 41357-41364.

increase in reporter activity compared with the activation induced when only vector sequence was cotransfected with the promoter.

Gel shift assays that distinguish whether a protein is free or bound into a complex showed that the PHD domains of AIRE homodimers and tetramers bound T-box (TTATTA) or G-box (ATTGGTTA) sequences, whereas AIRE monomers bound no DNA sequences. The formation of oligomers is likely to be essential for AIRE activation. The process may be stimulated by phosphorylation of AIRE by cyclic AMP (cAMP)-dependent protein kinase A and/or protein kinase C. Finally, interaction between AIRE and cAMP response element-binding (CREB) protein suggests a link to the control of cAMP-mediated signaling within cells by AIRE.

*AIRE* is expressed in a subset of epithelial cells in the thymic medulla and corticomedullary junction, in medullary thymocytes derived from monocytes, and within lymphocytes and neutrophilic granulocytes of the lymph nodes and spleen. Among circulating cells, it is expressed most prominently in CD14-positive monocytes and dendritic cells.

Mouse studies suggest that thymic expression of *Aire* (the mouse homologue of human *AIRE*) is dependent on signals from surrounding thymocytes. Prenatal expression of *Aire* in thymic medullary epithelium is normally recognized by day E14, a time after lymphocytes have started committing to the  $\alpha\beta$  subset of TCR lineage. This expression is not observed in Tg $\epsilon$ 26 mice, a transgenic line characterized by early thymocyte developmental arrest and consequent disruption of thymic architecture. In contrast, RAG<sup>null</sup> mice experience late thymocyte developmental arrest. Their thymic architecture is normal, and they exhibit normal *Aire* expression. Furthermore, *Aire* expression is dependent on the expression of the RelB transcription factor. RelB is a member of the NF- $\kappa$ B family and is required for the normal development of the thymic dendritic cells and the medullary-activated stromal cells that express a surface molecule that binds antibody 29. It is these 29-positive cells that predominantly produce *Aire*. RelB-deficient transgenic mice are deficient of 29-positive stromal cells, do not express *Aire*, and have abnormal thymic architecture and negative selection. However, absent *Aire* expression is not merely the result of 29-positive epithelial cell hypoplasia. The thymi of the Tg $\epsilon$ 26 mice contain 29-positive epithelial cells but do not express *Aire*.

Elegant studies in congenic mice with specific deficiencies of negative or positive selection provide functional evidence of a link between *Aire* production

and intrathymic negative selection of T lymphocytes. Therefore, our understanding of normal *AIRE* function suggests that it is a transcription factor whose expression is activated at a relatively late stage of thymic ontogeny by signals from developing thymocytes. The absence of AIRE is associated with abnormal negative selection. However, a causal relationship with this functional defect has not been proven.

Nearly 50 mutations of AIRE have been reported worldwide in patients with APS-1. Founder effects are apparent. Of 67 mutant alleles in Finnish patients, 55 (82%) had the most common error, R257X, in which a substitution of thymidine for cytosine resulted in an arginine being changed to a stop codon in exon 6. The resulting 256-amino acid translation product had no function. This mutation was also common among Russian patients, but among studies of other continental European populations, only 15 of 47 alleles (32%) had the R257X mutation. In the United Kingdom, a 13-bp deletion in exon 8 that encoded a functionless truncated protein accounted for 20 of 25 cases (80%), and only 1 allele contained the R257X sequence. The exon 8 mutation was also the one reported most frequently in North America, where it was found in 21 of 42 cases (50%). In Norway, the 13-bp deletion was the most common mutation found in a series of 20 patients, but interestingly, it was detected in at least four different haplotypic backgrounds. Founder effects were very apparent in Italian and Iranian populations. Among Sardinians, 10 of 11 mutant alleles demonstrated a mutant R139X allele in exon 3, and a missense mutation in exon 2 (W78R) was reported to be frequent among APS-1 patients living in a restricted area of southern Italy. All affected Iranian Jews who have been studied expressed a homozygous Y85C missense mutation in exon 2.

The total number of patients with APS-1 who have been genotyped is small; therefore, it has been impossible to perform adequately powered analyses of the relationship between genotype and phenotype. Such endeavors are made very complex by the intrafamilial heterogeneity of phenotypes and the inability to determine the eventual phenotype in young patients who, in time, may develop other features. Nonetheless, it is of interest that the missense mutation among Iranian Jews that allows production of a full-length protein is associated with a clinical spectrum that includes lower rates of mucocutaneous candidiasis and adrenocortical failure than is the case with other populations.

Functional assays with mutant *AIRE* constructs have shed some light on normal molecular function.

The nuclear pattern of AIRE labeling in wild-type monkey kidney cells became predominantly cytoplasmic, rather than nuclear, when the cells were transiently transfected with the common Finnish mutant gene. Two other missense mutations that interfered with PHD domain formation similarly disrupted nuclear staining, but the introduction of the Y85C mutation of Iranian Jews did not change the nuclear AIRE labeling pattern from that seen in the wild type.

A peptide comprising AIRE amino acids 101 to 141 contains a consensus nuclear localization signal. Its transient transfection into monkey kidney cells permitted transport of endogenously produced green fluorescent protein into the nucleus. C-terminal peptides that did not contain this 41-amino acid sequence were also normally transported to the nucleus, suggesting the possibility of another undefined transport sequence.

Disruption of the PHD domains is reported to reduce transcription activation, and mutant homogeneously staining region (HSR) sequences (Fig. 1) may prohibit dimer formation and, therefore, preclude promoter activation.

Perhaps the most important aid for delineating the role of AIRE in the pathogenesis of APS-1 will be strains of Aire knockout mice. The phenotype of an Aire-deficient mouse strain was reported to include many features of APS-1, including atrophy of the adrenal cortex, ovaries, and thymus. Hepatitis was common, and autoantibodies were detected against multiple organs. A second group also demonstrated organ-selective mononuclear cell infiltrates in multiple organs and circulating autoantibodies in another Aire knockout mouse model. The group further proposed that low expression of selected self-antigens by thymic medullary epithelial cells was associated with the autoimmune phenotype.

It is not clear whether genetic influences apart from AIRE influence the progress of APS-1. Various genes in the class I and class II HLA complex have been implicated, but none has been observed consistently. One observation of interest is the increased frequency of DQB1\*0602 in patients with APS-1 compared with that in control populations. This same allele is associated with protection from type 1 diabetes.

### **Immunology**

APS-1 includes an unusual mixture of conditions suggesting that multiple defects exist in the control of immune responses. Whereas a high incidence of mucocutaneous candida suggests immune deficiency, most of the other component diseases arise from unbridled responses against self-targets. However, the

apparent paradox may be resolved when one considers that antigen-specific tolerance is established by an active process that requires interaction between T lymphocytes and self-antigen.

The best experimental evidence for disrupted tolerance in APS-1 is the presence of autoantibodies against multiple organs. These immunoglobulins arise before organ failure is clinically apparent and frequently become undetectable after diagnosis, presumably indicating that the autoimmune response has completely eliminated its target. In most autoimmune diseases, the autoantibodies are not the primary effectors of target organ damage, but they do serve as markers of immune autoreactivity. Based on this premise, the requirement of B lymphocytes in immune-mediated pancreatic beta cell destruction was questioned. Nonobese diabetic (NOD) mice that normally develop diabetes spontaneously did not, however, become hyperglycemic if their B-lymphocyte compartment was genetically eliminated. Diabetes was restored in these knockout mice after their bone marrow was reconstituted with wild-type B-lymphocyte precursors. Studies suggest that the antigen-presenting function of the grafted B lymphocytes helps to establish organ destruction that is caused by T lymphocytes and macrophages.

Originally, organ-specific autoantibodies were detected by overlaying patient sera on normal frozen tissue sections and then identifying specific immunoglobulin binding to the tissue with fluorescently labeled anti-immunoglobulin second antibodies. The method is labor intensive, highly complex, and somewhat inconsistent between runs. It also lacks the sensitivity of more modern techniques.

Adrenocortical failure in patients with APS-1 features autoantibodies that bind selectively to normal adrenal cortex. The presence of a distinct population of circulating autoantibodies in patients with adrenocortical failure that also bind steroid-producing cells of gonads and placenta is associated with gonadal failure. The antigens recognized by these autoantibodies have been identified, and as a result, improved immunoassays are now available, albeit on a limited basis. Targets include the cytochromes P450c21 (21-hydroxylase), P450c17 (17-hydroxylase), and P450scc (side chain cleavage) as well as the neurotransmitter synthetic enzyme aromatic L-amino acid decarboxylase (AADC). In some but not all laboratories, adrenocortical failure is more closely associated with autoantibodies against P450c21 and AADC, whereas gonadal failure is more strongly linked with responses to P450scc and P450c17. Apart from their association with APS-1-related adrenal gland failure,

autoantibodies against AADC are seen in approximately 90% of APS-1 patients who have either hepatitis or vitiligo.

When interpreting autoantibody results (or tests of other immune phenomena), it is important to consider the particular patient under study. Positive results in patients with isolated autoimmune endocrinopathy do not necessarily carry the same importance as does similar positivity in patients with polyorgan disease. The poor predictive value of GAD<sub>65</sub> autoantibodies for diabetes in patients with APS-1 is a good example. Another involves the autoimmune hepatitis that arises in some 15% of patients with APS-1. It is clinically and histologically indistinguishable from other forms of chronic active hepatitis. Hepatic autoantigens in APS-1 patients with chronic hepatitis include four P450 cytochromes: 1A1, 1A2, 2A6, and 2B6. Autoantibodies against P450 1A2 are specific for APS-1-related hepatitis, and antibodies recognizing P450 2A6 are uncommon outside of the APS-1 setting but do not necessarily correlate with clinically detectable hepatitis. Furthermore, no immunity against autoantigens typically observed in other forms of autoimmune hepatitis has been reported in patients with APS-1. Female gender and genes in the HLA complex, including DRB1\*0301 (or a gene close to it), are other determinants of isolated forms of hepatitis that are not important in APS-1.

The calcium sensor in the parathyroid gland was reported to be a target of autoantibodies in the sera from 6 of 17 patients (35%) with APS-1 and hypoparathyroidism and in 8 of 8 patients with isolated adult-onset idiopathic hypoparathyroidism. No parathyroid autoantigens were immunoprecipitated from the sera of 72 euparathyroid controls who either were healthy or had an autoimmune disease affecting just one organ. The limited reports of autoantibodies against parathyroid antigens in patients with APS-1 have been difficult to reproduce. Furthermore, tissue samples are rarely available for direct microscopic or biochemical scrutiny because chronically affected glands become atrophic and cannot be found at postmortem examination.

The susceptibility to candida infection in APS-1 has not been explained satisfactorily. Patients are able to generate spontaneous humoral responses against candida antigens. Standard immunological techniques have not convincingly defined an abnormality in the maturation or function of T lymphocytes of patients with APS-1. A disturbing association with oropharyngeal squamous cell carcinoma is becoming apparent, further suggesting the possible existence of a subtle but functionally important defect in immune

surveillance. Local effects of chronic mucocutaneous candida could also be important in this susceptibility. The significance of defective immunoglobulin production, including immunoglobulin A (IgA) deficiency in patients with APS-1, is unclear.

### Environment

Infectious agents harbor antigens that are similar enough to human self-components that they can induce cross-reactive autoimmunity that persists long after the germ has been eliminated from the host. This process of antigen mimicry is important in causing rheumatic heart disease and some HLA-B27-positive forms of connective tissue disease. Microbes could also induce autoimmunity by infecting immune cells directly, disrupting target tissues that then release sequestered antigens to the extracellular space, or altering the soluble immune milieu.

Alternatively, the "hygiene hypothesis" suggests that infections may prevent autoimmunity. In general, autoimmune diseases are becoming more frequent among affluent populations worldwide. It has been proposed that improved infection control has reduced rates of cyclic immune stimulation and regulation. It is further reasoned that this reduced immune activity provides a permissive environment for autoimmunity among genetically predisposed individuals.

A modifying role of environmental agents in APS-1 may account for members of the same kindred experiencing remarkably different disease courses; however, specific data to support the hypothesis are lacking. A female bias for APS-1 has been reported, and the definite female predisposition for gonadal failure could reflect hormonal effects on immune function. Anatomical and genetic effects could also account for these phenomena. The *Aire*<sup>-/-</sup> mouse should provide a good experimental model for assessing the effects of environment on APS-1.

### Clinical Aspects of APS-1

Patients with undiagnosed APS-1 are frequently assessed many times before the diagnosis is established. Recurrent hypocalcemic, hypoglycemic, or hyponatremic seizures might be misdiagnosed as epilepsy, or the primary cause of altered metabolism might not be sought once the metabolic derangement is corrected. Patients with undiagnosed adrenocortical failure may receive extensive and uninformative assessments of the gastrointestinal tract because of anorexia, weight loss, malaise, and/or abdominal discomfort. In a Norwegian report, 7 of 17 patients

who were diagnosed with APS-1 in the midst of a metabolic crisis were misdiagnosed at earlier medical visits for a manifestation of the condition. The author's North American experience suggests that this problem is not limited to Norway.

The diagnosis of APS-1 should be fairly straightforward in children once the first clinical manifestation is recognized. Among 43 children who had at least a 2-year history of mucocutaneous candidiasis and who did not have a demonstrable defect in T-lymphocyte function, 19 (44%) had an associated autoimmune endocrinopathy, with adrenal cortex and thyroid diseases being most frequent. Thus, candida infections that are prolonged, severe, or unresponsive to routine treatment merit an evaluation of immune function. If that proves to be normal, serum calcium and phosphorus concentrations and autoantibodies against adrenal, gonadal, gastric parietal cell, thyroid, and pancreatic islet autoantigens should be examined. If signs or symptoms suggesting the presence of specific organ failure arise, functional tests of that organ should also be performed. Annual monitoring of calcium concentrations and adrenal, steroidal cell, and gastric parietal cell autoantibodies should be considered in children who initially have no evidence of autoimmunity. It is hoped that these recommendations will be revised when assays for autoantibodies against defined autoantigens become widely available.

Hypoparathyroidism is uncommon in children outside of the neonatal period and should prompt routine monitoring for polyglandular autoimmunity unless another cause for parathyroid dysfunction is diagnosed. Children who present with isolated autoimmune adrenocortical failure are more likely to have isolated disease (or perhaps APS-2) than to have APS-1. A number of historical features (outlined later in the article) will help to identify those with the more common forms of disease. Recall that APS-1 is autosomal recessive; therefore, family history should carefully seek consanguinity as a clue to its diagnosis.

Patients with component diseases of APS-1 will benefit from genetic testing of *AIRE* when it becomes commercially available. A negative result will not exclude APS-1, but a positive result will confirm the need to monitor gland function.

Despite appropriate hormone and vitamin replacement therapy, patients with APS-1 often fare poorly. Pernicious anemia is often misdiagnosed for many months due to its nonspecific presentation. Acquired hyposplenism and cholethiasis were reported in one American series. Issues of self-esteem arise when hypogonadism, alopecia, and/or vitiligo become

important. Of even more concern are severe, poorly understood complications that interfere with quality of life, can be life threatening, and respond poorly to treatment. These include malabsorption, hepatitis, oropharyngeal carcinoma, progressive idiopathic myopathy, Sjögren syndrome, and red cell hypoplasia.

## TYPE 2 AUTOIMMUNE POLYGLANDULAR SYNDROME

APS-2 is defined by the presence of autoimmune adrenocortical failure and either type 1 diabetes or any type of autoimmune thyroiditis. As outlined in [Table 1](#), other organs are frequently targeted by the immune system in patients with this condition.

### Epidemiology

The prevalence of APS-2, from studies in the United Kingdom and Western Europe, is estimated to be 1.4 to 5.0 per 100,000 population. Type 1 diabetes is most often the first disease to be identified. Whereas adrenocortical failure is typically the last feature of APS-1 to be diagnosed, it is more frequently the first component disease diagnosed in patients with APS-2. The peak age of patients at the time of their diagnosis of APS-2 is between 20 and 40 years. Females are affected about twice as often as are males, with the bias being more obvious among patients who are affected by thyroiditis. Component diseases that frequently coexist with thyroiditis in the absence of adrenal gland disease (e.g., vitiligo, atrophic gastritis/pernicious anemia) also show a female bias. It is not established whether these associated conditions are independently influenced by gender or whether the bias is related to their association with thyroiditis. Gonadal insufficiency occurs predominantly in females with APS-2, but at a lower rate than that seen in patients with APS-1.

### Etiology

#### Genetics

Higher concordance for APS-2 among monozygotic twins than among dizygotic twins attests to the genetic influences on the condition. Polygenic inheritance with genetic heterogeneity (a similar phenotype resulting from distinct genotypes), gene interactions, variable penetrance, and environmental influences all complicate the APS-2 phenotype.

The HLA-DRB1\*0301/DQB1\*0201 haplotype is associated with a high risk of APS-2. This is expected because isolated autoimmune adrenocortical failure,

Table 1 Characteristics of APS-1 and APS-2

	APS-1	APS-2
Definition	Two of three diseases: Mucocutaneous candidiasis Hypoparathyroidism Adrenocortical failure	Adrenocortical failure plus at least one of the following: Type 1 diabetes mellitus Autoimmune thyroiditis
Associated conditions	Hypogonadism (~45%) Alopecia (~40%) Ectodermal dysplasia (e.g., teeth, nails, cornea) Atrophic gastritis/Pernicious anemia (~15%, young onset) Vitiligo (~15%) Hepatitis (~15%) Malabsorption (~15%) Thyroiditis (~15%) Type 1 diabetes (~5%) Hyposplenism Cholelithiasis	Hypogonadism (<5%) Alopecia (rare) Atrophic gastritis/Pernicious anemia (rare, adult onset) Vitiligo (~5%) Myasthenia gravis Sjögren syndrome
Usual first manifestation	Mucocutaneous candidiasis	Type 1 diabetes
Prevalence	1 in 9000 to 1 in 80,000	1.4 to 5.0 per 100,000
High-risk populations	Iranian Jews Finns Sardinians	Females
Genetics	Autosomal recessive AIRE gene	Polygenic HLA DR3
Age of onset	First decade	Third or fourth decade

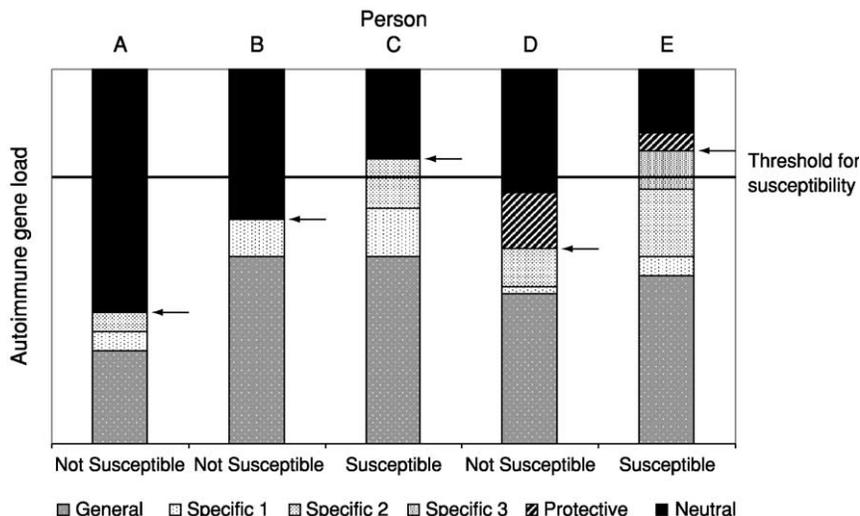
type 1 diabetes, and thyroiditis share this association. Early reports suggested that APS-2, like diabetes, was associated with the DRB1\*04/DQB1\*0302 haplotype. It was subsequently determined that this effect partitioned with the presence or absence of diabetes. The high-risk haplotype was not seen at a higher frequency in APS-2 patients with disease restricted to the adrenal and thyroid glands.

Although much more information is available about the genetics of the individual component diseases of APS-2, little else is known in the strict context of polyglandular autoimmunity. A number of inferences can be drawn from what is known. Recall first that APS-2 was originally defined by a clinical phenotype that had to include adrenocortical failure. Other combinations of autoimmune gland failures that were similar to APS-2 but did not include adrenocortical failure either were relegated to a heterogeneous group of disease clusters, called APS-3, or were left unclassified. Most evidence developed since the original description in 1980 suggests that the genetic predisposition to APS-2 and its component diseases arises from a convergence of autoimmune susceptibility genes. Therefore, from a genetic perspective,

the mandatory inclusion of adrenal insufficiency to distinguish APS-2 may be somewhat arbitrary.

Our group has proposed that polyglandular autoimmunity is genetically influenced by three major categories of genes: (1) antigen/organ nonspecific, (2) antigen/organ specific, and (3) protective (Fig. 2). The first category of genes encodes proteins that provide an internal immunological milieu that permits autoimmune processes in general to initiate or persist. One or more genes associated with the DR3 haplotypes listed previously represents an example. In addition to APS-2, associations with DR3 have been reported for myasthenia gravis, autoimmune hepatitis, rheumatoid arthritis, and systemic lupus erythematosus.

The second category of genes encodes proteins that provide absolute or relative organ specificity to the autoimmune responses. The HLA-DRB1\*04 association with type 1 diabetes in APS-2, but not adrenocortical failure or thyroiditis, is an example of this phenomenon. To date, variations in the number of tandem repeat sequences in the upstream untranslated region of the insulin gene are associated with diabetes but no other autoimmune disease. The third category of autoimmune genes confers protective effects.



**Figure 2** A model for polygenic inheritance of susceptibility to autoimmune disease. Each bar represents a portion of one individual's genome. The combined genetic load of general and specific autoimmune genes (arrow) in person A is below the threshold where disease susceptibility occurs. Person B has more general susceptibility genes than does person A. Genetic interactions cause some augmentation of the influence of the specific autoimmune gene, but the combined effects are still inadequate to confer susceptibility to clinical disease. Person C has an additional specific autoimmune gene (specific gene 2) that interacts with specific autoimmune gene 1. The combined effects now exceed the threshold of susceptible, and this person may develop autoimmune disease if required environmental influences also act. Person D has a protective gene that reduces the activity of specific gene 1 and some general autoimmunity genes. Gene interactions cause a slight reduction in the action of specific gene 2. This person's genetic load is below the threshold, and there is no susceptibility to autoimmunity. Person E has three specific genes. In addition to promoting autoimmunity directly, the third gene modulates the effects of other susceptibility and protective genes. The total genetic load is above the susceptibility threshold.

Protection from type 1 diabetes conferred by DQB1\*0602 is an example.

This model oversimplifies the true pathological processes causing APS-2. For example, the CTLA-4 locus may encode a protein that provides general pro-autoimmune effects, but only in certain circumstances. The locus encodes a T-lymphocyte protein that binds most avidly to CD80 (B7-1) on antigen-presenting cells, and this cell-to-cell signal contributes prominently to immunoregulatory responses. Polymorphisms of CTLA-4 are reported to be associated with susceptibility to type 1 diabetes, Graves' disease, Hashimoto's thyroiditis, Wegener's granulomatosis, systemic lupus erythematosus, and celiac disease but not multiple sclerosis. In addition, the DQB1\*0602 gene powerfully exerts protection from type 1 diabetes unless the patient has APS-1 or APS-2. Furthermore, DQB1\*0602 is a susceptibility gene for the autoimmune attack of the central nervous system in multiple sclerosis. This difference may reflect different avidities of binding between the HLA-antigen complexes and the TCRs of pathogenic clones. It is also possible that the molecular and biochemical processes that promote one self-reactive response in one disease may inhibit or have no effect on another disease.

It is also important to recall that genes may exert control on portions of the autoimmune response but not on the entire process. Experiments with congenic mouse strains have identified genetic intervals that control (1) insulinitis but not progression to diabetes and (2) autoantibody production but not progression to overt lupus nephritis. In humans, inheritance of a propensity to make autoantibodies against thyroid gland components appears to be inherited in a dominant fashion from affected mothers. However, progression to overt thyroiditis is not transmitted as completely.

The genetic data further support the hypothesis that polyorgan failure in APS-2 is not the result of a single process that generalizes to multiple organs, as may be the case with APS-1. More likely, multiple autoimmune processes with both distinctive and overlapping components are coexpressed. The NOD mouse model of diabetes and sialoadenitis provides an example of diverse mechanisms causing autoimmunity simultaneously in the same animals. When the interleukin-4 (IL-4) gene was disrupted in NOD mice, the rate of spontaneous diabetes and islet infiltration by immune cells was unchanged compared with the wild type. In contrast, salivary gland infiltration by

leukocytes persisted in the IL-4 knockout mice but was not associated with xerostomia.

### Immunology

The epidemiology and genetics of APS-2 and its component diseases are similar (with the exception of gonadal failure). Therefore, the immunological processes underlying the diseases may overlap significantly. The immunology of APS-2 specifically has not been studied as well as it has for the individual component diseases. Autoantibodies appear to bind the same target antigens, irrespective of whether the autoimmune response occurs in isolation or as part of APS-2 (Table II). Immune cell infiltrates have been reported in virtually all target organs, including adrenal cortex, islets, thyroid, and developing ovarian follicles. In patients with vitiligo, small clusters of immune cells are seen around melanocytes of skin biopsied along the edges of the lesions. Alopecia is associated with mononuclear immune cell infiltrates in and around the hair bulbs.

Premature ovarian failure associated with either APS-2, autoimmune adrenocortical failure, or asymptomatic adrenal autoantibody positivity has been distinguished from that occurring in the presence of

autoimmune conditions that do not affect the adrenal gland. Autoantibodies against the P450c17 and P450scc cytochrome enzymes are rarely observed in patients with idiopathic premature gonadal failure who also have disease involving thyroid, islet, and/or gastric parietal cells. Microscopically, lymphocytic oophoritis appears almost exclusively in conjunction with adrenal gland autoimmunity. The association of autoimmune oophoritis and adrenocortical failure may prove to be invaluable in understanding the organ specificity of autoimmune disease.

Changes of cellular immune function in thyroiditis and type 1 diabetes have been investigated intensively. Altered function of antigen-presenting cells, regulatory and responder lymphocyte populations, cytokine environment, and production of mediators of cell destruction all have been reported. Whereas experimental animal models have been explored in great detail, causal phenomena from studies of humans have not been demonstrated.

### Environment

Associations of the component diseases of APS-2, particularly diabetes and thyroiditis, have been investigated intensively, yet few agents have been associated

Table II Defined Autoantigens in APS-1, APS-2, and Component Diseases

Organ	APS-1	APS-2	Isolated component diseases
Adrenal cortex	P450c21	P450c21	P450c21
	AADC	?P450c17	?P450c17
	?P450c17	?P450scc	?P450scc
	?P450scc		
Gonads	P450c17	P450c17	None to steroid-producing cells
	P450scc	P450scc	
Islet	GAD <sub>65</sub>	GAD <sub>65</sub>	GAD <sub>65</sub>
	GAD <sub>67</sub>	IA-2	IA-2
Vitiligo	Tyrosinase		Insulin
	AADC		Tyrosinase <sup>a</sup>
			TRP-2
Thyroid	Thyroid peroxidase	Thyroid peroxidase	Thyroid peroxidase
	Thyroglobulin	Thyroglobulin	Thyroglobulin
	TSH receptor	TSH receptor	TSH receptor
Gastric parietal cells	H <sup>+</sup> ,K <sup>+</sup> ATPase	H <sup>+</sup> ,K <sup>+</sup> ATPase	H <sup>+</sup> ,K <sup>+</sup> ATPase
	Intrinsic factor	Intrinsic factor	Intrinsic factor
Liver	P450 1A2		P450 2D6
	P450 2A6		Asialoglycoprotein receptor
	AADC		
Parathyroid	Calcium sensor		Calcium sensor

Note. GAD, glutamic acid decarboxylase; AADC, aromatic-L-amino acid decarboxylase; TSH, thyroid-stimulating hormone; TRP-2, tyrosinase-related protein-2.

<sup>a</sup>Many vitiligo patients had other autoimmune endocrinopathy but did not have APS-1 or APS-2.

with disease unequivocally. A proven causal association exists between congenital rubella infection and diabetes. Hypothetical precipitants of type 1 diabetes have included other viruses (e.g., Coxsackie B4 or other enteroviruses), diet components (e.g., cow milk proteins), routine childhood immunizations, and chemicals (e.g., nitrosamines in smoked meat). Of these, the relationship between enterovirus infection and type 1 diabetes is most compelling. Although proof that enteroviruses can initiate in human islet autoimmunity is lacking, convincing data suggest that enteroviral infections at least accelerate the process. Patients with newly diagnosed type 1 diabetes carry enteroviral RNA in their serum more frequently than would be expected. In addition, seroconversion for islet autoantibodies among young children has been temporally associated with recent enterovirus infections.

Thyroiditis has been associated with environmental agents such as dietary iodine, *Yersinia enterocolitica*, and Epstein–Barr virus. Autoimmune adrenocortical failure has no reported associations with environmental triggers. Thus, there are no known environmental exposures that completely account for the spectrum of polyglandular autoimmunity. However, most of the proposed environmental agents for organ-specific autoimmune diseases are commonly encountered. It remains plausible that various autoimmune diseases coexist given that multiple disease-specific environmental exposures have occurred.

### Clinical Aspects of APS-2

It can be difficult to distinguish patients with APS-2 from those with a single autoimmune condition when they present with their first organ failure. About half of patients with autoimmune adrenocortical failure either have or will develop a second autoimmune condition. Therefore, it is obvious that autoantibodies against thyroid, islets, steroid-producing organs, and gastric parietal cells should be monitored. The presence of organ-specific autoantibodies should prompt close clinical and laboratory testing for failure of the targeted organ.

Type 1 diabetes is frequently the first sign of APS-2, and occasionally thyroiditis can precede other components. Autoantibodies against other organs are far less prevalent in patients with either of these two conditions than in those with adrenocortical failure. Therefore, laboratory investigations are not cost-effective, and close attention must be paid to the medical history and physical examination. A family history of polyglandular autoimmunity should prompt periodic antibody screening, as should the

appearance of vitiligo or alopecia (areata, totalis, or universalis). The onset of idiopathic premature ovarian failure in patients with thyroiditis or diabetes is cause for concern about the risk of future adrenocortical failure, especially if the P450 cytochrome autoantibodies are present.

Patients with type 1 diabetes, especially females, have at least a 20% risk of developing thyroid autoantibodies. Monitoring for thyroid disease among patients with type 1 diabetes is usually considered as standard care. This may be accomplished with either sensitive thyrotropin (TSH) or thyroid peroxidase autoantibody assays. Patients who have clinical or laboratory evidence of both islet and thyroid autoimmunity should be monitored clinically and serologically for other autoimmune conditions. Parietal cell autoantibodies will be positive in these individuals more frequently than will adrenal or steroidal cell autoantibodies.

Important clinical clues to the presence of adrenocortical failure in patients with other autoimmune conditions include chronic fatigue despite adequate thyroid hormone replacement, skin crease pigmentation that does not have the texture of acanthosis nigricans in patients with insulin resistance, and improved blood glucose control in patients with diabetes who have not changed their treatment or daily routines. Finally, it is also worth recalling that the institution of thyroid hormone replacement in hypothyroid patients who have undiagnosed adrenocortical failure can precipitate an adrenal crisis.

### CONCLUSION

The two forms of APS include groups of autoimmune diseases that coexist at rates beyond those expected by chance. Despite the endocrinological emphasis in the syndrome names, disease is not restricted to endocrine organs. APS-1 and APS-2 have distinct epidemiologies and genetics, but investigators have been less successful at identifying specific abnormalities of immune function or environmental precipitants of autoimmunity. To minimize morbidity and mortality, it is important for clinicians to anticipate which patients with an autoimmune disease are likely to develop a second organ failure. Patients with APS-1 often become ill during their second or third decade of life and may develop any of a number of severe illnesses that can prove to be fatal. The prognosis of APS-2 is somewhat better, but patients with diabetes are at risk for its chronic complications. Further research into the nature of the autoantigens, immune tolerance, immune surveillance, the AIRE transcription factor, and genes

controlling susceptibility and resistance to autoimmunity is of great importance in improving the lives of these patients.

### See Also the Following Articles

Adrenal Insufficiency • Diabetes, Type 1 • Graves' Disease • Hashimoto's Disease • Hypoparathyroidism • Premature Ovarian Failure • Thyroid Autoimmunity

### Further Reading

Ahonen, P., Myllarniemi, S., Sipila, I., and Perheentupa, J. (1990). Clinical variation of autoimmune polyendocrinopathy–candidiasis–ectodermal dystrophy (APECED) in a series of 68 patients. *N. Engl. J. Med.* **322**, 1829–1836.

- Betterle, C., Dal Pra, C., Mantero, F., and Zanchetta, R. (2002). Autoimmune adrenal insufficiency and autoimmune polyendocrine syndromes: Autoantibodies, autoantigens, and their applicability in diagnosis and disease prediction. *Endocrine Rev.* **23**, 327–364.
- Liston, A., Lesage, S., Wilson, J., Peltonen, L., and Goodnow, C. C. (2003). Aire regulates negative selection of organ-specific T cells. *Nat. Immunol.* **4**, 350–354.
- Muir, A., and She, J.-X. (1999). Advances in the genetics and immunology of autoimmune polyglandular syndrome II/III and their clinical applications. *Ann. Med. Int. (Paris)* **150**, 301–312.
- Perheentupa, J. (2002). APS-I/APECED: The clinical disease and therapy. *Endocrinol. Metab. Clinics North America* **31**, 295–320.
- Pitkanen, J., and Peterson, P. (2003). Autoimmune regulator: From loss of function to autoimmunity. *Genes Immunol.* **4**, 12–21.
- Schatz, D. A., and Winter, W. E. (2002). Autoimmune polyglandular syndrome: II. Clinical syndrome and treatment. *Endocrinol. Metab. Clinics North America* **31**, 339–352.



# Autonomic Nervous System, Aging and

K. E. Hall and John W. Wiley

University of Michigan Healthcare System, Ann Arbor, Michigan, United States

## Glossary

**age-associated** Describing a condition that is more prevalent in a specific age segment of the population, specifically advanced age.

**apoptosis** Programmed cell death or “cell suicide”; a genetically mediated series of events by means of which cells actively trigger their own destruction.

**homeostasis** In an organism, the fact or process of maintaining the internal environment in a balanced state that promotes health and survival, by means of various sensory, adaptive, and control mechanisms that react to internal and external conditions to produce this balance.

**oxidative stress** Excessive production of, and/or impaired removal of, oxidants (chemicals that withdraw electrons from other chemicals) with a concomitant decrease in the reducing capacity of cells; associated with various age-related diseases and degenerative conditions.

As humans age, many body functions begin to deteriorate, including the function of various neuronal systems. The peripheral nervous system in the body is divided into those nerves that affect voluntary movement (the voluntary system) and those that regulate the function of internal organs such as the heart and gastrointestinal tract (the autonomic system). Significant alterations in the function of the autonomic nervous system appear to occur with aging.

## INTRODUCTION

Deciphering the mechanisms of physiological aging versus the consequences of disease is an ongoing challenge for gerontologists. In general, autonomic nervous system (ANS) function is well preserved in otherwise healthy individuals through the seventh decade of life. However, there are some notable changes in neuromuscular function of the cardiovascular and gastrointestinal (GI) ANS that accompany advanced age. These changes have been reported in

human and animal models. Age-dependent autonomic dysfunction could involve loss of neurons and/or altered neuronal or smooth muscle function. This article explores age-associated changes in the sympathetic and parasympathetic innervation of the cardiovascular system and the intrinsic innervation in the GI tract (enteric nervous system [ENS]). Studies performed on human subjects are emphasized, and animal research is discussed where the results provide insights regarding potential mechanisms and future directions in this field.

## AGE-ASSOCIATED CHANGES IN CARDIOVASCULAR FUNCTION

The ANS is an important factor implicated in cardiovascular pathophysiology such as hypertension, myocardial ischemia, and cardiac arrhythmias.

## AGING AND THE SYMPATHO-ADRENAL SYSTEM

The sympathetic nervous system (SNS) plays an essential role in the maintenance of physiological homeostasis in general and arterial blood pressure in particular. This appears to be the case both under basal (resting) conditions and in response to acute stress. Postganglionic sympathetic neurons innervating the heart and resistance vessels help to control cardiac output, arterial blood pressure, and regional vascular conductance, thereby ensuring the proper perfusion of vital organs. SNS stimulation of adrenalin (epinephrine) release from the adrenal medulla contributes significantly to the regulation of cardiovascular function and regulates energy metabolism. The SNS also has a key role in the regulation of internal body temperature. Impairment of SNS function has been implicated in the development of a variety of common medical conditions, including hypertension, congestive heart failure, sudden cardiac death, insulin resistance (metabolic) syndrome, and obesity. Thus, understanding the mechanisms of

SNS dysfunction in these conditions is of major clinical significance.

The current literature supports the view that basal SNS activity increases with advancing age in healthy adult humans. Elevation in SNS activity appears to be region specific, targeting skeletal muscle and the gut rather than other organs innervated by the SNS such as the kidney. The SNS tone of the heart is increased, due primarily to reduced neuronal noradrenalin reuptake. In contrast, basal and stress-associated adrenalin secretion from the adrenal medulla is reduced markedly in human aging. Despite decreased secretion, plasma adrenalin concentrations are relatively unchanged due to a concomitant reduction in plasma clearance of adrenalin. The mechanisms underlying these age-associated changes in sympatho-adrenal function have not been established. Evolving research suggests that the increase in basal peripheral SNS activity with age is associated with elevated forebrain noradrenergic activity. These studies support the hypothesis that increased central nervous system sympathetic activity may underlie peripheral increases in SNS tone noted in studies of aged humans. The situation in other animal models appears to be rather different. In particular, whereas reduced peripheral baroreflex inhibition appears to be an important primary change with aging in animal studies, baroreceptor function in aged humans appears to be relatively preserved.

Reduced arterial compliance is a well-described feature of aging. It is less clear whether neuromuscular function of the venous system is altered by aging. Age-associated changes in venous compliance have been difficult to document, although some studies in humans have shown that aging reduces venous compliance and capacity. The constrictor response of the human dorsal hand vein to the  $\alpha$ -adrenoceptor agonists noradrenalin and phenylephrine is unchanged. However, aging is associated with a reduced dilator response to the  $\beta$ -adrenoceptor agonist isoprenalin, and this may have clinical implications in patients who require these agents for blood pressure support. In summary, adult human aging is associated with significant changes in SNS physiological function and regulation that likely have clinical implications.

### AGING AND PARASYMPATHETIC REGULATION OF THE CARDIOVASCULAR SYSTEM

In contrast to the lack of effect of aging on human sympathetic baroreceptor function, significant

decreases in cardiovagal baroreflex sensitivity (BRS) have been described. Diminution in cardiovagal BRS is associated with impaired regulation of arterial blood pressure and, in the presence of myocardial ischemia, with enhanced risk of ventricular tachyarrhythmias and sudden death. A sedentary lifestyle is associated with a reduction in cardiovagal BRS, even in healthy individuals. This effect can be documented as a decreased R-R interval variability in the electrocardiogram (ECG). It has been hypothesized that these changes in the cardiovagal BRS are related to diminished compliance of the large elastic arteries (aorta and carotid) in which the arterial baroreceptors are located. Decreased cardiovagal baroreceptor sensitivity in aged individuals can be ameliorated by exercise, with improvement noted in aged individuals who exercise regularly as compared with sedentary age-matched adults.

### AGING, GENDER, AND ANS REGULATION OF CARDIOVASCULAR FUNCTION

Gender differences in the incidence and clinical course of a range of cardiovascular states are also well recognized. Both short- and long-term prognoses after myocardial infarction are worse for women than for men. However, women with nonischemic cardiomyopathy have improved survival as compared with age-matched men. In addition to the well-known difference in age of presentation of coronary heart disease (CHD), women with CHD are more likely to suffer from autoimmune phenomena such as Raynaud's disease and to experience dysautonomic symptoms such as presyncopal or syncopal episodes. Therefore, an appreciation of gender differences in the structure and function of the ANS is important to a full understanding of a number of common and important clinical presentations.

Cardiac responses to the sympathetic agonist isoprenalin were similar in older males and females but greater in younger males than in younger females. This implies that women may undergo a relatively greater age-dependent decline in autonomic function. A similar observation was made for muscle contractile response to SNS activity. Under 50 years of age, muscle response to sympathetic nerve activity was significantly greater in men than in women, but no differences between the genders were noted for older individuals.

## AGE-ASSOCIATED CHANGES IN GASTROINTESTINAL FUNCTION

Much of the clinical literature suggests that as healthy humans age, the GI tract functions adequately. Problems with swallowing (dysphagia) and gastroesophageal reflux disease (GERD) are relatively common complaints in the elderly, and some studies have shown an age-related loss of enteric neurons in the human esophagus. However, when objective criteria are assessed, actual age-related changes in esophageal physiology appear to be minimal. In fact, quantitative manometry reveals that esophageal contractile wave amplitudes and velocity and duration of lower esophageal sphincter relaxation in healthy individuals over 70 years of age are not different from those in younger individuals. Nevertheless, there is evidence of reduced amplitude of peristaltic contractions in the lower esophagus of the elderly, consistent with the observation that esophageal clearance after gastroesophageal reflux is impaired in older individuals. In general, the rates of gastric emptying and small intestinal transit are unchanged with age in humans. However, there may be subtle alterations in control of gastric and small bowel neuromuscular function that render aged individuals more susceptible to the effects of superimposed disease (e.g., diabetes) or anticholinergic side effects of medications. Medications, in particular, are a potent source of GI problems in older patients and should be reviewed in patients presenting with impaired GI function.

Evidence that colonic transit is reduced in the elderly is mixed. Some reports document increased colonic transit time in aged healthy adults, whereas others show no difference when compared with younger controls. Some of these discrepancies may be attributed to the presence of chronic illness with systemic effects and concomitant use of multiple medical therapies with GI side effects.

## AGING EFFECTS ON GASTROINTESTINAL AUTONOMIC INNERVATION AND FUNCTION

Morphological data from human tissues suggest that the total number of colonic neurons in the myenteric plexus may decrease by as much as one-third with aging. Age-related changes in the human colonic submucosal plexus in advanced age have not been reported. Unfortunately, there is a lack of morphological data concerning the aging human ENS compared with that obtained in animal studies.

Therefore, many conclusions concerning the effect of aging on neuromuscular function in the GI tract are based on nonprimate animal studies that correlate age-related impairment in colonic neuromuscular function with decreased neuronal number and function. The number of cholinergic neurons, acetylcholine release, and smooth muscle response all are decreased in colon preparations from aged rats as compared with youthful controls. Decreased numbers of neurons immunoreactive for nitric oxide have been demonstrated in the colon of rat models of aging, and limited data indicate that this may also be the case in humans. Nitric oxide-containing neurons are essential for normal coordination of motility and inhibit spasmodic nonpropagating contractions. Thus, a significant decrease in this myenteric neuronal population with age could explain the increase in colonic spasticity and constipation in geriatric-aged patients. A separate, but equally important, issue is the effect of aging on extrinsic sympathetic and parasympathetic innervation to the GI tract. Neuroanatomical tracing studies have failed to show neurodegenerative changes in motor or sensory vagal fibers that project to the proximal gut (stomach and small intestine) in rats. Some animal studies suggest that vagal motor innervation of the distal GI tract declines with advanced age and may correlate with reports of region-specific loss of enteric neurons.

Advanced age may be associated with selective impairment of neuronal function without significant loss of neurons. In the peripheral, prevertebral, and paravertebral sympathetic ganglia of aged humans and rodents, terminal axonal dystrophy and remodeling of synapses occur rather than neuronal cell loss. Because these neurodegenerative changes are documented to affect only some neurons, this suggests that a subset of adult postganglionic sympathetic neurons is more sensitive to neurodegenerative effects of aging. One of the characteristics of the aging-resistant postganglionic sympathetic neurons is that they appear to have an increased capacity to take up neurotrophic factors (e.g., nerve growth factor [NGF]). This may be a compensatory response to decreased release of neurotrophins at the nerve terminal and cell body (soma) documented in aged animal models. There are little data on the effect of aging on the differential responsiveness of neuronal subpopulations to distinct neurotrophins such as glial-derived neurotrophic factor (GDNF) and brain-derived neurotrophic factor (BDNF). Data in younger animals indicate that this kind of differential sensitivity to neurotrophins may be important in neuronal survival because BDNF pretreatment enhances the survival-promoting effects of

NGF. The role of advanced age in the expression and function of these neurotropic factors in the ANS is an active area of investigation. In summary, the observations outlined here suggest that the effects of advanced age on the ENS are likely to be region specific and more pronounced in the distal GI tract.

## POSSIBLE MECHANISMS UNDERLYING PHYSIOLOGICAL AGING IN THE NERVOUS SYSTEM

As the preceding discussion indicates, two fundamental questions have emerged as critical issues surrounding aging of the nervous system. Does loss of neurons occur naturally with aging (i.e., senescent loss)? If so, is physiological neuronal senescence a process that begins with neurodegeneration and culminates with neuronal death? Neuronal death appears to occur via at least two distinct pathways. Neurons are capable of undergoing programmed death (apoptosis) when the apoptotic enzyme cascade is activated by appropriate conditions. These include infection, injury, or as a normal feature of nervous system development *in utero*. A second distinct self-destruct program involving the long processes (axons) that precedes the death of the soma has also been described. This second mechanism of “die-off” is more prevalent than apoptosis in many neurodegenerative diseases associated with aging; thus, it is more relevant to a patient’s clinical course. The relative contributions of these two pathways in normal aging is unclear and is a focus of current research. There is considerable interest in identifying genes that may be directly linked to physiological aging and assessing how environmental factors influence the function of these genes and their products. A number of laboratories are examining the potential contribution of oxidative stress as a contributing factor to the diminution in neuronal function observed with advanced age.

## See Also the Following Articles

Aging and Longevity of Human Populations • Aging and the Male Reproductive System • Aging, Animal Models for • Aging, Immunology and • Body Weight, Body Composition, and Aging • Functional Genomics of Aging • GI Hormones Outside the Gut: Central and Peripheral Nervous System • Neuroendocrine System and Aging • Oxidative Stress and Aging • Stress, Aging, and Central Nervous System Interactions

## Further Reading

- Cowen, T. (2002). Selective vulnerability in adult and ageing mammalian neurons. *Autonomic Neurosci.* **96**, 20–24.
- Finch, C. E. (1990). “Longevity, Senescence, and the Genome.” University of Chicago Press, Chicago.
- Gershon, M. D., Kirchgessner, A. L., and Wade, P. R. (1994). Functional anatomy of the enteric nervous system. In “Physiology of the Gastrointestinal Tract” (L. R. Johnson, ed.), 3rd ed., pp. 381–422. Raven Press, New York.
- Hall, K. E. (2002). Aging and neural control of the GI tract: II. Neural control of the aging gut—Can an old dog learn new tricks? *Am. J. Physiol. Gastrointestinal Liver Physiol.* **283**, G827–G832.
- Hall, K. E., and Wiley, J. W. (1998). Neural injury, repair, and adaptation in the GI tract: I. New insights into neuronal injury—A cautionary tale. *Am. J. Physiol. Gastrointestinal Liver Physiol.* **274**, G978–G983.
- Lee, C. K., Weindruch, R., and Prolla, T. A. (2000). Gene-expression profile of the ageing brain in mice. *Nat. Genet.* **25**, 294–297.
- Mattson, M. P. (2001). Molecular biology of the aging nervous system. In “Pathology of the Aging Human Nervous System” (S. Duckett and J. C. de la Torre, eds.), pp. 16–41. Oxford University Press, New York.
- Phillips, R. J., and Powley, T. L. (2001). As the gut ages: Timetables for aging of innervation vary by organ in the Fischer 344 rat. *J. Comp. Neurol.* **434**, 358–377.
- Raff, M. C., Whitmore, A. V., and Finn, J. T. (2002). Axonal self-destruction and neurodegeneration. *Science* **296**, 868–871.
- Schmidt, R. E. (2002). Age-related sympathetic ganglionic neuropathology: Human pathology and animal models. *Autonomic Neurosci.* **96**, 63–72.
- Wade, P. R. (2002). Aging and neural control of the GI tract: I. Age-related changes in the enteric nervous system. *Am. J. Physiol.* **283**, G489–G495.



# Auxology, Childhood

M. A. Preece

University College London, London, United Kingdom

## Glossary

**anthropometry** The science and technology of measurement of the human body.

**centile (or percentile)** An arbitrary division of a range of variability. For example, the 25th centile for height at age 6 years would define the height at which 25% of the healthy 6-year-old population would be at or below.

**consonance** The ordered progression of events in a predetermined order, e.g., the events of puberty.

**cross-sectional** Growth data collected on a single occasion from children of many different ages.

**longitudinal** Growth data collected on at least two occasions from each child of a study group.

**A**uxology is the science and study of physical growth and development. It derives from a Greek root, *αυξειν* (*auxein*), a term that was first used in France in 1919 and subsequently entered English usage in the 1970s. It encompasses many aspects of physical anthropology, genetics, and physiology.

## GROWTH IN HEIGHT

### The Human Growth Curve

Most human dimensions follow a similar pattern of growth throughout childhood and height may be used as a typical example. The height curve of an individual boy is shown in Fig. 1: Fig. 1a shows the height attained or distance curve and Fig. 1b shows the height velocity curve. The distance curve represents all the height accumulated over time until the latest measurement and is therefore relatively insensitive to recent changes in the child's circumstances. In contrast, the velocity curve represents growth over a clearly defined and relatively short period of time, conventionally 1 year.

The height velocity curve shows a trend for a decelerating growth rate from birth to maturity, but with

two interruptions to this trend. The first is the smaller of the two and is not seen in all individuals; it occurs between the ages of 6 and 8 years and is called the midgrowth or midchildhood spurt. There is little difference between sexes except that it may occur approximately 6 months earlier in girls. The mechanism of this spurt is unknown although its timing approximates the initial rise in the secretion of adrenal androgens.

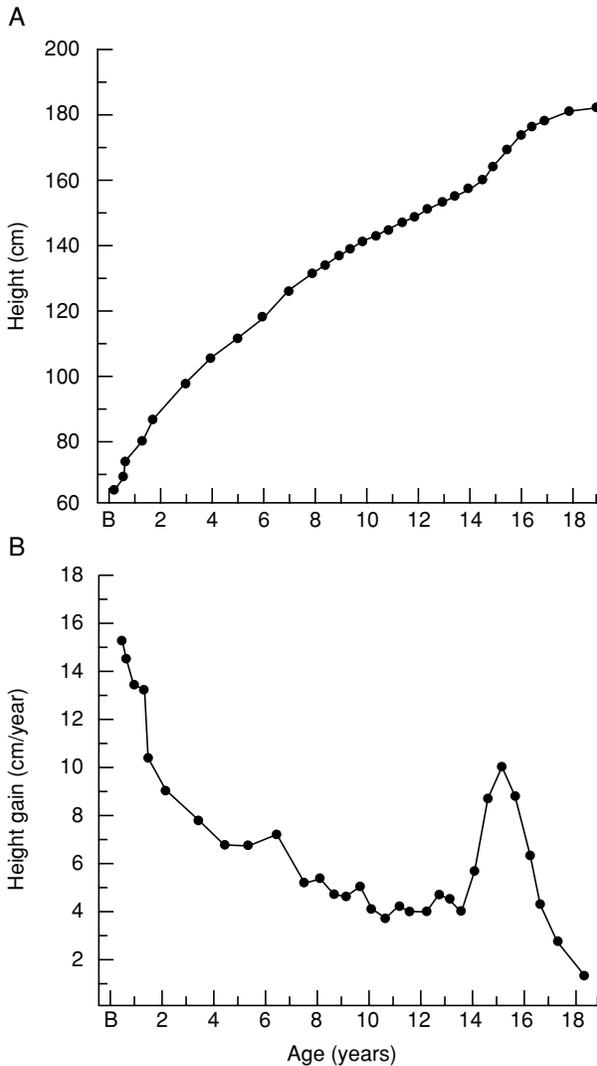
A second and much greater interruption of the downward velocity trend is the adolescent growth spurt, which is illustrated for the average British girl and boy in Fig. 2. Before the onset of the adolescent growth spurt, there is a negligible difference in velocity between the sexes, but the girls' growth spurt starts approximately 2 years before the boys and for approximately 4 years girls are, on average, taller than boys. However, by the time they reach maturity, males are 14 cm taller than females and this difference is almost entirely due to events at puberty. Males grow for 2 years longer before puberty at an average rate of 5–6 cm/year, with the remainder accruing from the rather more intense male growth spurt.

### National Growth Reference

The preceding descriptions concern only the average child, in terms of both size and the timing of events such as the adolescent growth spurt. There is, however, considerable variation in both. The magnitude of variation in size at a given age is illustrated in Fig. 3, which shows the growth reference for height of boys in the United Kingdom. Although the variation in size seems very straightforward, this figure obscures the variation in timing (or tempo, as this aspect is usually termed). This problem is of sufficient clinical importance to justify detailed consideration.

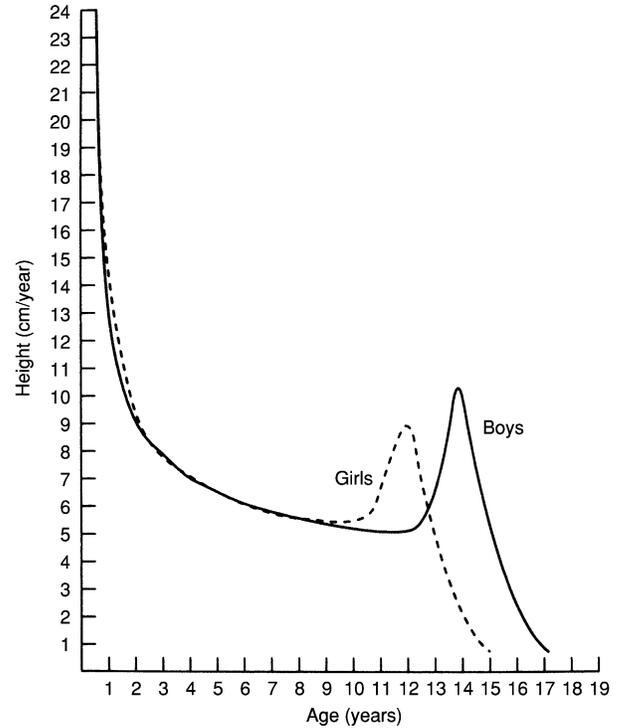
### Tempo of Growth

Consider a boy who will ultimately reach a mature height that is exactly average for the population. If he



**Figure 1** (A) The growth curve for height of a single boy from birth to maturity. (B) The same data as for (A), but expressed as height gain or velocity. Reprinted from Preece, M., Ch. 7.1.1 of “The Oxford Textbook of Endocrinology and Diabetes” (2002) (Wass and Shalet, eds.), by permission of Oxford University Press.

is also exactly average in tempo, then he will follow the mean growth curve, start the adolescent growth spurt at 12 years, reach the peak at 14 years, and reach mature height at approximately 18 years. However, if he is an early developer with fast tempo, then he will always be rather further along his personal growth curve at any given age than average children. He will therefore seem tall at each stage until maturity, which he reaches early, say 16 years. In exact contrast, the late developing boy, also destined for average mature height, will seem short until maturity. These patterns for early and late development are illustrated in Fig. 4. A similar

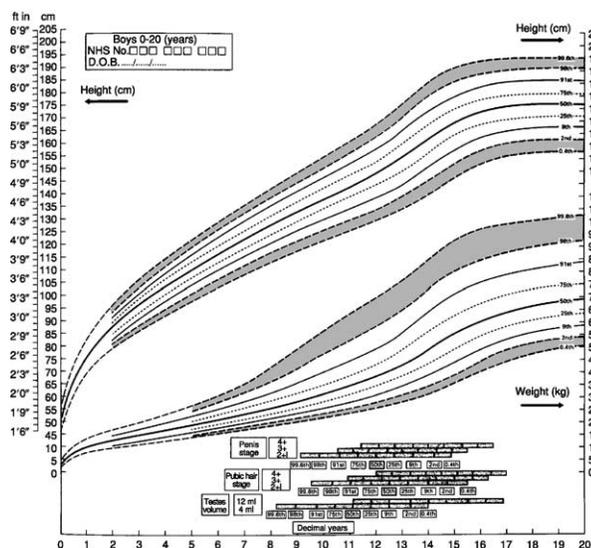


**Figure 2** Median growth velocity curves of British girls and boys whose peak growth spurt is reached at an average age (12 and 14 years, respectively). Reprinted from Preece, M., Ch. 7.1.1 of “The Oxford Textbook of Endocrinology and Diabetes” (2002) (Wass and Shalet, eds.), by permission of Oxford University Press.

process occurs for children with potential for all levels of mature height so that variation in height at any age before maturity may be a combination of varying adult height potential and variation in tempo.

### Population Height Standards

The construction and role of growth standards have previously been reviewed in many texts. Much controversy exists concerning the relative merits of cross-sectional and longitudinal growth references. The former uses data from a large number of children, each of which is measured on one occasion. In contrast, pure longitudinal standards are based on a smaller number of children measured on multiple occasions over many years. It has been argued that the latter are more appropriate for clinical use because they reflect the correct shape of an individual growth curve. In contrast, cross-sectional curves will tend to distort the pattern of individual growth at approximately the age of puberty because of the variability of

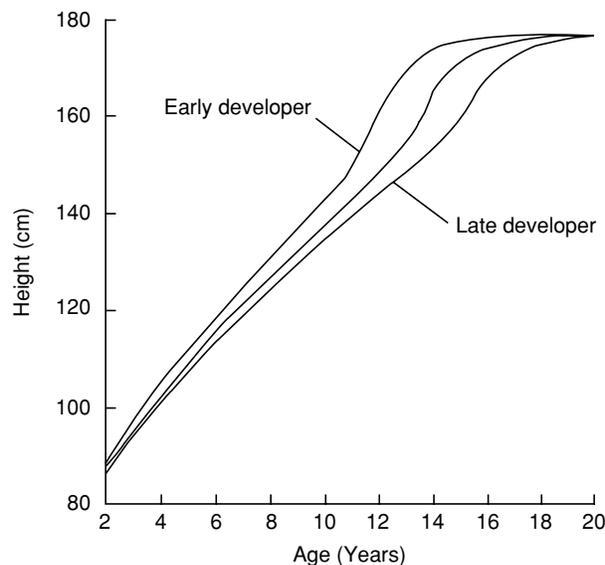


**Figure 3** Growth reference charts for boys in the United Kingdom. Obtained from Harlow Printing Co., Tyne and Wear, UK.

timing of the adolescent growth spurt among individuals. However, the analysis of longitudinal data to provide individual-type growth standards requires a number of assumptions that are not really tenable: a typical assumption is that the growth spurt has essentially the same shape in all children of the same sex, varying only in timing and intensity. A further disadvantage is that the length of time needed to collect longitudinal data makes regular revision difficult. In actual practice, it makes little difference which type of standard is used as long as it is remembered that individuals will probably cross centile lines upward or downward at the time of puberty, needing careful interpretation by the clinician. **Figure 3** shows the latest British height chart for boys, which is derived from cross-sectional data from a nationally representative sample of British children. The centile lines shown are the 0.4th, 2nd, 9th, 25th, 50th, 75th, 91st, 98th, and 99.6th, which are selected to allow a regular distance between adjacent centiles of 0.67 standard deviations. Children whose height falls below the 0.4th centile represent the shortest 4 per 1000 of the population and require immediate investigation. Those between the 0.4th and 2nd centile need careful follow-up and those above the 2nd centile a less intense evaluation.

### Height Velocity Standards

**Figure 5** shows an example of standards for height velocity for boys. The data are based on pure longitudinal

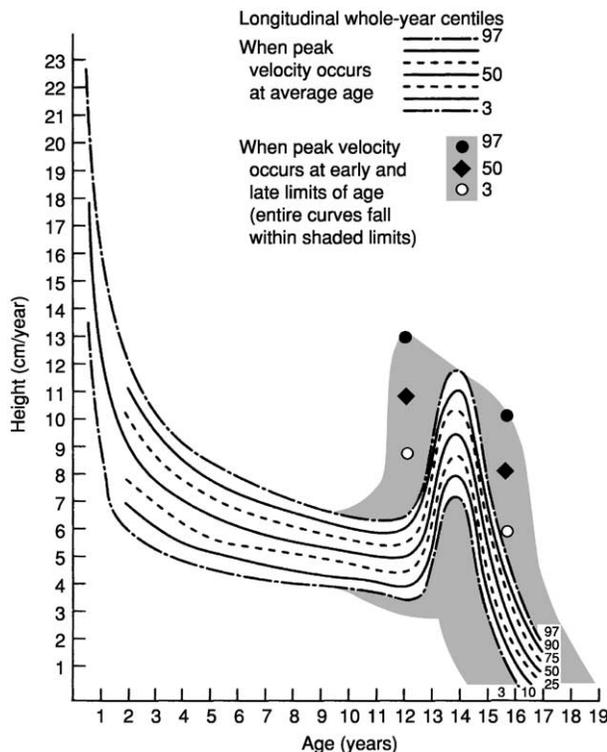


**Figure 4** Schematic representation of the height growth curves of three boys who all have the same mature height, but who are, respectively, 2 years early, average, and 2 years delayed in tempo. Reprinted from Preece, M., Ch. 61 of "Clinical Endocrinology" (1998) (A. Grossman, ed.), 2nd ed. Blackwell Science, Oxford, UK, with permission.

observations and the extremes of normality are indicated by centiles. It is important to make three qualifications. In the first place, these standards are based on data collected before 1966 and might not represent modern children. However, in the nationally representative sample of modern-day children, the *shape* of the average growth curve has changed very little except in the first 2 years of life. This indicates that height velocity at any age over 2 years is also unchanged and in the absence of any more modern data these standards remain approximately valid.

The second qualification is that it must be remembered that these normal values and the position of the various centiles are based on velocity measurements taken over a whole year. This is because if shorter intervals are used, measurement errors assume greater importance and because there are substantial variations in growth velocity during any single year. It is important to interpret cautiously growth velocities in patients when they are based on periods of measurements of less than 1 year.

A final point is that centiles for height velocity do have a slightly different meaning than do height centiles. On average, over a year or more any individual child who is growing normally should have a height velocity somewhere near the 50th centile for the



**Figure 5** Height velocity growth reference for boys in the United Kingdom; see text for explanation. Reprinted from Preece, M., Ch. 61 of “Clinical Endocrinology” (1998) (A. Grossman, ed.), 2nd ed. Blackwell Science, Oxford, UK, with permission.

population. If a child is to maintain his position on, say, the 25th centile for height, he should grow somewhere near the 40th centile for velocity. The child persistently growing at a 10th centile velocity will steadily fall further and further behind his peer group and soon become distinctly short.

In Fig. 5, there are two large gray areas flanking the adolescent growth spurt. The centiles for the latter are those for a child whose growth spurt peaks at an average age, i.e., 14 years for a boy. There is substantial variation in this age, which may be 2 years earlier or later in healthy children; the gray-shaded areas take this into account. In an individual case, the growth spurt should occur within the gray area with a shape similar to the centile centered at 14 years. Note that earlier growth spurts tend to be more intense and later growth spurts less so.

## PUBERTY

### The Events of Puberty

There are many skeletal and soft tissue events that occur during puberty that are important in the development in differences in shape between the adult male

and female. Delineation of these events requires a standardized method of describing progress through puberty and almost universally the systems in use depend on the criteria formalized by Tanner. These depend on rating breast and pubic hair development on a 5-point scale for girls and external genitalia and pubic hair on similar scales for boys. Axillary hair is rated on a 3-point scale for both sexes. Details of the stages are available in many standard texts and will not be repeated here.

In much the same way as variation in height for age was discussed, variation in the age of attainment of different pubertal stages should also be considered. Standards for attainment of puberty stages are usually included on the stature charts, as illustrated in Fig. 3. The designations PH2, PH3, etc., represent the time when attainment of these stages first becomes apparent, though the instant at which this occurs is never observable; the doctor can only see that a child has attained PH2 but not yet PH3. Therefore, the subject is described as being in PH2+ and the standards give the centiles for age of *being in* that stage. Quite arbitrarily, the centiles are ordered such that lower centiles relate to delayed pubertal progress and higher centiles to advanced progress.

### Tempo and the Consonance of Puberty

In general, the events of puberty, involving skeletal, soft-tissue, and secondary sexual characteristics, unfold in an orderly and highly integrated manner. The whole process may be advanced or delayed by several years, but still maintains an internal order that closely follows that of the typical child. Whenever this consonance is disturbed, for example, with an absent growth spurt but normal pubertal progression, then pathology must be suspected. In an otherwise healthy child, this pathology will usually be endocrinological.

## PRINCIPLES OF AUXOLOGICAL ANTHROPOMETRY

### Introduction

In clinical endocrinology, the most useful auxological measurements are height (or supine length in those who cannot stand), sitting height (or crown-rump length), weight, and skinfolds (especially triceps and subscapular). Sitting height is a preferable measurement for the assessment of trunk-to-limb proportions as opposed to the use of measurement of span or upper and lower segments. The first of these is vulnerable to positional problems as the measurement is

significantly affected by rotation of the shoulder girdle and the second depends on accurate placement of the pubic symphysis, which can be difficult, especially in the obese.

The basic rules for making auxological measurements should be the same as for any other endocrine parameters such as hormone measurements. There must be adequately trained technicians using appropriate equipment in a reserved area. Quality assurance is every bit as important although this usually takes the form of repeat measurements on a series of children at appropriate intervals and analysis by appropriate statistical methods.

### Body Composition

Convenient assessment of body composition would be of great clinical value. Direct measurements are difficult or even impossible to obtain in young children and thus indirect methods must be used. In general, these fall into one of three categories: isotope dilution methods, electrophysiological measurements, and anthropometry. Details of these methods and their application are beyond the scope of this article, but they are covered in a number of other sources. The anthropometric methods are among the least accurate, but have the advantage of low cost and simplicity. Generally, they depend on the use of various combinations of height, weight, skinfold thickness, and upper-arm circumference. The techniques for these additional measurements are described briefly.

#### Weight

Weight is an often neglected measurement because of its complex make-up and vulnerability to short-term fluctuations. However, it has the great advantage of being measurable to a high degree of precision and reproducibility if a few basic precautions are taken. The patient should be weighed either nude or in the minimum of standard clothing, which can be corrected for, if necessary. A modern electronic scale should be used.

#### Skinfold Measurements

These measurements are prone to many errors and, of the techniques described here, are the most difficult. However, they are the most convenient method for assessment of subcutaneous fat; extrapolations to total body fat and other measures of body composition are less certain and require care. The two most useful sites for measurement are the triceps and subscapular

areas and a suitable instrument is the Harpenden skinfold caliper

#### Upper-Arm Circumference

This is one of the most straightforward measurements to make and is used widely in nutritional anthropometry. It reflects both subcutaneous fat and muscle bulk, which can be relevant in anthropometric nutritional assessment. The measurement is taken midway between the acromion and the olecranon, while the child is positioned with the arm completely relaxed and extended at the side. The tape must be inextensible and in a plane perpendicular to the long axis of the limb; the tissues must not be compressed.

#### Body Mass Index

Although not strictly a body composition measurement, body mass index (BMI) is increasingly being used as a proxy measurement for obesity or underweight states. It is calculated as  $BMI = [\text{weight (kg)}] / [\text{height (meters)}^2]$ . A number of different national reference charts exist to aid in interpretation.

## INTERPRETATION OF GROWTH DATA

### Reproducibility

#### Sources of Error

The basic requirement of any measuring instrument is that it gives precise and reproducible readings. However, the simple provision of an adequately mounted and well-designed instrument, though essential, does not necessarily solve all the problems associated with taking measurements. The technique of measurement is vital and many errors are introduced by bad practices at this stage. Attention to detail is probably the most important method of limiting error to an irreducible minimum. The errors that are left are intrinsic in the measurement of children, particularly in the younger age groups.

#### Reliability

Good anthropometrists have standard errors of measurement ( $s_{\text{meas}}$ ) ranging from 1.5 to 2.5 mm. This means that a technician with an  $s_{\text{meas}}$  of 2.0 mm (determined, say, over subjects of an age range from 3.0 to 18.0 years) obtains a value for height that 95% of the time is within 4 mm above and 4 mm below the true value.  $s_{\text{meas}}$  can be determined by taking a series of duplicate measurements and calculating the difference  $d$  for each of the subjects. Provided there is no consistent change from first to second occasion (and

there should not be), the  $s_{\text{meas}} = SD_d/\sqrt{2}$ , where  $SD_d$  is the standard deviation of  $d$ .

When the precision of measurements of growth velocity is considered, the problems are much greater. The errors of each of the component measurements are as described above, but the absolute change between them may be quite small and the total error assumes a far greater proportion of the actual measurement.

### Interpretation of Growth and Its Changes

A frequent problem arises as to the best methods of expressing growth variables (height will be taken as a typical example) and of analyzing the changes. While children are within the normal range, transformation of the height to its centile is perfectly acceptable and is often the easiest way to explain the situation to the child and family. However, centile values do not lend themselves to easy statistical analysis and are clearly useless when the subjects are below the lowest recorded centile.

In analyzing growth data, especially in children outside the normal range, the use of standard deviation scores (SDS) is more appropriate. In principle, any variable can be so transformed, provided the mean value and the standard deviation for each age and sex are known and the variable assumes a Gaussian distribution. The width of the age band used for calculating SDS should be quite narrow, ideally 6 months for children between the ages of 2 and 10 years and 3 months for children under 2 years and for those between the ages of 10 and 16 years. The SDS is calculated from the relationship  $SDS = (X - X_i)/SD_i$ , where  $X$  is the actual measurement, and  $X_i$  and  $SD_i$  are the mean and standard deviation for the  $i$ th age band of the appropriate sex. The SDS can be treated as a variable independent of age and sex. In a normal population, the mean SDS will have a value of 0 and a standard deviation of 1; the 95% confidence limits will be  $-2$  and  $+2$  SDS. This transformation is made more powerful by extension to data with a non-Gaussian distribution, such as weight.

When SDS are applied to a population of subjects with delayed growth at approximately the time of normal puberty, a problem can arise. As subjects in the normal population of adolescents commence their growth spurt (e.g., males at 12 years), the mean height increases rapidly. This is an inappropriate mean to use for calculating a patient's SDS, if he is prepubertal, and if SDS are calculated, they will be artificially more negative than they should be. Some researchers advocate using the appropriate mean for skeletal age. In

simple pubertal delay, this is probably acceptable, although it should be remembered that the link between skeletal age and the timing of puberty is not exact. In many chronic diseases, especially when corticosteroids are used, this relationship is even weaker and skeletal age seldom resolves the problem. The best approach would be to create values for the mean heights of children with various degrees of growth delay. Using Swedish data, Johan Karlberg has attempted a formal development of this approach and created charts that go some way toward providing this type of curve, but only within the variation seen in healthy children.

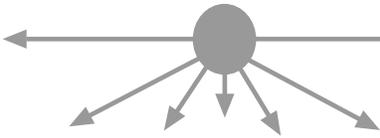
There is a further caution to be exercised with velocity measurements, whether expressed as centimeters per year or as SDS. All the reference data are based on velocities measured over a whole year; strictly, this should apply in the case of test subjects, as shorter time periods will be associated with larger relative errors and therefore larger SD at any particular age. In many situations, whole-year velocities are not practical, particularly for children with complex diseases where the course of the illness is unstable. In this case, shorter time intervals may be required, but interpretation of results requires caution. Periods of less than 6 months are probably valueless.

### See Also the Following Articles

Body Composition During Growth • Body Proportions • Growth, Normal Patterns and Constitutional Delay • Intra-uterine Growth Retardation • Postnatal Normal Growth and Its Endocrine Regulation • Puberty: Physical Activity and Growth • Short Stature and Chromosomal Abnormalities • Skeletal Development During Childhood and Adolescence

### Further Reading

- Cameron, N. (1984). "The Measurement of Human Growth." Croom Helm, London, UK.
- Cameron N. (ed.). (2002). "Human Growth and Development." Academic Press, San Diego, CA.
- Cole, T. J., Freeman, J. V., and Preece, M. A. (1998). British 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood. *Statist. Med.* **17**, 407–429.
- Dattani, M. T., and Preece, M. A. (1998). Physical growth and development. In "Forfar and Arneil's Textbook of Pediatrics" (A. G. M. Campbell and N. McIntosh, eds.), 5th ed., pp. 349–380. Churchill Livingstone, Edinburgh, UK.
- Preece, M. A., and Cox, L. A. (1997). Growth and bone age. In "Assessment Methodology in Orthopaedics" (P. B. Pynsent, J. C. T. Fairbank, and A. J. Carr, eds.), pp. 167–182. Butterworth Heinemann, Oxford, UK.



# Bardet-Biedl Syndrome

Philip L. Beales

University College London, London, United Kingdom

## Glossary

**brachydactyly** Short broad fingers or toes.

**hydrometrocolpos** An enlarged fluid-filled uterine cavity, often secondary to a nonpatent hymen.

**nyctalopia** Night blindness.

**orthologue** One of a set of homologous genes in different species.

**postaxial** Lateral to the mid-axis of the limb.

**proband** The affected index case in a pedigree.

**syndactyly** Pronounced webbing between digits.

**B**ardet-Biedl syndrome is an inherited disorder that includes the following cardinal features: retinal dystrophy, obesity, polydactyly, cognitive impairment, hypogonadism, and renal anomalies.

## INTRODUCTION

### Eponyms

During recent times, a good deal of confusion has arisen over the naming of what has come to be called Bardet-Biedl syndrome (BBS). Laurence-Moon-Biedl syndrome, Laurence-Moon-Bardet-Biedl syndrome, and (more recently) Bardet-Biedl syndrome all have been applied. In 1866, Laurence and Moon described members of an English family with retinitis pigmentosa (RP), mental retardation, obesity, and male hypogonadism. More than 55 years later, Bardet and Biedl reported cases with the triad of retinal dystrophy, polydactyly, and obesity. Since the 1980s, it has been suggested that the condition described by Laurence and Moon is distinct from BBS in that they did not report the presence of polydactyly in their family and that affected members of this family later went on to develop spastic lower limb signs, thereby giving birth to a separate entity, the Laurence-Moon

syndrome (LMS). The label “Bardet-Biedl syndrome” is the current preference despite very little evidence to support this division. Mutational analyses will resolve this issue in time.

### Prevalence

BBS has been described in most populations, although with differing frequencies. In populations of North European descent, rates range from 1 in 100,000 to 1 in 160,000. Among certain Arab populations, the syndrome is twice as common (1 in 65,000), and it is at least 10-fold more frequent in the Newfoundland population (1 in 17,500) owing to a founder effect.

## NATURAL HISTORY AND DIAGNOSIS

Despite recent advances in the genetics of BBS, diagnosis remains clinical for the most part. The phenotype is emerging in that often the only discernible feature at birth may be postaxial polydactyly (up to 30% do not have polydactyly). Most affected infants are of normal birthweight and have normal vision until 7 or 8 years of age, when nyctalopia may ensue. A steady decline in visual acuity progresses over the subsequent two decades of life, with most patients being registered as fully blind by the time they reach their 20s. The cardinal features of BBS are a rod-cone dystrophy, central obesity, dystrophic extremities (including postaxial polydactyly), varying degrees of cognitive impairment, male hypogonadism (and female genital tract anomalies), and structural and functional renal abnormalities. Four of six features have been considered enough for a positive diagnosis. However, in recognition of many secondary features and to facilitate early diagnosis, modified criteria have been proposed; either four primary features (of the six cardinal signs) or three primary and two secondary features should be present (Table I). Among the secondary features are diabetes mellitus (type 2), hypertension, speech and behavioral abnormalities,

**Table 1** Diagnostic Criteria for Bardet–Biedl Syndrome**Primary features**

Retinal dystrophy  
 Polydactyly  
 Obesity  
 Learning disabilities/Mental retardation  
 Male hypogonadism  
 Renal anomalies

**Secondary features**

Speech disorder/delay  
 Strabismus/cataracts/astigmatism  
 Brachydactyly/Syndactyly  
 Developmental delay  
 Nephrogenic diabetes insipidus  
 Ataxia/Poor coordination/Imbalance  
 Diabetes mellitus  
 Dental anomalies  
 Heart disease (congenital)/Hypertension  
 Hepatic fibrosis

*Source:* Beales, P. L., Elcioglu, N., Woolf, A. S., Parker, D., and Flinter, F. A. (1999). New criteria for improved diagnosis of Bardet–Biedl syndrome: Results of a population survey. *J. Med. Genet.* 36, 437–446.

*Note.* Four primary features or three primary plus two secondary features should be present.

endocrinopathies, atresia ani, Hirschsprung disease, and gait anomalies. There appears to be as much intrafamilial variation as there is interfamilial variation in the expression of the clinical features, and no fastidious phenotype–genotype correlations can be claimed.

## MAJOR CLINICAL FEATURES

### Ophthalmologic Findings

For the majority of patients, visual acuity is normal at birth and throughout infancy. Most children will attend school with near-perfect vision, only to experience a gradual decline over the ensuing two to three decades. Despite the presence of early visual disturbance such as nyctalopia, fundoscopic changes remain minimal with some vessel attenuation. The classical “bony corpuscular” distribution characteristic of RP is usually a late manifestation; therefore, normal fundoscopy should not be regarded as synonymous with normal function. Electroretinogram (ERG) remains the cornerstone of investigation but is often difficult to perform in very young children and may be normal in children under 7 years of age. Extraretinal features, such as horizontal nystagmus, high myopia,

**Figure 1** Postaxial polydactyly in the foot.

strabismus, diabetic retinopathy, cataracts, and glaucoma, should not be overlooked.

### Dystrophic Extremities

Polydactyly is almost always postaxial in distribution, ranging from a fully formed and functional extra digit to a vestigial nubbin (Fig. 1). Likewise, the number of limbs involved is variable. Probably even more common is the presence of brachydactyly (more often in the feet than in the hands) in the majority of patients, and syndactyly frequently presents between the second and third toes.

### Obesity

Weight gain is significant in BBS, usually beginning within the first 2 years of life, following weaning. There is little evidence to support an abnormality in basal metabolic rate, and a combination of excess intake with potentially reduced activity levels may contribute to this aspect of the phenotype. Body mass index (BMI) in BBS adults usually lies between 30 and 39, with a minority constantly above 40. Adiposity is generalized.

### Cognitive Impairment

There is a range of conflicting reports regarding the degree of intellectual impairment prevailing in BBS. It is clear that there is a great deal of variation, with a minority displaying severe mental retardation and the majority showing some form of learning difficulty to the extent that special educational arrangements are required. There are anecdotal reports of autistic tendencies in some BBS children, and behavioral traits

(e.g., obsessiveness, disinhibition) abound but have not been objectively corroborated. A study assessing behavior and IQ in 21 BBS children reported that only 3 children (17%) had full-scale IQs in the average range (>79) and that 11 (52%) fell within the mental retardation range (<70); nonetheless, the majority were mildly intellectually impaired.

### Genital and Reproductive Anomalies

Hypogonadism among males has been well described and may also be accompanied by infertility. The case among female patients is less clear, although there are many examples of pregnancy leading to viable birth. A number of reports of genital malformations suggest that these are a common manifestation of BBS and include hypoplastic fallopian tubes, uterus, and ovaries; partial and complete vaginal atresia; septate vagina; duplex uterus; hematocolpos; persistent urogenital sinus; vesico-vaginal fistula; absent vaginal orifice; and absent urethral orifice. Hydrometrocolpos is one of the three features of McKusick-Kaufman syndrome (MKS) in which the underlying gene (*MKKS/BBS6*) has been shown to harbor mutations common to both MKS and BBS.

### Renal Dysfunction

Until the 1980s, the renal component of BBS had been reported only infrequently, although a high frequency of structural abnormalities was observed at postmortem.

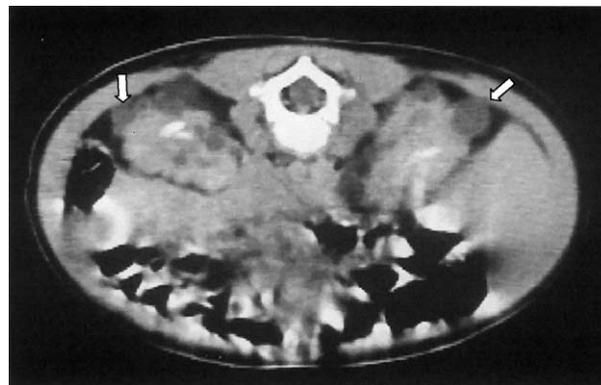
#### Structural Changes

Renal dysplasia can be present without clinical evidence of renal disease; thus, careful imaging of the urinary tract is warranted. Among the renal structural anomalies encountered are cystic spaces communicating with the collecting system, parenchymal cysts (Fig. 2), persistence of fetal lobulation, small kidneys with parenchymal scarring, prolonged nephrogram, and blunting and clubbing of the calyces. Persistent fetal lobulation has been described in isolation, and it has been suggested that this may reflect a defect in maturation.

Ultrasound scanning and intravenous pyelography are the primary investigations of choice.

#### Functional Renal Impairment

Chronic renal failure has been reported in several studies and ranges from 30 to 100% of the cohort described. Many end-stage patients have undergone



**Figure 2** Abdominal computed tomography (CT) scan depicting polycystic kidneys (arrow) in a young male with BBS.

renal transplantation successfully, although no studies of long-term follow-up have been reported.

### Minor Diagnostic Features

The recent recognition of several less well-characterized features has helped to secure a diagnosis particularly in young patients, where retinal degeneration might not have emerged at the time of examination.

### Endocrine Disturbance

One study of 38 patients reported that 2 were insulin dependent, 4 were maintained on oral hypoglycemic agents, and 6 were managed with diet alone. The researchers concluded that 25% of BBS patients were diabetic by 35 years of age and that 50% were diabetic by 55 years of age. The underlying cause is unknown, and there have been no reports of abnormal pancreas at postmortem. It is likely that diabetes mellitus is secondary to insulin resistance.

Other reported endocrinopathies include hypothyroidism, hypopituitarism, and gonadotropin and testosterone deficiency. Diabetes insipidus is common and is mainly nephrogenic in origin.

### GENETICS

BBS has emerged as one of the most heterogeneous monogenic conditions known, with evidence for at least eight loci: *BBS1* on 11q13, *BBS2* on 16q21, *BBS3* on 3p12, *BBS4* on 15q22.2–q23, *BBS5* on 2q31, *BBS6* on 20p12, *BBS7* on 4q27, and *BBS8* on 14q32 (all mapped loci, with the exception of *BBS3* and *BBS5*, have been cloned as of September 2003). However, the cloning of the first BBS locus, *BBS6*,

followed by mutational analyses on a large multiethnic cohort implied that some mutations do not conform to a traditional model of autosomal recessive disease transmission.

### Multiallelic Inheritance in BBS

Following the identification of two BBS loci, *BBS2* and *BBS6*, sequence analyses showed that, in some families, a total of three mutations in two genes are necessary for pathogenesis, thereby establishing a digenic triallelic model of disease transmission and suggesting that BBS might be a useful model to study oligogenic traits.

### BBS Proteins

The search for conceptual homologues using corresponding nucleotide and amino acid sequence from each of the BBS genes has yielded only modest similarities. The closest orthologue to *BBS6/MKKS* is that of the archaeobacterial thermosome or chaperonin-containing T-complex polypeptide (CCT), members of the group of type II chaperonins. These are thought to have a role in facilitating the correct folding of nascent peptides into their functional configuration, and perhaps the BBS phenotype results from misfolding of *BBS6* peptides.

*BBS1*, -2, -4, and -7 all are representative of novel peptides without any chaperonin homology, but many have domains with multiple tetratricopeptide (TPR) repeats that are probably important as for protein-protein interactions. *BBS4* has similarity to O-linked N-acetyl glucosamine transferase (OglcNac).

### Clinical Genetics

#### Recurrence Risks

For the majority of families with BBS probands, the recurrence risk for another affected child will be that for an autosomal recessive disorder, that is, 1 in 4. The risk that any sibling may be a carrier is 2 in 3. Less than 10% of all cases exhibit multiallelic inheritance. If such a case is proven, then the recurrence risk must be modified accordingly; that is, in a triallelic pedigree, the risk of further affected offspring will theoretically fall to 1 in 8, although such risks have yet to be substantiated.

### Prenatal Diagnosis

Until recently, prenatal diagnosis was inaccurate and limited to ultrasound scanning for potential anomalies such as postaxial polydactyly and renal anomalies in those families at risk. With the advent of screening, *in utero* mutation detection can be offered at amniocentesis or by chorionic villus sampling provided that the responsible mutation(s) has been identified in the parents. Nevertheless, given the large number of genes that require screening (not to mention those that have not yet been cloned), it is not surprising that most diagnostic laboratories do not offer this service.

### See Also the Following Article

Kallmann's Syndrome and Idiopathic Hypogonadotropic Hypogonadism

### Further Reading

- Badano, J. L., Ansley, S. J., Leitch, C. C., *et al.* (2003). Identification of a novel Bardet-Biedl syndrome protein, *BBS7*, that shares structural features with *BBS1* and *BBS2*. *Am. J. Hum. Genet.* **72**, 650-658.
- Beales, P. L., Elcioglu, N., Woolf, A. S., Parker, D., and Flinter, F. A. (1999). New criteria for improved diagnosis of Bardet-Biedl syndrome: Results of a population survey. *J. Med. Genet.* **36**, 437-446.
- Green, J. S., Parfrey, P. S., Harnett, J. D., *et al.* (1989). The cardinal manifestations of Bardet-Biedl syndrome, a form of Laurence-Moon-Bardet-Biedl syndrome. *N. Engl. J. Med.* **321**, 1002-1009.
- Katsanis, N., Ansley, S. J., Badano, J. L., *et al.* (2001). Triallelic inheritance in Bardet-Biedl syndrome, a Mendelian recessive disorder. *Science* **293**, 2256-2259.
- Katsanis, N., Beales, P. L., Woods, M. O., *et al.* (2000). Mutations in *MKKS* cause obesity, retinal dystrophy, and renal malformations associated with Bardet-Biedl syndrome. *Nat. Genet.* **26**, 67-70.
- Katsanis, N., Lupski, J. R., and Beales, P. L. (2001). Exploring the molecular basis of Bardet-Biedl syndrome. *Hum. Mol. Genet.* **10**, 2293-2299.
- Mykytyn, K., Braun, T., Carmi, R., *et al.* (2001). Identification of the gene that, when mutated, causes the human obesity syndrome *BBS4*. *Nat. Genet.* **28**, 188-191.
- Mykytyn, K., Nishimura, D. Y., Searby, C. C., *et al.* (2002). Identification of the gene (*BBS1*) most commonly involved in Bardet-Biedl syndrome, a complex human obesity syndrome. *Nat. Genet.* **4**, 435-438.
- Nishimura, D. Y., Searby, C. C., Carmi, R., *et al.* (2001). Positional cloning of a novel gene on chromosome 16q causing Bardet-Biedl syndrome (*BBS2*). *Hum. Mol. Genet.* **10**, 865-874.
- Schachat, A. P., and Maumenee, I. H. (1982). Bardet-Biedl syndrome and related disorders. *Arch. Ophthalmol.* **100**, 285-288.
- Slavotinek, A. M., Stone, E. M., Heckenlively, J. R., *et al.* (2000). Mutations in *MKKS* cause Bardet-Biedl syndrome. *Nat. Genet.* **26**, 15-16.



# Baroreceptor Responses

Edward M. Stricker and Alan F. Sved  
*University of Pittsburgh, Pittsburgh, Pennsylvania, United States*

## Glossary

- bradycardia** Decreased heart rate.
- hypotension** Reduced arterial blood pressure.
- hypovolemia** Reduced blood volume.
- tachycardia** Increased heart rate.

Given the importance of the cardiovascular system, it is not surprising that elaborate mechanisms have evolved to promote the stability of blood volume and arterial blood pressure. These mechanisms include well-known physiological responses that restore blood volume and blood pressure after blood loss with or without accompanying decreases in arterial blood pressure. Those responses include adaptive cardiovascular responses, autonomic reflexes, endocrine secretions, and ingestive behaviors motivated by thirst and salt appetite.

## INTRODUCTION

Animals must be able to detect changes in blood volume and blood pressure for any compensatory responses to occur. Insight into the location of the receptors and their properties comes from a consideration of how changes in blood volume and blood pressure affect the cardiovascular system. For example, after a moderate hemorrhage, the blood loss is not distributed uniformly throughout the cardiovascular system but instead is borne nearly entirely on the venous side of the circulation. Thus, the veins, which are very distensible, collapse around the remaining blood and thereby help to reduce the impact of the blood loss on venous blood pressure. Together with an appropriate increase in heart rate, arterial blood volume and blood pressure undergo little change. Thus, it makes sense that receptors are situated to monitor blood volume on the venous side of the circulation. In contrast, after severe hemorrhage, the effects of blood loss may extend to the arterial side of the circulation and reduce arterial blood volume and blood pressure. Thus, it also makes sense that

additional receptors are situated to monitor blood pressure on the arterial side of the circulation. In fact, the receptors that monitor the stretch of venous and arterial blood vessels are known to be located at key sites in the cardiovascular system. One such site is at the junction of the inferior vena cava and the right atrium. Another site is the aortic arch through which arterial blood leaving the heart is carried. A third site is the carotid artery through which blood flows to the brain.

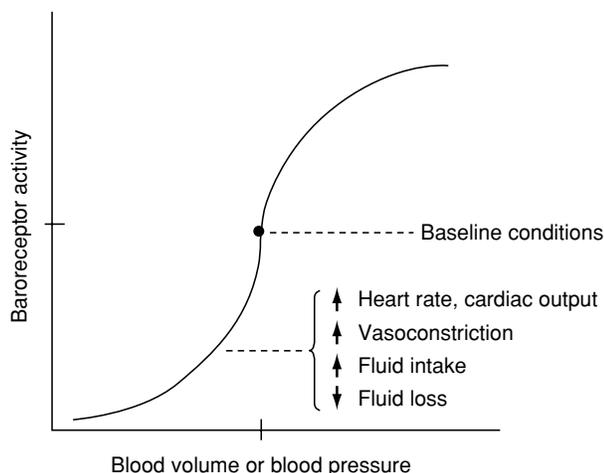
The receptor cells in question are called “baroreceptors,” which literally means that they are responsive to changes in pressure. However, the sensory endings actually are stretch receptors, located on the outside walls of the blood vessels and cardiac atria, that respond to changes in the conformation of the vessel walls much like other stretch receptors on the outside walls of the stomach and bladder detect the fullness of those organs. Thus, following moderate hemorrhage, the stretch receptors on the low-pressure side of the circulation decrease their firing rate as the loss in blood volume decreases the distension of the right atrium and adjacent vena cava. Following a more severe hemorrhage, the stretch receptors on the high-pressure side of the circulation also detect the loss in arterial blood pressure as a decrease in the distension of the proximal aorta and the carotid sinus.

## RESPONSES

Neural afferents from these low- and high-pressure baroreceptors project as part of the cranial nerves IX and X to the nucleus of the solitary tract (NST) in the caudal brainstem. From there, complex neural circuitry enables the variety of responses that occur in response to decreases in blood volume or arterial blood pressure (Fig. 1). The most prominent responses are discussed in the following subsections.

### Cardiac Reflexes

Changes in heart rate help to compensate for changes in blood volume and blood pressure. For example, an



**Figure 1** Baroreceptor-evoked responses to hypovolemia and hypotension. The activity of veno-atrial baroreceptors and the activity of arterial baroreceptors are related to changes in blood volume and blood pressure, respectively. Neural and endocrine baroreceptor reflex responses to decreases in blood volume or blood pressure produce increases in heart rate and cardiac output, increases in vasoconstriction, and decreases in fluid excretion by the kidneys. In addition, baroreceptor input to the brain elicits ingestive behaviors to increase body fluid volume. In general, increases in blood volume and blood pressure elicit baroreceptor-evoked responses that are the converse of those seen with decreases in these parameters.

increase in heart rate occurs during hypovolemia or hypotension and is one of the well-known “baroreceptor reflexes.” This response appears to reflect the three autonomic effects that result from baroreceptor input to the NST. The first effect involves the control of sympathetic neural activity; this accelerates heart rate and increases cardiac contractility when stimulated during volume contraction or decreases in arterial pressure. The second is the converse effect on parasympathetic neural activity; this also causes tachycardia. The third involves stimulation of the adrenal medulla and the release into the systemic circulation of epinephrine; this has a prominent effect on beta-adrenergic receptors on the heart and thereby further stimulates increased heart rate. The multiple controls of heart rate ensure that these responses are rapid and precise.

### Vasoconstriction

Another baroreceptor response to hemorrhage involves the release of norepinephrine from sympathetic postganglionic fibers that innervate vascular smooth muscle, resulting in activation of the alpha-adrenergic receptors on the arterioles that carry blood to various tissues, most prominently abdominal viscera and

muscle, and constriction of those vessels. Vasoconstriction at these sites causes arterial blood to be pooled in the central circulation, thereby enabling perfusion of the major organs to be maintained while compromising blood flow to less critical tissue. Activation of sympathetic neural input to veins causes venoconstriction, thereby increasing venous return of blood to the heart and cardiac output.

### Renin–Angiotensin System

Sympatho-adrenal activity stimulates renin secretion from the kidneys, beginning a cascade of biochemical events that result in the generation of the peptide hormone angiotensin II, the most potent pressor agent available to animals in response to blood loss. Briefly, renin is a proteolytic enzyme that is synthesized in renal juxtaglomerular cells and is released in response to increased activation, either by sympathetic nerves or by circulating epinephrine, of beta-adrenergic receptors that are located on the cell walls. Renin acts by cleaving four amino acids from a 14-amino acid protein (angiotensinogen) that is synthesized in the liver and is always present in the circulation. The resultant 10-amino acid protein (angiotensin I) then quickly loses 2 more amino acids through the action of an enzyme that converts angiotensin I to angiotensin II. This enzyme is heavily concentrated in the lungs, an arrangement that allows the hormone to be delivered into the arterial blood that leaves the heart (as opposed to renin, which enters the general circulation through the renal veins and has no direct vasomotor effects).

Circulating angiotensin II acts directly on angiotensin II receptors in vascular smooth muscle to support blood pressure. In addition, it acts in the adrenal cortex to stimulate secretion of the mineralocorticoid aldosterone, which promotes urinary sodium ( $\text{Na}^+$ ) retention (and, secondarily, fluid retention) as well as urinary potassium ( $\text{K}^+$ ) excretion; this is especially useful when blood loss is associated with tissue damage. Angiotensin II also acts on angiotensin receptors in the brain, specifically in the subfornical organ (SFO). The SFO, which is situated dorsal to the third cerebral ventricle, lacks a blood–brain barrier and, therefore, can detect variations in blood levels of the hormone. Neural circuits from the SFO then appear to stimulate central systems that increase sympatho-adrenal activity. They also stimulate secretion of vasopressin, another hormone with pressor properties, from the posterior lobe of the pituitary gland.

## Vasopressin

The release of pituitary vasopressin is stimulated by neural projections from the NTS to the supraoptic and paraventricular nuclei of the hypothalamus in response to decreased baroreceptor input during hemorrhage. Vasopressin levels in blood of only 6 to 8 pmol (i.e., 6–8 pg/ml) are required for maximum antidiuresis. Effective pressor levels of vasopressin are substantially higher (50–60 pg/ml), and those elevated levels are obtained following hemorrhage or arterial hypotension. Pituitary secretion of the other neurohypophyseal hormone, oxytocin, parallels the secretion of vasopressin in response to these stimuli. The contribution of oxytocin to the regulation of blood volume and arterial blood pressure in rats was revealed recently by observations that it constituted yet another significant stimulus of renin secretion.

## Pituitary–Adrenal Axis

Neural projections from the NST to the paraventricular nuclei of the hypothalamus also stimulate release of corticotropin-releasing factor (CRF), and this elicits secretion of adrenocorticotrophic hormone (ACTH) from the anterior lobe of the pituitary gland. ACTH, in turn, stimulates the adrenal cortex to increase its synthesis and release of two steroid hormones: aldosterone and the glucocorticoid cortisol. Cortisol improves vascular reactivity to the catecholamines and other general responses to stress.

## Ingestive Behavior

In addition to these physiological responses, hypovolemia is known to stimulate thirst and (in rats but not humans) salt appetite. Remarkably, the resultant ingestion of water and sodium chloride (NaCl) solution occurs in amounts equivalent to a fluid mixture that is isotonic to plasma. Baroreceptor input to the brainstem is presumed to provide the main signal of thirst under hypovolemic conditions, although angiotensin II is a known dipsogen and presumably contributes stimulation as well. The signal of salt appetite is less clear and is controversial, although an important role of angiotensin II seems to be apparent. Thirst (but not salt appetite) also is stimulated by arterial hypotension in rats, but the induced water intake is mediated exclusively by angiotensin II, and baroreceptors do not appear to provide a direct neural signal.

In addition to these various responses stimulated by cardiac and arterial baroreceptors, it may be noted that renin secretion is stimulated by “renal baroreceptors.”

Despite its name, this latter mechanism does not resemble the other baroreceptors in that neural input to and from the central nervous system is not involved. Instead, a decrease in renal perfusion, associated with systemic hypovolemia or arterial hypotension, reduces the fullness of the distensible renal arterioles. The cells that synthesize and secrete renin line the walls of those blood vessels; therefore, decreased renal perfusion produces conformational changes in the cells, and this stimulates renin secretion. Finally, it has been proposed that decreased renal perfusion also provides a neural signal of hypovolemia and arterial hypotension, mediated by renal afferents projecting to the spinal cord; this might contribute additional stimulation of the compensatory physiological and behavioral responses.

## CONCLUSION

Many adaptive responses occur during hypovolemia and arterial hypotension that appear to involve signals originating in the activity of cardiovascular baroreceptors. These responses include several chemical agents that independently provide vasoconstriction and support of arterial blood pressure, hormone secretions that promote urinary conservation of water and Na<sup>+</sup>, and the sensations of thirst and salt appetite that lead to water and NaCl consumption. The precise stimulus of each response is not always certain, but it may involve neural signals, bloodborne signals, or both. This redundancy of stimuli is a common feature of homeostatic regulatory systems; it is adaptive because it allows compensation for damage- or disease-induced impairments of individual components of the system. On the other hand, this redundancy makes it very difficult to identify all of the contributing signals and effector systems that participate in homeostasis as well as their importance relative to one another.

Although this article has focused on the baroreceptor responses seen after reductions in blood volume and blood pressure, it is important to recognize that many, but not all, of these same variables are influenced by increases in blood volume and blood pressure. For example, a reflexive bradycardia is signaled by low-pressure baroreceptors during hypervolemia and by high-pressure baroreceptors during hypertension. These effects result when sympathetic fibers to the heart are inhibited while parasympathetic fibers are stimulated. Other baroreceptor responses to hypervolemia and hypertension are generally analogous, although some inconsistencies are evident. Thus, vasopressin secretion is decreased when blood volume is expanded but not when blood pressure is elevated.

Conversely, thirst is decreased when arterial blood pressure is elevated acutely but not when blood volume is expanded. In short, baroreceptors mediate a variety of adaptive autonomic, endocrine, and behavioral responses to both decreases and increases in blood volume and blood pressure.

### See Also the Following Articles

Aldosterone in Congestive Heart Failure • Atherogenesis • Atherosclerosis • Carbenoxolone • Catecholamines • Tissue Renin-Angiotensin-Aldosterone System

### Further Reading

- Gauer, O. H., and Henry, J. P. (1963). Circulatory basis of fluid volume control. *Physiol. Rev.* **43**, 423–481.
- Stricker, E. M., and Sved, A. F. (2000). Thirst. *Nutrition* **16**, 821–826.
- Stricker, E. M., and Sved, A. F. (2002). Controls of vasopressin secretion and thirst: Similarities and dissimilarities in signals. *Physiol. Behav.* **77**, 731–736.
- Stricker, E. M., and Verbalis, J. G. (2002). Fluid intake and homeostasis. In “Fundamental Neuroscience” (L. R. Squire, F. E. Bloom, J. L. Roberts, M. J. Zigmond, S. K. McConnell, and N. C. Spitzer, eds.), 2nd ed., pp. 991–1009. Academic Press, San Diego.



# Bartter's Syndrome

I. Lenga and K. S. Kamel  
 University of Toronto, Toronto, Ontario, Canada

## Glossary

**hypocalciuria** A low rate of calcium excretion in the urine, usually defined as the excretion of less than 0.1 mmol of calcium per millimole of creatinine in the urine.

**hypokalemia** A serum potassium level that is lower than 3.5 mmol/L.

**hypomagnesemia** A serum magnesium level that is lower than 0.7 mmol/L.

**metabolic alkalosis** An arterial blood pH that is higher than 7.40 due to an elevated plasma bicarbonate concentration.

**B**artter's syndrome and Gitelman's syndrome are autosomal recessive tubulopathies characterized by hypokalemia with renal potassium wasting, metabolic alkalosis, and normal or low blood pressure despite hyperreninemia and secondary hyperaldosteronism.

## INTRODUCTION

More than 40 years ago, Bartter's syndrome and Gitelman's syndrome of inherited hypokalemic metabolic alkalosis were described. Yet despite numerous studies and hundreds of publications, their pathophysiology remained enigmatic. Recently, the recognition of distinct phenotypes within this heterogeneous group of patients has permitted the application of powerful molecular genetic methods to define the primary basis of these disorders. Although these findings have significantly improved our understanding of renal physiology and may have implications for the development of novel therapeutic agents, a number of questions remain to be answered. This article presents a concise overview of the topic, highlighting areas where progress has been made and also pointing out issues that remain to be explored.

## EPIDEMIOLOGY

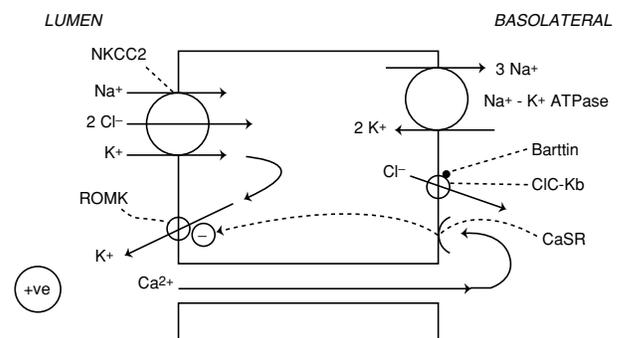
Bartter's and Gitelman's syndromes are rare diseases that exhibit autosomal recessive inheritance. There is no known racial predilection, and both genders seem to be affected equally. Gitelman's syndrome is more common, although its true prevalence is unknown.

## MOLECULAR LESIONS

### Bartter's Syndrome

The fundamental defect in Bartter's syndrome is tubular dysfunction in the thick ascending loop of Henle (TAL) (Fig. 1).  $\text{Na}^+$  reabsorption in this nephron segment occurs via the electroneutral furosemide-sensitive  $\text{Na}^+/\text{K}^+/2 \text{Cl}^-$  cotransporter (NKCC2). Recycling of  $\text{K}^+$  through an apical membrane  $\text{K}^+$  channel (ROMK) provides the  $\text{K}^+$  for continued operation of the transporter and also generates the positive luminal voltage that drives the paracellular reabsorption of  $\text{Na}^+$  and of the divalent cations,  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$ .  $\text{Cl}^-$  exits the cells via a basolateral  $\text{Cl}^-$  channel,  $\text{ClC-Kb}$ .

Loss of function mutations have been described in these three ion channels. Another variant, type IV



**Figure 1** Molecular lesions causing Bartter's syndrome. Schematic presentation shows transport mechanisms in the thick ascending limb of the loop of Henle. Defects at five sites that can lead to diminished reabsorption of  $\text{Na}^+$  and  $\text{Cl}^-$ , and thereby Bartter's syndrome, are illustrated.

Bartter's syndrome, has also been described. It is caused by mutation in BSND, the gene that encodes for a protein called barttin, an essential  $\beta$ -subunit of the ClC-Kb channel. It is associated with sensorineural deafness because barttin is also expressed in the  $K^+$ -secreting dark cells of the inner ear.

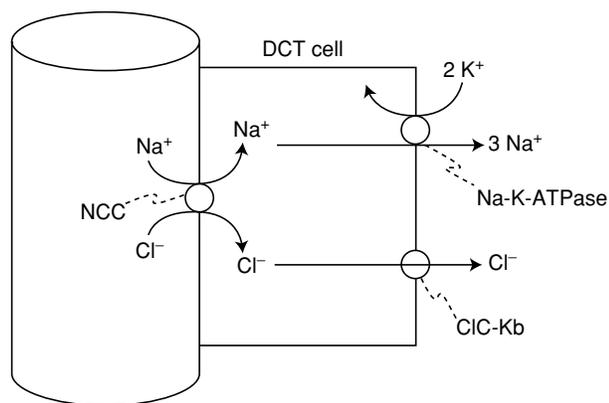
$Ca^{2+}$  binds to the calcium-sensing receptor (CaSR) expressed in the basolateral membrane on the cells of the TAL. This leads to the generation of an arachidonic acid metabolite that inhibits ROMK and hence causes a loop diuretic-like effect. A Bartter's-like syndrome has also been described in severe cases of autosomal dominant hypocalcemia. The genetic defect in this case is an activating mutation (gain of function) in the gene encoding for CaSR.

### Gitelman's Syndrome

Gitelman's syndrome is usually caused by a loss of function mutation in the thiazide-sensitive sodium chloride cotransporter (NCC) in the distal convoluted tubule (DCT) (Fig. 2). In addition, mutations in other transporters in the DCT can indirectly cause functional impairment of the NCC, leading to similar phenotypes. Mutations have been documented in the ClC-Kb of the DCT. Mutations in the  $\gamma$ -subunit of the basolateral Na-K-ATPase have been found in isolated patients with hypomagnesemia and hypocalciuria.

### CLINICAL FEATURES

The phenotypic expression of Bartter's syndrome is highly variable. Types I and II tend to present during the neonatal period, whereas type III typically



**Figure 2** Molecular disorders causing Gitelman's syndrome. The most common defect causing Gitelman's syndrome is a mutation involving NCC, but others include a lesion in the  $\gamma$ -subunit of the Na-K-ATPase or in  $Cl^-$  channels on basolateral membranes in this nephron segment (ClC-Kb).

presents during infancy or childhood. Polyuria and polydipsia with clinical extracellular fluid (ECF) volume contraction are very common. Muscle weakness and cramping are nearly universal. Other complaints may include constipation, vomiting, and salt craving. There may be growth retardation and neurocognitive impairment to a variable degree. Nephrocalcinosis occurs in some patients but is notably absent in type III Bartter's syndrome. Sensorineural hearing loss is a unique feature of type IV Bartter's syndrome.

Neonatal Bartter's syndrome is the most severe form of the condition. Maternal polyhydramnios and premature labor are common due to excessive urine production *in utero*. Infants have characteristic triangular facies, with a drooping mouth, large eyes, and large pinnae. Polyuria, polydipsia, failure to thrive, and growth retardation are common findings.

Gitelman's syndrome presents during late childhood or even during adulthood, and at times it is discovered incidentally during routine blood work. Patients often complain of muscle cramps and may also present with tetany. Chondrocalcinosis (calcium pyrophosphate deposition disease) is a recognized association.

### BIOCHEMICAL FEATURES AND PATHOPHYSIOLOGY

#### Fluid and Electrolytes

The TAL is responsible for 30% of NaCl reabsorption in the nephron. Failure of this cotransporter leads to the ECF volume depletion that typifies Bartter's syndrome. The reabsorption of NaCl in this segment also serves to generate the high tonicity of the renal interstitium, providing the osmotic force for water reabsorption in the medullary-collecting duct. Patients with Bartter's syndrome fail to concentrate their urine maximally in response to exogenous vasopressin. The salt wasting and urine-concentrating defect manifest symptomatically with polyuria and polydipsia.

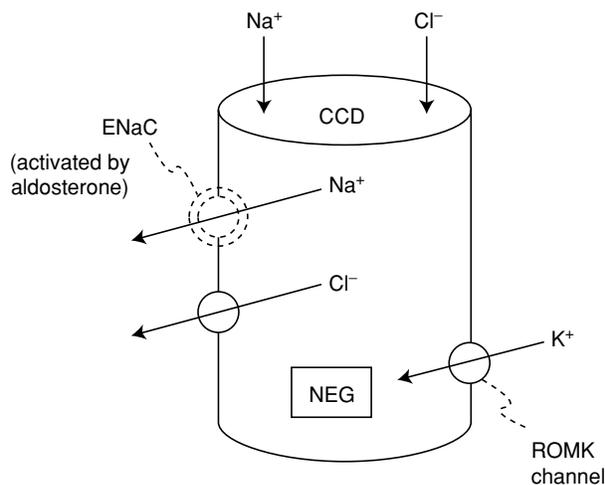
The NCC accounts for only 5% of NaCl reabsorption and has no role in establishing the osmolality of the renal interstitium. Consequently, patients with Gitelman's syndrome have only a mild degree of ECF volume contraction and no urine-concentrating defect. Nevertheless, chronic hypokalemia may impair urine-concentrating ability.

Hypokalemia in these syndromes results from increased urinary potassium excretion. Excretion of potassium depends on its rate of secretion in the cortical collecting duct (CCD) and on the flow rate in this nephron segment.

$$K^+ \text{ Excretion} = [K^+]_{\text{CCD}} \times \text{Flow Rate}_{\text{CCD}}$$

In the presence of vasopressin actions, flow in the terminal CCD is determined by the rate of delivery of osmoles to the CCD. Inhibition of NaCl reabsorption in the TAL or DCT increases the delivery of osmoles to the CCD, leading to a higher rate of flow in the terminal CCD. Secretion of  $K^+$  by principal cells in the CCD requires a lumen-negative voltage and open  $K^+$  channels (Fig. 3). The lumen-negative voltage is generated by electrogenic reabsorption of  $Na^+$ . In Bartter's and Gitelman's syndromes, there is increased delivery of  $Na^+$  and  $Cl^-$  to the CCD. The high levels of serum aldosterone due to ECF volume contraction open luminal epithelial sodium channels (ENaC) in the CCD. If the capacity for reabsorption of  $Na^+$  exceeds that for  $Cl^-$ , a negative luminal voltage is generated and drives the secretion of  $K^+$  via luminal  $K^+$  channels (Fig. 3).

Metabolic alkalosis is a feature of both syndromes. Hypokalemia stimulates ammoniogenesis and generation of bicarbonate by the cells of the proximal convoluted tubule. Most of the filtered  $HCO_3^-$  is reclaimed in the proximal tubule by secretion of  $H^+$  via the  $Na^+/H^+$  exchanger (NHE-3). The activity of NHE-3 is stimulated by angiotensin II and intracellular acidosis. In patients with Bartter's and Gitelman's syndromes, ECF volume contraction stimulates the release of



**Figure 3**  $K^+$  secretion in the distal nephron. Secretion of  $K^+$  by principal cells in the CCD requires (1) an open ROMK channel and (2) a lumen-negative voltage (NEG). The latter is generated when  $Na^+$  is reabsorbed via ENaC at a rate faster than the accompanying anion,  $Cl^-$ . In patients with Bartter's and Gitelman's syndromes, there is increased delivery of  $Na^+$  and  $Cl^-$  to the CCD. A lumen-negative voltage is generated if the capacity for  $Na^+$  reabsorption via ENaC (stimulated by aldosterone) exceeds that for  $Cl^-$ .

angiotensin II. Furthermore, during hypokalemia, a shift of  $K^+$  from the intracellular fluid to the ECF in exchange for  $H^+$  results in intracellular acidosis.

## Divalent Cations

The perturbations in divalent cation metabolism in these disorders are less well understood. In the TAL, recycling of  $K^+$  through ROMK channels generates a positive luminal charge that drives paracellular  $Mg^{2+}$  and  $Ca^{2+}$  reabsorption.  $Ca^{2+}$  and  $Mg^{2+}$  are reabsorbed in the DCT, although the mechanisms involved are not fully understood. Hypercalciuria is expected in patients with Bartter's syndrome. Surprisingly, and despite the fact that a large proportion of filtered  $Mg^{2+}$  is normally reabsorbed in the TAL, hypomagnesemia is not a common finding in these patients. On the other hand, patients with Gitelman's syndrome are usually hypomagnesemic. Hypocalciuria is also a characteristic finding in patients with Gitelman's syndrome.

A unifying hypothesis to explain these findings focuses on the DCT being the final critical nephron site for reabsorption of  $Mg^{2+}$  but not of  $Ca^{2+}$ . Although calcium reabsorption does occur in the DCT, the final critical nephron site for its reabsorption is the connecting segment (CS). In Bartter's syndrome, the DCT function is up-regulated to compensate for less reabsorption of NaCl in the LOH. The enhanced function of the DCT increases  $Mg^{2+}$  reabsorption and maintains magnesium homeostasis. On the other hand, enhanced  $Ca^{2+}$  reabsorption in both the DCT and the CS is still insufficient to overcome the increased delivery of  $Ca^{2+}$  from the LOH, and hypercalciuria results. Alternatively, patients with Gitelman's syndrome are likely to have decreased DCT cell mass due to chronic low activity of the transporters in this nephron segment. Thus, magnesium wasting with hypomagnesemia occurs, underscoring the DCT in the final nephron site for magnesium reabsorption. DCT dysfunction leads to enhanced delivery of NaCl to the CS, with up-regulation of its function. Because calcium reabsorption in the TAL in Gitelman's syndrome is intact, the enhanced  $Ca^{2+}$  reabsorption at the CS renders the urine hypocalciuric.

## Prostaglandin $E_2$

Prostaglandin  $E_2$  is hypersecreted in patients with Bartter's syndrome. In fact, neonatal Bartter's syndrome, in which  $PGE_2$  levels are particularly elevated,

has been called the hyperprostaglandin E<sub>2</sub> syndrome. However, this term is inappropriate because the prostaglandin synthesis in this condition is a secondary phenomenon resulting from the persistent ECF volume contraction. In Gitelman's syndrome, where ECF volume is less contracted, PGE<sub>2</sub> synthesis is normal. The effects of high PGE<sub>2</sub> levels are thought to be fever, secretory diarrhea, and osteopenia.

## DIAGNOSIS

Bartter's and Gitelman's syndromes must be differentiated from other causes of hypokalemic metabolic alkalosis. Primary hyperaldosteronism and Liddle's syndrome can be distinguished from these syndromes by the presence of hypertension, absence of ECF volume contraction, and suppressed plasma renin activity. Surreptitious vomiting can often be identified through a careful history or, more objectively, by the presence of a low urine chloride concentration ( $\leq 20$  mEq/L). Diuretic abuse may present a distinct challenge. As mentioned earlier, the NKCC2 is directly inhibited by loop diuretics such as furosemide, and its chronic use will present a clinical and biochemical profile indistinguishable from Bartter's syndrome. Similarly, thiazide diuretics inhibit the NCC and will generate a clinical picture identical to Gitelman's syndrome. This diagnosis can often be made if a spot urine sample reveals low sodium and chloride concentrations ( $\leq 20$  mEq/L), reflecting physiological salt avidity during periods when the diuretic effects have worn off. Urine samples with abundant Na and Cl should be assayed for diuretics prior to making a diagnosis of Bartter's or Gitelman's syndrome.

## TREATMENT

Because the ion channel defects in these patients cannot be corrected, treatment must be directed as the secondary hormonal and electrolyte derangements in these syndromes. Correction of hypokalemia in these patients is rather difficult even with large K<sup>+</sup> supplements. Hypomagnesemia may be an important factor in the enhanced kaliuresis in some patients with Gitelman's syndrome. Again, correction of hypomagnesemia with oral magnesium is usually difficult and also limited by gastrointestinal side effects. K<sup>+</sup>-sparing diuretics (e.g., amiloride, spironolactone) may help to conserve K<sup>+</sup>. A common clinical observation is that

even high doses of amiloride fail to curtail the excessive kaliuresis in patients with Bartter's and Gitelman's syndromes. Part of the explanation of this diminished effect could be related to the higher flow in the CCD due to inhibition of NaCl reabsorption in upstream nephron segments. This higher volume delivery to the CCD lowers the concentration of ENaC blockers, diminishing their effectiveness. A potential concern using these agents in patients with Bartter's and Gitelman's syndromes is that they may aggravate the salt wasting in these patients. This may become evident, for example, if dietary salt intake is decreased or if there is a nonrenal loss of NaCl.

ACE inhibitors have been used with variable success to reduce levels of angiotensin II and aldosterone, with hypotension being a major concern. In Bartter's syndrome, prostaglandin synthesis can be reduced with cyclo-oxygenase inhibitors such as indomethacin. Inhibition of PGE<sub>2</sub> attenuates salt wasting and hypokalemia and minimizes the systemic symptoms of prostaglandin excess. Caution is advised in using these agents during the neonatal period because acute renal failure and patent ductus arteriosus are documented complications. The prolonged use of these drugs also is discouraged due to the potential for chronic renal dysfunction.

## See Also the Following Articles

Hyperreninemia • Ion Channels • Mineralocorticoids and Mineralocorticoid Excess Syndromes

## Further Reading

- Birkenhager, R., Otto, E., Schurmann, M. J., Vollmer, M., Ruf, E. M., Maier-Lutz, I., Beekmann, F., Feketa, A., Omran, H., Feldmann, D., Milford, D. V., Jeck, N., Konrad, M., Landau, D., Knoers, N. V., Antignac, C., Sudbrak, R., Kispert, A., and Hildebrandt, F. (2001). Mutations of BSND causes Bartter syndrome with sensorineural deafness and kidney failure. *Nat. Genet.* **29**, 310–314.
- Gamba, G. (1999). Molecular biology of distal nephron sodium transport mechanisms. *Kidney Int.* **56**, 1606–1622.
- Halperin, M. L., and Kamel, K. S. (1998). Potassium. *Lancet* **352**, 135–142.
- Jeck, N., Konrad, M., Peters, M., Weber, S., Bonzel, K. E., and Seyberth, H. W. (2000). Mutations in the chloride channel gene, CLCNKB, leading to a mixed Bartter–Gitelman phenotype. *Pediatr. Res.* **48**, 754–758.
- Kamel, K. S., Oh, M. S., and Halperin, M. L. (2002). Bartter's, Gitelman's, and Gordon's syndromes: From physiology to molecular biology and back, yet still some unanswered questions. *Nephron* **92**, 18–27.

- Kamel, K. S., Oh, M. S., and Halperin, M. L. (2003). Treatment of hypokalemia and hyperkalemia. In "Therapy in Nephrology and Hypertension: A Companion to Brenner and Rector's 'The Kidney'" (H. R. Brady and C. S. Wilcox, eds.), pp. 349–363. W. B. Saunders, Philadelphia.
- Kurtz, I. (1998). Molecular pathogenesis of Bartter's and Gitelman's syndromes. *Kidney Int.* **54**, 1396–1410.
- Meij, I. C., Koenderink, J. B., van Bokhoven, H., Assink, K. F. H., Groenestege, W. T., de Pont, J., Bindels, R. J. M., Monnens, L. A. H., van den Heuvel, L. P. W. J., and Knoers, N. V. A. M. (2000). Dominant isolated renal magnesium loss is caused by misrouting of the Na<sup>+</sup>, K<sup>+</sup>-ATPase gamma-subunit. *Nat. Genet.* **26**, 265–266.
- Simon, D. B., and Lifton, R. P. (1998). Ion transporters mutations in Gitelman's and Bartter's syndromes. *Curr. Opin. Nephrol. Hypertension* **7**, 43–47.
- Watanabe, S., Fukumoto, S., Chang, H., Takeuchi, Y., Hasegawa, Y., Okazaki, R., Chikatsu, N., and Fujita, T. (2002). Association between activating mutations of calcium-sensing receptor and Bartter's syndrome. *Lancet* **360**, 692–694.



# Beckwith-Wiedemann Syndrome (BWS)

Juan F. Sotos

Ohio State University and Children's Hospital, Columbus, Ohio, United States

Jesús Argente

Universidad Autónoma de Madrid and Hospital Universitario Infantil Niño Jesús, Spain

## Glossary

**cryptorchidism** A developmental abnormality in which the testes fail to descend normally into the scrotum and remain in the abdomen.

**hyperplasia** An increase in the size of a tissue or an organ due to an increase in the number of its constituent cells.

**hypoglycemia** A condition in which the concentration of glucose in the blood is abnormally low; can cause symptoms such as a cold sweat, piloerection, tremulousness, hypothermia, headache, irritability, confusion, hallucinations, bizarre behavior, convulsions, and coma.

**macroglossia** An abnormally large tongue.

Beckwith-Wiedemann syndrome was initially described independently by J. B. Beckwith in 1963 and H. R. Wiedemann in 1964. It is associated with prenatal and postnatal overgrowth, with the most characteristic features occurring at birth: omphalocele or umbilical defects, macroglossia, and gigantism (Table I and Fig. 1).

## INTRODUCTION

Up to half (30–50%) of children with Beckwith-Wiedemann syndrome (BWS) may have severe persistent hypoglycemia, beginning during the first days of life, from hyperinsulinism as a result of pancreatic islet cell hyperplasia. The hypoglycemia usually subsides by 4 months of age. Other features include neonatal polycythemia, prominent eyes with relative infraorbital hypoplasia, prognathism, large fontanelles, capillary nevus in the central forehead and eyelids, unusual creases in the lobule of the external ear, and indentations or marks in the posterior rim of the helix. The head is relatively small and may have a protruding occiput. There is visceromegaly with

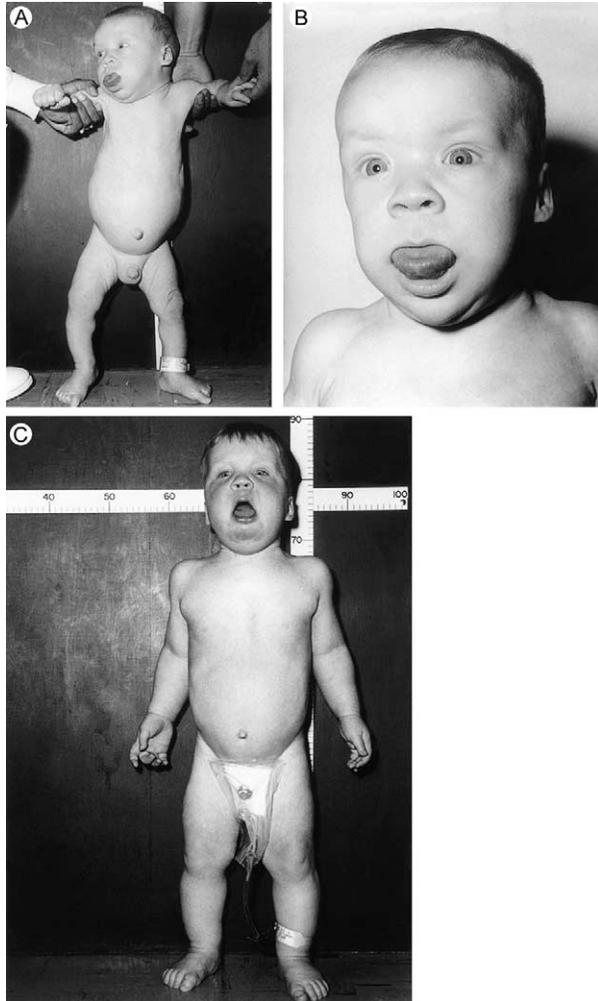
enlargement of the liver, kidneys, pancreas, and (at times) cardiomegaly; renal medullary dysplasia; fetal adrenal cortical cytomegaly; and interstitial cell hypoplasia of the gonads. Gastrointestinal malrotation and cryptorchidism are also seen. In 1983, Wiedemann reported hemihypertrophy in 12.5% of cases.

There is an increased incidence of malignant tumors in children with BWS. In 1983, Wiedemann reported that of 388 children, 29 (7.4%) developed 32 neoplasms, with the most frequent being Wilms' tumors and adrenal cortical carcinoma. Other tumors could be nephroblastoma, hepatoblastoma, and rhabdomyosarcoma. Nearly half (49%) of the children with neoplasms had partial or complete hemihypertrophy. Tumors usually occur during the first 5 years and rarely occur after 7 years of age.

Table I Beckwith-Wiedemann Syndrome Characteristics

<b>Overgrowth</b>
Prenatally and postnatally <sup>a</sup>
Advanced bone age
<b>Craniofacial</b>
Macroglossia <sup>a</sup>
Prominent eyes
Capillary nevus, forehead
Earlobe creases
<b>Malformations</b>
Omphalocele <sup>a</sup>
Renal medullary dysplasia
Hemihypertrophy (12.5%)
<b>Other</b>
Hypoglycemia—severe <sup>a</sup> (nesidioblastosis)
Adrenocortical cytomegaly <sup>a</sup> (consistently)
Interstitial cell hypoplasia of the gonads
<b>Risk of malignant tumors<sup>a</sup> (7–10%)</b>

<sup>a</sup>Most characteristic.



**Figure 1** (A, B) A 4 1/2-month infant with Beckwith–Wiedemann syndrome showing macroglossia, umbilical hernia, prominent calf muscles, and overgrowth. The length is 67.3 cm (26 1/2 inches), and the weight is 7.7 kg (both >75 percentile). (C) At 15 months, the height is 87.5 cm (34 1/2 inches), and the weight is 14.6 kg (both >95 percentile). Macroglossia is still evident. The infant did not have hypoglycemia.

## GROWTH

Children with BWS are large at birth, with increased muscle mass, thick subcutaneous tissue, and advanced bone maturation. Birthweights average 4 kg (8 pounds 13 ounces). The mean birthlength for males is greater than the 95 percentile and for females is on the 75 percentile, increasing to above the 95 percentile by 2 years. Growth velocity is usually above the 90 percentile until 4 to 6 years of age and is normal thereafter. Affected children reach an average height of 2.5 standard deviation score (SDS) at or after puberty, with weights between the 75 and 95 percentiles. In 1989,

W. G. Sippell and associates reported adult heights exceeding parental heights by 13.2 cm (5.2 inches) on the average. One girl with a predicted adult height of 190 cm (6 feet 3 inches) was treated with a high dose of estrogens.

## FREQUENCY

BWS is not rare. Several hundred cases have been reported. The prevalence is not known but is estimated to be approximately 1 in 14,000 persons. Approximately 85% of the cases are sporadic and 15% are inherited. There have been many families reported to have members affected in two or three generations, transmitted vertically, suggesting autosomal-dominant inheritance with incomplete penetrance.

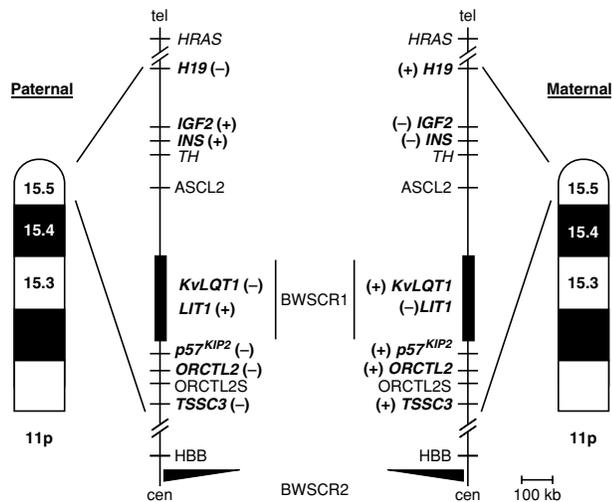
## ETIOLOGY

BWS is a complex and heterogeneous genetic disorder resulting from alterations of the expression of imprinted genes, involved in growth and cell cycle control, in the 11p15 chromosomal region. In the epigenetic phenomenon of imprinting, the DNA of the two alleles of a gene is differentially modified in such a manner that only one parental allele, parent specific for each gene, is normally expressed.

The 11p15 region harbors a cluster of imprinted genes (Fig. 2). IGF-2, insulin, and LIT1 normally show paternal expression, whereas H19, p57<sup>KIP2</sup>, and KvLQT1, except for some tissue variation, are predominantly maternally expressed. H19 is closely linked to IGF-2, and p57<sup>KIP2</sup> is separated by 700 kb from the IGF-2. The KvLQT1 gene spans much of the interval between p57<sup>KIP2</sup> and IGF-2.

Based on many studies, it would appear that the phenotype of BWS and tumor predisposition are caused by the balance between increased expression of paternally expressed growth-promoting genes and loss of maternally expressed growth-inhibiting genes. Despite the overwhelming number of studies, the etiology remains elusive.

The loss of imprinting of the maternal IGF-2 gene, with resulting biallelic expression, is one of the most common molecular defects found in patients with BWS without chromosomal abnormalities (Table II). In a proportion of these patients, loss of maternal IGF-2 imprinting is associated with complete suppression of maternal H19 expression. However, in a significant number of cases, IGF-2 shows biallelic expression, even though H19 expression and methylation are normal. This indicates that there



**Figure 2** Map of the chromosome 11p15.5 region showing a cluster of imprinted genes, indicated in bold. (+) indicates genes that are expressed, and (-) indicates genes that are imprinted. BWSCR1 and BWSCR2 indicate Beckwith–Wiedemann syndrome critical translocation regions. See text for explanation. Modified from R. Weksberg and C. Shuman (2001).

must be an alternative H19-independent pathway by which allele-specific IGF-2 expression is established or maintained. Overexpression of IGF-2 can also result from paternal uniparental disomy (UPD) of chromosome 11 or from duplications of the paternal 11p15 region associated with trisomy of 11p. In each case, UPD appears to result from a postzygotic event resulting in mosaicism for segmental paternal isodisomy containing the 11p15 region.

IGF-2 is an important growth factor for fetal growth. Transgenic experiments in mice have shown

that loss of the functional paternal IGF-2 allele results in 40% prenatal growth retardation, overexpression of IGF-2 can result in most of the symptoms of BWS, and targeted disruption of IGF-2 receptor (which is involved in degradation of IGF-2) results in overgrowth. It is of interest that overgrowth in BWS is in organs rich in IGF-2. In addition, IGF-2 imprinting is lost in sporadic Wilms' tumors, suggesting that IGF-2 overexpression is responsible for overgrowth and development of tumors. Biallelic expression of IGF-2, however, would not appear to be sufficient for tumor development because many patients with BWS do not develop embryonal tumors.

The contribution or role of the loss of function of maternally expressed genes that act as growth or tumor suppressors (e.g., H19,  $p57^{KIP2}$ , KvLQT1) by methylation defects, mutations, translocations, or inversions is difficult to assess. Alterations in any of these suppressors could cause BWS, and the phenotypic spectrum might depend on which maternally expressed gene is mutated.

In a 2002 study of 92 patients, there was a significantly higher frequency of altered DNA methylation of H19 in patients with cancer and of LIT1 in patients with midline abdominal wall defects and macrosomia. Paternal UPD of 11p15 was associated with hemihypertrophy, cancer, and hypoglycemia.

H19 is closely linked to IGF-2 and encodes for a biologically active, nontranscribed mRNA that may function as a tumor suppressor. Biallelic expression of IGF-2, by disruption of maternal IGF-2 imprinting, has been shown in mice as a result of a maternally inherited targeted deletion of the H19 gene. Also, biallelic expression of IGF-2 associated with methylation

**Table II** Chromosome 11p15.5 Region: Genetic Abnormalities Observed in Beckwith–Wiedemann Syndrome—Allelic Expression of IGF-2 Gene

	Father	Mother	Frequency (percentage)	Inheritance/Sporadic
Normally	Active	Inactive (imprinted)		
Beckwith–Wiedemann syndrome				
Paternal disomy (UPD)	Active-Active	—	10–20	Sporadic
Paternal 11p15 duplications (trisomy of 11p)	Active-Active	Inactive	<1	Sporadic or inherited
Normal karyotype IGF-2 loss of imprinting	Active	Active	~50	Unknown
Maternal H19 methylation	Active	Active	5–10	Unknown
Maternal 11p15 translocations, inversions, and KvLQT1 disruption	Active	Active	<1	Inherited or sporadic
Maternal $p57^{KIP2}$ mutations	Active	?	5–20	Usually inherited
Maternal LIT1 loss of imprinting	Active	Usually inactive	~50	Unknown

*Note.* See text for explanation.

and silencing of the maternal, normally expressed H19 gene has been found in BWS and Wilms' tumors, suggesting that H19 might directly regulate IGF-2 expression. Some studies, however, have demonstrated uncoupled expression of IGF-2 and H19 in patients with BWS and some tumors, suggesting that biallelic expression of IGF-2 can be independent of H19 control.

p57<sup>KIP2</sup> (CDKN1C) is a member of a family of cyclin-dependent kinase (cdk) inhibitors that negatively regulates cell proliferation and is potentially a negative regulator of fetal growth and a tumor suppressor. Only the maternal allele is expressed in both mice and humans. Mice with targeted disruption of the p57<sup>KIP2</sup> have a number of abnormalities (omphalocele, renal medullary dysplasia, and adrenal cortical hyperplasia and cytomegaly) that overlap with, but do not completely resemble, BWS.

In a study of 9 patients with BWS, two were heterozygous for mutations in the p57<sup>KIP2</sup> gene. In one of the cases, the missense mutation was transmitted from the proband's carrier mother. The expression of the maternal allele was mutant, and the repressed paternal allele was normal. In an analysis of 40 unrelated BWS patients, only 5% showed mutations of p57<sup>KIP2</sup>. In 1 case, the mutation was transmitted to the proband's mother, who was also affected, from the maternal grandfather. A 1997 study found a maternally transmitted coding mutation in the cdk inhibitor domain of the KIP2 gene in one of five cases of BWS. No somatic mutations in KIP2 were found in 12 primary Wilms' tumors, suggesting that KIP2 is a BWS gene but is not a Wilms' tumor suppressor. The manner by which loss of p57<sup>KIP2</sup> function results in the BWS phenotype is not yet clearly understood.

The KvLQT1 (KCNQ1) gene spans much of the interval between P57 and IGF-2 and is paternally imprinted and maternally expressed in most tissues, with the exception of the heart. Each of the break points in chromosomal rearrangements (inversions or translocations) in BWS disrupts this gene. The KvLQT1 gene product forms part of a potassium channel, and mutations in the gene are known to cause at least two cardiac arrhythmia syndromes. Patients with BWS with disruption of this gene have biallelic IGF-2 expression. It has been proposed that this gene may serve as an imprinting center for the 11p15 region and that disruption may affect transcription and DNA replication of distant genes such as IGF-2.

LIT1 (KCNQ1-AS) is an antisense transcript that is normally expressed from the paternal allele and that lies within the KvLQT1 (KCNQ1) gene. Loss of

imprinting of the maternal LIT1 gene with biallelic expression of LIT1 is the most frequent abnormality in BWS, accounting for 40 to 50% of patients, and is not linked to loss of imprinting of IGF-2.

TSSC3 and ORCTL2 were recently identified imprinted genes that show preferential maternal expression. Neither of them has yet been implicated with BWS.

Despite the overwhelming number of observations, the etiology of BWS remains elusive.

## DIAGNOSIS AND DIFFERENTIAL DIAGNOSIS

The BWS diagnosis is based on the clinical manifestations that have been described. Some of the phenotypic changes are somewhat similar to those of infants of diabetic mothers. Other disorders to consider would be hypothyroidism (because the enlarged tongue) and glycogen storage disease type I (because of the hepatomegaly and hypoglycemia), but they should be easy to distinguish.

Reports have indicated that, in many cases, BWS may be milder than one would expect from other published reports, and some cases may have only one or a few signs of BWS. The hemihypertrophy could be a good diagnostic clue in detecting the syndrome in relatives of affected patients.

Hemihypertrophy is associated with an increased tumor rate of 3% and a spectrum of tumors similar to those seen in BWS. Hemihypertrophy may represent BWS with minimal clinical findings, but it is seen with other syndromes such as Proteus syndrome and neurofibromatosis.

Patients with BWS may be difficult to distinguish from patients with Simpson–Golabi–Behmel syndrome (SGBS). Patients with SGBS have some features similar to those with BWS such as omphalocele, macrosomia at birth, visceromegaly, ear lobe creases, renal dysplasia, neonatal hypoglycemia, and a risk of embryonal tumor, including Wilms' tumor. Distinguished characteristics of SGBS are a cleft lip and palate, cardiac defects, polydactyly, vertebral and rib anomalies, hypoplastic or absent index fingernails, and X-linked recessive transmission.

Perlman syndrome is characterized by macrosomia, nephromegaly, neonatal hypoglycemia, mental retardation, and a high incidence of bilateral Wilms' tumor, usually occurring during the first year of life. It is not known whether this syndrome represents a separate entity or is actually a subtype of BWS with autosomal recessive inheritance.

## NATURAL HISTORY AND PROGNOSIS

The main problems with BWS are those encountered during the neonatal period with apnea, cyanosis, respiratory and feeding difficulties, polycythemia, and (particularly) hypoglycemia. The clinical findings tend to become less distinctive with age. The visceromegaly regresses, and there is a tendency to attain normal facial features. The other risk is the development of malignant neoplasms in 7 to 10% of patients.

A number of patients die during the neonatal period. Affected individuals who survive infancy generally are healthy and usually mentally normal. Some may have mild to moderate mental deficiency that may be related to severe hypoglycemia during the first few months.

Nephromegaly during infancy and early childhood is a risk factor for the development of Wilms' tumor in BWS. Infants with nephromegaly should be considered at greatest risk for Wilms' tumor, and screening may be best targeted at this group.

Fully 85% of the cases are sporadic and 15% could be inherited. Patients with mutations in the maternal 11p15 region are at risk for inheritance. Patients with maternal p57<sup>KIP2</sup> mutations and 11p15 translocations and inversions have a risk as high as 50%. Patients with paternal UPD have a low risk because the UPD is believed to arrive from a postzygotic somatic recombination event.

## TREATMENT

Detection and treatment of hypoglycemia are most important for surviving and for preventing neurological damage. The excessive rate of growth for the first few years and the tall stature during childhood and adolescence require no treatment unless an excessive adult height is predicted, particularly in females.

Muscle hypertrophy characteristic of macroglossia results in malocclusion, and surgical reduction of the tongue during early life should be considered.

Regular follow-up for possible tumor development is needed, and routine ultrasonography of the kidneys is mandatory, because Wilms' tumor and adrenocortical carcinoma are the most frequent neoplasms.

Abdominal ultrasound every 3 months to the age of 6 years is recommended. A baseline abdominal magnetic resonance imaging study during the first year of life is also advisable. For early detection of hepatoblastoma, alpha-feto protein testing should be done periodically up to 6 years of age.

## See Also the Following Articles

Congenital Adrenal Hyperplasia, Prenatal Diagnosis and Therapy • Congenital Lipoid Adrenal Hyperplasia • Hypoglycemia • Undescended Testes

## Further Reading

- Beckwith, J. B. (1963). Extreme cytomegaly of the adrenal fetal cortex, omphalocele, hyperplasia of the kidneys and pancreas, and Leydig-cell hyperplasia: Another syndrome? Abstract, Western Society for Pediatric Research, Los Angeles, November 11, 1963.
- Beckwith, J. B. (1969). Macroglossia, omphalocele, adrenal cytomegaly, gigantism, and hyperplastic visceromegaly. *Birth Defects: Original Article Series* **5**, 188–196.
- DeBaun, M. R., Niemitz, E. L., McNeil, E., et al. (2002). Epigenetic alterations of H19 and LIT1 distinguish patients with Beckwith–Wiedemann syndrome with cancer and birth defects. *Am. J. Hum. Genet.* **70**, 604–611.
- Lee, M. P., DeBaun, M. R., Mitsuya, K., et al. (1999). Loss of imprinting of a paternally expressed transcript, with antisense orientation to KvLQT1, occurs frequently in Beckwith–Wiedemann syndrome and is independent of insulin-like growth factor II imprinting. *Proc. Natl. Acad. Sci. USA* **96**, 5203–5208.
- Li, M., Squire, J. A., and Weksberg, R. (1998). Overgrowth syndromes and genomic imprinting: From mouse to man. *Clin. Genet.* **53**, 165–170.
- Sippell, W. G., Partsch, C. J., and Wiedemann, H. R. (1989). Growth, bone maturation, and pubertal development in children with the EMG-syndrome. *Clin. Genet.* **35**, 20–28.
- Sotos, J. F. (1997). Overgrowth: Section VI—Genetic syndromes and other disorders associated with overgrowth. *Clin. Pediatr.* **36**, 155–173.
- Weksberg, R., and Shuman, C. (2001). Beckwith–Wiedemann syndrome. In "Management of Genetic Syndromes" (S. B. Cassidy, and J. E. Allanson, eds.), pp. 49–69. Wiley–Liss, New York.
- Wiedemann, H. R. (1964). Complexe malformatif familial avec hernie ombilicale et macroglossie: Un "syndrome nouveau"? *J. Genet. Hum.* **13**, 223.
- Wiedemann, H. R. (1983). Tumors and hemihypertrophy associated with Wiedemann–Beckwith syndrome. *Eur. J. Pediatr.* **141**, 129.



# Benign Prostatic Hyperplasia (BPH)

Jaspreet S. Sandhu and Alexis E. Te

Weill Medical College of Cornell University, Cornell, New York, United States

## Glossary

**cystoscopy** An endoscopic procedure used to evaluate the urethra, prostate (in males), and bladder. It is often used to rule out bladder cancer or other bladder/urethral pathology as a cause of hematuria or irritative voiding symptoms.

**lower urinary tract symptoms** A symptom complex caused by lower urinary tract dysfunction consisting of obstructive and irritative symptoms. It is associated with, but not pathognomonic of, benign prostatic hyperplasia.

**prostate-specific antigen** A normal product of the prostate that is secreted via prostatic ducts during semen emission. Under pathologic conditions, such as prostate cancer, infection, or benign prostatic hyperplasia, it may be elevated in serum.

**transition zone** The part of the prostate most likely to be pathologically enlarged due to benign prostatic hyperplasia. It includes the periurethral part of the prostate and is therefore involved in outflow resistance to the bladder.

**transrectal ultrasound** An imaging modality used to evaluate the size and structure of the prostate. It is usually used to help guide prostate biopsies.

**urodynamics** A test that measures pressure within the bladder at different volumes and flow rates during voiding. It is useful in complex voiding dysfunction cases to determine the etiology of dysfunction.

**B**enign prostatic hyperplasia (BPH) is a pathologic process that consists of androgen-dependent increases in the number of epithelial and stromal cells in the periurethral area of the prostate. BPH contributes to, but is not the sole cause of, lower urinary tract symptoms in aging men.

## INTRODUCTION

BPH is a pathologic process characterized by an increase in the number of cells in the periurethral area

of the prostate. This process contributes to the development of lower urinary tract symptoms (LUTS) in aging men that were classically described as “prostatism.” LUTS are symptoms arising from lower urinary tract dysfunction that are further subdivided into obstructive (urinary hesitancy, straining, weak stream, terminal dribbling, prolonged voiding, incomplete emptying) and irritative (urinary frequency, urgency, nocturia, urge incontinence, small voided volumes) symptoms. BPH is a histologic diagnosis that is but one cause of LUTS. Autopsy series have shown that no men less than 30 years of age have the disease, whereas more than 50% of men greater than 60 years of age have histologic evidence of the disease. This prevalence reaches almost 100% in the ninth decade. BPH is, as its name implies, a benign process, with symptoms being the most common cause for treatment. However, untreated BPH can lead to severe complications including urinary retention, obstructive uropathy, azotemia, intractable hematuria, and urinary tract infections.

## PATHOPHYSIOLOGY

### Etiology

The etiology of BPH is multifactorial and not definitively established. BPH is an overgrowth of the periurethral area (transition zone) of the prostate that is dependent mainly on androgens, particularly dihydrotestosterone (DHT). 5 $\alpha$ -Reductase is responsible for the conversion of testosterone to DHT, which has five times more potency for its receptor than testosterone. DHT binds to androgen receptor and the complex is primarily responsible for the stimulation of growth factors that influence prostate cell division and growth. DHT is responsible for maintaining the balance between cell proliferation and cell death and therefore prostate growth. Elevated levels of DHT, along with hypothesized hormonal imbalances, result in BPH. A lack of DHT results in a lack of prostate growth and resultant BPH or prostate cancer as

demonstrated in a cohort of patients with type 2 5 $\alpha$ -reductase deficiency reported by Imperato-McGinley.

## Natural History

The natural history of BPH is progression in the majority of patients. Progression results in increased prostate size, worsening of symptoms, deterioration in urinary flow rate, increased risk of acute urinary retention, and increased risk of surgery for BPH. Some patients do not progress, although in longitudinal studies the risk of BPH-related surgery and of acute urinary retention increases as patients age. Determining which patients will progress continues to be an area of intense study.

## DIAGNOSIS

BPH is a difficult diagnosis because of the many causes of LUTS, including bladder dysfunction, bladder cancer, urethral stricture disease, and prostatitis. A strategy for the management of BPH is to follow objective signs after an initial evaluation and intervene when symptoms become moderate to severe or when other clinical indications apply.

## Initial Evaluation

Initial evaluation for men presenting with LUTS consists of a detailed history and physical exam, including a digital rectal exam and a neurological exam. Oral intake history, including intake amounts of caffeine and alcohol, should be obtained. Medications affecting urinary volume, such as diuretics and stool-bulking agents, and those affecting voiding, such as antihistamines and decongestants, should be noted. Systemic diseases, such as diabetes, congestive heart failure, and neurological or neoplastic diseases, should also be noted.

## Objective Signs

### Symptoms

LUTS symptoms are well known. The American Urological Association (AUA) has created a validated, reproducible index that is designed to determine disease severity and determine response to therapy. This 35-point scale has questions about both obstructive and irritative symptoms. An AUA symptom score of 0–7 is considered mild, 8–19 is moderate, and 20–35 is severe.

### Urodynamic Parameters

Formal urodynamic testing, including bladder volume, detrusor pressure, and urinary flow, is the gold standard for the diagnosis of LUTS associated with obstruction. However, there is no clear consensus on the role of urodynamics in the routine evaluation of men with BPH. There is a clear role for urodynamics in complex cases. Maximum flow rate using a simple uroflowmeter is an easy parameter to use for following men with BPH for progression. Postvoid residual urine (PVR) is also an easy parameter to document; however, there are no guidelines for its use and there is no proven correlation between PVR and BPH.

### Transrectal Ultrasound

Transrectal ultrasound (TRUS) is an important imaging modality, particularly as a guide for directed biopsies in suspected prostate cancer cases. TRUS gives an accurate evaluation of prostate size and might be useful in determining transition zone size; however, prostate volume is not diagnostic of BPH. This imaging modality is most useful in BPH treatment selection, particularly for minimally invasive therapy.

### Prostate-Specific Antigen

Prostate-specific antigen (PSA) is an important serum marker for prostate cancer. It is also elevated, albeit less so, in men with BPH. Although some evidence exists for the use of PSA in following men with BPH, PSA should be obtained in these men for the purpose of prostate cancer screening. This is because the population at risk for BPH is the same one at risk for prostate cancer.

### Cystoscopy

Although a large prostate is evident on cystoscopy, it is not routinely indicated in the diagnosis and management of BPH. Cystoscopy has a role in the evaluation of hematuria and prior to invasive therapy.

## Indications for Treatment

Symptom relief is the most common reason for treatment of BPH. The most common form of therapy for mild to moderate symptoms is medical therapy. Surgical therapy is usually reserved for medical failures. There are, however, some clear indications for surgical therapy including frequent urinary tract infections, bladder stones, azotemia, urinary retention, and intractable hematuria.

## TREATMENT

### Medical Therapy

The initial management for BPH includes watchful waiting or medical therapy, especially in those patients with mild or moderate symptoms and no indications for surgical intervention. Medical therapy consists of phytotherapy, alpha-blocking agents, and 5 $\alpha$ -reductase inhibitors.

Phytotherapy consists of common plant extracts and herbs, such as *Serenoa repens* (saw palmetto) and *Pygeum africanum* (African plum). Although phytotherapy is touted as being effective, the active ingredients and therapeutic dosages are unknown for most extracts, the mechanism of action is poorly understood, and the safety profile has not been adequately studied with clinical trials. Therefore, phytotherapy is not often prescribed by urologists or other medical professionals.

Alpha-blocking agents are the first line of prescribed medical therapy. Their efficacy has been documented by multiple clinical trials and they have a low risk of morbidity. These agents reduce symptoms equally well in obstructed and nonobstructed patients and they generally increase flow rate by approximately 2–3 ml/s compared to placebo. Nonselective alpha-blocking agents are particularly useful in hypertensive patients because of their effect on lowering blood pressure. Common alpha-blockers include phenoxybenzamine, prazosin, terazosin, doxazosin, and alfuzosin. The alpha-blocker tamsulosin has the advantage of being prostate-selective and not needing titration.

5 $\alpha$ -Reductase inhibitors, such as finasteride and dutasteride, have been shown to prevent the progression of BPH over time. They decrease the prostate volume by approximately 20% through reduction of DHT stimulation. These agents have a slower onset of action and should be used only in patients with large prostates (>40 g). Finasteride has been studied extensively and has shown mild decreases in symptoms and mild increases in flow rate, along with a reduction in prostate volume. The risk of BPH-related surgery and the risk of acute urinary retention were reduced by 55 and 57%, respectively, in those patients on this drug. Studies with dutasteride produced similar results. Combination therapy with alpha-blocking agents and 5 $\alpha$ -reductase inhibitors has shown greater efficacy than either agent alone in patients with large prostates.

### Minimally Invasive Surgical Therapy

Minimally invasive surgical therapy consists of office-based therapies, such as transurethral needle ablation of the prostate, transurethral microwave therapy, or

interstitial laser coagulation. Each of these therapies can be performed with minimal anesthesia and has shown significant improvement in symptom scores and flow rates. However, the durability of these therapies is still being assessed and each has a significant retreatment rate (approximately 15–20% in most series). Minimally invasive surgical therapies are a good option for those patients who are poor surgical candidates or who are concerned about surgical morbidities.

### Transurethral Surgery

Transurethral resection of the prostate (TURP) has traditionally been the gold standard for comparison of treatments for BPH. TURP is associated with durable significant relief of symptoms and increase in flow rates with excellent 10-year results including low reoperation rates. TURP is associated with a low morbidity rate, but the morbidities that do occur are significant and include dilutional hyponatremia and excessive blood loss requiring transfusions or reoperation. Variations of TURP have gained prominence due to the decreased amount of anesthesia required and the decreased rates of complications.

Transurethral vaporization of the prostate is associated with less blood loss than TURP with similar efficacy. Transurethral incision of the prostate, useful for select patients with small-volume prostates, is associated with decreased anesthesia requirements and less operative time. Transurethral laser prostatectomy, performed by using the neodymium, holmium, or potassium-titanyl-phosphate lasers, has an efficacy similar to that of TURP with decreased morbidity and decreased anesthesia requirements. Long-term efficacy data on these newer transurethral techniques have not yet been reported.

### Open Surgery

Open prostatectomy remains an option for those patients with large-volume prostates (>80 g) or those with concomitant bladder pathology, such as bladder stones or diverticula. Open prostatectomy has had excellent durable results with approximately 98% of patients experiencing relief of symptoms and increases in flow rate. However, it is associated with the morbidities and complications expected for pelvic surgery, such as pulmonary embolus, bleeding, and infection, as well as incontinence and impotence.

## SUMMARY

BPH, an overgrowth of the periurethral transition zone of the prostate, is very prevalent in elderly men.

It is associated with significant symptoms and morbidity. Although the definitive diagnosis of BPH is not easy to make, there exist validated tools to help clinicians with the diagnosis. Most cases of BPH are progressive and dependent on androgens. Effective therapies for the management of BPH-related symptoms and complications, including medical and surgical options, exist and continue to evolve.

### See Also the Following Articles

Aging and the Male Reproductive System • Prostate Cancer

### Further Reading

- Imperato-McGinley, J., Guerrero, L., Gautier, T., and Peterson, R. E. (1974). Steroid 5 $\alpha$ -reductase deficiency in man: An inherited form of male pseudohermaphroditism. *Science* **186**, 1213–1215.
- Kirby, R., McConnell, J. D., Fitzpatrick, J. M., Roehrborn, C. G., and Boyle, P. (1996). "Textbook of Benign Prostatic Hyperplasia." Isis Medical Media, Oxford, UK.
- McConnell, J. D., Barry, M. J., Bruskewitz, R. C., *et al.* (1994). In "Clinical Practice Guideline No. 8: Benign prostate hyperplasia: Diagnosis and Treatment." AHCPR Publication 94-0582. U. S. Department of Health and Human Services, Agency for Health Care Policy and Research, Rockville, MD.
- Partin A. W., Rodriguez, R., Roehrborn, C. G., McConnell, J. D., Lepor, H., Lowe, F. C., Fitzpatrick, J. M., Mebust, W. K., Han, M., Alfert, H. J., and Partin, A. W. (2002). Benign prostatic hyperplasia. In "Campbell's Urology" (P. C. Walsh, A. B. Retic, E. D. Vaughan, and A. J. Wein, eds.), 8th ed., Section VI, pp. 1235–1434. Saunders, Philadelphia, PA.
- Weiss, R. E., and Fair, W. R. (1997). "Management of Prostate Diseases," 2nd ed. Professional Communications Inc., New York.



# Biological Rhythms

Gregory E. Demas

*Indiana University, Bloomington, Indiana, United States*

Lance J. Kriegsfeld

*Columbia University, New York, New York, United States*

## Glossary

**biological rhythm** Any recurrent endogenous cycle (behavioral or physiological) that persists in constant conditions in the absence of geophysical or environmental temporal cues.

**circadian rhythm** Endogenous rhythm that is approximately 1 day in length.

**circannual rhythm** Endogenous rhythm that is approximately 1 year in length.

**entrainment** Synchronization of an internal endogenous rhythm by a recurring external time cue such as the day–night cycle.

**free running** Expression of a biological rhythm in the absence of any environmental temporal cues.

**photoperiod** Amount of daylight in a given day; day length.

**seasonal rhythm** Yearly rhythm that is exogenously regulated by geophysical cues.

**suprachiasmatic nucleus** The primary “biological clock” located in the anterior hypothalamus of the brain that coordinates rhythmic and photoperiodic information via inputs and outputs of the nervous and endocrine systems.

A biological rhythm is any recurrent endogenous cycle (behavioral or physiological) that persists in the absence of geophysical or environmental temporal cues but is normally synchronized (entrained) with a period that approximates that of the geophysical cycle.

## INTRODUCTION

As most people have experienced throughout their lifetimes, timing is everything. The ability to modify one’s behavior or physiology on a regular basis, be it daily, monthly, or yearly, is arguably one of the most

adaptive responses of life and occurs in virtually all organisms, ranging from single-celled algae to complex vertebrates, including humans. But why have biological rhythms evolved, and what functions do they serve? All physiological systems require proper coordination and synchronization with one another to ensure that a wide variety of rhythmic events occur during an optimal time of the day or year. For example, for the body to prepare appropriately to digest a meal, numerous physiological processes must occur in anticipation of eating. Many species must also predict the time of year to coordinate breeding activities. These daily and yearly changes require critical coordination among numerous internal processes necessary to maintain homeostasis. Furthermore, for biological rhythms in behavior and physiology to be coordinated with the appropriate time, these rhythms need to be synchronized to external time cues within the environment. The means by which intrinsic rhythms are generated by a clock in the brain, and how these rhythms are synchronized to local time, is the primary focus of this article.

## HISTORICAL OVERVIEW

The importance of coordinated timing of biological events has been known since the dawn of civilization. For example, written records dating back to Alexander the Great in the 4th century B.C. noted the daily movements of flower petals. However, the study of biological timing as an empirical science is in its relative infancy. Chronobiology, the scientific study of biological rhythms, formally began in 1729 with a brief communication by Jean Jacques d’Ortous de Mairan, who noted that the daily leaf movements of a heliotropic plant persisted in complete darkness and in the absence of any other geophysical cues. However, it was not until the 1950s, based on the pioneering work of noted “clock watchers” Colin

Pittendrigh, Serge Daan, and Jergen Aschoff, that a firm scientific understanding of the characteristics and formal properties of biological rhythms began to take shape.

## TYPES OF RHYTHMS

Since these early discoveries, a wide variety of biological rhythms have been identified in both vertebrate and invertebrate species, and these rhythms are typically categorized by the length of a single cycle. For example, circadian rhythms are daily rhythms that are approximately 24 h and mimic the geophysical cycles of day and night. The alternating pattern of activity followed by periods of rest (sleep) that are observed in most vertebrate species is the most salient example of a circadian rhythm. However, biological rhythms need not be 24 h in length. Ultradian rhythms are rhythms that occur more frequently than 24 h. For example, many hormones (e.g., cortisol) and enzymatic reactions display rhythms only a few hours in length. In contrast, infradian rhythms are rhythms longer than 24 h but shorter than 1 year. Cycles of reproductive endocrine activity, such as menstrual cycles in humans and estrous cycles in nonhuman animals, are examples of this kind of rhythm. Finally, some species display circannual rhythms, with cycle lengths of approximately 1 year that persist in the absence of environmental influences. Circannual rhythms commonly involve yearly cycles in breeding activity and reproductive endocrinology. For example, some species of birds and rodents display yearly fluctuations in gonadal mass and changes in reproductive hormones (e.g., testosterone, estrogen), and these fluctuations persist when the animals are maintained within constant conditions of the laboratory. Closely related to circannual rhythms are seasonal rhythms, which also approximate 1 year. However, unlike true circannual rhythms, seasonal rhythms are generated in response to environmental cues and do not persist in the absence of these cues.

Although many kinds of biological rhythms have been identified, the majority of research within the field of chronobiology has focused on circadian and circannual/seasonal rhythms. As will be seen later, the neuroendocrine mechanisms responsible for generating circadian rhythms are also intimately involved in generating seasonal rhythms. Because the circadian clock is necessary for seasonal rhythms, the discussion that follows outlines the means by which circadian rhythms are generated and maintained and then the mechanisms by which seasonal rhythms are generated. The

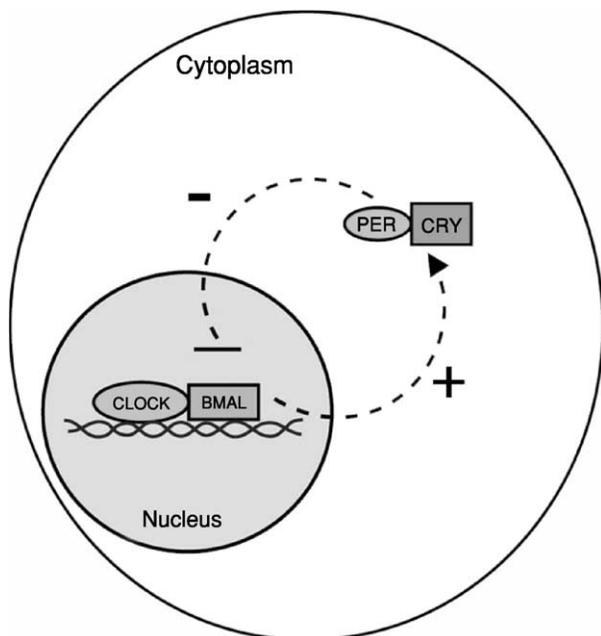
article concludes with a discussion of various endocrine disorders that have been tied to abnormalities in circadian and circannual/seasonal rhythms.

## CIRCADIAN RHYTHMS

Circadian comes from the Latin *circa* (meaning “around”) and *diem* (meaning “a day”). Virtually all measurable physiological and behavioral responses display circadian rhythms. There are circadian rhythms in everything from heart rate and liver metabolism to attention and speed of decision making and reaction time. It is now known that all of these rhythms are controlled by a bilateral nucleus at the base of the brain. This brain region was identified in 1972 as an area of the hypothalamus called the suprachiasmatic nucleus (SCN). Lesions that destroy the SCN result in an abolition of all rhythmic processes from daily rhythms in hormone secretion to the sleep–wake cycle. As will be seen later, SCN lesions can also abolish seasonal rhythms.

Further evidence that the SCN is the circadian clock comes from studies showing that rhythmicity can be restored in SCN-lesioned animals by transplanting SCN tissue from a donor animal into the brains of animals with their own SCN destroyed. This procedure restores circadian rhythms in these formerly arrhythmic animals, demonstrating that one can literally remove the circadian clock from one animal and provide it to another animal whose own clock is not functioning. Importantly, the rhythm is restored with the period of the donor, demonstrating that the restoration of circadian behavior is a property of the transplant rather than restoring function in the host animal’s brain. Finally, the SCN can be removed from an animal and maintained in a culture dish, and it will continue to show circadian rhythms in neural firing rate. These studies are important because they show that circadian rhythms are contained within the SCN and are not simply being driven by temporal input to this brain region.

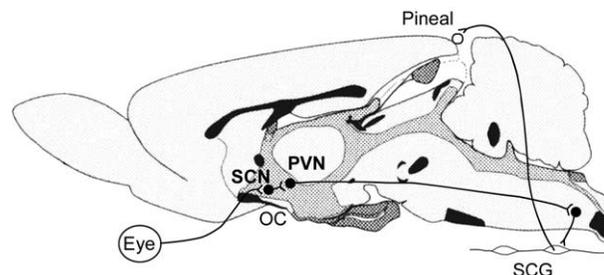
Knowledge of where the primary biological clock is located within the brain set the stage for investigations of the SCN at the cellular level to determine what makes the clock “tick.” To draw an analogy between the SCN and a watch, if one opened the back of a watch, a large number of gears could be seen. One might know that the gears are all working together to allow the watch to function properly. However, to fully understand the watch’s internal mechanism, one would have to carefully remove one gear at a time to see how it is connected to the next gear, and so on,



**Figure 1** Basic molecular model of the intracellular circadian feedback loop. The process begins in the cell nucleus when the CLOCK and BMAL proteins dimerize to drive the transcription of the *per* and *cry* genes. In turn, *per* and *cry* are translocated to the cytoplasm and translated into their respective proteins. Throughout the day, the PER and CRY proteins build within the cell cytoplasm. When levels of PER and CRY reach a threshold, they form a heterodimer, feed back to the cell nucleus, and negatively regulate CLOCK:BMAL-mediated transcription of their own genes. This feedback loop takes approximately 24 h, thereby leading to an intracellular circadian rhythm.

until all of the gears were taken out. In much the same way, one can study individual cells within the SCN to see how all of the cells (approximately 10,000) in a unilateral SCN work together to keep time. Unlike the gears of a watch, individual cells within the SCN are capable of showing their own independent circadian rhythms. When one begins to look at other cells, one can see that they too show rhythms, but each cell has its own intrinsic period. So, like the question of how all of the gears of a watch work together to keep accurate time, one question of interest to circadian biologists is how all the cells of the SCN work together to produce a rhythm of 24 h.

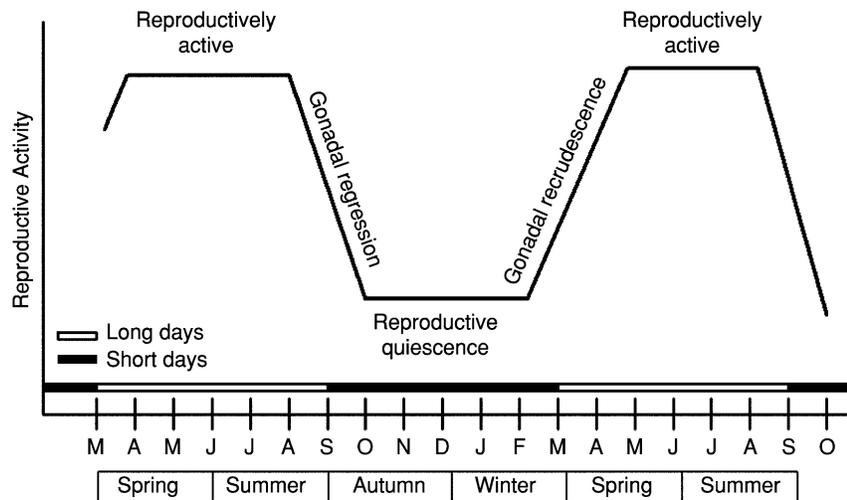
The knowledge that individual cells of the SCN can produce a rhythm on their own suggested that the complete machinery for producing a rhythm is contained within each cell of the SCN. This information allowed researchers to begin investigating the cellular mechanisms responsible for rhythm generation. Over the past 10 years or so, enormous progress has been made uncovering the cellular and molecular



**Figure 2** Neural pathway by which environmental light reaches the suprachiasmatic nucleus (SCN) and pineal gland to regulate circadian entrainment and regulation of melatonin secretion. Light enters the retina and is transmitted to the SCN located directly above the optic chiasm (OC). The SCN then reaches the pineal through a multisynaptic pathway, first synapsing in the paraventricular nucleus of the hypothalamus (PVN). From here, fibers proceed through the medial forebrain bundle to the spinal cord. At this point, fibers actually exit the central nervous system and synapse onto the superior cervical ganglion (SCG) of the sympathetic nervous system. From here, sympathetic nerve fibers terminate in the pineal, where they regulate melatonin secretion.

mechanisms responsible for the generation of circadian rhythms within a cell. It is now known that these rhythms are produced by a feedback loop within the cell that takes approximately 24 h to complete. The process begins in the cell's nucleus when two proteins called CLOCK and BMAL bind to one another and drive the transcription of messenger RNA (mRNA) of the *period* (*per*) and *cryptochrome* (*cry*) genes. The mRNA for these genes then migrates to the cytoplasm of the cell and is translated into PER and CRY proteins. Over the course of the day, these proteins build up within the cytoplasm, and when they reach high enough levels, they bind to one another (PER:CRY). These newly formed PER:CRY dimers then feed back to the nucleus, where they bind to the CLOCK:BMAL protein complex to turn off their own production. Although numerous other clock genes and their protein products have been identified, this regulatory loop provides the core molecular clock mechanism in mammals (Fig. 1).

Interestingly, over the past several years, the same clock genes (e.g., *clock*, *bmal*, *per*, *cry*) that are found in cells of the SCN have been identified in other brain areas as well as in many other peripheral tissues in the body. However, these “clocks” in other parts of the body are different from those found in the SCN. If these clocks are removed from the body, their endogenous rhythms will cycle for only a few days before dampening, whereas those present in the SCN will continue to cycle indefinitely. These findings suggest that the SCN is the master clock that is required to communicate with other subordinate



**Figure 3** Typical pattern of seasonal/photoperiodic changes in reproductive physiology in a long-day breeding rodent. Reproductive activity is maximal during the long days of summer, but reproductive regression occurs as the days get shorter during the autumn. The reproductive system maintains a period of quiescence (inactivity) during the short days of winter. As the day length increases during the spring, reproductive activation (recrudescence) occurs anew.

clocks so that they can continue to keep time. There are many reasons why these subordinate clocks might be needed in tissues other than the SCN. These clocks may be necessary for tissues to anticipate upstream signals stimulating production of specific factors. For example, the cellular machinery necessary to produce a hormone in response to stimulation could require hours of preparation within the cell. However, if the tissue has its own clock, it can turn on this machinery in advance of stimulation and be prepared to produce the hormone much faster.

The preceding findings summarize what is known about the physiological functioning of the SCN. However, for a clock to be adaptive for an organism, the clock needs to coordinate all of these rhythms with local time in the environment. In mammals, this synchronization is accomplished via a direct projection, the retinohypothalamic tract (RHT), from the eye to the SCN (Fig. 2). Unlike rod and cone photoreceptors in the retina that are responsible for black/white and color vision, respectively, the photoreceptors that communicate with the SCN are ganglion cells that are directly responsive to light because they contain a photopigment called melanopsin. Essentially, these melanopsin-containing ganglion cells project to the SCN to inform the SCN of the time in the environment. Because the SCN is synchronized (entrained) to local time, it can prepare the brain and body well in advance of the time that rhythmic processes need to occur. For example,

the SCN can stimulate the pancreas to produce insulin before one eats lunch. In this way, one's body can be prepared to digest the food before one even eats.

Now that the intrinsic rhythm within the SCN is synchronized to local time, this "time-stamped" rhythmic information needs to be communicated all over the brain and body to coordinate the timing of thousands of physiological and biochemical processes. For example, the SCN needs to communicate with brain areas involved in sleep to coordinate sleep with nighttime. The SCN accomplishes this communication using both neural and humoral signals. Researchers can determine neural communication from the SCN relatively easily by using chemicals to trace the "wiring" from the SCN to its targets. Through the use of tracing techniques, it is now known that the SCN communicates neurally to a vast array of brain targets as well as to targets in the periphery such as the liver and heart. It is also known that the SCN regulates some rhythmic functions through humoral signals. For example, if the fibers exiting the SCN in an animal are surgically cut, some circadian rhythms persist. However, because diffusible signals can essentially travel anywhere in the brain and body, the means by which these neurochemicals communicate is much more difficult to investigate experimentally. To date, two neurochemicals found in neurons of the SCN, transforming growth factor- $\alpha$  (TGF- $\alpha$ ) and prokineticin-2, have been implicated

as diffusible factors that may be used by the SCN to control locomotor behavior in rodents. However, determining where specifically in the brain these chemicals act to regulate locomotor behavior requires further investigation.

## CIRCANNUAL/SEASONAL RHYTHMS

Just as animals experience large fluctuations in physiology and behavior throughout a single day, many organisms experience pronounced changes across the seasons of the year. Recall that circannual rhythms are endogenously driven rhythms that approximate 1 year. Such rhythms have been demonstrated in several birds and a few mammalian species (e.g., ground squirrels, sheep). In contrast to true circannual rhythms, the majority of species display seasonal rhythms in behavior and physiology in response to predictable changes in the environment such as day length. Although seasonal rhythms are not endogenously regulated, they still produce pronounced changes in physiology and behavior throughout the year. Much of the research conducted in the area of seasonality has focused on seasonal changes in breeding and reproductive endocrinology.

Many animals are faced with potentially large seasonal fluctuations in environmental conditions, including changes in day length, ambient temperature, rainfall, humidity, food availability, and social interactions. These potentially large fluctuations can dramatically affect the world in which an animal lives. For many animals, the energy required for thermoregulation is high during the winter, a time when energy availability is typically low. Because of this “energetic bottleneck,” many animals have evolved specific physiological and behavioral adaptations to cope with winter conditions. For example, some animals may migrate or reduce activity (e.g., hibernate, enter torpor) when food availability is low and energetic demands are high. Other animals restrict breeding to specific periods of the year (e.g., April–August) (Fig. 3) because breeding at inappropriate times can potentially compromise the survival of both the parents and their offspring. But how does an animal determine the optimal time of year?

As discussed previously, a wide variety of proximate environmental factors can be used in the timing of seasonal breeding, and all of these factors fluctuate seasonally. However, the majority of scientific studies suggest that animals rely mainly on day length as a primary cue with which to estimate the time of year. Most environmental cues other than day length vary

in relatively unpredictable ways throughout the year. For example, although winter is usually associated with lower ambient temperatures compared with summer in the Northern Hemisphere, an unseasonably warm “Indian summer” may occur during the autumn/winter or a cold wave may roll through during the summer. Furthermore, food and water may be available to animals only sporadically throughout the year. In contrast to these seasonal factors, day length information is relatively “noise free” and can be used to coordinate energetically expensive activities (e.g., breeding) to coincide with adequate energy availability. By relying on just two pieces of information, the absolute period of daylight and the direction of change across time, animals can determine the precise time of year. For many species, breeding occurs during the long days of summer, whereas short “winter-like” days result in complete regression of the reproductive system as well as in changes in body weight and fur thickness. Interestingly, some species (e.g., sheep) actually breed during the winter. As would be predicted, the effects of day length on reproductive physiology and behavior are generally reversed in these species because their lengthy gestation time allows for birth during summer. The ability of an animal to use seasonal changes in day length to coordinate physiological and behavioral adaptations is commonly referred to as photoperiodic time measurement (PTM).

How is day length information interpreted physiologically in mammals? As with circadian rhythms, seasonal changes in day length are detected by the photosensitive retinal ganglion cells within the eye that send projections to the SCN (Fig. 2), suggesting that seasonal rhythms are regulated to some extent by the circadian system. In fact, abnormalities within the circadian system that “speed up” the circadian clock also appear to accelerate the rates of seasonal responses. The SCN projects multisynaptically to the pineal gland, where the photoperiodic signal is transduced into an endocrine message. Specifically, neural activation of the pineal gland results in the production of the indolamine hormone melatonin. During prolonged periods of darkness (e.g., night), melatonin is synthesized and secreted in abundance, whereas exposure to daylight will immediately suppress the nocturnal synthesis of melatonin. The resulting rhythmic secretion of melatonin, which is intimately tuned to the presence or absence of environmental light, provides the biochemical “signal” for photoperiod in mammals. In other words, the long nights and short days characteristic of winter enable the pineal gland to synthesize melatonin for a longer duration than that which occurs during the short nights and long

days of summer. The increased duration of melatonin release during short days, as compared with long days (i.e., winter vs summer), is responsible for inducing many seasonal changes, including gonadal regression and the cessation of breeding activities. Interestingly, surgical removal of the pineal gland (which removes the endogenous source of melatonin) prevents photoperiodic responses in nearly all mammalian species studied to date. Furthermore, it is the duration of nocturnal melatonin secretion, and not the total amount of melatonin, that appears to provide important photoperiodic information, with longer durations (e.g., > 12 h) associated with short days (i.e., winter) and shorter durations (e.g., < 8 h) associated with long days (i.e., summer).

Following secretion by the pineal during nighttime, melatonin travels to target tissues via the bloodstream, where it can bind to specific receptors located in either the brain or the periphery. Three types of melatonin receptors have been identified: Mel<sub>1a</sub>, Mel<sub>1b</sub>, and Mel<sub>1c</sub>. Mel<sub>1a</sub> and Mel<sub>1b</sub> receptors have been identified in a wide range of species, whereas the Mel<sub>1c</sub> receptor occurs exclusively in invertebrates. Furthermore, melatonin receptors have been identified in a wide variety of brain sites, including the SCN of the hypothalamus, the thalamus, and the pars tuberalis (a structure within the pituitary gland that is involved in regulating endocrine responses), suggesting that these brain sites may be important targets for the actions of melatonin. In addition, melatonin receptors have been isolated in several peripheral tissues, including the spleen, gonads, and adipose tissue.

The modulation of seasonal responses involves the effect of melatonin on brain mechanisms regulating peripheral responses as well as more direct actions of the hormone on organs and glands in the periphery. For example, in the context of seasonal breeding in rodents, melatonin can act on the hypothalamus to regulate the release of gonadotropin-releasing hormone (GnRH), a critical hormone in the endocrine axis regulating reproductive physiology and behavior. Thus, during short winter-like days, prolonged melatonin secretion can exert a negative influence on GnRH, resulting in reduced secretion of the reproductive hormones luteinizing hormone (LH) and follicle-stimulating hormone (FSH) from the anterior pituitary. Decreases in LH and FSH, in turn, act at the level of the gonads (e.g., testes, ovaries) to decrease sex steroid hormones (e.g., testosterone, estrogen). Reductions in these hormones lead to regression of the reproductive system and cessation of reproductive behavior, typical of winter. As

the days grow longer as spring approaches, the brain becomes insensitive to long durations of melatonin. The inhibitory influence of this hormone on the reproductive endocrine axis and reproductive physiology and behavior is restored the following winter.

## CLINICAL SIGNIFICANCE OF BIOLOGICAL RHYTHMS

As mentioned previously, circadian and seasonal rhythms are critical for organisms to survive in a changing environment. To maintain homeostasis, hundreds of physiological and biochemical processes need to be coordinated temporally with both the environment and relative to one another within the body. As a result, normal behavioral, physiological, and psychological functioning is dependent on the body's timing systems.

For example, one particular type of clinical depressive disorder, seasonal affective disorder (SAD), is of particular interest to chronobiologists. SAD is a cyclic illness characterized by recurrent episodes of winter depression alternating with periods of normal affect during the summer. Because of the seasonal component of SAD, it has been suggested that the shortened day lengths of winter may be part of the etiology of this disorder. In fact, exposure to bright lights during early morning or early evening results in marked improvements in clinical symptoms of SAD; therefore, bright light phototherapy has become the treatment of choice for SAD. Although the precise causes of SAD are still not known, abnormal phase shifts in one's circadian rhythms and abnormal patterns of melatonin secretion both have been suggested. Phototherapy may act to shorten the duration of melatonin secretion by suppressing its production early in the morning. Alternatively, phototherapy may act to change the phase of circadian rhythms in numerous physiological factors that may contribute to this seasonal depression.

In addition to seasonal changes in mood, many organisms, including humans, display pronounced seasonal fluctuations in disease and death. Although some of these are due to seasonal changes in the pathogen (e.g., viruses, bacteria), there are also potentially large fluctuations in one's immune system. These changes in immunity likely contribute to the seasonal fluctuations in disease, although researchers are just beginning to understand the physiological mechanisms underlying these changes. As with other seasonal adjustments, seasonal changes in immunity

are regulated by changes in day length. As animals use photoperiod to time changes in reproduction, this same signal can be used to organize the timing of immune alterations throughout the seasons. Because most human societies use artificial lighting yet still experience seasonal changes in disease and death, the mechanisms regulating seasonal changes in human immune function are likely different and require further study.

### See Also the Following Articles

Circadian Rhythms: Hormonal Facets • Melatonin • Pineal Gland

### Further Reading

- Bronson, F. H., and Heideman, P. D. (1994). Seasonal regulation of reproduction in mammals. *In* "The Physiology of Reproduction" (E. Knobil and J. D. Neill, eds.), 2nd ed., pp. 541–584. Raven Press, New York.
- Kriegsfeld, L. J., LeSauter, J., Hamada, T., Pitts, S. M., and Silver, R. (2002). Circadian rhythms in the endocrine system. *In* "Hormones, Brain, and Behavior" (D. Pfaff, *et al.*, eds.), pp. 33–91. Academic Press, San Diego.
- Moore-Ede, M. C., Fuller, C. A., and Sulzman, F. M. (1994). "The Clocks that Time Us." Harvard University Press, Cambridge, MA.
- Refinetti, R. (2000). "Circadian Biology." CRC Press, New York.
- Takahashi, J. S., Turek, F. W., and Moore, R. Y. (2001). *In* "Handbook of Behavioral Neurobiology," Vol. 12: "Circadian Clocks." Kluwer, New York.



# Bisphosphonates

H. Fleisch

University of Bern, Bern, Switzerland

## Glossary

**bisphosphonates** Synthetic compounds characterized by a P–C–P group that are used in various industrial applications and more recently in clinical medicine to decrease bone resorption. The main indications are osteoporosis, tumor bone disease, and Paget's disease.

**osteolytic tumor bone disease** Condition in which tumors of various origins induce bone destruction, which can lead to fractures, pain, and hypercalcemia.

**osteoporosis** Disease characterized by a decrease in bone mass and a deterioration in its architecture and quality, resulting in a greater risk of fracture.

**Paget's disease** A localized and progressive disorder of bone, characterized by increased bone remodeling, bone hypertrophy, and abnormal bone structure.

Bisphosphonates, formerly called diphosphonates, are compounds characterized by two C–P bonds. If the two bonds are located on the same carbon atom, resulting in a P–C–P structure, the compounds are called geminal bisphosphonates; in clinical practice, these are referred to simply as bisphosphonates.

## PRECLINICAL

### Chemistry

Bisphosphonates are analogues of pyrophosphate that contain a carbon instead of an oxygen atom. The P–C–P structure allows a great number of possible variations, especially by changing the two lateral chains on the carbon atom. Many bisphosphonates have been investigated in animals and humans with respect to their effect on bone. Alendronate, clodronate, etidronate, ibandronate, pamidronate, risedronate, tiludronate, and zoledronate are commercially available in various countries for use in human bone disease (see Fig. 1).

Each bisphosphonate has its own physicochemical and biological characteristics. Thus, each compound

must be considered on its own, with respect to both its use and its toxicology. The P–C–P bonds of the bisphosphonates are stable to heat and most chemical reagents and are completely resistant to enzymatic hydrolysis. The bisphosphonates have a strong affinity for metal ions, including calcium.

### Actions

#### Physicochemical Effects

The physicochemical effects of bisphosphonates are very similar to those of pyrophosphate. Thus, they inhibit the formation and aggregation and slow the dissolution of calcium phosphate crystals. These effects are related to the marked affinity of these compounds for solid-phase calcium phosphate, to the surface of which they bind strongly. This property is the basis for the use of these compounds as skeletal markers in nuclear medicine and the basis for their selective pharmacological effect on bone.

#### Inhibition of Bone Resorption

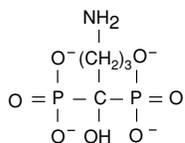
The main effect of the pharmacologically active bisphosphonates is to inhibit bone resorption.

#### In Vitro

Bisphosphonates block bone resorption induced by various means in cell and organ culture. In the former, they inhibit the formation of pits by isolated osteoclasts cultured on mineralized substrata. In organ culture, they decrease the destruction of bone in embryonic long bones and in neonatal calvaria. This inhibition is present whether or not resorption is stimulated. At present, results indicate that the effect of all the stimulators of bone resorption, such as parathyroid hormone, 1,25(OH)<sub>2</sub> vitamin D, and prostaglandins, as well as the products of tumor cells, has been inhibited by bisphosphonates.

#### In Intact Animals

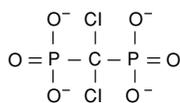
In growing rats, bisphosphonates can block the degradation of both primary and secondary trabeculae,



(4-Amino-1-hydroxybutylidene)-bis-phosphonate

**Alendronate\***

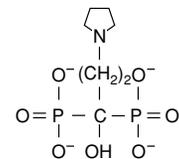
*Abiogen; Merck sharp & Dohme*



(Dichloromethylene)-bis-phosphonate

**Clodronate\***

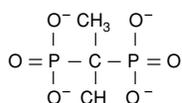
*Abiogen; Astra  
Leiras; Rhône-Poulenc Rorer; Roche*



[1-Hydroxy-3-(1-pyrrolidiny)-propylidene]bis-phosphonate

**EB-1053**

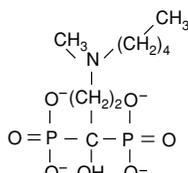
*Leo*



(1-Hydroxyethylidene)-bis-phosphonate

**Etidronate\***

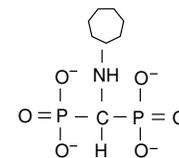
*Abiogen; Procter & Gamble*



[1-Hydroxy-3-(methylpentylamino)propylidene]bis-phosphonate

**Ibandronate\***

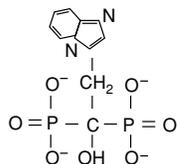
*Roche*



[(Cycloheptylamino)-methylene]bis-phosphonate

**Incadronate**

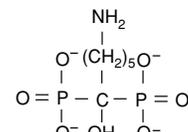
*Yamanouchi*



[1-Hydroxy-2-imidazo-(1,2-a)pyridin-3-ethylidene]bis-phosphonate

**Minodronate**

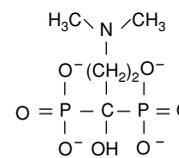
*Yamanouchi-Hoechst*



(6-Amino-1-hydroxyhexylidene)bis-phosphonate

**Neridronate**

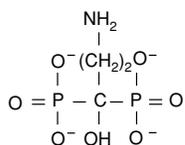
*Abiogen*



[3-(Dimethylamino)-1-hydroxypropylidene]bis-phosphonate

**Olpadronate**

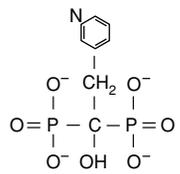
*Gador*



(3-Amino-1-hydroxypropylidene)bis-phosphonate

**Pamidronate\***

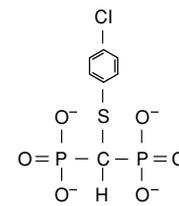
*Gador; Novartis*



[1-Hydroxy-2-(3-pyridinyl)ethylidene]bis-phosphonate

**Risedronate\***

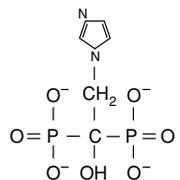
*Procter & Gamble*



[[4-Chlorophenyl]thio]methylene]bis-phosphonate

**Tiludronate\***

*Sanofi*



[1-Hydroxy-2-(1H-imidazole-1-yl)ethylidene]bis-phosphonate

**Zoledronate\***

*Novartis*

**Figure 1** Chemical structure of bisphosphonates investigated for their effect on bone in humans. Asterisks indicate commercially available bisphosphonates. Reprinted from Fleisch (2000), with permission.

thus arresting the modeling and remodeling of the metaphysis. The latter therefore becomes club-shaped and radiologically denser than normal. This effect is often used as an experimental assay to estimate the potency of new compounds (Schenk test). The inhibition of bone resorption by bisphosphonates has also been documented using  $^{45}\text{Ca}$  kinetic studies and hydroxyproline excretion, as well as by other means. The effect occurs within 24–48 h and is therefore slower than that of calcitonin. The decrease in resorption is accompanied, at least in the growing animal, by a positive calcium balance and an increase in the mineral content of bone and in bone mass. This is possible because of an increase in intestinal calcium absorption. The increase is, however, smaller than expected because after a certain time, bone formation also decreases, so that the main effect of bisphosphonates is a reduction in bone turnover.

#### *In Animals with Experimentally Induced Bone Resorption*

Bisphosphonates can also prevent an experimentally induced increase in bone resorption. Thus, they impair resorption induced by many bone resorbing agents, such as parathyroid hormone, 1,25(OH) $_2$  vitamin D, and retinoids, the effect of the last agent having been used to develop a powerful and rapid screening assay for new compounds.

Bisphosphonates also inhibit bone loss induced by different procedures to induce experimental osteoporosis, such as immobilization, ovariectomy, administration of corticosteroids, or lactation combined with a low-calcium diet. When not given in excess, bisphosphonates also have a positive effect on mechanical characteristics both in normal animals and in various experimental osteoporosis models. This effect seems to be due to alterations in bone mass, architecture, and quality.

Finally, bisphosphonates also inhibit bone resorption induced experimentally by implantation of various tumor cells. *In vivo*, three types of models have been investigated, namely, the subcutaneous implantation of tumor cells leading to tumoral hypercalcemia, their intracardiac injection inducing osseous metastases, and their implantation next to bone-inducing local erosion. In all three models, the tumoral effect is diminished by bisphosphonates. This effect is the basis for their use in tumor-related bone disease. It used to be thought that bisphosphonates had no direct effect on tumor cells. However, recent results show that *in vitro* they inhibit the adhesion and spreading of tumor cells and induce the apoptosis of tumor cells. They may also act indirectly by decreasing, as a result of the lower bone resorption, the release from bone of

cytokines that stimulate tumor cell multiplication, decreasing the tumor burden in the skeleton.

#### *Relative Activity of Bisphosphonates in Inhibiting Bone Resorption*

The activity of bisphosphonates on bone resorption varies greatly from compound to compound. For etidronate, the dose required to inhibit resorption is relatively high; in the rat, the dose is greater than 1 mg/kg parenterally per day. This dose is very near that which impairs normal mineralization. One of the aims of bisphosphonate research has therefore been to develop compounds with a more powerful anti-resorptive activity, without a stronger inhibition of mineralization. This has proven to be possible. Compounds have been developed that are up to 10,000 times more powerful than etidronate in the inhibition of bone resorption in experimental animals without being more active at inhibiting mineralization.

At present, the structural requirements for activity are only partially defined. The length of the aliphatic carbon chain is important. The presence of a hydroxyl group bound to the carbon atom at position 1, the presence of a nitrogen atom, or the presence of cyclic substituents increases activity.

#### *Mechanisms of Action*

There is no doubt that the action *in vivo* is mediated through mechanisms other than the physicochemical inhibition of crystal dissolution, as was initially postulated, namely, by acting on the osteoclast. Indeed, both the number and the morphology of the latter are altered with signs of apoptosis and alterations in the cytoskeleton and the ruffled border. Four mechanisms are likely to be involved: (1) inhibition of osteoclast recruitment; (2) possible inhibition of osteoclastic adhesion; (3) shortening of the life span of osteoclasts due to earlier apoptosis; and (4) inhibition of osteoclast activity. The first three will lead to a decrease in the number of osteoclasts, which is usually seen after treatment. The fourth will lead to inactive osteoclasts. These effects are made possible by the uptake of these compounds by the osteoclasts during the resorption process, a process favored by the fact that the bisphosphonates also deposit preferentially under the osteoclasts, where they can attain very high concentrations, in the range of  $10^{-4}$  M or higher.

Very recently, the cellular mechanism was partially unraveled. It was found that nitrogen-containing bisphosphonates can inhibit the mevalonate pathway, by inhibiting farnesyl pyrophosphate synthase and possibly some other enzymes in this pathway. This leads to a decrease in the formation of isoprenoid

lipids, such as farnesyl- and geranylgeranylpyrophosphates. These are required for the posttranslational prenylation (transfer of fatty acid chains) of proteins, including the GTP-binding proteins Ras, Rho, Rac, and Rab, which are important for many cell functions, including cytoskeletal assembly and intracellular signaling. Therefore, N-containing bisphosphonates will induce a series of changes leading to decreased activity, probably the main effect, and to earlier apoptosis in several cell types, including osteoclasts. In osteoclasts, the lack of geranylgeranylpyrophosphate appears to be responsible for the effects.

In contrast, some non-nitrogen-containing bisphosphonates that closely resemble pyrophosphate, such as clodronate, etidronate, and tiludronate, can be incorporated into the phosphate chain of ATP-containing compounds so that they become nonhydrolyzable. The new P–C–P-containing ATP (AppCCl<sub>2</sub>p) analogues impair cell function by decreasing mitochondrial function through an inhibition of ADP/ATP translocase. This may lead to apoptosis and cell death. Thus, the bisphosphonates can be classified into two major groups with different modes of action.

Bisphosphonates also have some effects on osteoblasts, such as decreasing their apoptosis and stimulating them to secrete an inhibitor(s) of osteoclast recruitment. The role *in vivo* of these effects is not yet clear.

#### **Inhibition of Mineralization**

Bisphosphonates inhibit calcification *in vivo* very efficiently. Thus, they prevent experimentally induced calcification of many soft tissues, both when given parenterally and when given orally. However, they also inhibit, at similar doses, the mineralization of normal calcified tissues. The amount required to have this effect is similar for the various bisphosphonates, in contrast to bone resorption, where the different compounds vary greatly in their activity. The inhibition of calcification can lead to fractures and to an impairment of fracture healing. The propensity to inhibit mineralization of normal bone has hampered the therapeutic use of bisphosphonates in ectopic calcification. It is probable that the effect on mineralization is due to a physicochemical mechanism of crystal growth inhibition following the tight binding of the compound to the mineral surface.

#### **Pharmacokinetics**

The bisphosphonates on which data have been published appear to be absorbed, stored, and excreted

unaltered in the body. Therefore, these bisphosphonates seem to be nonbiodegradable, at least with respect to their P–C–P bond.

#### **Intestinal Absorption**

The bioavailability of an oral dose of a bisphosphonate in animals as well as in humans ranges between less than 1 and 10%, probably because of their low lipophilicity, which prevents transcellular transport, and their high negative charge, which prevents paracellular transport. Bioavailability shows great inter- and intraspecies variation. Absorption is substantially diminished when the drug is given with meals, especially in the presence of calcium and iron. Therefore, bisphosphonates should never be given at mealtimes and never together with milk or dairy products or with calcium or iron supplements.

#### **Distribution**

The half-life of circulating bisphosphonates is short, on the order of minutes to hours. Some 20–80% of the absorbed bisphosphonate is taken up very rapidly by bone and the remainder is rapidly excreted in the urine. This rapid uptake by bone means that the soft tissues are exposed to bisphosphonates for only short periods, explaining why practically only bone is affected *in vivo*. The areas of deposition were generally thought to be mostly those of bone formation. This property is used to measure areas of high bone turnover in nuclear medicine by means of <sup>99m</sup>Tc-linked bisphosphonates. However, bisphosphonates have also recently been found to accumulate under the osteoclasts.

Once deposited in the skeleton and covered under new layers of bone, the bisphosphonates will be released to a large extent only when the bone in which they were deposited is resorbed. The half-life in bone bisphosphonates is therefore very long; for humans, it can be over 10 years. However, there is no indication that the bisphosphonates in the skeleton usually have any pharmacological activity as long as they are buried. But this may not be true at the sites of high turnover, where large amounts of the drug are deposited focally, for example, in patients with bone metastases or with Paget's disease. This would explain why a single administration of a bisphosphonate can be active for long periods of time in these conditions.

#### **Renal Clearance**

The rate of renal clearance of bisphosphonates is high. When it is taken into account that they have only partial ultrafilterability, the rate of renal clearance

can be, at least in animals, higher than that of inulin, indicating active secretion.

## CLINICAL

### Paget's Disease

#### **Clinical Features**

Paget's disease is a localized and progressive disorder of bone, with a genetic predisposition, presenting itself usually after the age of 40, and characterized by increased bone remodeling, bone hypertrophy, and abnormal bone structure. It is in some countries the second most common metabolic bone disease after osteoporosis.

For as yet unknown reasons, bone turnover becomes abnormally increased at certain sites in the skeleton; the cause is probably a slow paramyxovirus infection. The first event is a marked elevation of bone resorption, followed by a compensatory increase in formation. This leads to a mosaic of lytic and sclerotic lesions and local deformations of the skeleton.

The condition is most commonly asymptomatic. The most common complaint is pain, which may be very severe, and bone deformity. Laboratory findings reflect increased bone turnover. Thus, the indices of both bone formation, reflected by increased total and bone-specific serum alkaline phosphatase activity, and resorption, as assessed today by urinary pyridinium cross-links, especially the C-terminal and N-terminal cross-linked telopeptides of type I collagen, are elevated. Scintigraphy with  $^{99m}$ technetium-labeled compounds, in which the lesions show up as hot spots, is of great diagnostic value.

#### **Treatment**

Treatment should be offered to all symptomatic patients as well as to asymptomatic patients with involvement of skeletal areas that have the potential to give rise to complications, such as the skull, vertebral bodies, long bones, and areas near major joints. Practically the only treatment, apart from bisphosphonates, is calcitonin. Calcitonin has, however, some drawbacks, so that the bisphosphonates are now the therapy of choice in Paget's disease.

#### **Effects of Bisphosphonate Therapy**

All bisphosphonates that have been tested thus far have proven to be active in decreasing bone turnover. The difference between the compounds lies in their potency and in their adverse event profile. Most clinical studies in Paget's disease deal with alendronate, clodronate, etidronate, pamidronate, risedronate, and tiludronate. Other bisphosphonates

showing clinical efficacy include ibandronate, neridronate, olpadronate, and zoledronate. The effects observed are qualitatively very similar, but differ with respect to the rapidity with which they develop, the absolute reduction of turnover, and the duration of effect.

Both bone resorption and formation are decreased. The effect on resorption precedes the effect on formation, suggesting that, as in animals, the decrease in formation is secondary, due to the coupling between the two processes. This time differential has a practical consequence in that markers of bone resorption are a better way to assess the acute effect of treatment than alkaline phosphatase. The latter will give useful information only after approximately 4 weeks and sometimes even longer. The decrease in bone turnover can be accompanied by a small decrease in serum ionized calcium and an elevation of serum parathyroid hormone (PTH). The last two results are often seen when patients with very active osteolysis are treated with a powerful inhibitor of bone resorption. In order to avoid the secondary increase in PTH, it is strongly advised that the bisphosphonates be administered, especially in patients with severe disease or elderly people, together with 0.5–1.0 g calcium and 400–800 units of vitamin D per day.

Morphological studies confirm the reduction in turnover. The number of osteoclasts is diminished, but the virus-like cellular inclusions in the nuclei and the measles-type viral antigens in the osteoclasts persist unmodified. The bone formed during treatment with bisphosphonates often returns to a lamellar organization, in contrast to the woven bone formation typical of this disease.

Bone pain usually decreases and can disappear completely, except when it is due to arthritic changes. Neurological spinal syndromes can be improved and elevated cardiac output can be normalized. The pathological uptake of radioactive technetium is markedly decreased and some of the radiological alterations may be improved.

The effect of treatment is monitored by one of the biochemical markers of bone turnover. Serum alkaline phosphatase is currently the simplest and least expensive marker for bone formation. Collagen I pyridinium cross-linked telopeptides are becoming increasingly used for bone resorption. The aim of the treatment is to decrease bone turnover to the normal range. Indeed, a relationship appears to exist between the decrease in turnover obtained during treatment and the duration of the effect. If a normalization of turnover is not possible with a specific bisphosphonate, even after longer therapy, a more potent compound

should be used if available. Patients with more active forms of the disease require larger total amounts of bisphosphonates. The plateau is usually reached within 3–6 months for alkaline phosphatase in the case of oral administration and is faster for the parameters of bone resorption. Treatment is then discontinued until the indices of bone turnover start to increase again or symptoms recur. An increase above the normal range, or if the latter was not attained, an increase of at least 25%, seems to be a reasonable guideline for resuming treatment. However, the decrease in bone turnover, as well as the other improvements, can also last for a long time, often many years, after the discontinuation of treatment. In some patients, resistance to the bisphosphonate may develop with time. The cause of this resistance is unknown. Sometimes it can be overcome by increasing the dose; in other cases, it can be overcome by switching to another bisphosphonate, a more potent one, if possible.

#### *Treatment Regimens for Widely Available*

##### *Commercial Compounds*

- *Alendronate*: Oral administration of 40 mg daily for 6 months is the dose recommended by the manufacturer.

- *Clodronate*: A daily oral dose of 1600 mg *per os* for 6 months is used in some countries.

- *Etidronate*: This compound was the first to be used in this disease but is practically not administered any longer, because the required doses to normalize bone turnover also induce an inhibition of mineralization.

- *Pamidronate*: Intravenous infusions of 60–90 mg given once are usually used. Treatment can be repeated after 6 months if the result is not satisfactory.

- *Risedronate*: A dose of 30 mg daily for 2 months, followed by another course 6 months later, if necessary, is recommended by the manufacturer.

- *Tiludronate*: Daily oral administration of 400 mg for 3–6 months is used in some countries, but this treatment is less efficient than some of the above-mentioned regimens.

#### **Conclusion**

Bisphosphonates can normalize bone turnover in most patients and improve the clinical signs and symptoms. The improvements can last for many years after discontinuation of therapy. They are today the drug of choice for treatment of Paget's disease.

## **Osteolytic Tumor-Induced Bone Disease**

### **Clinical Features**

Osteolytic tumor-induced bone disease is a condition in which tumors of various origins, mostly breast, lung, prostate, and multiple myeloma, induce bone destruction, which can lead to fractures, pain, and hypercalcemia with its symptomatology. The osteolysis occurs either through local invasion or at distance by secreting into the bloodstream bone-resorbing products, mostly PTH-related protein (PTHrP). Often they work through both mechanisms. The mechanism of local resorption of bone involves the secretion of an array of bone resorbing cytokines, including PTH-related protein, by the tumor cells or by other cells that are stimulated by said tumor cells. If generalized bone destruction is accompanied by hypercalcemia, the term humoral hypercalcemia of malignancy is used. The hypercalcemia is then due to both increased bone destruction and increased renal tubular reabsorption, both being stimulated by PTHrP. The fact that hypercalcemia is produced by various mechanisms has therapeutic consequences. Because only bone resorption is influenced by bisphosphonates, plasma calcium will be less influenced by these compounds in patients in which nonosteolytic mechanisms, that is renal reabsorption, are prominent. This is the case, for example, in humoral hypercalcemia of malignancy.

The parameters usually followed for diagnosis and during the treatment of tumor-induced bone disease are primarily pain, fractures, and osteolytic foci, as well as plasma calcium, urinary calcium, and pyridinium cross-links or their peptides. The radiophysical techniques used include X-rays, scintigraphy, computed tomography, and magnetic resonance imaging.

### **Treatment**

Treatment of tumor-induced bone disease is of course primarily the usual cancer therapy, such as excision of the primary tumor, chemotherapy, hormonal therapy, or radiotherapy. Hypercalcemia in the acute stage is first treated with fluid repletion. Thus, the bisphosphonates should by no means replace the latter treatments, but should be used only as a supportive treatment to improve the quality of life. They are effective in most tumors, including breast, lung, myeloma, and prostate, at decreasing hypercalcemia as well as decreasing the occurrence of skeletal complications, such as fractures, pain, and new osteolytic lesions.

### Effects of Bisphosphonate Therapy

Bisphosphonates are currently the drugs of choice for the treatment of tumoral hypercalcemia in patients in whom cancer therapy and fluid repletion are not effective. They effectively reduce and often normalize hypercalcemia. With intravenous administration, the effect starts to be clinically significant after 2–3 days; normocalcemia is obtained after 3–5 days and the full effect is obtained after approximately 1 week. The effect is more pronounced in patients in whom hypercalcemia is completely or largely a result of bone resorption only, as is the case, for example, in myeloma. Hypercalciuria is greatly decreased by treatment in all types of tumor. The markers of bone resorption, urinary pyridinium cross-links and their C- and N-terminal telopeptides, are markedly decreased. The most practical procedure for treating tumoral hypercalcemia is to monitor plasma calcium and to resume treatment when hypercalcemia resumes. The duration of the effect after discontinuation of treatment is variable, usually less than 1 month. It seems to depend on the potency of the bisphosphonate, the total dose administered, and the type of bone disease. In patients with focal involvement, resorption appears to be inhibited for a longer time than in those with humoral hypercalcemia.

Bisphosphonates also decrease the occurrence of skeletal-related events, such as bone pain, pathological fractures, vertebral collapse, spinal cord compression, the requirement of radiation and bone surgery, as well as the development of new osteolytic foci, leading sometimes to a marked improvement in the quality of life. The onset of these effects is relatively slow. The efficacy is lost after discontinuation of the drug. Recent results with clodronate suggest that bisphosphonates may possibly also diminish in some cases the formation of bone metastases in patients who have not developed them yet. This effect, if confirmed, would open the possibility of using these compounds in patients who have not yet developed skeletal metastases.

### Treatment Regimens for Widely Available Commercial Compounds

From the available data, there is no indication that there are any fundamental differences in the qualitative effect of the various bisphosphonates in tumor-induced bone disease. There is, however, a great difference in their quantitative effect. If resistance to the treatment develops, either increasing the dose or switching to another, preferably more potent, bisphosphonate may solve the problem.

- *Alendronate*: This compound is commercially available only in Japan for hypercalcemia of malignancy at a dose of 10 mg, or if necessary 20 mg, administered in intravenous infusions.

- *Clodronate*: For hypercalcemia, one infusion of 1500 mg in no less than 500 ml is usually used. Sometimes oral treatment of 1600 mg is administered daily as maintenance therapy. The latter is especially used as long-term therapy for skeletal complications of tumor bone disease.

- *Etidronate*: Among the various bisphosphonates used, the effect of etidronate, which is the least potent, is the least satisfactory and the compound is used very little today.

- *Ibandronate*: This compound, administered as a single intravenous (iv) infusion of 2–4 mg, is used for hypercalcemia.

- *Pamidronate*: One infusion of pamidronate is usually sufficient to normalize plasma calcium in hypercalcemia. The dose can be adapted to the degree of hypercalcemia and is usually between 60 and 90 mg. Treatment is repeated when plasma calcium rises again. In order to reduce or prevent skeletal morbidity, infusions of 90 mg monthly are usually recommended.

- *Zoledronate*: Infusion of 4 mg iv of this compound has a more powerful effect on hypercalcemia than 90 mg iv of pamidronate. Zoledronate is commercially available for tumor-induced bone disease at this dose, infused intravenously every 3–4 weeks.

### Conclusion

In addition to classic anti-neoplastic treatments, bisphosphonates are today the drugs of choice for treating tumor-induced bone disease. By inhibiting bone resorption, bisphosphonates correct hypercalcemia, reduce pain, prevent the development of new osteolytic lesions, and reduce the occurrence of fractures, and as a consequence improve the quality of life.

### Non-Tumor-Induced Hypercalcemia

Only a few data are available concerning the use of bisphosphonates in hypercalcemias due to causes other than malignant bone disease. Most of the reports are on hyperparathyroidism. The effects seem to be limited and of short duration, also because part of the elevation in plasma calcium is due to increased reabsorption in the kidney. The compounds that are most frequently used are pamidronate and clodronate.

## Osteoporosis

### Clinical Features

Osteoporosis is a disease characterized by a decrease in bone mass and a deterioration in the architecture of the bones, which leads to an enhanced fragility of the skeleton and, therefore, to a greater risk of fracture. It is defined as being present in women when the bone mineral density is more than 2.5 standard deviations below that of the young female adult (*T*-score). It is a very common disorder that will become even more common with increased life expectancy. It is also frequent in men, although less so than in women. Its main cause is the continuous loss during life of both cancellous and cortical bone, which is exacerbated in women after menopause. The second contributory factor is failure to achieve adequate peak bone mass during adolescence. The causes of these changes are not yet clear, although genetic factors are involved, at least for the latter factor.

The clinical manifestations of osteoporosis are fractures, often occurring spontaneously or after minimal trauma, and their consequences. Osteoporosis is diagnosed and assessed quantitatively by techniques that measure bone mineral density (BMD), most commonly dual X-ray absorptiometry. Chemical analyses cannot be used to diagnose osteoporosis. Markers of bone turnover, however, are useful to determine bone turnover and consequently to identify those patients who are likely to be losing bone rapidly and to follow the effect of treatment.

### Treatment

Many remedies are used to prevent and treat osteoporosis. Practically all of them aim to decrease bone resorption. Probably the treatment that was the most often used and the most effective for osteoporosis, apart from the bisphosphonates, was estrogen replacement after menopause. However, it was shown recently that estrogen replacement can increase the risk of breast cancer and of vascular insults when given over a longer time. Calcitonin is sometimes used, but parenteral administration sometimes has unpleasant side effects and the nasal form has a relatively weak effect on BMD and fracture incidence. Calcium can also decrease bone turnover and diminish bone loss in certain conditions. A daily intake of 1–1.5 g is recommended in the adult, especially during lactation and in the elderly. Vitamin D should be present in sufficient amounts and is generally recommended as a supplement for elderly people. Today only very few compounds are able to increase bone formation, namely, fluoride and PTH. Fluoride has not been shown to

decrease the occurrence of fractures. However, PTH appears very promising, since it increases BMD and decreases the occurrence of fractures.

### Effects of Bisphosphonate Therapy

Most of the studies on the effects of bisphosphonate therapy have been performed with alendronate, etidronate, and risedronate. Many well-controlled studies proved the efficacy of bisphosphonates in preventing skeletal loss, actually increasing skeletal calcium as assessed by BMD, and in decreasing bone turnover. However, BMD is not necessarily a true reflection of actual bone mass, since it is influenced not only by the amount of bone, but also by the degree of mineralization of the bone present, which increases when bone turnover is decreased. Bisphosphonates are active in white, Asian, and black osteoporotic women. They are also effective in the elderly and healthy women, as well as in men. They prevent and even partially reverse the bone loss in glucocorticoid-treated patients and are therefore a standard therapy in patients receiving this drug over a longer time. When given together with hormone replacement therapy or raloxifene, they have a greater effect on bone mineral density than either treatment given alone. All the bisphosphonates induce a marked decrease in bone turnover when given in doses effective on bone mineral density. Both bone formation and bone resorption are decreased. Both alendronate and risedronate decrease by approximately one-half the occurrence of vertebral and nonvertebral fractures in osteoporotic patients. This effect is due both to the increase in BMD and to the decrease in bone turnover.

After discontinuation of short-term treatment, bone turnover returns to pretreatment values within months and bone loss appears to resume, although at a later time. This reversal occurs more slowly and later if the bisphosphonates have been given over a longer period, such as years.

### Treatment Regimens for Widely Available

#### Commercial Compounds

- *Alendronate*: The dose recommended by the manufacturers is 10 mg orally daily and 5 mg in Japan. Since this compound has a similar effect on BMD when given once weekly at 70 mg, the weekly regimen is used today in countries where this regimen is commercially available.
- *Etidronate*: The regimen recommended by the manufacturer is 400 mg daily orally for 2 weeks every 3 months.
- *Risedronate*: The recommended regimen is 5 mg daily orally or 35 mg once weekly.

**Conclusion**

Bisphosphonates stop bone loss, increase BMD, and decrease bone turnover in various types of osteoporosis in women and men. This leads to a decrease in vertebral and nonvertebral fractures. Bisphosphonates are an important addition to the therapeutic modalities available for the treatment and prevention of osteoporosis.

**Heterotopic Calcification and Ossification**

Etidronate is the only bisphosphonate investigated thus far for the prevention of ectopic calcification and ossification. Although it appears to be useful in some cases, this has not yet proven by double-blind studies. Furthermore, the effective dose is the same as that which inhibits normal mineralization, which makes its use difficult. Etidronate is commercially available in some countries at an oral dose of 20 mg/kg initially for 2–4 weeks before the intervention, followed by 10 mg/kg for 10–12 weeks to prevent heterotopic ossifications following a hip prosthesis implantation and after spinal lesions with immobilization.

**Other Diseases**

Bisphosphonates can be useful treatments in other diseases. Among these diseases, osteogenesis imperfecta, fibrous dysplasia of bone, and reflex sympathetic dystrophy syndrome are the best known.

**Adverse Events**

As was the case in animals, studies in humans have revealed only a few important adverse events. The most frequently seen are the following: Rapid injection led to renal failure, probably because of the formation of a solid phase of bisphosphonate in the blood, which is then held back in the kidney. No further events of that kind have been observed since bisphosphonates, when given intravenously, are administered by slow infusion in a large volume of fluid, usually 250–500 ml. Bisphosphonates may sometimes induce a certain degree of hypocalcemia, which is usually clinically irrelevant. An exception may be the association with aminoglycoside antibiotics with which very severe hypocalcemia can occur, so that the two drugs should not be administered together. Oral administration of bisphosphonates, especially those containing a nitrogen atom, can be accompanied by digestive tract disturbances. The latter can be substantially reduced by taking the drug with enough fluid and by not reclining after the intake.

They appear to be further reduced by administering the compounds at the same total dose weekly instead of daily. The intravenous administration of N-containing compounds can induce a transient pyrexia of usually 1–2°C, accompanied by flu-like symptoms. There is an increase in serum C-reactive protein, so that the events resemble an acute-phase response. As yet, no negative consequences of these episodes have been described. Finally, compounds with little efficiency and which must be administered in higher doses, such as etidronate, can inhibit normal skeletal mineralization. This can happen at doses of etidronate that are greater than 800 mg daily.

**Contraindications**

At present, only a few contraindications have been described for the bisphosphonates. The question is often raised as to whether these compounds can be administered for renal failure. Because they are cleared from the blood to a large extent by the skeleton, there is no theoretical reason to deny bisphosphonates in patients with moderate renal failure. However, plasma levels are likely to be higher, so that the dose should possibly be reduced. Fracture healing or new orthopedic implants are not contraindications to the use of bisphosphonates, provided they are not given in doses that inhibit mineralization. In view of the potential for gastrointestinal effects of bisphosphonates during oral administration, this mode of administration should be used with caution in patients with inflammatory gastrointestinal and esophageal conditions and with esophageal dysmotility. Finally, bisphosphonates should not be given during pregnancy and lactation.

**Future Prospects**

The bisphosphonates represent an important development in the treatment of bone diseases and it is possible that this is only the beginning of a new era of therapy. Many issues are still unresolved. For example, it is not yet known whether the optimal regimen for the various compounds available has been found. Will there be an advantage in the future to combine bisphosphonates with other inhibitors of resorption or with a stimulator of bone formation? In view of the persistence of bisphosphonates in the body, which is a concern for some, it might be considered to interrupt treatment for osteoporosis for a certain time. It may be possible in the future to devise drugs that have effects similar to those of the bisphosphonates, but are metabolically broken down.

Possibly bisphosphonates can be of use in diseases other than bone diseases, such as arthritis, dental disease, and tumors in general. Finally, bisphosphonates may be used as carriers for drugs to be brought to the bone or to other calcified tissues.

### See Also the Following Articles

Bone Remodeling, Dynamics of • Bone Structure • Hypercalcemia and Hypercalcemia Treatment • Osteoporosis, Overview • Paget's Disease of Bone

### Further Reading

- Body, J.-J. (ed.) (2000). "Tumor Bone Diseases and Osteoporosis in Cancer Patients: Pathophysiology, Diagnosis, and Therapy." Dekker, New York.
- Fleisch, H. (2000). "Bisphosphonates in Bone Disease: From the Laboratory to the Patient," 4th ed. Academic Press, San Diego, CA.
- Kanis, J. A. (1998). "Pathophysiology and Treatment of Paget's Disease of Bone," 2nd ed. Martin Dunitz, London.
- Marcus, R., Feldman, D., and Kelsey, J. (eds.) (2001). "Osteoporosis." Academic Press, San Diego, CA.



# Body Composition During Growth

Kenneth J. Ellis

Baylor College of Medicine and Texas Children's Hospital, Houston, Texas, United States

## Glossary

**body cell mass** The cellular mass of the body containing the oxygen-exchanging, potassium-rich, glucose-oxidizing, work-performing tissue.

**body composition** The mass of the human body can be described in terms of its chemical, molecular, cellular, physiological, and anatomical makeup.

**bone mineral z-score calculator** Normalizes the measured bone mineral content for an individual child with a pediatric reference population based on age, gender, ethnicity, and height.

**reference model** Multicomponent model of body composition used to describe changes in the relative proportions of bone, muscle, visceral tissues, and fat stores during normal growth.

**G**rowth is not only a change in height but also a change in the relative proportions of bone, muscle, water, visceral tissues, and body fat that comprise body weight.

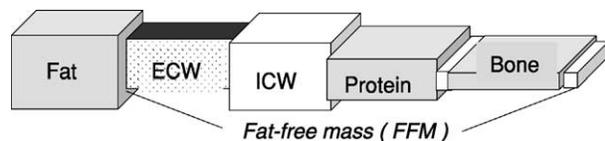
## INTRODUCTION

Based on an extensive review of the scientific literature, a task group of the International Commission on Radiological Protection developed a comprehensive model of the chemical and compositional makeup of the young adult male body, called Reference Man. During growth, however, not only does the human body increase in size but also there are changes in the relative proportions of bone, muscle, visceral tissues, and fat stores. To describe these changes, Samuel Fomon and colleagues developed the concepts of the Reference Infant, Reference Child, and Reference Adolescent models of body composition. These models have proven to be very practical on a population basis for estimating the nutritional requirements needed in support of normal growth. There has been an increasing clinical interest in body composition, and

the Fomon models have often been used as the reference or standard for examining abnormal growth. Although these reference models have been of great value for establishing pediatric nutritional requirements, there were several limitations in the modeling process that the authors acknowledged could constrain their utilization for the assessment of the individual child. These included the paucity of reliable pediatric body composition data 30 years ago, which necessitated the pooling of data from different populations, and the extrapolation or interpolation of missing values at many ages. Also, there were insufficient data to adequately develop models suitable for minority or non-Caucasian children. This article provides a contemporary update to the basic Fomon models using body composition data as reported by Nancy Butte and colleagues for infants and by Kenneth Ellis and colleagues for older children and adolescents from a multiethnic population.

## CONSTRUCTION OF A REFERENCE MODEL

The modeling scheme originally developed by Fomon and colleagues was used to describe the changes in body composition during growth for contemporary populations. This approach was chosen to allow for a direct comparison with Fomon's models. The basic components of this model are illustrated in Fig. 1. To construct this model, three separate *in vivo* measurements are needed: total body potassium (TBK), total body water (TBW), and total body bone mineral content (BMC). In order to obtain estimates for the subcategories of the fat-free mass (FFM), the following set of assumptions were used: (i) intracellular and extracellular potassium concentrations of 150 and 4 mEq/liter, respectively; (ii) intracellular and extracellular mineral concentrations of 9.0 and 9.4 g/liter, respectively; (iii) nitrogen content of protein is 160 g/kg; (iv) lean tissue accretion during growth contains 2.17 mEq potassium per gram of nitrogen; and (v) glycogen



**Figure 1** Multicompartment model of body composition. Body weight = fat mass + fat-free mass (FFM), where total body water = extracellular water (ECW) + intracellular water (ICW).

makes up 0.6% of the FFM. Based on these assumptions and the measurements of TBK, TBW, and BMC, the size of the subcategories of FFM can be derived using the following set of equations:

$$\text{Extracellular water (ECW)} = 1.027 \times \text{TBW (liter)} - 0.00685 \times \text{TBK (mEq)}$$

$$\text{Intracellular water (ICW)} = \text{TBW (liter)} - \text{ECW (liter)}$$

$$\text{Protein (kg)} = 0.002881 \times \text{TBK (mEq)}$$

$$\text{Glycogen (kg)} = 0.006 \times \text{FFM (kg)}$$

$$\text{Nonosseous mineral} = [9.0 \times \text{ICW} + 9.4 \times \text{ECW}] \times 10^{-3}$$

$$\text{Osseous mineral (kg)} = \text{BMC}$$

Fat-free mass was defined as the sum of the TBW, protein, mineral, and glycogen compartments. Using the Fomon modeling approach, the body's fat mass (FM) was calculated as body weight minus FFM.

## BODY COMPOSITION MEASUREMENTS

The review article by Kenneth Ellis provides a comprehensive description of the various body composition techniques that are used to assess the living human body. For the assay of TBW, the deuterium dilution method is most frequently used. For infants and toddlers, urine samples are collected at baseline and 3–5 h after administration of an oral dose of the nonradioactive tracer. For older children, plasma is the preferred fluid sample taken at the baseline and postdose time points. Measurement of the deuterium content of the fluid sample can be obtained using gas-isotope-ratio mass spectrometry. During the 3- to 5-h equilibration period, some of the deuterium will exchange with nonaqueous hydrogen in the body, such that the deuterium dilution space has to be corrected by approximately 4% to obtain an accurate measure of TBW. The body water assay technique has a reported precision of 1 or 2%.

The total body potassium (TBK) measurement is based on a unique nuclear property of natural potassium. A small percentage of natural potassium exists in a radioactive ( $^{40}\text{K}$ ) state, such that approximately 200 gamma rays per minute are emitted per gram potassium. The energy of these gamma rays is sufficiently high (1.46 MeV) that most of this signal can be detected external to the body. This is accomplished using an instrument called a whole-body counter. Typical counting times are less than 15 min, with a reported precision of  $\pm 1$ –3% depending on the subject's body size and the design of the counter.

The measurement of BMC of the whole body can be obtained using dual-energy X-ray absorptiometry (DXA). This technology did not exist at the time the original Fomon models were developed. For the DXA measurement, a narrowly collimated X-ray beam is passed through the body, and the relative intensity of the transmitted beam is measured on the opposite side of the body using an array of detectors. Two different energies of X rays are used to separate the attenuation response due to the difference in density between bone and soft tissue. Newer DXA instruments can scan the total body in a few minutes. There is a small exposure to X rays resulting from the scan, but it is well within the variation in the natural background levels throughout the United States. The precision for whole-body BMC measurements is reported to be  $\pm 1\%$  for older children and adolescents, increasing to approximately  $\pm 3\%$  for infants. An important feature of the whole-body DXA measurement is that estimates of the body's FM and nonfat lean tissue mass can also be obtained. However, in this article, FM is estimated using only the model developed by Fomon (i.e., FM = body weight – FFM).

## BODY COMPOSITION REFERENCE DATA

The tabulated and graphical results presented in this article are derived from the data of Butte and colleagues for infants and Ellis and colleagues for older children and adolescents. Stepwise multiple regression analysis (MINITAB, Version 12) was used to identify the best set of anthropometric parameters that could be used to predict body composition. The general linear model with the analysis of variance was used to test for possible ethnic differences in these relationships.

Butte *et al.* reported longitudinal body composition data for 72 healthy infants obtained at 0.5, 3, 6, 9, 12, 18, and 24 months of age. Both bottle-fed and

**Table I** Body Length, Weight, TBW, and TBK for Infants during the First 2 Years of Life<sup>a</sup>

Age (months)	Weight (kg)	Length (cm)	TBW (liters)	TBK (mEq)
Boys				
0.5	3.76 (0.44)	52.5 (1.7)	2.80 (0.32)	163 (18)
3	6.33 (0.68)	61.2 (1.8)	3.54 (0.36)	232 (26)
6	8.04 (0.81)	67.9 (1.7)	4.54 (0.51)	304 (35)
9	9.13 (0.86)	72.2 (1.8)	5.34 (0.54)	377 (39)
12	10.03 (1.01)	76.1 (2.1)	5.86 (0.56)	426 (50)
18	11.43 (1.12)	82.6 (2.0)	6.69 (0.61)	510 (44)
24	12.46 (1.17)	87.6 (2.8)	7.21 (1.03)	587 (55)
Girls				
0.5	3.64 (0.44)	52.0 (1.8)	2.67 (0.28)	154 (23)
3	6.03 (0.59)	60.7 (1.7)	3.34 (0.42)	212 (23)
6	7.60 (0.66)	66.5 (1.9)	4.20 (0.47)	274 (39)
9	8.62 (0.72)	71.0 (1.8)	4.89 (0.56)	341 (37)
12	9.50 (0.83)	75.3 (2.2)	5.40 (0.58)	404 (42)
18	10.94 (1.08)	82.0 (2.2)	6.26 (0.68)	480 (53)
24	12.02 (1.19)	87.7 (2.6)	6.97 (0.88)	551 (67)

<sup>a</sup>Data from Butte *et al.* (2000). Data presented as mean (SD). TBW, total body water; TBK, total body potassium.

breast-fed infants were included in the study, and the infants were Caucasian, African American, or Mexican American. The mean values for body length and weight, TBW, and TBK are listed in [Table I](#) for each

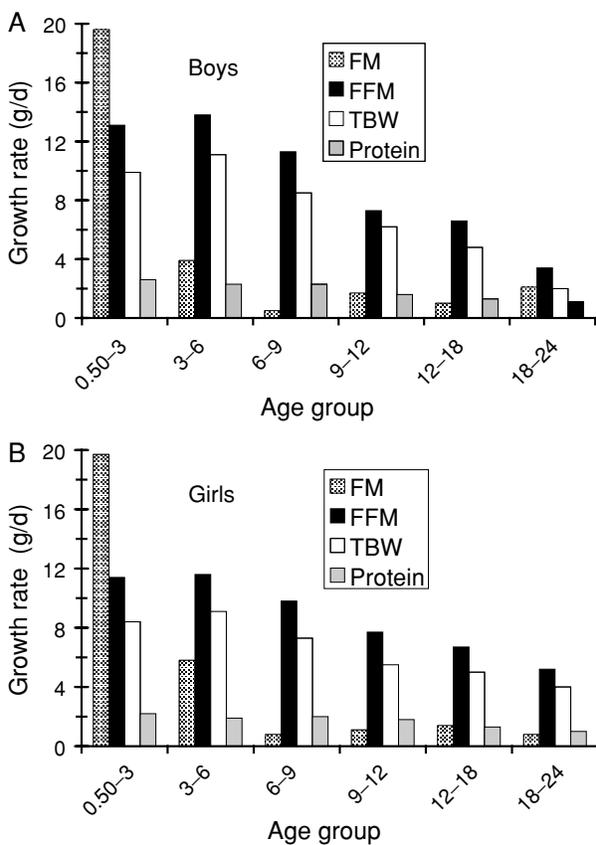
gender. The corresponding values for FM, body fatness, and FFM and its subcategories are listed in [Table II](#). It is also interesting to note that these same researchers published body composition data for infants grouped by their primary feeding method during early life. As one would expect, the amount of protein, osseous mineral, and intracellular water increased with age, whereas the total hydration of FFM and its extracellular water content decreased. Butte *et al.* used these data to calculate the average daily incremental growth rates for FM, FFM, TBW, protein, and glycogen. The mean rates are illustrated in [Fig. 2](#). The deposition rates tended to be highest during the first 6–12 months of life, followed by a 12-month period of relatively constant accretion. For example, in boys, the average protein accretion was approximately  $2.3 \pm 0.8$  g/day for the first year of life, followed by a slower rate of approximately  $1.3 \pm 0.6$  g/day for the next 12 months. Overall, the accretion patterns for girls were very similar to those observed for boys, although the average rates were approximately 10% lower for each of the FFM subcategories.

Most of the studies in older children and adolescents that have included body composition data have been cross-sectional in design. In one study, Ellis and colleagues reported the mean body composition values for Caucasian, African American, and Mexican American children. The weight, height, TBW, TBK, and BMC for each ethnic group are shown in [Table III](#)

**Table II** FM, Percentage FM, FFM, Protein, BMC, ECW, and ICW for Infants during the First 2 Years of Life<sup>a</sup>

Age (months)	FM (kg)	%FM	FFM (kg)	Protein (kg)	BMC (g)	ECW (liters)	ICW (liters)
Boys							
0.5	0.44 (0.32)	11.4 (8.0)	3.35 (0.40)	0.47 (0.04)	67 (12)	1.73 (0.15)	1.04 (0.10)
3	1.91 (0.39)	30.2 (4.0)	4.37 (0.43)	0.66 (0.05)	109 (9)	2.06 (0.18)	1.48 (0.12)
6	2.32 (0.50)	29.1 (4.7)	5.63 (0.60)	0.86 (0.06)	146 (11)	2.61 (0.20)	1.93 (0.14)
9	2.34 (0.64)	25.7 (5.2)	6.71 (0.67)	1.08 (0.07)	188 (7)	2.93 (0.23)	2.43 (0.21)
12	2.56 (0.59)	25.6 (4.0)	7.40 (0.72)	1.21 (0.08)	215 (22)	3.15 (0.30)	2.73 (0.21)
18	2.83 (0.94)	24.5 (6.0)	8.55 (0.76)	1.47 (0.09)	265 (9)	3.37 (0.32)	3.32 (0.21)
24	3.10 (0.56)	25.4 (4.7)	9.13 (1.06)	1.66 (0.11)	310 (18)	3.26 (0.40)	3.77 (0.27)
Girls							
0.5	0.52 (0.34)	14.2 (9.0)	3.12 (0.45)	0.42 (0.04)	66 (25)	1.66 (0.16)	0.93 (0.11)
3	1.90 (0.41)	31.5 (5.6)	4.11 (0.48)	0.62 (0.06)	99 (29)	1.96 (0.21)	1.37 (0.14)
6	2.44 (0.40)	32.0 (4.4)	5.21 (0.57)	0.80 (0.07)	135 (36)	2.42 (0.27)	1.78 (0.18)
9	2.47 (0.46)	28.8 (5.0)	6.12 (0.66)	0.98 (0.08)	171 (42)	2.69 (0.29)	2.20 (0.20)
12	2.62 (0.50)	27.6 (4.3)	6.88 (0.68)	1.16 (0.08)	213 (48)	2.81 (0.30)	2.61 (0.21)
18	2.87 (0.63)	26.3 (4.4)	7.99 (0.80)	1.37 (0.10)	256 (56)	3.14 (0.34)	3.11 (0.22)
24	3.05 (0.46)	25.4 (4.7)	8.99 (1.10)	1.56 (0.09)	297 (63)	3.48 (0.32)	3.53 (0.22)

<sup>a</sup>Data from Butte *et al.* (2000). Data presented as mean (SD). FM, fat mass; FFM, fat-free mass; BMC, bone mineral content; ECW, extracellular water; ICW, intracellular water.



**Figure 2** Average daily accretion rates for fat mass (FM), fat-free mass (FFM), total body water (TBW), and protein mass during the first 2 years of life for boys and girls. Adapted from Butte *et al.* (2000).

for girls and Table IV for boys. Using the Fomon model, the corresponding values for FM and each of the subcategories of the FFM were also reported and are given in Table V for girls and Table VI for boys.

The data presented in Tables I–VI provide a contemporary update of body composition for infants, children, and adolescents based on the reference models originally developed by Fomon more than 30 years ago. A major advantage of the current data is that the measurements of body potassium, body water, and bone mineral mass were obtained for each of the subjects. Furthermore, subjects at all ages were examined for the contemporary population, thus eliminating the need to extrapolate for missing ages.

Compared to Fomon's reference models, there are no significant differences in height for the Caucasian children, whereas body weights, on average, are substantially higher, especially for older children and adolescents. These differences can be attributed to a clear increase in body fatness relative to that of the

original Fomon population. There are also differences in the general growth patterns for the protein and bone mineral categories. However, bone mineral mass was directly measured using the DXA technology for the contemporary population, whereas it had to be extrapolated from a measurement of the radius for the Fomon models.

Some of the differences in the body composition categories between the Fomon and contemporary models have not been fully resolved as true physiological changes or as methodological-induced artifacts. The results of the contemporary models, however, provide an updated reference suitable for assessing the current status of today's children, especially if there are conditions such as disease that may have altered normal growth. This assessment can be accomplished by comparing the patient's values with those of healthy children of similar age from the reference population. Many childhood diseases, however, often contribute to stunted or delayed growth, which is also frequently accompanied by altered body composition. Therefore, if one performs only age-matched comparisons, the effects of the smaller physical size of the patient may not be adequately taken into consideration. A better assessment in these cases can be obtained when there is also an adjustment made for any height deficiency. Thus, several prediction equations, using multiparameter regression analysis, have been calculated using the data for the contemporary reference population. These equations are presented in Table VII for children in the 5- to 18-year age range. In all cases, height was the primary anthropometric-based predictive parameter, with adjustments for gender, age, and ethnicity. Inclusion of the Tanner stage was also tested, but it did not add substantially to the predictive power of the models, and it may not always be available for the patient. Body weight was not used since it consists of the sum of FFM and FM and therefore would introduce a bias into the prediction results for extremes of either category.

In summary, the tabulated data presented here provide a contemporary update of body composition reference values from infancy through adolescents, illustrating the changes that can be expected during normal growth for different ethnic populations. In the case of bone, these models have been further refined to construct a z-score calculator for the DXA measurement of bone mineral density. Calculators are readily available to the pediatric clinical and research community on the Internet at [www.bcm.tmc.edu/bodycomplab](http://www.bcm.tmc.edu/bodycomplab). Future expansion of this Web site will include similar z-score calculators for the body's protein, water, and potassium composition.

Table III Body Weight, Height, TBW, TBK, and BMC for Girls<sup>a</sup>

Age (years)	Weight (kg)	Height (cm)	TBW (liters)	TBK (g)	BMC (g)
African American (black)					
5-7	25.9 (8.1)	120.5 (8.8)	15.1 (3.5)	43.7 (10.0)	724 (203)
8-10	38.3 (12.6)	138.5 (11.3)	21.3 (5.6)	63.4 (16.2)	1181 (433)
11-13	43.3 (18.3)	148.1 (18.3)	24.6 (7.4)	79.5 (11.7)	1399 (433)
14-16	63.1 (15.7)	158.3 (5.1)	31.7 (6.3)	96.6 (13.4)	2076 (289)
17-19	62.9 (13.6)	162.0 (7.0)	32.9 (5.2)	101.9 (14.4)	2179 (419)
European American (white)					
5-7	21.6 (3.9)	114.9 (5.1)	12.8 (1.8)	36.9 (4.8)	573 (94)
8-10	30.5 (9.9)	132.7 (8.6)	17.6 (4.0)	52.6 (13.3)	907 (221)
11-13	42.2 (14.5)	150.3 (9.3)	24.8 (4.9)	77.1 (21.1)	1431 (541)
14-16	59.3 (9.9)	162.1 (3.7)	31.1 (2.4)	94.5 (8.6)	2179 (285)
17-19	62.4 (9.7)	165.1 (6.7)	32.3 (5.5)	100.0 (11.8)	2167 (279)
Mexican American (Hispanic)					
5-7	24.0 (5.7)	117.9 (7.4)	13.4 (2.5)	38.5 (6.9)	639 (127)
8-10	35.4 (9.9)	135.4 (8.4)	18.9 (4.2)	56.1 (11.6)	967 (206)
11-13	55.6 (8.5)	153.5 (6.5)	26.6 (4.1)	84.4 (9.5)	1639 (235)
14-16	56.3 (12.6)	156.4 (5.7)	26.8 (5.2)	85.8 (12.2)	1858 (339)
17-19	64.9 (17.5)	163.0 (9.2)	31.2 (5.9)	93.1 (16.0)	2013 (320)

<sup>a</sup>Data from Ellis *et al.* (2000).

Data presented as mean (SD). TBW, total body water; TBK, total body potassium; BMC, bone mineral content.

Table IV Body Weight, Height, TBW, TBK, and BMC for Boys<sup>a</sup>

Age (years)	Weight (kg)	Height (cm)	TBW (liters)	TBK (g)	BMC (kg)
African American (black)					
5-7	26.0 (6.1)	121.4 (6.9)	16.1 (3.1)	46.9 (7.5)	763 (161)
8-10	34.8 (9.3)	137.4 (9.1)	21.1 (4.9)	65.1 (14.1)	1139 (277)
11-13	55.3 (15.3)	158.0 (9.7)	33.3 (8.5)	104.8 (25.8)	1840 (451)
14-16	79.6 (15.1)	175.7 (7.5)	48.2 (6.6)	161.5 (18.8)	2908 (408)
17-19	80.5 (12.6)	177.7 (8.5)	48.0 (6.8)	172.4 (23.8)	3237 (543)
European American (white)					
5-7	22.4 (3.4)	118.7 (7.3)	14.9 (2.3)	41.9 (6.4)	671 (158)
8-10	33.7 (10.6)	136.9 (8.7)	20.2 (3.6)	60.4 (13.1)	1027 (264)
11-13	51.4 (13.7)	155.8 (9.3)	29.9 (7.2)	91.5 (18.8)	1611 (386)
14-16	63.3 (13.0)	169.6 (6.4)	43.1 (5.8)	126.9 (23.8)	2279 (501)
17-19	73.1 (13.5)	178.7 (6.6)	46.5 (7.5)	152.4 (20.7)	2753 (425)
Mexican American (Hispanic)					
5-7	23.9 (5.2)	117.6 (6.5)	14.3 (2.7)	41.3 (6.9)	658 (134)
8-10	35.1 (9.5)	135.4 (7.1)	18.6 (3.4)	59.5 (11.6)	984 (171)
11-13	52.0 (11.7)	154.3 (10.0)	29.4 (6.6)	92.9 (20.1)	1578 (393)
14-16	64.2 (13.0)	169.6 (6.3)	38.3 (5.9)	128.6 (19.9)	2312 (344)
17-19	70.0 (14.5)	172.7 (7.3)	41.2 (5.6)	142.3 (20.8)	2579 (380)

<sup>a</sup>Data from Ellis *et al.* (2000).

Data presented as mean (SD). TBW, total body weight; TBK, total body potassium; BMC, bone mineral content.

Table V ECW, ICW, Protein, Nonosseous Mineral, and Fat Mass for Girls<sup>a</sup>

Age (years)	ECW (liters)	ICW (liters)	Protein (kg)	NonOss (g)	FM (kg)
African American (black)					
5-7	7.8 (2.1)	7.4 (1.7)	3.20 (0.71)	139 (32)	7.9 (5.6)
8-10	10.9 (3.7)	10.3 (2.7)	4.58 (1.1)	196 (52)	12.7 (7.6)
11-13	13.6 (1.8)	13.5 (2.6)	5.86 (0.87)	249 (34)	15.7 (13.5)
14-16	15.3 (4.4)	16.4 (2.3)	7.12 (1.0)	291 (58)	27.5 (10.4)
17-19	16.4 (4.2)	17.0 (2.4)	7.51 (1.06)	306 (53)	21.8 (9.4)
European American (white)					
5-7	6.5 (1.2)	6.0 (0.92)	2.66 (0.39)	116 (18)	5.7 (1.9)
8-10	8.5 (2.2)	8.7 (2.2)	3.88 (0.98)	162 (37)	8.8 (6.1)
11-13	11.9 (2.7)	12.9 (3.0)	5.69 (1.56)	227 (45)	10.8 (7.5)
14-16	15.1 (2.3)	16.0 (1.3)	6.96 (0.63)	286 (22)	21.8 (9.8)
17-19	15.6 (4.5)	16.6 (2.0)	7.36 (0.86)	296 (53)	21.3 (6.9)
Mexican American (Hispanic)					
5-7	7.2 (2.1)	6.4 (1.2)	2.82 (0.50)	124 (26)	7.4 (3.4)
8-10	9.8 (2.6)	9.3 (1.9)	4.13 (0.85)	174 (39)	12.7 (6.0)
11-13	12.8 (3.0)	13.8 (1.8)	6.14 (0.77)	246 (38)	22.3 (6.7)
14-16	13.9 (4.8)	14.0 (1.9)	6.26 (0.89)	251 (57)	21.3 (7.4)
17-19	15.7 (4.1)	15.6 (2.7)	6.76 (1.2)	277 (55)	27.9 (13.9)

<sup>a</sup>Data from Ellis *et al.* (2000).

Data presented as mean (SD). NonOss, nonosseous mineral; FM, Fat mass; ECW, extracellular water; ICW, intracellular water.

Table VI ECW, ICW, Protein, Nonosseous Mineral, and Fat Mass for Boys<sup>a</sup>

Age (years)	ECW (liters)	ICW (liters)	Protein (kg)	NonOss (g)	FM (kg)
African American (black)					
5-7	7.6 (1.7)	7.8 (1.3)	3.5 (0.6)	124 (23)	5.8 (3.8)
8-10	10.1 (3.4)	11.0 (2.8)	4.8 (1.0)	194 (45)	7.8 (6.6)
11-13	15.4 (4.8)	17.6 (4.7)	7.7 (1.9)	303 (79)	12.8 (9.1)
14-16	21.3 (4.4)	26.9 (3.7)	11.9 (1.4)	442 (61)	14.8 (9.7)
17-19	19.7 (4.3)	28.3 (4.6)	12.7 (1.7)	440 (66)	13.6 (7.3)
European American (white)					
5-7	7.6 (1.2)	7.1 (1.0)	3.1 (0.5)	135 (18)	4.2 (1.8)
8-10	10.0 (1.8)	10.5 (2.0)	4.5 (1.0)	188 (33)	7.7 (4.9)
11-13	14.8 (4.1)	15.1 (3.5)	6.7 (1.4)	275 (66)	13.8 (10.4)
14-16	20.3 (3.7)	22.8 (3.2)	9.4 (1.8)	396 (53)	13.0 (7.5)
17-19	21.0 (4.1)	25.5 (4.3)	11.2 (1.5)	427 (69)	14.0 (9.0)
Mexican American (Hispanic)					
5-7	7.3 (1.8)	7.0 (1.2)	3.0 (0.5)	135 (22)	6.0 (2.8)
8-10	8.7 (2.3)	9.8 (1.7)	4.4 (0.9)	183 (38)	10.9 (5.1)
11-13	14.2 (3.4)	15.9 (3.2)	6.8 (1.5)	284 (67)	14.9 (6.4)
14-16	16.9 (3.4)	21.4 (3.4)	9.5 (1.5)	386 (66)	15.1 (8.9)
17-19	17.9 (3.8)	23.5 (3.4)	10.5 (1.5)	409 (65)	14.6 (11.4)

<sup>a</sup>Data from Ellis *et al.* (2000).

Data presented as mean (SD). ECW, extracellular water; ICW, intracellular water; NonOss, nonosseous mineral; FM, Fat mass.

Table VII Anthropometric-Based Prediction Equations for TBW, Protein, BMC, Glycogen, and FFM<sup>a</sup>

		R <sup>2</sup> (%)	SEE
Male	TBW (liters) = 0.744 × age + (6.05 + 0.346 × ethnicity) × Ht <sup>3</sup> + (0.346 × ethnicity - 1.08)	90.4	4.28
Female	TBW (liters) = (6.57 - 0.334 × ethnicity) × Ht <sup>3</sup> + 3.02	87.0	3.25
Male	Protein (kg) = 0.227 × age + (1.473 - 0.091 × ethnicity) × Ht <sup>3</sup> - 1.091	92.5	0.94
Female	Protein (kg) = 0.064 × age + (1.30 - 0.053 × ethnicity) × Ht <sup>3</sup> + 0.426	90.2	0.64
Male	BMC (g) = 58.2 × age + (381 - 26.6 × ethnicity) × Ht <sup>3</sup> - 412	92.1	251
Female	BMC (g) = 40.6 × age + (374 - 10.9 × ethnicity) × Ht <sup>3</sup> - 238	86.3	270
Male	Glycogen (g) = 3.07 × age + (50.8 - 3.16 × ethnicity) × Ht <sup>3</sup> - 0.44	83.2	42
Female	Glycogen (g) = 1.98 × age + 57.4 × Ht <sup>3</sup>	78.9	42
Male	FFM (kg) = 1.11 × age + (7.81 - 0.521 × ethnicity) × Ht <sup>3</sup> - 2.75	91.8	5.2
Female	FFM (kg) = (8.78 - 0.413 × ethnicity) × Ht <sup>3</sup> + 3.31	88.4	3.9

<sup>a</sup>Applicable only for the 5- to 18-year age range. TBW, total body water; BMC, bone mineral content; FFM, fat-free mass; Ht, height (m). Ethnicity code: -1, black; 0, white; +1, Hispanic.

## Acknowledgments

This work was supported by the U.S. Department of Agriculture, Agricultural Research Service (USDA/ARS) under Cooperative Agreement No. 58-6250-6001 with Baylor College of Medicine. This work is a publication of the USDA/ARS Children's Nutrition Research Center, Department of Pediatrics, Baylor College of Medicine, and Texas Children's Hospital. The contents of this publication do not necessarily reflect the views or policies of the USDA, nor does mention of trade names, commercial products, or organizations imply endorsement.

## See Also the Following Articles

Auxology, Childhood • Body Proportions • Growth, Normal Patterns and Constitutional Delay • Postnatal Normal Growth and Its Endocrine Regulation • Puberty: Physical Activity and Growth • Sexual Maturation, Female • Sexual Maturation, Male • Skeletal Development During Childhood and Adolescence

## Further Reading

Butte, N. F., Hopkinson, J. M., Wong, W. W., Smith, O. B., and Ellis, K. J. (2000). Body composition during the first 2

years of life: An updated reference. *Pediatr. Res.* **47**, 578-585.

Ellis, K. J. (2000). Human body composition: In vivo methods. *Physiol. Rev.* **80**, 649-680.

Ellis, K. J., and Shypailo, R. J. (1993). Whole-body potassium measurements independent of body size. In "Human Body Composition: In Vivo Methods, Models, and Assessment" (K. J. Ellis and J. Eastman, eds.), pp. 371-375. Plenum, New York.

Ellis, K. J., Shypailo, R. J., Pratt, J. A., and Pond, W. G. (1994). Accuracy of dual-energy X-ray absorptiometry for body composition measurements in children. *Am. J. Clin. Nutr.* **60**, 660-665.

Ellis, K. J., Shypailo, R. J., Abrams, S. A., and Wong, W. W. (2000). The reference child and adolescent models of body composition: A contemporary comparison. *Annu. N. Y. Acad. Sci.* **904**, 374-382.

Fomon, S. J., Haschke, F., Ziegler, E. E., and Nelson, S. E. (1982). Body composition of reference children from birth to age 10 years. *Am. J. Clin. Nutr.* **35**, 1169-1175.

Haschke, F. (1989). Body composition during adolescence. In "Body Composition Measurements in Infants and Children" (W. J. Klish and N. Kretschmer, eds.), pp. 76-83. Ross Laboratories, Columbus, OH.

Snyder, W. S., Cook, M. J., Nasset, E. S., Karhausen, L. R., Howells, G. P., and Tipton, I. H. (1984). "Report of the Task Group on Reference Man," International Commission on Radiological Protection No. 23. Pergamon, New York.



# Body Proportions

William J. M. Gerver

University of Maastricht, Maastricht, The Netherlands

## Glossary

**mastoid process** Cone-shaped extremity of the temporal bone, extending forward and downward just posterior to the acoustic opening.

**occiput** The back part of the head.

The growth process that transforms a newborn into an adult implies that there is not only an increase in height but above all a constant change in body proportions due to differing rates and times of growth of different parts of the body.

## INTRODUCTION

In 1943, Medawar gave a clear description of the transformation that occurs during growth, showing the head–trunk–legs gradient, which indicates that the development of the head is advanced over that of the trunk and the development of the trunk is advanced over that of the legs during the maturation process. This clearly shows that the growth of different parts of the body does not occur simultaneously or at the same speed.

In a child, this transformation can be defined by the interrelationship of a number of linear measurements, such as sitting height to full height, or head circumference to crown–rump length. The ease with which one can equate the human body's shape with an accurate estimate of human age—whether the compactness of a 5 year old, the coltishness of a young teenager, or the dimensions of a mature adult—tends to be taken for granted. Nevertheless, this feat is remarkable. Even more extraordinary is the precision with which the human eye registers any deviation from the norm.

The mathematical expression of these perceptions for scientific purposes is a highly complex task with imperfect results. Vital information is lost. It is possible, however, to use a reference population to

organize the complexity of what even the untrained eye perceives naturally.

Knowledge of the body proportions gives insight into the natural process of maturation and any disturbance will be visible in abnormal body proportions. A principal distinction can be made between a child with stunted growth affecting the whole body or affecting just some body parts. Such an observation leads to different conclusions about the origin of some growth deviation.

In this article, an overview is given of the different measurements: the techniques of taking the measurements, the normal values based on a Dutch study of 2001, and the methods of interpretation. Of the 20 measurements taken in the Dutch study, the most important are selected for discussion here.

## THE MEASUREMENTS

### Length and Height

In pediatrics, the curve representing a child's increase in height with age is used as one of the indicators of the child's health. The child's height is compared to that of a reference population in order to judge whether the child is atypical in terms of the height distribution of this population. Child and reference population must share the same geographical and socioeconomic background. Also, parental or familial height should be taken into consideration to assess the relative contribution of genetic and other factors to the child's height at any particular age.

Since growth is a dynamic process, one isolated measurement in time is meaningless for judging growth. Repeated measurements are needed to calculate the growth velocity. Knowledge of growth velocity is particularly relevant when dealing with a sick child. Depressed growth velocity may indicate the severity of an illness, whereas normalization of growth velocity or catch-up growth may signal recuperation.

Height is the distance between the top of the head and the sole of the foot. This distance is the sum of

the height of the skull, the length of the spine, and the length of the lower extremities. Each of these different parts of the body has its own age-dependent growth velocity. Therefore, height is the result of the sum of different values, which are not linearly related in time.

### **Method of Measurement**

The distance between the top of the head and the sole of the foot is called length when an individual is measured lying down and height when the individual is measured standing upright.

Height is measured using a stadiometer. The stadiometer comprises a rigid vertical backboard and a horizontal headboard running free, perpendicular to the backboard and without cross-play. The top of the head must be in contact with the headboard. A 0.5 kg weight is placed on the headboard. This serves the purpose of flattening the child's hair and frees the physician's hands so as to enable him to keep the child in the correct upright position.

To measure standing height, the subject's shoes and socks are removed. The child is placed so that his heels, buttocks, and shoulders are in contact with the vertical plane of the stadiometer. The child's feet must be flat against the floor while either ankles or knees remain in contact. The child's head is held in the "Frankfurt plane": the lower borders of the orbits are in the same horizontal plane as the external auditory meati. The measurement is taken while a gentle upward pressure is exerted on the mastoid processes so that the child is fully extended. In older children, stretching is achieved by telling them to breath deeply.

To measure supine length, a measuring table resembling a stadiometer is used but on a horizontal plane. The headboard is fixed and the footboard is movable. Both the headboard and the footboard must be large enough to ensure that both the top of the head and the feet are in contact with them. To measure supine length, two people are necessary: one holds the infant's head in the vertical Frankfurt plane while at the same time another individual keeps the child's legs straight. The child's shoulders and buttocks must be in contact with the table. Supine length is measured in the first 2 years of age.

### **Crown Rump Length, Sitting Height, and Subischial Leg Length**

As stated earlier, height defined as the distance between the top of the head and the sole of the foot is

the result of the sum of different values, which are not linearly related in time. Therefore, the first step is to split height into its different components.

Generally, sitting height is taken as one of the components and compared to height. The relationship between sitting height and height is often expressed as an age-dependent ratio. The use of ratios can be misleading for two reasons: first, two ratios may be equal, although their nominators and denominators are not; and second, when a change in the nominator automatically leads to a change in the denominator, the change in the ratio will be even more misleading. Therefore, a more straightforward approach is to simply plot height against sitting height and an even clearer one is to use sitting height and subischial leg length as variables for comparison, since a change in sitting height will automatically induce a change in height but not a change in subischial leg length.

### **Method of Measurement**

The distance between the top of the head and the buttocks is called sitting height when the measurement is taken of a child sitting upright and is called crown rump length when an infant is measured lying down.

Sitting height is measured using a sitting height table. The table comprises a rigid vertical backboard and a horizontal headboard running free perpendicular to the backboard and without cross-play. The surface of the headboard must be in contact with the top of the head. A 0.5 kg weight is placed on the headboard. This weight flattens the child's hair and also frees the physician's hands so as to be able to keep the child in the correct position. The child must be in the sitting position with his feet on a footrest so that his full weight is on his buttocks. Insofar as is possible, arching of the back is avoided by gently applying upward pressure to the mastoid processes. In older children, stretching of the back is achieved by asking them to breathe deeply. The child's head is held in the Frankfurt plane: the lower borders of the orbits are in the same horizontal plane with the external auditory meati.

The recommended instrument for measuring crown rump length is similar to the stadiometer used for supine length. The headboard is fixed and the footboard is movable. Headboard and footboard must be large enough to ensure that the most protruding points of both head and buttocks are in contact with the boards. In order to measure this length, an assistant holds the infant's head in the vertical

Frankfurt plane while the physician holds the child's legs at a 90° angle with the table. When this is achieved, the footboard is pressed against the buttocks.

Subischial leg length is defined as the arithmetic difference between height and sitting height or between supine length and crown rump length.

### Head Circumference

The head circumference is routinely measured in newborn infants since it correlates well with skull volume. Skull volume is highly correlated with gestational age, body weight, and body length. Since in intrauterine growth retardation, the brain is less affected than the weight and the length, the extent of the discrepancy between these measurements in the newborn will be an indicator of the severity of the retardation.

Because of the fast growth velocity of the head circumference, especially during the first year of life, its measurement provides important information about the general condition of the child. In full-term healthy newborns, the head circumference increases approximately 1 mm per day initially.

#### Method of Measurement

To measure head circumference, a fiberglass-reinforced tape of nonstretchable material is used. The tape is placed around the head at the most protruding points of occiput and forehead. In younger children, the tape is placed just above the brow ridges. The tape is placed gently so as to leave no marks after removal.

### Arm Span

The measurement of the limbs so as to describe body proportions is an important tool when evaluating development. Arm span is the most common way to measure upper limb length. However, one must realize that by measuring the span of the outstretched arms from the tips of the longest fingers, the distorting information of the width of the trunk is added.

#### Method of Measurement

To measure arm span, a measuring rod can be used. The distance between the tips of the stretched middle fingers is measured while the subject stands with his arms fully extended.

### APPLICATION

To judge the body proportions of an individual, the following steps must be taken.

First, for each measurement, the standard deviation score,  $z(t)$ , is calculated by

$$z(t) = \frac{x(t) - \mu(t)}{\sigma(t)}$$

where  $x(t)$  denotes the individual's body measurement score at age  $t$  (e.g., height),  $\mu(t)$  denotes the population mean at age  $t$ , and  $\sigma(t)$  denotes the corresponding standard deviation.

The  $z$ -scores are given in the reference tables. Since the distribution of the reference population is skewed, two different SD values are calculated. In the reference tables, they are given as +1 SD and -1 SD, the SD value above and below the  $P_{50}$  (popular mean), respectively. Calculating a  $z$ -score or SD score, one needs to use the +1 SD if the individual's measurement is above the  $P_{50}$  of the reference group and -1 SD if the individual's measurement is below the  $P_{50}$  of the reference group.

Second, the typicality or atypicality of an individual can be calculated by means of the squared distance  $D^2$ , which follows a  $\chi^2$  distribution if normal children are considered.

The squared distance is calculated as follows:

$$D^2 = [1 - r^2(b, b')]^{-1} [z_b^2 + z_{b'}^2 - 2z_b z_{b'} r(b, b')]$$

where  $r$  denotes the correlation coefficient of the two measurements  $b$  and  $b'$  according to Table I.

$D^2 \leq 5.991$  indicates that the considered individual is typical or normal, whereas  $D^2 \geq 5.991$  suggests that something is wrong with size and/or shape.

Size and shape can be separated by  $z_b + z_{b'}$  as characteristic for size and  $z_b - z_{b'}$  as characteristic for shape. This leads to the statement that size is atypical if

$$|z_b + z_{b'}| / [2 + 2r_t(b, b')]^{1/2} \geq 1.960$$

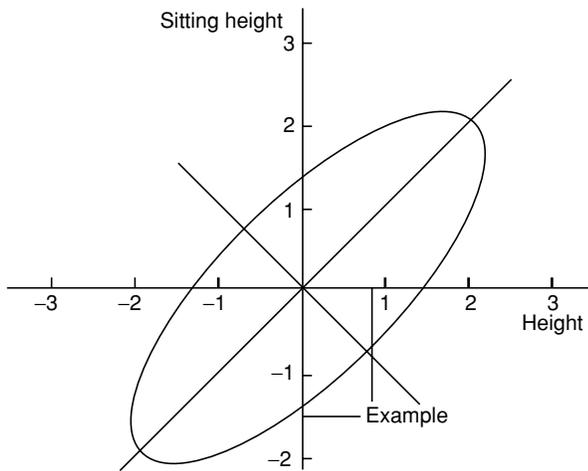
and shape is atypical if

$$|z_b - z_{b'}| / [2 - 2r_t(b, b')]^{1/2} \geq 1.960$$

Table I Correlation Coefficient of Pairs of Measurements of Body Proportions According to Sex

	Boys	Girls	Mean
Height/arm span	0.87	0.85	0.86
Height/sitting height	0.82	0.82	0.82
Length/head circumference	0.36	0.36	0.36

Note. These correlations are almost independent of age for children between the ages of 3 and 17 years.



**Figure 1** The 95% confidence limits of the pair of  $z$ -scores of height and sitting height. The example is the patient mentioned in the text.

Graphically, one can display the atypicality of a particular individual by constructing an ellipse, which comprises 95% of the pairs of scores (Fig. 1).

**Example**

If a boy of 6.5 years has a height of 134.6 cm and a sitting height 63.5 cm, the corresponding  $z$ -scores for height and sitting height are  $z_b = 0.92$  and  $z_{sb} = -1.5$ .

These scores considered separately do not indicate that something is wrong with this child. The difference in sign, however, is somewhat alarming because the correlation between the measurements is positive.

Using the value 0.82 for the correlation coefficient between height and sitting height (Table I), one obtains

$$|z_b - z_{b'}|/[2 + 2r(b, b')]^{1/2} \geq 1.960 = 16.05$$

$$\geq \chi^2_{2,05} = 5.991$$

with the consequence that the two scores considered together indicate atypicality. This is interesting because, as discussed earlier, separate considerations did not indicate sufficient evidence of abnormality.

If one considers the size and shape characteristics based on the  $z$ -scores, then one obtains

$$|z_b + z_{b'}|/[2 + 2r_t(b, b')]^{1/2} = \frac{-0.58}{(2 + 1.64)^{1/2}} \approx -0.2$$

and

$$|z_b - z_{b'}|/[2 - 2r_t(b, b')]^{1/2} = \frac{2.42}{(2 - 1.64)^{1/2}} \approx 4$$

indicating that size is normal, whereas shape is abnormal, because the absolute value of the shape characteristic, 4, is larger than 1.960.

**CONCLUSION**

Pediatricians are often confronted with children of all ages who are referred for growth problems, such as growth retardation or dimorphism. Growth standards are therefore needed and one of the incentives of this article was to provide data that will give the possibility of discerning growth variation according to typical or atypical patterns. In other words, the child's growth deviation concerns only size or shape or both. It is customary to define shape by one or more ratios of measurements, such as height to sitting height. It should be realized that a simple ratio of two measurements is only one characteristic for shape. Moreover, when studying the typicality of particular individuals, it should be kept in mind that two ratios may be equal although there exist great differences between their nominators and denominators. Hence, one should not ignore size. This problem can be dealt with by defining typicality or atypicality by means of the squared distance  $D^2$ .

If  $D^2 > \chi^2$ , the child is regarded as atypical but the cause of this atypicality can be discussed only by reconsidering the original scores  $b$  and  $b'$  or their standardized equivalents  $z_b$  and  $z_{b'}$ . If the size/shape representation based on  $z_b + z_{b'}$  and  $z_b - z_{b'}$  is used, then the interpretation of a significant result is immediate. In the example given here, a score of  $-0.2$  was obtained for the standardized size characteristic and a score of approximately 4 was obtained for shape. Size is satisfactorily normal but shape is too large, which means that  $b$  is too large and/or  $b'$  is too small, when considered in relation to each other.

**See Also the Following Articles**

- Body Composition During Growth • Growth, Normal Patterns and Constitutional Delay • Postnatal Normal Growth and Its Endocrine Regulation • Puberty: Physical Activity and Growth • Skeletal Development During Childhood and Adolescence

**Further Reading**

Gerver, W. J. M., and de Bruin, R. (2001). "Paediatric Morphometrics: A Reference Manual," 2nd extended ed. Datayse University Press, Maastricht, The Netherlands.  
 Medawar, P. B. (1943). The shape of the human being as a function of time. *Proc. R. Soc. London Ser. B* 132, 133-141.



# Body Weight, Body Composition, and Aging

Lyndon J. O. Joseph and Alice S. Ryan

University of Maryland School of Medicine and Baltimore VA Medical Health Care System, Baltimore, Maryland, United States

## Glossary

**body mass index (BMI)** A practical measure of an individual's weight in relation to height. It is used to define overweight and obesity.

**fat-free mass (FFM)** Encompasses all of the body's nonfat tissues, including the skeleton, water, muscle, connective tissue, and organ tissues.

**osteoporosis** A condition characterized by low bone mass (density) and microarchitectural deterioration of bone tissue, with a consequent increase in bone fragility and susceptibility to fracture.

**sarcopenia** A condition characterized by an involuntary age-related decline in lean body mass, primarily due to the loss of skeletal muscle, that affects functional capacity and strength of older adults.

**visceral fat (intra-abdominal fat)** A portion of the internal fat in the abdominal cavity lining the intestinal tract.

One of the fastest growing segments of the population is individuals >65 years of age. This age group, which currently accounts for ~15% of the population, is expected to grow to between 19 and 25% by 2025. Increased body fat and loss of bone mineral density (BMD) and muscle mass are defining characteristics of the aging process. These changes in body composition occur as a result of normal aging, have a detrimental effect on health status, and have substantial economic consequences on the health care system. Obesity is associated with an increased prevalence of comorbidities, including cardiovascular disease, type 2 diabetes mellitus, hypertension, dyslipidemia, and other metabolic diseases. The decline in skeletal muscle mass is associated with weakness, functional disability, frailty, and morbidity, whereas the decrease in BMD increases the risk of bone fractures and ultimately results in high rates of disability, morbidity, and mortality in the elderly. This article discusses the classification and prevalence of overweight and obesity and the changes that occur during the aging process, with emphasis on body weight, fat mass, and fat-free mass and its constituents of skeletal muscle mass, total body water, and bone. These changes in body composition are also described in context with metabolic disease states.

## CLASSIFICATION OF OVERWEIGHT AND OBESITY

Body mass index (BMI), a practical measure of an individual's weight in relation to height, is used to define overweight and obesity. A BMI between 25 and 29 kg/m<sup>2</sup> defines overweight in adults, whereas obesity is defined as a BMI  $\geq 30$  kg/m<sup>2</sup>. These criteria are based on epidemiologic evidence that shows a strong association between a BMI >25 kg/m<sup>2</sup> and an increased incidence of cardiovascular disease, type 2 diabetes mellitus, hypertension, and dyslipidemia, which affect mortality and morbidity. Although BMI is simple to obtain, it does not differentiate between fat mass and lean muscle mass. Therefore, it can overestimate body fat in persons who are very muscular and underestimate body fat in persons who have excess fat and reduced muscle mass but normal body weight. Thus, other methods, such as hydrodensitometry, bioelectric impedance, and dual-energy X-ray absorptiometry, are used to determine body composition in research settings.

## PREVALENCE OF OVERWEIGHT AND OBESITY

The National Center for Health Statistics examined the prevalence and trends of overweight U.S. adults from 1960 to 2000 via the National Health Examination Survey I and National Health and Nutrition Examination Surveys (I–III, continuous). The prevalence of overweight and obesity increased significantly during the 40-year period from 43% in 1960–1962 to 64.5% in 1999–2000. Most of this increase is attributable to a dramatic increase in the prevalence of obesity (BMI >30 kg/m<sup>2</sup>) that remained stable at ~15% for the first three survey periods (1960–1980) and then increased dramatically to 30.5% in 1999–2000. The increased prevalence of overweight and obesity in adults during this time period was evident in both genders, across all races and ethnicities, and across all age groups (Table I).

Table 1 Prevalence of Overweight and Obesity by Age, Sex, and Racial/Ethnic Group: United States, 1999–2000

Sex	Age (years) <sup>a</sup>	Sample size			Prevalence of overweight (BMI ≥25) (%)			Prevalence of obesity (BMI ≥30) (%)					
		All <sup>b</sup>	Non-Hispanic white	Non-Hispanic black	Mexican American	All <sup>b</sup>	Non-Hispanic white	Non-Hispanic black	Mexican American	All <sup>b</sup>	Non-Hispanic white	Non-Hispanic black	Mexican American
Both sexes	≥20	4115	1831	794	1105	64.5	62.3	69.6	73.4*	30.5	28.7	39.9*	34.4
	Men	≥20	2043	946	374	538	67.2	67.4	60.7	74.7	27.5	27.3	28.1
Women	20–39	666	276	125	184	60.5	61.0	52.6	67.5	23.7	22.0	27.4	30.4
	40–59	595	262	127	157	70.0	69.9	63.9	79.1	28.8	28.5	29.9	27.0
	≥60	782	408	122	197	74.1	74.3	69.1	79.6	31.8	34.3	26.4	29.7
	≥20	2072	885	420	567	61.9	57.3	77.3*	71.9	33.4	30.1	49.7*	39.7
Women	20–39	640	249	140	180	54.3	49.0	70.8*	61.6	28.4	24.4	46.2*	30.6
	40–59	653	249	141	193	66.1	61.0	81.5*	79.3	37.8	34.2	53.2*	48.5
	≥60	779	387	139	194	68.1	65.8	81.7*	77.5	35.0	33.3	50.2*	41.0

<sup>a</sup>Estimated prevalences for ages ≥20 years were age standardized by the direct method to the 2000 Census population using age groups 20–39, 40–59, and ≥60 years.

<sup>b</sup>Includes racial/ethnic groups not shown separately.

\*Significantly different from non-Hispanic whites ( $p < 0.05$ , with Bonferroni correction).

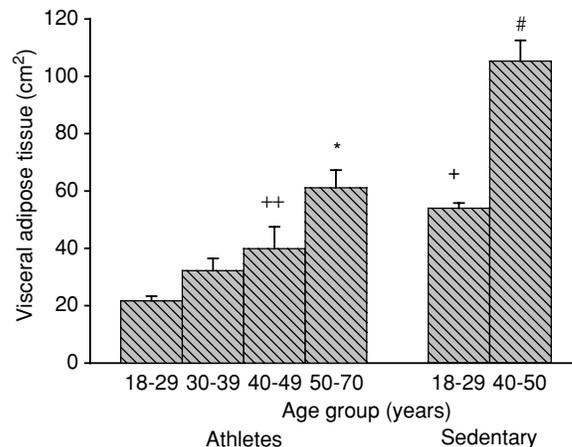
Reprinted from Flegal *et al.* *J. Am. Med.* **288**(14), 1723–1727 (2002).

## CHANGE IN BODY WEIGHT AND FAT MASS WITH AGING

In healthy people, body weight increases gradually from early adulthood until the fifth or sixth decade of life and then remains stable until age 65–70. During the period of weight gain, body weight increase occurs at a rate of 0.39–0.45 kg/year for women and 0.36–0.41 kg/year for men, irrespective of race. The slightly higher rate for women suggests that the transition from pre- to postmenopausal status affects the amount of weight gain. After age 60, body weight decreases slowly at a rate of 0.05–0.36 kg/year, depending on race and gender. The rate of decrease is usually higher in women than in men, regardless of ethnicity. Moreover, the rate of body weight decrease is highest in African American women and lowest in African American men. The variation in body weight regulation suggests that a nonlinear pattern of weight change occurs through the aging process.

The pattern of weight gain is generally consistent between cross-sectional and longitudinal studies and is characterized by a greater increase of fat than lean mass accounting for the increase in body weight. The relationship between age and body fat is curvilinear, with the greatest increase in body fat occurring in middle-aged persons and smaller increases occurring in both the young and the elderly. That is, there is a pattern of increasing fat mass with age until 50–69 years, followed by a slow decline in fat mass in older age groups (>70 years). Irrespective of the age group, 7–31% of the variation in the increase in fat mass (absolute or relative) in the adult population is accounted for by age.

The increase in weight with age is accompanied by an age-related redistribution of fat to the abdominal region. Sedentary women in their fifth decade of life have an average 1.5- to 3-fold higher visceral and subcutaneous abdominal fat area compared to their younger counterparts and athletes of similar age (Fig. 1). Similarly, older women athletes in the fifth and sixth decades of life have a 2- or 3-fold higher visceral fat area compared to younger athletes despite similar BMI and percentage body fat (Fig. 1). Therefore, there is an accumulation of fat in the abdominal region during the aging process irrespective of physical activity status. However, maintaining an active lifestyle partially negates the consequences of aging by dramatically reducing the amount of abdominal fat deposition. The ability to alter the accumulation of fat in the abdominal region is of clinical significance when one considers that intra-abdominal fat areas >100–110 cm<sup>2</sup> increase the risk



**Figure 1** Intra-abdominal adipose tissue area determined by computed tomography in women athletes and sedentary women by age group. Significant differences ( $p < 0.05$ ): \*, 18- to 29-year-old women athletes vs 50- to 70-year-old women athletes; #, 18- to 29-year-old sedentary women vs 40- to 50-year-old sedentary women; ++, 40- to 49-year-old women athletes vs 40- to 50-year-old sedentary women; +, 18- to 29-year-old women athletes vs 18- to 29-year-old sedentary women. Adapted from Ryan *et al.* *Am. J. Physiol.* 271, E916–E921.

for developing cardiovascular disease and other metabolic diseases. The age-associated increase in visceral adiposity may contribute to the increase in triglyceride and total cholesterol concentrations and glucose intolerance with age.

An increase in the infiltration of fat around and within skeletal muscle occurs with aging and adversely affects glucose and lipoprotein metabolism. Computed tomography scans document a twofold increase in midhigh subcutaneous fat between pre- and postmenopausal sedentary women. There is also a twofold increase in low-density lean tissue, a marker of intramuscular fat. Approximately 27–33% of the variation in the increase in midhigh subcutaneous fat and low-density lean tissue areas in the midhigh region is accounted for by age. Aging studies on fat infiltration in leg muscles in men are lacking, but presumably a similar increase in intramuscular fat with age occurs. The contribution of intramuscular fat to the metabolic dysfunction associated with the insulin resistance syndrome may help explain racial differences in glucose metabolism.

## CHANGE IN FAT-FREE MASS AND ITS CONSTITUENTS WITH AGING

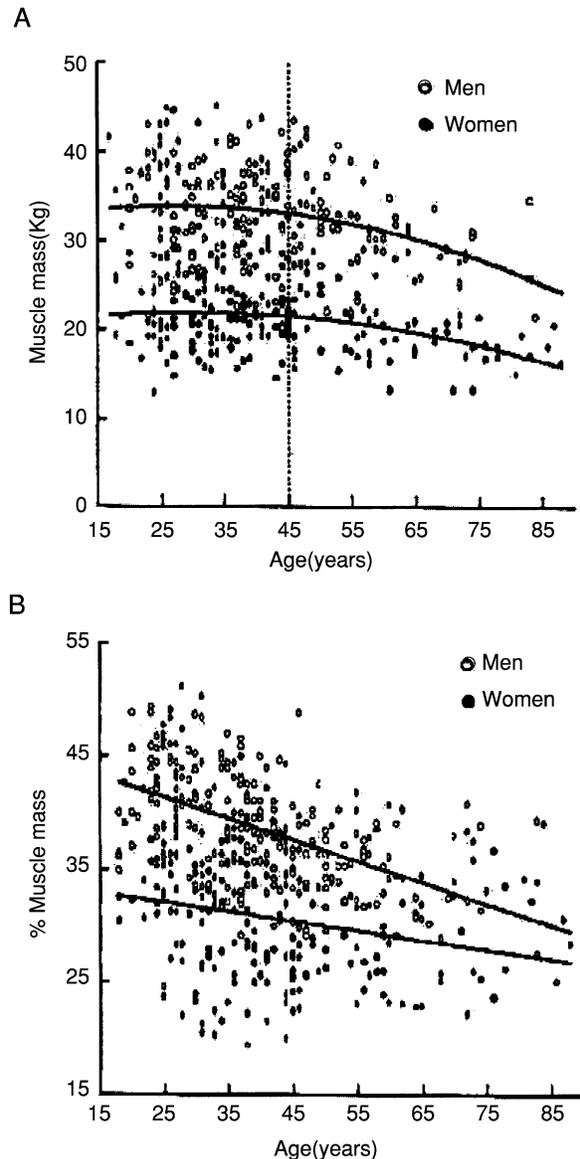
As body weight and fat mass increase with aging, total fat-free mass (FFM) and its constituents (skeletal

muscle mass, body cell mass, total body water, and bone mineral mass) gradually decrease. Peak FFM is reached in the mid-thirties and then progressively declines. In females, FFM remains relatively stable until approximately age 50, after which the decline in FFM occurs at a slower rate than in males. The average loss of FFM is approximately 16% between ages 25 and 70 in both men and women at a rate of  $\sim 0.16$  kg/year. Since skeletal muscle accounts for more than half ( $\sim 55\%$ ) of total FFM, the decrease in skeletal muscle mass between ages 20 and 70 is slightly lower than the loss of FFM at 10–15%, with the rate of decline greater in men (0.8–1.9 kg/decade) than in women ( $\sim 0.4$ –1.1 kg/decade). Although FFM starts to decrease during the third decade, skeletal muscle mass is preserved until the fifth decade, with a noticeable decrease in absolute skeletal muscle mass occurring at  $\sim 45$  years in both men and women (Fig. 2).

The involuntary age-related decline in FFM, primarily due to the loss of skeletal muscle, is called sarcopenia, and it affects functional capacity and strength of older adults. The prevalence of sarcopenia varies from 6 to 24% in persons younger than 70 years of age to  $>50\%$  in persons older than 80 years of age, depending on the definition and measure of muscle mass. Sarcopenia is associated with a three- or four-fold increase in functional impairment, disabilities, and falls in the elderly.

The loss of skeletal muscle mass is strongly associated with a loss of body water because a large proportion of skeletal muscle ( $\sim 75$ – $80\%$ ) is water. Total body water (TBW) accounts for approximately 80% of FFM at birth. In young adults, TBW comprises approximately 72% of FFM. Thus, a loss of body water occurs until maturity but remains relatively constant throughout adulthood and middle age. On average, TBW is lower in females than in males. Losses of body water occur after age 70 in females and even more so in the very old. The decline in TBW suggests a change in the hydration of the fat-free compartment (increased with normal aging), although definitive conclusions are lacking.

In addition to the losses of FFM, skeletal muscle mass, and TBW with age, the loss of bone mass is consistently documented. Peak bone mineral mass is reached at age 20–30 years, followed by a progressive decline. By age 70, spinal and femoral neck bone mineral density (BMD) are diminished by



**Figure 2** (A) Relationship between whole body skeletal muscle mass and age. (B) Relationship between relative skeletal muscle mass (body mass/skeletal muscle mass) and age. Adapted from Janssen *et al.*, *J. Appl. Physiol.* **89**, 81–88 (2000).

approximately 20 and 25%, respectively. Furthermore, the rate of bone loss varies with site and may be greater in areas with more trabecular bone than in those with predominantly compact bone. Total body mineral may decline at a slower rate than the decline observed in specific sites. In women, a more dramatic loss of bone mass occurs during menopause. The rate of BMD loss is greater among perimenopausal women compared to pre- and postmenopausal women and is site specific. Longitudinal studies estimate the rate of premenopausal BMD loss at 0.7–1.3% per year

at the lumbar spine and 0.2–0.3% per year at the femoral neck. In contrast, the rate of BMD loss for perimenopausal women is 2 or 3%/year at the lumbar spine and 0.6–1% per year at the femoral neck. The estimated BMD loss at the lumbar spine and the femoral neck is 1.3–1.5% and 1–1.4% per year, respectively, in postmenopausal women, with the fastest rate of bone loss occurring immediately after menopause.

The rate of osteoporotic fracture risk is higher in women than in men, such that the loss of BMD in the spine in men is two-thirds the rate in women and that at the femoral neck is one-half the rate in women. Despite these differences, an age-related increase in fractures associated with osteoporosis is evident in men as well. The slower rate of bone loss results in a lower prevalence of osteoporosis in men (~6%) compared to women (~20%), as well as an incidence of hip and vertebral fractures in men older than 65 years that is approximately half that of similar-aged women.

## CONCLUSION

The changes in body weight and composition associated with aging have major public health implications. The increase in total and abdominal obesity usually observed during midlife is associated with increased risk for developing cardiovascular disease, type 2 diabetes, hypertension, dyslipidemia, and other metabolic diseases at an estimated cost of more than \$100 billion per year to the health care system. More important, approximately 300,000 deaths per year may be attributable to obesity. Osteoporotic fractures cost \$17 billion per year due to acute hospitalization and subsequent rehabilitation and, like obesity, increase morbidity and mortality (e.g., there is a 12% mortality rate during the first year after a hip fracture). However, structured programs that emphasize lifestyle behavior changes, such as proper nutrition, increased regular physical activity, and intentional weight loss, reduce the risk for developing metabolic and cardiovascular diseases. Moreover, caloric restriction to induce weight loss of 5–10%, aerobic exercise alone, and weight loss in combination with an aerobic exercise training program reduce the incidence of developing type 2 diabetes and improve the cardiovascular risk profiles of obese individuals. Aerobic and resistive exercise training programs >6 months

duration maintain or increase muscle and bone mass and may prevent the progression to sarcopenia and osteoporosis. Because epidemiologic studies document that the fastest growing segment of the population is individuals >65 years of age and an increase in the average life span, structured behavioral programs that include nutrition and exercise should be recommended to alter the adverse body composition changes observed with aging.

## See Also the Following Articles

Aging: Muscle • Bone Mass Measurement • Bone Structure • Caloric Restriction, Aging and Oxidative Stress • Obesity, Treatment of • Osteoporosis in Older Men • Osteoporosis in Older Women

## Further Reading

- Burger, H., de Laet, C. E. D. H., van Daele, P. L. A., Weel, A. E. A. M., Witteman, J. C. M., Hofman, A., and Pols, H. A. P. (1998). Risk factors for increased bone loss in an elderly population. The Rotterdam study. *Am. J. Epidemiol.* **147**, 871–879.
- Flegal, K. M., Carroll, M. D., and Ogden, C. L. (2002). Prevalence and trends in obesity among US adults, 1999–2000. *J. Am. Med. Assoc.* **288**(14), 1723–1727.
- Fujiwara, S., Fukunaga, M., Nakamura, T., Chen, J. T., Shiraki, M., Hashimoto, T., Yoh, K., Nakamura, T., Mizunuma, H., Tomomitsu, T., Kasagi, F., Masunari, N., and Orimo, H. (1998). Rates of change in spinal bone density among Japanese women. *Calcif. Tissue Int.* **63**, 202–207.
- Gallagher, D., Ruts, E., Visser, M., Heshka, S., Baumgartner, R. N., Wang, J., Pierson, R. N., Pi-Sunyer, F. X., and Heymsfield, S. B. (2000). Weight stability masks sarcopenia in elderly men and women. *Am. J. Physiol.* **279**, E366–E375.
- Hughes, V. A., Frontera, W. R., Roubenoff, R., Evans, W. J., and Fiatarone-Singh, M. A. (2002). Longitudinal changes in body composition in older men and women: Role of body weight change and physical activity. *Am. J. Clin. Nutr.* **76**, 473–481.
- Janssen, I., Heymsfield, S. B., Wang, Z., and Ross, R. (2000). Skeletal muscle mass distribution in 468 men and women aged 18–88 yr. *J. Appl. Physiol.* **89**, 81–88.
- Mazess, R. B., and Barden, H. (1999). Bone density of spine and femur in adult white females. *Calcif. Tissue Int.* **65**, 91–99.
- Ryan, A. S., and Nicklas, B. J. (1999). Age-related changes in fat deposition in mid-thigh muscle in women: Relationships with metabolic cardiovascular disease risk factors. *Int. J. Obesity Relat. Metab. Disorders* **23**, 126–132.
- Ryan, A. S., Nicklas, B. J., and Elahi, D. (1996). A cross-sectional study on body composition and energy expenditure in women athletes during aging. *Am. J. Physiol.* **271**, E916–E921.
- Sheehan, T. J., DuBrava, S., DeChello, L. M., and Fang, Z. (2003). Rates of change for black and white Americans over a twenty year period. *Int. J. Obesity Relat. Metab. Disorders* **27**, 498–504.



# Bone Mass Measurement

Paul D. Miller

University of Colorado Health Sciences Center, Denver, Colorado, United States

## Glossary

**least significant change (LSC)** The minimal change necessary between two measurements to be significant rather than a measurement error as defined by the precision error of the measurement. In bone densitometry, the LSC is defined by the *in vivo* precision error of repeated measurements multiplied by the confidence level desired to be certain that a change in bone mineral density (BMD) is real rather than due to a measurement error.

**prevalence** The percentage of the population being studied that is affected by a particular disease at a given time. In bone densitometry, prevalence is usually applied to the percentage of the population above or below a certain standard deviation cut-point from the mean BMD of the reference population.

**risk** The possibility of loss, injury, disadvantage, or destruction. In bone densitometry, risk is used to define either the relative or absolute risk of fracture at any given BMD value at a particular age.

The clinical application of bone densitometry is one of the advances in the field of osteoporosis that has led to increased patient awareness of this increasingly prevalent disease. Bone densitometry has made it possible for clinicians to diagnose osteoporosis before the first fracture has occurred and to predict risk for fracture in postmenopausal women, men, and patients receiving glucocorticoids. It can be used as a surrogate marker to follow the efficacy of therapies and to examine those patients that might be osteoporosis-specific therapeutic non-responders. Hence, there are three reasons clinicians perform bone mineral density (BMD) measurements: diagnosis using the World Health Organization criteria for osteoporosis, fracture risk prediction, and monitoring the natural progression of diseases that effect BMD or monitoring the therapeutic response to osteoporosis-specific treatments. This article examines each of these applications of bone mass measurements and provides principles of clinical application.

## DIAGNOSIS

In 1994, the World Health Organization (WHO) selected a bone mineral density (BMD) cut-point for defining the prevalence of osteoporosis in the Caucasian postmenopausal female population. The main intent of the WHO committee was to assess the prevalence of low bone mass in the population and relate bone mass to the anticipated lifetime fracture risk in order to advise health policymakers of the potential medical and economic burden of osteoporotic fractures in this population. The use of standard deviation (SD) scores (*t* score) rather than absolute BMD in grams per centimeter squared was decided upon because of the known different absolute BMD calibrations that exist between the BMD measuring devices. A SD score mitigates some of the variance in absolute BMD measured by different manufacturer devices.

The WHO cut-point of  $t = -2.5$  or lower used for the diagnosis of osteoporosis is based on a close association between prevalence at this cut-point and lifetime fracture risk of hip fractures or all fractures (hip, vertebrae, forearm, humerus, and pelvis). At the femoral neck, as assessed by the young, normal reference population database, 16% of postmenopausal women 50 years of age or older have a score of  $-2.5$  SD or less, and the lifetime fracture risk after age 50 in postmenopausal Caucasian women is 16%. In addition, approximately 30% of this population has a score of  $-2.5$  SD or less as measured at the hip, spine, and wrist, which approximates the lifetime fracture risk of all or global fractures after age 50. However, the relationship between low BMD and increased fracture risk is a gradient and not a threshold. Thus, the WHO created a separate classification, osteopenia, to recognize this gradient of risk and to help clinicians assess additional risk factors besides the level of BMD that may lead to increased fracture risk. In this regard, in the National Osteoporosis and Risk Assessment (NORA) dataset, although the 1-year fracture rates and relative risk for fracture/SD

reduction in BMD were highest at  $t$  scores of  $-2.5$  SD or lower, the largest number of fractures in NORA, partially related to the larger sample size of the population with  $t$  scores between  $-1.0$  and  $-2.5$  SD, was seen in this osteopenic category.

Low BMD in the postmenopausal population assessed by any BMD device and by any database is predictive of an increased fracture risk if the value is low; this relationship is also seen in the multiethnic U.S. population. However, for various reasons,  $t$  scores obtained by peripheral devices may not always be as low as  $t$  scores determined from central DXA devices. The major reason for underdetection of osteopenia or osteoporosis is inconsistent young, normal reference population databases: Peripheral devices seem to underestimate the prevalence so that less people may be detected than would be by central DXA. The prevalence of WHO osteoporosis ( $<-2.5$  SD) in the NORA dataset of postmenopausal women 50 years of age or older averaged 7% for all four peripheral devices combined; half of the prevalence was determined by measuring only the femoral neck. Thus, although a low peripheral BMD predicts an increase in fracture risk, any  $t$  score can only be predictive if the value is low. Thus, a normal peripheral BMD should be followed up with a central BMD measurement if the patient has additional risk factors that would lead the clinician to believe that there could be a low central DXA BMD.

Discrepancies between  $t$  scores from various BMD devices are well recognized, and they also exist between different skeletal sites using central DXA technology. Different peripheral devices yield different prevalences of WHO-defined osteoporosis or osteopenia. The development of a standardized young, normal reference population database would do much to mitigate  $t$  score discrepancies and unify the WHO diagnosis between central and peripheral devices. In addition, it has been shown that using a consistent young, normal database mitigates the  $t$  score discrepancies that exist between different devices (heel ultrasound, heel DXA, and hip DXA).

When a clinician must interpret different  $t$  scores obtained by different machines for a single patient, it is difficult to determine whether the patient has osteoporosis. Until the  $t$  score discrepancies can be resolved, the clinician must base his or her opinion on the evidence that the probability of a  $t$  score lower than  $-2.5$  SD at the hip when the heel ultrasound measurement is higher than  $-1.0$  SD is low ( $<10\%$ ), and the probability of a hip  $t$  score being  $<-1.0$  SD if a heel  $t$  score is 0 is also low. On the other hand, if the heel ultrasound is  $<-1.0$  SD, the likelihood that a hip

score is  $<-2.5$  SD is high (70%). In the absence of central DXA access, clinical decisions about risk assessment and interventions may be considered for patients 50 years of age or older if a peripheral  $t$  score is  $<-1.0$  SD. Alternatively, if the physician is uncertain about a particular  $t$  score value and needs confirmation, central DXA should be performed, if available. Although there are approximately 15,000 central DXA devices in the United States, many urban elderly people or rural people of all age ranges have limited access to these "gold standard" devices.

The National Osteoporosis Foundation and the International Society for Clinical Densitometry (ISCD) have undertaken a joint project, called the  $t$ -score equivalence, to try to equate different  $t$  scores from different BMD devices. The fundamental concept behind this project is to select a device-specific  $t$  score cutoff that predicts the same 5-year hip fracture risk as that predicted from the femoral neck  $t$  score at  $-2.5$  SD as determined by the National Health and Nutrition Examination Survey III (NHANES III) reference population database. NHANES III is the only consistent database in existence. It exists only for hip DXA measurements and is the default database for all central DXA manufacturers. Most of the data for  $t$  score equivalence are derived from the Study of Osteoporotic Fractures (SOF), one of the largest prospective epidemiological fracture studies in postmenopausal women in the United States. In SOF, at the age of 70 years, the 5-year hip fracture risk as determined at a  $t$  score of  $-2.5$  SD at the femoral neck is 5%. Using the heel ultrasound device manufactured by Hologic, a manufacturer-specific database derived a  $t$  score at the same age of  $-2.0$  SD. Hence, if a physician had only the Sahara heel ultrasound available and the patient (age specific) had a  $t$  score as determined on the Sahara that predicted the same hip fracture risk as if the patient had a hip DXA performed at a  $t$  score of  $-2.5$  SD, this specific heel  $t$  score would be the value for the diagnosis of osteoporosis (i.e., for the 70 year-old example,  $-2.0$  SD by ultrasound and  $-2.5$  SD by femoral neck). In this way, the physician with a specific peripheral device can use a different  $t$  score cutoff point for the diagnosis of osteoporosis because it predicts an equal risk as a  $-2.5$  SD by the femoral neck, as assessed by the NHANES III database. This short-term solution to the  $t$  score discrepancy problem is supported by the scientific community and the Food and Drug Administration (FDA) regulatory device division.

Despite some limitations, the  $t$  score equivalency project offers the best short-term answer to the clinical

application of the  $t$  score discrepancy problem. Abandonment of the  $t$  score has been suggested for all BMD devices and skeletal sites with the exception of the hip  $t$  score; all other skeletal sites and BMD/ultrasound technologies should be used only for fracture risk assessment. If we abandon the  $t$  score, what will replace it? Absolute BMD measurements in grams per centimeter squared also differ between BMD devices even for the same skeletal site and the same region of interest. This is true for both peripheral devices and central DXA devices. For example, the average BMD of the axial skeleton (L2–L4) in a healthy 20-year-old Caucasian women is 1.25 g/cm<sup>2</sup> by Lunar-GE and 1.00 g/cm<sup>2</sup> using a different device. These different BMD values are related to the different calibrations of the two manufacturers. Since there are so many devices that yield inconsistent absolute BMD values, there is no standardized BMD for all devices related either to prevalence or to fracture risk. However, there is a standardized BMD for the three central DXA devices for the spine and hip. This was developed by performing duplicate BMD measurements on 100 postmenopausal women and by the use of a linear regression to derive a mathematical equation that reduces much of the absolute BMD between the central DXA manufacturers. In the central DXA computer printout reports, the BMDs should automatically be calculated at the bottom of the page so the clinician can compare absolute BMDs between manufacturers. Although a standardized BMD helps in the serial monitoring of patients, there are no data relating BMD to fracture risk prediction. Furthermore, there are no BMDs for the wrist by central DXA or any of the multiple peripheral devices. After the universal, standardized database is completed for both the young, normal reference population database and the older population with and without spine and hip fractures on all FDA-approved devices, it will be possible to calculate a common standardized BMD value for different devices, with this standardized BMD number linked to fracture risk prediction.

## PREDICTION

Both central and peripheral BMD devices can predict an increased risk for fracture. Risk approximately doubles for each standard deviation reduction in BMD in the postmenopausal population. However, the risk associated with low BMD is very dependent on the age of the patient. Risk is far greater as age increases. In fact, the risk for hip fracture is

approximately 45-fold greater at age 80 than at age 50, even at the same BMD or  $t$  score. Although some of the increased risk seen with advancing age may be related to an increase in falls, there are bone quality changes with aging that render a bone more susceptible to fracture at equivalent BMDs. Clinicians cannot capture these qualitative features with BMD devices.

The best documented application of peripheral technologies is fracture risk prediction. The first prospective observation documenting the ability of BMD technology to predict fracture risk was made in 1988 when forearm BMD was shown to predict an increased risk for nonspinal fractures in postmenopausal women. This study was followed by a large meta-analysis that documented the ability of multiple technologies, both peripheral and central, to predict an increased risk for vertebral, nonvertebral, and hip fracture risk. It has been suggested that the femoral neck is a more robust skeletal site for predicting the risk of hip fracture (RR/SD = 2.4). Although this suggestion is based on the SOF head-to-head hip/heel ultrasound device study, the hip fracture risk that was predicted in NORA using wrist or heel DXA was also 2.4, and the receiver operating characteristic (ROC) curves of the two peripheral devices used in NORA match the ROC curves seen with femoral neck DXA obtained from SOF. NORA was not a head-to-head comparison with hip DXA. Nevertheless, it appears that low peripheral device values in the postmenopausal population are powerful predictors of hip fracture risk in untreated postmenopausal women.

What about risk prediction in men and women of other ethnic groups? Data from head-to-head trials comparing fracture rates in Caucasian men and women indicate that both genders have the same absolute hip fracture risk at the same absolute BMD at the femoral neck and that both have the same 10-year absolute hip fracture risk at the same  $t$  score when the  $t$  score is derived from the female NHANES III reference hip database. These head-to-head comparisons, whether based on absolute BMD or SD scores, suggest that there may be no Caucasian gender differences in fracture rates, although there are clear differences in prevalence rates when calculating  $t$  score from a male vs female reference population database, as expected from the basic analysis of the  $t$  score equation. Small differences in the SD of the reference population database profoundly affect the subsequent  $t$  score calculation. Thus, when  $t$  scores are calculated from two different young, healthy reference population databases, they

invariably differ for the same patient. As mentioned previously, the only consistent young, normal reference population database for different BMD measurement devices is the NHANES III database for the central DXA hip.

The percentage of men with  $t$  scores  $< -2.5$  SD at the femoral neck when the  $t$  scores are calculated from a male database is 6% and that from the female database is 4%. There are few differences between the two databases with regard to hip prevalence. Thus, when considering only the hip, the small differences in prevalence between the two genders using NHANES III databases will not miss many men at risk. On the other hand, when the prevalence of osteoporosis by WHO criteria is determined in men from a male or a female database by using the spine, wrist, and hip measurements, it is 19% when the  $t$  score is calculated from a male database and 6% when calculated from a female database. Thus, when  $t$  scores are calculated from a female database, men are underdiagnosed if the clinician examines multiple skeletal sites. Therefore, the number of men determined to be at risk for fracture is underestimated when all three central DXA-measured skeletal sites are combined if a female reference population database is used. It is for this reason that the ISCD Position Development Conference recommended that gender-specific reference databases be used to calculate  $t$  scores. Head-to-head trials examining volumetric BMD (that takes bone size into account) and areal BMD assessed by DXA (that does not take bone size into account) with fracture outcomes are required to determine if there are differences in fracture risk between the genders as a function of bone size. Preliminary data, however, suggest that men may fracture at the same volumetric BMD as women.

In contrast to the equal prevalence of vertebral fractures between men and women, a longitudinal study of incident vertebral fractures in Caucasian men vs women suggested that incident fractures are twice as frequent in women as men. The reason for this difference is unknown since there are only three known risk factors: age, BMD, and the prevalence of vertebral fracture.

The NHANES III reference database provides the only head-to-head ethnic comparison in which both  $t$  and  $z$  scores are calculated from both gender- and ethnic-specific databases for prevalence comparisons. Ethnic-specific databases were calculated from Caucasian, Hispanic, and African American male and female populations ranging in age from 20 to 80 years. There are no Asians populations in the NHANES III database. In this robust study,

there are clear prevalence differences when  $t$  or  $z$  scores from different ethnic groups are calculated from a non-ethnic-specific database. However, no fracture data correlate with the  $t$  or  $z$  scores from the NHANES III database. Hence, there are no data regarding whether there are differences in fracture rates as a function of the ethnicity-derived  $t$  scores from this valuable dataset.

The only multiethnic head-to-head prevalence and fracture data derive from NORA, in which all SD scores were calculated from the Caucasian female reference population database. The RR/SDs for global fractures are similar between Caucasians, American Hispanics, and Native Americans and lower for the American Asian and African American populations. On the other hand, the fracture rates across all ethnicities at  $t$  scores  $< -2.5$  SD are very similar. In the WHO-defined osteopenic category in NORA, American Asians had an unexpectedly low rate of fracture events. The reasons for this are unclear. However, because all of the risk calculations in NORA were done using a Caucasian female reference population database and fracture rates were not too dissimilar across ethnicities, at least for the U.S. multiethnic female postmenopausal population, it might be acceptable to calculate risk from a Caucasian female database, regardless of ethnicity. Native Chinese have a lower hip fracture risk (but similar vertebral fracture risk) despite having a much lower absolute BMD, even after adjusting for body mass index. It is possible that in the United States, lifestyle, nutrition, or even some elements of gene pool mix may result in non-Caucasian ethnic groups having the same fracture risk as Caucasians. Until there are ethnic-specific databases for all BMD manufacturers and prevalence data are linked to risk assessment, the ISCD suggests using ethnic-specific databases for  $t$  score calculation. However, if a physician uses a BMD device without any ethnic-specific database, the Caucasian database can be used for risk assessment.

## MONITORING

The surrogate endpoint markers for monitoring the efficacy of osteoporosis-specific agents are changes in BMD and changes in biochemical markers of bone turnover. Both of these surrogate markers have their proper place in clinical management, although neither is a perfect indicator of pharmacological response or nonresponse to therapeutic interventions to anti-resorptive agents. The prevention or reduction of

fractures is the ultimate goal of treatment. However, none of the osteoporosis-specific agents eliminate fracture risk. Hence, the occurrence of another (i.e., vertebral) fracture while on antiresorptive agents does not necessarily mean that the patient has not had a pharmacological effect of antiresorptive treatment. On the other hand, clinicians are uncomfortable about waiting for another fracture event in order to define treatment effectiveness; thus, the use of surrogate markers to assist therapeutic strategies is analogous to lowering blood pressure or cholesterol to measure treatment effect of antihypertensive or cholesterol-lowering agents.

The surrogate markers for osteoporosis-specific therapies are BMD and biomarkers of bone turnover (BCM); there is increasing evidence for the latter's clinical application. Since most data exist for changes in BMD and BCM with antiresorptive agents, this section emphasizes these two surrogate markers.

Although the FDA requires proof of a significant 3-year vertebral fracture reduction for registration of PMO osteoporosis-specific therapies, the FDA has approved both the once-a-week alendronate and the once-a-week risedronate formulations not based on any prospective-derived fracture data but on evidence that these weekly formulations increase BMD and reduce BCM to equivalent degrees as daily dosing. Hence, the FDA acknowledges that these surrogate markers for bisphosphonates cause improvement in bone strength and a reduction in fracture incidence. Since the once per week bisphosphonate registration has been done with no adequately powered fracture benefit, trust in these surrogate markers is the underpinning for confidence in drug effect. Thus, these FDA registrations are based on the assumption that daily dosing of bisphosphonates must be equivalent to weekly dosing. However, to the extent that the bone half-life may not reflect the bone biological, functional, bisphosphonate half-life, and because the bone binding and alteration in activation frequency of bisphosphonates differ between amino bisphosphonates, these assumptions may or may not be entirely valid.

In addition, there is evidence that bisphosphonates alter bone microarchitecture and, in part, their improvement in bone strength might be unrelated to their capacity to improve BMD or reduce BCM and bone turnover. In this regard, several studies have examined the relationship between the changes in BMD and/or BCM and the magnitude of reduction in either vertebral or nonvertebral incident fractures by robust meta-analysis of randomized, controlled clinical trials. In general, the meta-analyses are

in agreement that a portion (24–54%) of the risk reduction of antiresorptive agents can be attributed to the increase in BMD. Also, these meta-analyses are in agreement that there is an approximately 24% reduction in vertebral fracture risk at the intercept (no change in BMD), indicating that other non-BMD mechanisms play a role in the improvement in bone strength. The meta-analysis of the relationships between the changes in BMD and BCM due to antiresorptive agents and the reduction in nonvertebral fractures shows that all of the effects of antiresorptive agents that reduce the incident of nonvertebral fractures can be attributed to either the magnitude of increase in BMD or the magnitude of reduction in BCM of bone turnover. In this meta-analysis, risk reduction related to either the increase in BMD or the reduction in BCM could not be distinguished because the adjusted variances that the components contribute to risk reduction are so similar. Since the bisphosphonates are the only antiresorptive agents that show reductions in vertebral, nonvertebral, and hip fracture rates, and because they are the only agents that produce large increases in BMD in nonspine sites or a reduction in BCM that might be necessary to cause a reduction in the risk of nonvertebral fractures, the meta-analyses provide evidence to support their therapeutic effect at nonvertebral sites. By reducing bone turnover, there may be microarchitectural changes in bone that might not be reflected in BMD measurements that lead to increased bone strength. Alternatively, the areal BMD measurements that are routinely used in bone densitometry may underestimate the real magnitude of BMD increases. Areal BMD by DXA is a derived equation:  $BMD = BMC/area$ . The bone area may increase without any change or a small change in BMC; thus, the calculated BMD would decline, not necessarily due to any reduction in BMC but due to an increase in bone area. Evidence for this effect of bone area on calculated BMD derives from studies measuring changes in areal BMD by DXA vs quantitative computerized tomography (QCT), which measures the true bone mass and volumetric changes in bone mineral content. QCT changes were larger than DXA changes in this intermittent PTH study, even though BMC increased in both, suggesting that areal BMD measurements may underestimate the change in BMD. Nevertheless, areal BMD at both baseline and longitudinal over time reflects both the basal bone strength as it relates to fracture prediction and the improvement in bone strength as it relates to changes in BMD that are induced by antiresorptives.

## CONCLUSION

Serial BMD has come under scrutiny as a means of monitoring therapy in patients receiving antiresorptive agents. Skepticism is more a function of the performance of the DXA measurement as well as the clinical interpretation. For a clinician to know with certainty that a change in BMD between two measurements in an individual patient is real as opposed to an inherent measurement error of the DXA device or an error of the researcher performing the test, the *in vivo* precision error for the DXA site must be known. Performing daily phantom (*in vitro*) scanning is important in order to detect a “drift” in serial BMD values that may indicate the need to obtain manufacturer assistance to search for problems that affect machine performance, such as X-ray tube viability and software deterioration. However, phantoms do not move, whereas patients do. For a DXA facility to determine whether differences are due to the machine or biological, *in vivo* precision studies must be performed. The basic principle is to determine the *in vivo* coefficient of variation % (CV%) between multiple measurements, which allows the physician to determine whether the difference between two measurements is within or beyond the least significant change (LSC). If the independent precision error is not known for a DXA site, changes in BMD may be assumed to be significant (or nonsignificant) when in fact they may not be. For example, when the spine BMD increases slightly (~2%) and the total hip BMD decreases (~4%), DXA reports often indicate that the patient is nonresponding when in fact the precision error at the total hip is ~2%, requiring at least a 5.6% change in BMD to be significant. Unfortunately, when the BMD change is within the LSC but misinterpreted to be significant, many patients are often taken off therapy or their therapy is changed. The densitometry community has set a standard requiring a 95% confidence limit for determining whether a difference between two BMD measurements in an individual patient is significant. Using this standard, the change in BMD must be 2.77 times the CV% (precision error). Hence, *in vivo* precision studies must be performed to competently interpret BMD differences between two measurements.

None of the peripheral devices show changes in BMD between measurements in patients on antiresorptive therapy. This is not due to precision error differences between central and peripheral devices because these devices have similar errors. For reasons that are unclear, none of the peripheral technologies have shown BMD changes in any antiresorptive

clinical trials. In fact, there was a loss of forearm BMD in patients in an anabolic trial using teriparatide, which might be an artifact of the areal DXA measurement in which the bone area is increased by subperiosteal bone apposition. As measured by peripheral QCT in the teriparatide dataset, bone size increased, as did bone strength. Hence, with regard to monitoring BMD with teriparatide, areal (especially spine) DXA and QCT may be complimentary. However, one of the frustrating aspects of bone densitometry is the inability to monitor pharmacological therapy with peripheral devices, whereas they can be used quite appropriately to assess risk. Until improved methodology allows peripheral devices to be used for monitoring therapy, the physician has three choices when starting osteoporosis-specific therapy:

1. Do not monitor.
2. Use axial (central) DXA to obtain a baseline for monitoring since this is the preferred methodology for monitoring osteoporosis-specific therapies.
3. Use BCM of bone turnover to assess the bone biological response to the antiresorptive agent. Data suggest that if there is at least a 40% decline in urinary collagen cross-link markers 1–3 months after initiating a bisphosphonate, there is a greater likelihood that the 2-year follow-up repeat DXA will increase or at least not decline. In addition, data on the relationship between the magnitude of increase in BMD and the magnitude of decrease in BCM with the various antiresorptive agents suggest that the greater the decrease in BCM, the greater the risk reduction in vertebral as well as nonvertebral fractures. The strength of this relationship was as powerful as the strength of the relationship between the changes in BMD and the risk reduction. Hence, if a BCM declines adequately, a fracture-reduction benefit can be anticipated.

No professional scientific organization recommends substitution of BCM for DXA as a means of monitoring therapy. The two should be complimentary. The advantage of the 1- to 3-month BCM assessment is that it provides earlier feedback to the patient and the clinician: If it does decline below the LSC for BCM, then the following clinical assumptions are fair: The patient is taking the drug, the fastidiously absorbed oral bisphosphonates are being absorbed, and there is evidence for a bone biological effect.

Last, it is important to address the issues of “nonresponders” as measured by BMD and the principle of regression to the mean as it has been applied to clinical osteoporosis management. From clinical trials, it is assumed that the response rate

to bisphosphonates exceeds 90%. Clinic patients differ from clinical trial patients. Clinical trial patients are carefully selected. They have no secondary diseases affecting bone, nor are they taking other drugs that might alter bone metabolism. Clinical trial patients are highly motivated and have frequent compliance (i.e., pill-count) checks. The real-world response rate to osteoporosis-specific therapies is unknown.

Regression to the mean is an old statistical observation. It has been used in BMD testing to examine the change from pretreatment baseline in two pharmacological clinical trials of postmenopausal osteoporosis. This study stated that those patients treated for the first 2 years of these trials who either lost or gained bone during the first year subsequently either gained or lost bone during the second year, such that when the two data points were regressed to mean values, there were no differences. The study suggested that measuring BMD over time may not have meaning for those patients who gain BMD during the first year and lose it during the second year and those who lose it during the first year and gain it during the second year. One scientific flaw of these data is that it is a statistical fact that regression to the mean occurs in all biological measurements, especially at the extremes of a distribution, such that values that are high subsequently decline and those that are low subsequently increase due to regression to the mean in both treated and control groups. In addition, the control group data were not shown; in order to determine whether regression to the mean has any real meaning, the differences between the treated and nontreated groups at the regression point must be shown. Regression to the mean can never be applied to individual patients since it is always a group phenomenon seen at the extremes of distribution.

Serial BMD and increases in BMD while on osteoporosis-specific therapy are a valid surrogate marker of assessing improvement in bone strength. BMD increases account for a significant proportion but not all of the fracture risk reduction seen with antiresorptive agents.

BMD testing has allowed clinical application in diagnosis, risk prediction, and monitoring of disease or therapy. Without BMD availability, the field of osteoporosis would be relegated to guesswork. However, like all biological measurements, BMD testing at baseline and/or longitudinally is imperfect. Only competent clinical judgment and subsequent interpretation of bone mass measurement performance and data will allow this important quantitative technology to remain an important clinical tool.

## See Also the Following Articles

Bone Structure • Bone Turnover Markers • Osteoporosis in Older Men • Osteoporosis in Older Women • Osteoporosis, Overview

## Further Reading

- Binkley, N. C., Schmeer, P., Wasnich, R. D., and Lenchik, L. (2002). What are the criteria by which a densitometric diagnosis of osteoporosis can be made in males and non-Caucasians? *J. Clin. Densitom.* **5**, S19–S28.
- Bonnick, S. L. (2000). Monitoring osteoporosis therapy with bone densitometry: A vital tool or regression toward mediocrity? *J. Clin. Endocrinol. Metab.* **10**, 343–345.
- Bonnick, S. L., Johnston, C. C., Jr., Kleerekoper, M., Lindsay, R., Miller, P., Siris, E., and Sherwood, L. (2001). The importance of precision in bone density. *J. Clin. Densitom.* **4**, 105–110.
- Faulkner, K. G., VonStetten, E., and Miler, P. D. (1999). Discordance in patient classification using T scores. *J. Clin. Densitom.* **2**, 343–350.
- Hochburg, M. C., Greenspan, S., Wasnich, R. D., Miller, P. D., Thompson, D. E., and Ross, P. D. (2002). Changes in bone density and turnover explain the reductions in incidence of non-vertebral fractures that occur during treatment with antiresorptive agents. *J. Clin. Endocrinol. Metab.* **87**, 1586–1592.
- Hui, S. L., Slemenda, C. W., and Johnston, C. C., Jr. (1988). Age and bone mass as predictors of fracture in a prospective study. *J. Clin. Invest.* **81**, 1804–1809.
- Kanis, J. A., and Gleur, C. C. (2000). An update on the diagnosis and assessment of osteoporosis with densitometry. *Osteoporosis Int.* **11**, 192–202.
- Lenchik, L., and Watts, N. M. (2001). Regression to the mean: What does it mean? *J. Clin. Densitom.* **4**, 1–4.
- Looker, A. C., Wahner, H. W., Dunn, W. L., Calvo, M. S., Harris, T. B., Heyse, S. P., et al. (1995). Proximal femur bone mineral levels of US adults. *Osteoporosis Int.* **5**, 389–409.
- Marshall, D., Johnell, O., and Wedel, H. (1996). Meta-analysis of how well measurements of bone mineral density predict the occurrence of osteoporotic fractures. *Br. Med. J.* **312**, 1254–1259.
- Melton, L. J., III, Atkinson, E. J., O'Conner, M. K., O'Fallon, W. M., and Riggs, B. L. (1998). Bone densitometry and fracture risk in men. *J. Bone Miner. Res.* **13**, 1915–1923.
- Miller, P. D. (2002). Controversial issues in bone densitometry. In "Principles of Bone Biology" (J. P. Bilezikian, L. Raisz, and G. Rodan, eds.), 2nd ed., pp. 1587–1597. Academic Press, New York.
- Miller, P. D., and Bonnick, S. L. (1999). Clinical application of bone densitometry. In "Primer on the Metabolic Bone Diseases and Disorders of Mineral Metabolism" (M. Favus, ed.), 4th ed., pp. 152–159. Lippincott Williams & Wilkins, Philadelphia.
- Miller, P. D., Baran, D., Bilezikian, J. P., Greenspan, S., Lindsay, R., Riggs, B. L., and Watts, N. (1999). Practical clinical application of biomarkers of bone turnover. *J. Clin. Densitom.* **2**, 323–342.
- Miller, P. D., Zapaolowski, C., Kulak, C. A. M., and Bilezikian, J. P. (1999). Bone densitometry: The best way to detect osteoporosis and to monitor therapy. *J. Clin. Endocrinol. Metab.* **84**, 1867–1871.
- Miller, P. D., Siris, E. S., Barrett-Conner, E., Faulkner, K. G., Wehren, L. E., Abott, T. A., Chen, Y. T., Berger, M. L., Santora, A. C., and Sherwood, L. M. (2002). Prediction of fracture

- risk in postmenopausal Caucasian women with peripheral bone densitometry: Evidence from the National Osteoporosis Risk Assessment (NORA). *J. Bone Miner. Res.* **17**, 2222–2230.
- Orwoll, E. (2000). Assessing bone density in men. *J. Bone Miner. Res.* **15**, 1867–1870.
- Ross, P. D., Davis, J. W., Epstein, R. S., and Wasnich, R. D. (1998). Pre-existing fracture and bone mass predict vertebral fracture incidence in women. *Ann. Intern. Med.* **114**, 919–923.
- Siris, E. S., Miller, P. D., Barrett-Conner, E., Faulkner, K. G., Wehren, L. E., Abbott, T. A., Merger, M. L., Santora, A. C., and Sherwood, L. M. (2001). Identification and fracture outcomes of undiagnosed low bone mineral density in postmenopausal women: Results from the National Osteoporosis Risk Assessment (NORA). *J. Am. Med. Assoc.* **286**, 2815–2822.
- Wasnich, R. D., and Miller, P. D. (2001). Antifracture efficacy of antiresorptive agents are related to changes in bone density. *J. Clin. Endocrinol. Metab.* **85**, 231–236.



# Bone Remodeling, Dynamics of

Susan M. Ott

University of Washington, Seattle, Washington, United States

## Glossary

**activation frequency** The frequency at which a particular point on the bone surface will be reached by a basic multicellular unit.

**basic multicellular unit (BMU)** A group of osteoblasts and osteoclasts that resorb a volume of bone and replace it with new bone.

**bone mineral density** The weight of mineral per volume of bone tissue.

**bone tissue** The total material found within a bone, including the bone marrow, fat, mineralized bone, and unmineralized osteoid.

**bone turnover** An ambiguous term that is either the bone formation rate or the bone resorption rate and is meaningful only when these rates are equal.

**bone volume** The bone material, including mineralized bone and unmineralized osteoid.

**mineral apposition rate** The rate at which mineral is deposited at a discrete point on the bone surface; expressed in micrometers per day.

**mineralization density** The weight of mineral per bone volume.

**osteoid** Newly formed bone, composed primarily of lamellar collagen with small amounts of other bone proteins and glycoproteins, that forms a matrix that can accept deposition of hydroxyapatite crystals and that will become mineralized bone.

**wall thickness** The thickness (in three dimensions) of a completed volume of new bone, measured perpendicularly from the interface with the bone marrow.

**B**one remodeling is a physiological process in which bone is resorbed and replaced with new bone. It occurs in millions of discrete locations within the bone, with an overall effect of continuous bone renewal.

## INTRODUCTION

Bone is such a solid substance that an orthopedic operating room resembles a carpenter's workshop, with its saws, drills, screws, and hammers used to repair broken and damaged furniture. This reinforces the concept that bones are mechanical structures; indeed, one of the major functions of the skeleton is to support the rest of the body. The dynamic properties of the bone, however, are seldom considered, except to acknowledge that living bone, unlike the leg of a table, can repair a fracture. Actually, the bone is continually remodeling, responding to external stresses and internal signals. When bone remodeling is not in balance, osteoporosis results. Most pharmaceutical agents that improve bone strength act by altering the bone remodeling sequence.

## BASIC MULTICELLULAR UNIT

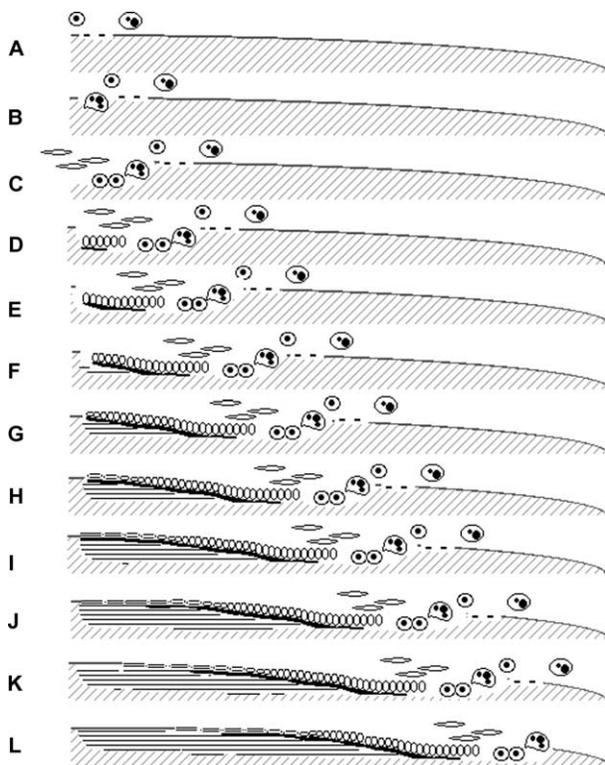
A fundamental property of bone remodeling is that it occurs in discrete locations and involves a group of different kinds of cells. This secondary level of organization, analogous to the nephron, was named the basic multicellular unit (BMU) by Harold Frost. Unlike a nephron, a BMU is not a permanent structure. In cancellous bone, the BMUs travel along the bone surface, spreading over an area. In the cortical bone, they tunnel through the bone, with the osteoclasts at the cutting edge and newly formed bone in the wake.

At any moment, ~20% of the cancellous bone surface is undergoing remodeling, and at any point on the surface remodeling will occur on average every 2–4 years. This rate is the activation frequency.

The cells in a BMU are osteoclasts and osteoblasts. The individual cells in a BMU have shorter active life spans than that of the BMU; osteoclasts and osteoblasts are formed from precursor cells and undergo

apoptosis, with replacement by new cells. Some of the osteoblasts differentiate into osteocytes or lining cells that remain in the newly formed bone. Other cells, including bone marrow stromal cells, bone marrow adipocytes, nerve cells, and vascular endothelial cells, can influence the BMU but are not actively resorbing or forming bone.

Each BMU performs its functions in the same sequence: origination and organization of the BMU, activation of osteoclasts, resorption of old bone, recruitment of osteoblasts, formation of new bone matrix, and mineralization (Fig. 1). Many of the details of this sequence have been elucidated by Michael Parfitt, and measurements of the timing of each phase were done by Erik Eriksen.



**Figure 1** The basic multicellular unit (BMU), moving along a cancellous surface. Each step represents approximately 10 days, and the BMU moves approximately 10  $\mu\text{m}$  each day. (A) Origination of BMU: lining cells contract to expose collagen and secrete RANK ligand. (B) Osteoclasts fuse into multinucleated cells that resorb cavity. (C) Mononuclear cells continue resorption, and preosteoblasts are stimulated to proliferate. (D) Osteoblast team forms at the bottom of the cavity and starts forming osteoid. (E) Osteoblasts continue forming osteoid (●) and previous osteoid starts to mineralize (horizontal line). (F–H) Osteoblasts continue formation and mineralization. (I and J) Osteoblasts begin to flatten. (K and L) Osteoblasts turn into lining cells; bone remodeling at initial surface (left of drawing) is now complete, but BMU is still advancing (to the right).

## SEQUENCE OF BONE REMODELING

### Origination

There is debate about which factors are the most important for initiation of new BMUs. Small cracks caused by microdamage will initiate a BMU. Signals from osteocytes or lining cells that have detected mechanical stress, local cytokines from marrow cells, or systemic hormones that regulate mineral metabolism may also initiate a BMU. Possibly, the BMUs commence at random.

Mori and Burr demonstrated an association between fatigue damage and intracortical remodeling. They anesthetized mature dogs and applied a cyclic load to the radius, which caused asymptomatic microscopic cracks. Eight days later, an identical load was applied to the opposite radius. Histologic sections showed an equal number of cracks on each side. However, at the site with the earlier load, there was a significant increase in resorption cavities that were adjacent to the cracks. The temporal design of the study demonstrated that the resorption occurred after the fatigue damage.

Origination occurs only once for each BMU at a quiescent surface of the cancellous bone or on a surface nearest a crack in the cortical bone. Lining cells change their shape from that of flat, epithelial-like cells to rounded cells, thereby exposing some of the collagen matrix. They also secrete collagenase to expose the bone mineral. These activated cells then produce RANK ligand, which binds to receptors on preosteoclasts and causes them to fuse and become mature multinucleated osteoclasts. In the cancellous bone, some of the lining cells appear to form a layer or canopy that separates the bone marrow from the resorbing bone.

### Activation

Activation is a continuing process that occurs at the cutting edge of the BMU, and as the BMU spreads, new surfaces undergo activation. The BMU “front” travels at a rate of approximately 10  $\mu\text{m}/\text{day}$  and is followed by osteoblasts and, in the cortex, a new blood vessel. The original osteoclasts undergo apoptosis while the BMU is still progressing; thus, new cells must replenish the dying ones. Where do the replacement osteoclasts come from? Systemic hormones, growth factors, and interleukins may enlarge the precursor pool, but systemic factors cannot localize the preosteoclasts to the cutting edge. The new vascular cells may determine when circulating osteoclasts

should enter the tissue. It is not clear how the replacement osteoclasts are formed from preosteoclasts, but this probably involves RANK ligand from either adjacent lining cells or osteoblasts within the BMU.

### Resorption

There are two phases of resorption. The first is the most rapid, carried out by multinucleated osteoclasts, and lasts ~8 days. Then a slower phase occurs that involves mononuclear cells and lasts ~34 days. Multinucleated osteoclasts are active for ~12 days and then undergo apoptosis. This process may be promoted by transforming growth factor- $\beta$  (TGF- $\beta$ ). The suicidal cells have been located at the junction between the resorbing surface and the reversal surface, which suggests that the process might also be involved in signaling new osteoblasts. Osteoclasts have been shown to excrete interleukin-6 (IL-6) and annexin-II, both of which may signal new osteoblasts. The depth of the eroded cavity is also linked to the life span of the active osteoclasts so that early apoptosis will result in a shallower eroded cavity.

During resorption, bone-derived growth factors are released, including TGF- $\beta$ , insulin-like growth factor (IGF), and fibroblast growth factor (FGF), which were deposited into the matrix by the previous generation of osteoblasts. Some, such as TGF- $\beta$ , may be activated by the acid environment caused by osteoclastic proton secretion. These growth factors (delayed autocrine factors) might account for the coupling between resorption and formation that is seen in normal situations, but direct evidence for this theory is lacking. High correlations between total skeletal resorption and formation may also be seen if the biochemical stimuli (e.g., IL-6) for origination of BMUs also participated in the recruitment of osteoblasts.

### Formation

After the maximum eroded depth has been achieved, there is a reversal phase that lasts ~9 days. During this phase, the osteoblasts converge at the bottom of the cavity. The osteoblasts derive from precursors in the marrow stromal cells, which can differentiate into either osteoblasts or adipocytes, depending on the transcription factor *Cbfa1*. The team of osteoblasts then begins to form osteoid. After 15 days, the osteoid begins to mineralize. The osteoblasts continue to form and to mineralize osteoid until the cavity is filled or nearly filled, which takes 124–168 days at any point on the surface.

At the bottom of the cavity, the new osteoblasts are plump and vigorous, have tall nuclei, and make a thick layer of osteoid. The cells then gradually flatten as they slow production, and finally they become quiescent lining cells. Some of the osteoblasts differentiate into osteocytes and remain in the matrix. The osteocytes may secrete inhibitory factors that slow the rate of bone formation as the resorbed cavity is nearly filled. As the BMU progresses, new osteoblasts are added, but only at the edge behind the osteoclasts.

The density of the osteoblasts at the formation site may vary. When the cells are more crowded, they are taller and narrower, and they collectively can make more osteoid than when there are fewer cells. Osteoporotic patients have the same rate of osteoid production per cell, but overall the wall thickness is decreased and the amount of newly formed bone is inadequate to fill the resorbed cavity. Some investigators think that bone formation is discontinuous (on and off), but studies with tetracycline labeling suggest that the osteoblast team does not take vacations.

### Mineralization

Mineralization begins ~15 days after osteoid has been formed. In most situations (except in the case of osteomalacia), the average rate of osteoid formation and the rate of mineralization are the same and are measured by tetracycline labels.

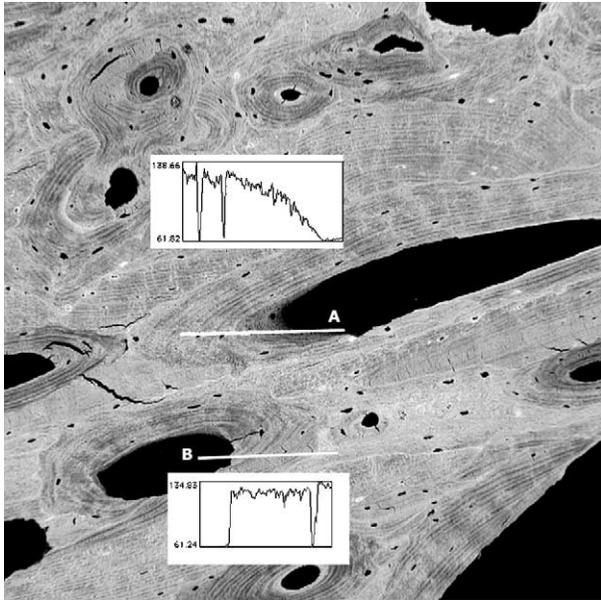
After the BMU has completely restored the bone volume, the mineralization density continues to increase as the crystals become more densely packed (Fig. 2). It is not known how long it takes to reach the maximum mineralization density, but indirect evidence suggests this takes approximately 3 years.

The mineralization density is related to the bone formation rate. When the bone formation rate is high, mineralization density is low. When the bone formation rate is low, more of the bone is older and more highly mineralized.

## CONTROL OF BONE REMODELING

### Mechanical Stress

The osteocytes and lining cells form a network that allows direct cellular communication. They are ideally located to detect mechanical stress. The osteocytes secrete factors such as IGF that probably lead to origination of a BMU.



**Figure 2** Electron backscattered image of iliac cortical bone showing two new osteons, with plot profiles taken along the white lines. The line A profile shows that density is low in the newly formed osteoid (dark gray) and gradually increases as the bone becomes older (toward the left). The line B profile shows an abrupt change from the Haversian canal space to the bone; this line crosses a surface of bone resorption. The downward spikes in the plot profiles represent the osteocyte lacunae. Near the end of line B (right), there is higher mineralization after line B crosses the cement line into an area of older bone, which appears whiter on the image.

### Growth Factors and Cytokines

Locally produced cytokines, such as IL-1,-6, and-11, can stimulate the lining cells and osteoblasts. Growth factors such as TGF- $\beta$ , FGF, bone morphogenic proteins, platelet-derived growth factors, and colony-stimulating factors can increase proliferation of osteoblasts or preosteoblasts. Osteoprotegerin, which is a decoy receptor for RANK ligand, is produced locally by osteoblasts and can inhibit the fusion of preosteoclasts.

### Hormones That Regulate Mineral Metabolism

Parathyroid hormone (PTH) and PTH-related protein can activate lining cells to secrete RANK ligand and originate a BMU. PTH may also affect osteoblast proliferation. When given as intermittent injections, it can bypass the bone remodeling system and stimulate bone formation directly on quiescent surfaces. PTH-related protein causes increased bone resorption and loss of bone during lactation.

Vitamin D can also activate lining cells to secrete RANK ligand. It increases the proliferation of preosteoclasts and may play a role in osteoblast recruitment. Indirectly, it is important in the maintenance of serum mineral levels.

Calcitonin receptors are present on osteoclasts, and when activated, the cells stop resorbing bone. The calcitonin gene-related protein appears to increase bone formation, but the mechanism is not clear.

Estrogen exerts a major effect on bone remodeling, possibly through different mechanisms. Estrogen can inhibit IL-6 production, which reduces bone resorption. The sensitivity of lining cells to mechanical stress depends on the estrogen levels. Estrogen also controls the timing of osteoclast apoptosis so that estrogen deficiency results in a longer life span of the osteoclasts.

### Vasculature

Each BMU is associated with a capillary. In cortical bone, the capillary grows along the excavated tunnel. On trabecular surfaces, small capillary-like structures are frequently seen adjacent to the osteoblasts. The cells that form these structures have characteristics that are unlike endothelial cells and resemble lining cells. In addition to supplying nutrients and a source of precursor cells, the vasculature may help direct the localization of the BMU.

### Nerves

Anatomic studies have documented a dense and intimate innervation of bone tissue. The fibers contain markers for neural tissue, both sensory fibers and sympathetic fibers. Nerve endings are seen in contact with bone cells. Glutamate is expressed in the fibers that are in proximity to bone cells, suggesting a potential role of glutamatergic innervation in the bone remodeling process.

Recent evidence has demonstrated a role for the central nervous system in the control of bone formation rates. Leptin injected into brains of mice causes a substantial decrease in bone formation rates at doses that do not affect body weight. These effects appear to be transmitted by the sympathetic nerves.

### Adipocytes

Adipocytes and osteoblasts derive from the same precursors, and factors such as lipid levels can favor differentiation into adipocytes, which will reduce the number of available osteoblasts. The adipocytes

also secrete leptin, but the ability of leptin to either stimulate or inhibit bone formation is still debated.

## BONE REMODELING IN OSTEOPOROSIS

When the bone resorption rate is higher than the bone formation rate, the bone volume will decrease, resulting in osteoporosis. The balance between formation and resorption is more important than the rate of either one alone. Thus, postmenopausal osteoporosis is a heterogeneous disorder, with either high or low bone formation rates. Menopause is associated with an increase in the origination of BMUs, possibly because estrogen deficiency results in increases in IL-6 and other cytokines. Also, longer osteoclast life span results in more bone resorption. This alone does not necessarily cause loss of bone mass, but each BMU does not completely replace as much bone as was resorbed.

With aging, the osteoblasts lose their ability to fill the resorbed spaces. This is shown by age-related decreases in the wall thickness. When bone porosity increases, the remaining bone accumulates microdamage at an exponential rate. A vicious cycle is begun: Bone mass decreases, so the remaining bone is subject to more fatigue damage that increases bone resorption, which may further weaken the bone and disrupt the osteocyte network.

Understanding bone remodeling is important for predicting the response to therapeutic agents used for osteoporosis. Antiresorptive agents (estrogen, bisphosphonates, risedronate, and calcitonin) all decrease bone resorption, but they also decrease bone formation. The more potent amino bisphosphonates decrease the tetracycline-labeled bone formation rate by ~95%. Small increases in bone volume occur while the BMUs that were forming bone continue to fill in the eroded cavities. Eventually, a new steady-state bone volume is reached, after which there is no further gain in bone volume because bone remodeling has been inhibited. Bone density will increase even after bone volume has reached a steady state because the mineralization density continues to increase. Women taking bisphosphonates have increasing bone density at the hip for ~3 years; thereafter, the bone density does not increase further. The long-term (>7 years) consequences of prolonged suppression of bone formation rates are not known, but potentially fatigue damage and microcracks could accumulate.

Intermittent PTH is an anabolic agent that increases both bone formation and resorption, but the balance is positive, with large increases of bone

density. Mineralization density decreases because a larger proportion of the bone is newly formed. Therefore, intermittent PTH has a completely different mechanism of action from that of antiresorbing agents. It is hoped that research on combinations or the sequential use of anabolic and antiresorbing agents will yield new treatment regimens that will prevent more fractures than the current treatments.

## OTHER MECHANISMS OF ALTERING BONE STRUCTURE

### Microcallus Formation

When trabeculae fracture, a microcallus can form that resembles the familiar "macrocallus" of a fractured bone. These microcallus formations in the spine increase after age 50, mostly in the lower thoracic and lumbar spine near the endplates. There are more in females than in males and even more in osteoporotic persons. These formations can account for up to 10% of the trabecular bone volume. Microcallus formations can allow creation of new trabeculae by forming bridges between existing trabeculae. They undergo resorption and modeling, eventually becoming mineralized and indistinguishable from trabecular bone. This mechanism of altering trabecular structure could be relatively important in patients with osteoporosis.

### Direct Activation

In some situations, bone formation can also occur along surfaces in the absence of previous resorption. With fluoride therapy, bone formation surfaces increase markedly, but the bone is woven and not normal lamellar bone. Beagles treated with aluminum show new bone formation in some, but not all, experimental conditions, but this phenomenon is not seen in humans, who develop osteomalacia when exposed to parenteral aluminum. Intermittent PTH causes increases in the bone formation rate that occur too soon to be explained by the bone remodeling cycle. Measurements of the distance from new bone edges to cement lines also suggest that previously quiescent surfaces start to form bone.

### Bone Arising from the Marrow Space

Patients with metastatic prostate cancer can develop osteoblastic lesions within the marrow spaces. First, spindle-shaped cells with characteristics of osteoblasts appear in the marrow spaces adjacent to the cancer

cells, and then these calcify and become woven bone. The bone eventually becomes osteosclerotic.

### Periosteal Bone Formation

The periosteal surfaces of cortical bone gradually expand throughout adult life. This compensates for loss of bone tissue because bone with a larger diameter has greater bending strength. Periosteal expansion occurs without prior bone resorption. In some cases, when there have been repetitive stresses on the bone, woven bone will form at the periosteal surface. This has been noted, for example, in young military recruits during basic training. The woven bone can be formed more rapidly than the lamellar bone formed during usual bone remodeling.

### See Also the Following Articles

Bone Mass Measurement • Bone Structure • Bone Turnover Markers • Fibroblast Growth Factor (FGF) • Insulin-like Growth Factors • Interleukin-6 • Osteoporosis in Older Men • Osteoporosis in Older Women • Osteoporosis, Overview • Parathyroid Hormone (PTH) • Platelet-Derived Growth Factor (PDGF) • Vitamin D

### Further Reading

- Boivin, G., and Meunier, P. J. (2002). Changes in bone remodeling rate influence the degree of mineralization of bone. *Connect. Tissue Res.* **43**, 535–537.
- Chavassieux, P. M., Arlot, M. E., Reda, C., *et al.* (1997). Histomorphometric assessment of the long-term effects of alendronate on bone quality and remodeling in patients with osteoporosis. *J. Clin. Invest.* **100**, 1475–1480.
- Ducy, P., Amling, M., Takeda, S., *et al.* (2000). Leptin inhibits bone formation through a hypothalamic relay: A central control of bone mass. *Cell* **100**, 197–207.
- Eriksen, E. F., Gundersen, H. J. G., Melsen, F., *et al.* (1984). Reconstruction of the formative site in iliac trabecular bone in 20 normal individuals employing a kinetic model for matrix and mineral apposition. *Metab. Bone Dis. Rel. Res.* **5**, 243–252.
- Hahn, M., Vogel, M., Amling, M., *et al.* (1995). Microcallus formations of the cancellous bone: A quantitative analysis of the human spine. *J. Bone Miner. Res.* **10**, 1410–1416.
- Mori, S., and Burr, D. B. (1993). Increased intracortical remodeling following fatigue damage. *Bone* **14**, 103–109.
- Parfitt, A. M. (1994). Osteonal and hemi-osteonal remodeling: The spatial and temporal framework for signal traffic in adult human bone. *J. Cell. Biochem.* **55**, 273–286.
- Parfitt, A. M., Villanueva, A. R., Foldes, J., *et al.* (1995). Relations between histologic indices of bone formation: Implications for the pathogenesis of spinal osteoporosis. *J. Bone Miner. Res.* **10**, 466–473.



# Bone Structure

Paul R. Odgren, Alison Gartland, April Mason-Savas, and Sandy C. Marks, Jr.<sup>†</sup>  
*University of Massachusetts Medical School, Worcester, Massachusetts, United States*

## Glossary

**compact bone** Solid, dense bone, such as that found in the shafts of the limb bones or the vertebral bodies (see trabecular bone), also called cortical bone when it occurs in the bone shaft, or lamellar bone due to its highly organized layers of collagen fibrils.

**diaphysis** The cylindrical shaft of a long bone composed of cortical bone and enclosing a marrow cavity.

**epiphysis** The region at the end of a long bone, often enlarged in a knob that is covered with articular cartilage and that transmits mechanical load from the joint to the rest of the bone.

**Haversian system** In large mammals, including humans, microscopic, concentric rings of bone organized around central canals containing small blood vessels.

**metaphysis** The region where a long bone flares between the cylindrical shaft to the broadened epiphysis that contains trabecular bone and is a site of metabolic activity; also called the primary spongiosa.

**osteoblast** Bone-synthesizing cell residing on the surface of bone and secreting type I collagen-rich protein matrix and vesicles that contain the mineral components of bone.

**osteoclast** Large, multinucleated bone-resorbing cell that forms by the fusion of mononuclear precursors of the monocyte lineage and attaches to bone, releasing degradative enzymes and absorbing the digested bone.

**osteocyte** Mature bone-forming cell that is completely surrounded by bone matrix, remaining in contact with the circulation and other bone cells via a dense network of cytoplasmic processes that permeate the surrounding bone.

**trabecular bone** Bone comprising an open network of fine spicules; also called spongy bone; occurring internally within skeletal elements, e.g., near the ends of the long bones and in the pelvis and vertebrae.

Contrary to common perception, bone is a complex and dynamic tissue throughout life. It plays critical roles in both mechanical support and mineral homeostasis, and its macroscopic and microscopic structure vary widely at different sites to reflect these functions. Specialized cell types cover all bone surfaces and fully inhabit the bone matrix. Principal cell types are osteoblasts, which secrete bone matrix, osteoclasts, which resorb bone, and osteocytes, which inhabit the matrix. Bone is a composite of protein and mineral. It derives tensile strength from a highly ordered protein matrix rich in type I collagen and compressive and shear strength from crystals of the hydroxyapatite form of calcium phosphate. Blood vessels permeate bone through a system of Haversian canals. Bone provides a microenvironment rich in growth factors and other molecules that participate in its complex metabolism and its interactions with other cells and tissues.

## INTRODUCTION

The structure of bone can be usefully considered in several different ways that reflect its variety of functions and constituents. Its precise form and composition vary with location, nature having selected optimal designs for different roles. First and most obviously, bone provides mechanical support and protection. It is a composite material, with tough, resilient protein fibers embedded in a hard and rigid mineral matrix (Fig. 1). Its great advantage over sophisticated human-designed composite materials is its ability to repair itself. Within a single skeletal element, there usually are dramatically different forms of bone, some, such as the cortical shaft, designed for load bearing and some, such as trabecular bone in the metaphysis, providing both a supporting network and a fractally large surface area for metabolic responses to the requirements of mineral homeostasis. Both of these types of bone are evident in the longitudinal section of the human tibia depicted in Fig. 2.

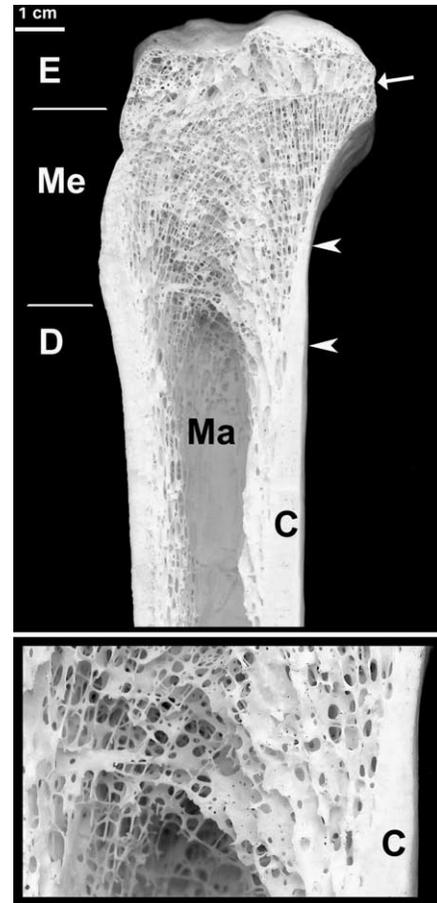
Given its obvious structural role, bone is not always appreciated as one of the more complex and dynamic

<sup>†</sup>Deceased.



**Figure 1** Composition of bone. Human fibula before (left) and after (right) extraction with (ethylenedinitrilo)tetraacetic acid (EDTA). EDTA removes the mineral by chelation and what remains is the tough, fibrous protein matrix, 90% of which is type I collagen. This demineralized bone matrix has very high tensile strength, greater than that of steel on a per weight basis. It has, however, little resistance to compressive or bending forces, enabling it to be tied in a knot. The calcium phosphate crystalline form (hydroxyapatite) that constitutes the mineral component has high compressive strength, but alone is brittle and has little resistance to tensile forces. Together they make the hard, somewhat flexible, composite material known as bone.

tissues in the body. Its cellular constituents include not only the bone-forming osteoblasts and the bone-resorbing, phagocytic osteoclasts, but also the matrix-embedded osteocytes, which maintain communication with one another and with the bone surface from deep within the solid bone matrix through a dense, interconnected network of submicrometer-sized canaliculi. Interactions between bone cells and those of cartilage, blood vessels, teeth, and cells that degrade and/or remodel the extracellular matrix during growth are regulated in part locally, in part systemically, and in part by mechanisms that are still under intensive investigation.



**Figure 2** Structure of a typical long bone (longitudinal section of an adult human tibia with all cellular material removed). Typically, a long bone can be divided into three regions. The epiphysis (E) of the tibia is the widened “knob” that articulates at the knee joint (the joint cartilage is removed in this specimen). The central shaft, or diaphysis (D), is formed by a cylinder of dense, high-strength cortical bone (C) that surrounds a marrow cavity (Ma). Between these two regions, the metaphysis (Me) is filled with a highly anastomosing network of bony trabeculae, or spicules, that both distributes mechanical load between the joint and the bone shaft and provides an extraordinarily large surface area for metabolic activity. Between the epiphysis and the metaphysis, one can see a boundary that is the remnant of the cartilage growth plate, all turned to bone in this adult specimen, indicated by the arrow. The arrowheads indicate the area shown at higher magnification in the bottom panel. The bottom panel shows in greater detail the meshwork of metaphyseal trabeculae. Note that as more trabeculae join the cortical bone shaft, thereby transferring greater mechanical load from the joint, there is a corresponding thickening of the cortex (C) to bear the extra weight.

Bone also contains pools of stem cells, one type of which is the bone-lining cell that covers all bone surfaces and responds to osteogenic signals by proliferating and maturing into matrix-secreting osteoblasts. In addition to the obvious protection bones provide

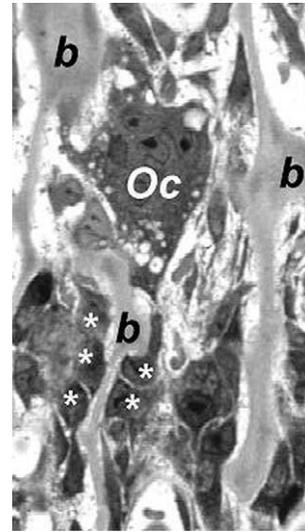
for the brain, spinal column, and thoracic organs, they enclose the soft and vulnerable hematopoietic marrow, providing homing signals for colonization by hematopoietic and stromal stem cells. Also, bone plays host to arteries, veins, and venous sinuses.

Bones also interface seamlessly with the tendons, ligaments, and cartilage of the joints to transduce muscular contractions into movement. Bones of the skull must join directly with one another, forming highly interdigitated sutures that in effect constitute a continuous, planar surface. Finally, the unique relationship between the jaws and the teeth results in a special form of joint, the tooth socket, with the jaw and tooth interlocked through the cementum and the periodontal ligament. The structure of bone, both macro- and microscopically, reflects all these functions and interactions.

### MACROSCOPIC STRUCTURAL FEATURES

There are three major subdivisions of the skeletal elements. The axial skeleton consists mainly of the vertebral column and ribs. The appendicular skeleton comprises the limbs and digits and their associated structures. The cranial skeleton consists of the skull, face, and jaws. Bones of the former two skeletal subdivisions are formed by a developmental process called “endochondral ossification,” in which a cartilage model, or anlage, is replaced by bone during development and growth, in a sense recapitulating the evolutionary path from a nonmineralized to a mineralized endoskeleton that characterizes all the vertebrates. In contrast, the flat bones of the skull, the mandible, and the clavicles are formed by a different developmental process in which bone-forming cells differentiate directly from the surrounding mesenchyme, a process called “intramembranous ossification.”

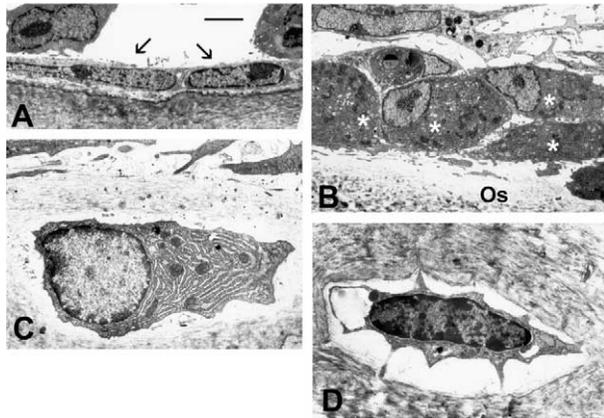
The overall structure of a typical long bone is shown in Fig. 2, a longitudinal section of an adult human tibia. The shaft of the bone, called the diaphysis, consists of a hollow cylinder of dense, or cortical, bone (C) surrounding a marrow space (Ma). The external surface of bone is covered by the periosteum and the internal surface (i.e., facing the marrow space) is covered by the endosteum. These are fibro-cellular layers not shown on the dried bone specimen in Fig. 2. At the end of the bone, one sees the widened region called the epiphysis (E), which supports the articular cartilage that forms the cushioned interface of the joint. The region where the bone narrows, or flares, from the epiphysis to the diaphysis is called the



**Figure 3** Dynamic balance of bone formation and resorption (light micrograph of 2-week-old rat tibia). This microscopic section shows a large, multinucleated osteoclast (Oc) actively resorbing trabecular bone spicules (b). Immediately adjacent to this resorptive activity, there are many plump, active osteoblasts secreting bone matrix, several of which are marked with asterisks. Although the overall activity level in this young, growing animal is extremely high, this simultaneous resorption and formation of bone continue throughout life.

metaphysis (Me). The metaphysis is a key structural and metabolic region that plays a crucial role in endochondral ossification and in mineral homeostasis. The metaphysis is occupied by cancellous, or spongy, bone, organized in a highly anastomosing, trabecular network of bone spicules that transmits forces between the joint and the shaft of the bone. This region of the bone is called the primary spongiosa. As each delicate trabecular arch joins the bone shaft, the cortical bone thickens to bear the additional load being transferred (see inset in Fig. 2). The crossing arches of bone trabeculae are reminiscent of human architectural solutions to the problem of transmitting the weight of a heavy roof to a building’s walls while leaving open interior spaces.

Of equal importance, the metaphysis provides signals and cell types required for the replacement of cartilage by bone during endochondral ossification in growing bones. When bones reach their adult size in humans and other large mammals, the layer of growth cartilage, called the physis or the epiphyseal growth plate, ceases its growth activity and becomes fully mineralized. The ossified remnant of the growth plate is visible as a boundary between the epiphysis and metaphysis in Fig. 2 (arrow).

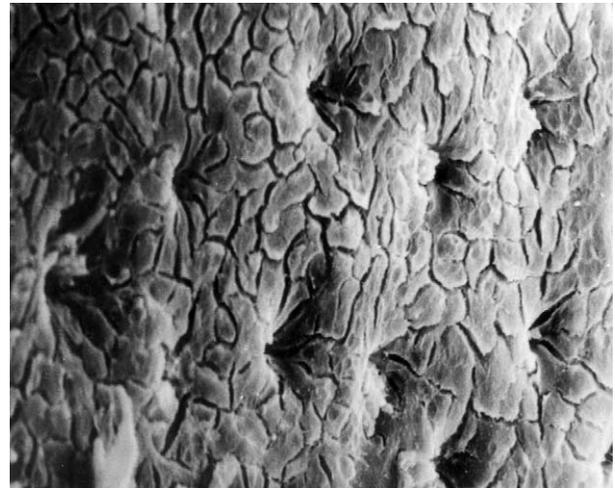


**Figure 4** Life cycle of bone-forming cells (transmission electron micrographs of rat tibia). (A) Quiescent bone-lining cells (arrows) cover the surface of bone. At the top is the marrow space with marrow cells at the right and left edges. Bone-lining cells can differentiate into active osteoblasts when stimulated. (B) Osteoblasts (asterisks) actively secreting bone matrix. They are roughly cuboidal and contain well-developed organelles needed to synthesize and secrete bone matrix. They lie upon the layer of osteoid (Os) they have just secreted. Part of one osteoblast at the right-hand edge of the image can be seen in the process of being entrapped within the bone matrix, thus becoming an osteocyte. (C) An osteogenic osteocyte is newly entrapped in bone and is still actively synthesizing matrix components: note the well-developed rough endoplasmic reticulum that fills the cytoplasm. Young osteocytes completely fill the lacuna that they inhabit. The lower margins of osteoblasts on the bone surface can be seen at the top of the image. (D) An osteolytic osteocyte. Late in their life cycle, osteocytes contract and also begin to degrade the surrounding bone matrix, creating unoccupied space within the lacuna. Some of the cytoplasmic projections that permeate the bone matrix can be seen around the cell's margin. Bar equals 3.2  $\mu\text{m}$  in A, 4.1  $\mu\text{m}$  in B, 1.5  $\mu\text{m}$  in C, and 1.7  $\mu\text{m}$  in D.

The high-magnification inset in Fig. 2 shows in detail the enormous surface area achieved by the metaphyseal bone spicules. As bone is the body's chief reservoir of calcium and phosphorus, the metaphysis is a site of constant metabolic activity. On microscopic examination, it is common to find both bone formation (by osteoblasts) and bone resorption (by osteoclasts) occurring side by side on the same trabecular surface (Fig. 3). This achieves an active dynamic balance that is poised to respond to environmental demands, maintaining mineral homeostasis over the long term without compromising mechanical needs.

## BONE CELLS

There are four principal bone cell types. Bone-lining cells are a pool of quiescent, fibroblastoid cells that cover all bone surfaces (Figs. 4A and 5). When called upon, they differentiate into osteoblasts (Fig. 4B) that secrete the proteinaceous bone matrix called osteoid

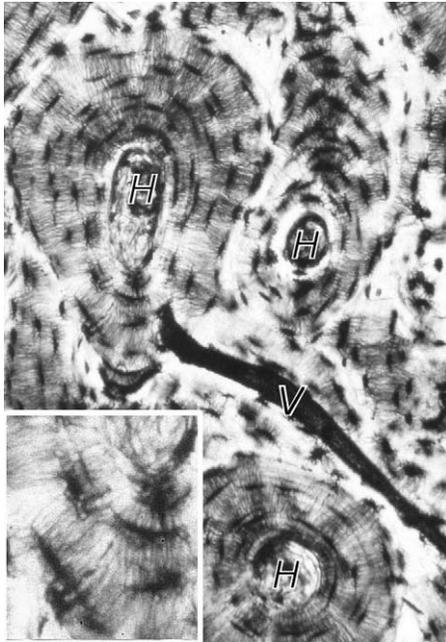


**Figure 5** Bone is completely clad in living cells (scanning electron micrograph of the endosteal surface of a rat tibia). For this image, the marrow was removed from the tibial marrow cavity, leaving the endosteum, i.e., a layer of potentially osteogenic bone-lining cells, intact. Note that 100% of the bone surface is covered by cells. The pits in the surface are points of entry for the blood vessels via the Volkmann's canals of the Haversian system. Individual cells are scale-like in appearance and are clearly bounded by dark margins.

and subsequently matrix vesicles that deliver the mineral component and initiate the precipitation of hydroxyapatite crystals in the matrix. Note that unlike the nomenclature used for some other cell systems, osteoblasts are fully differentiated and functioning cells. As the layer of bone matrix deposited by osteoblasts thickens, a subset of the osteoblasts becomes fully surrounded by the matrix. Once encased in bone, the cell is called an osteocyte (Figs. 4C and 4D). Osteocytes inhabit a lacuna within the matrix and maintain contact with other cells and with the bone surface via cytoplasmic processes throughout the bone matrix (Figs. 6 and 7). The fourth cell type is the osteoclast (Fig. 8), which differentiates via fusion of mononuclear cells of the monocyte/macrophage lineage. This process occurs on the bone surface and results in very large, multinucleated cells. Osteoclasts form a tight seal against the bone surface, acidify the subcellular compartment to dissolve the mineral, and secrete degradative enzymes. The solubilized bone is phagocytosed, broken down further, transported through the cell, and released from the free surface.

## INTRAMEMBRANOUS, OR "MEMBRANE," BONES

The intramembranous bones are defined by their non-endochondral developmental origin, forming directly

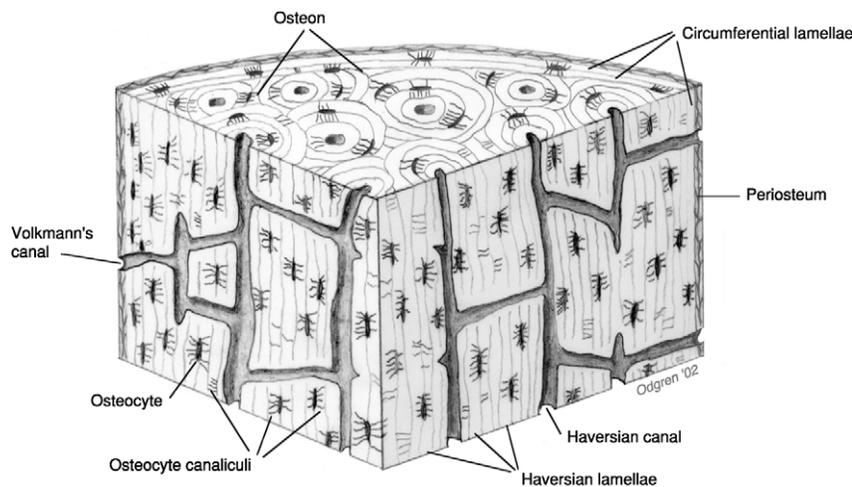


**Figure 6** Haversian systems in human cortical bone (transverse section of human cortical bone, polished and stained with India ink). Three Haversian canals (H) and their associated structures are visible in the micrograph at left (H). A transverse Volkmann's canal (V) is also evident. The concentric ring-shaped patterns are produced by the osteocytes surrounding the canals. Each Haversian canal, along with its accompanying concentric rings (Haversian system), is called an osteon. (Inset) The dense meshwork of canaliculi that interconnect the osteocytes to permit the transport of nutrients and signaling molecules from the blood vessels that occupy the canals throughout the entire bone volume.

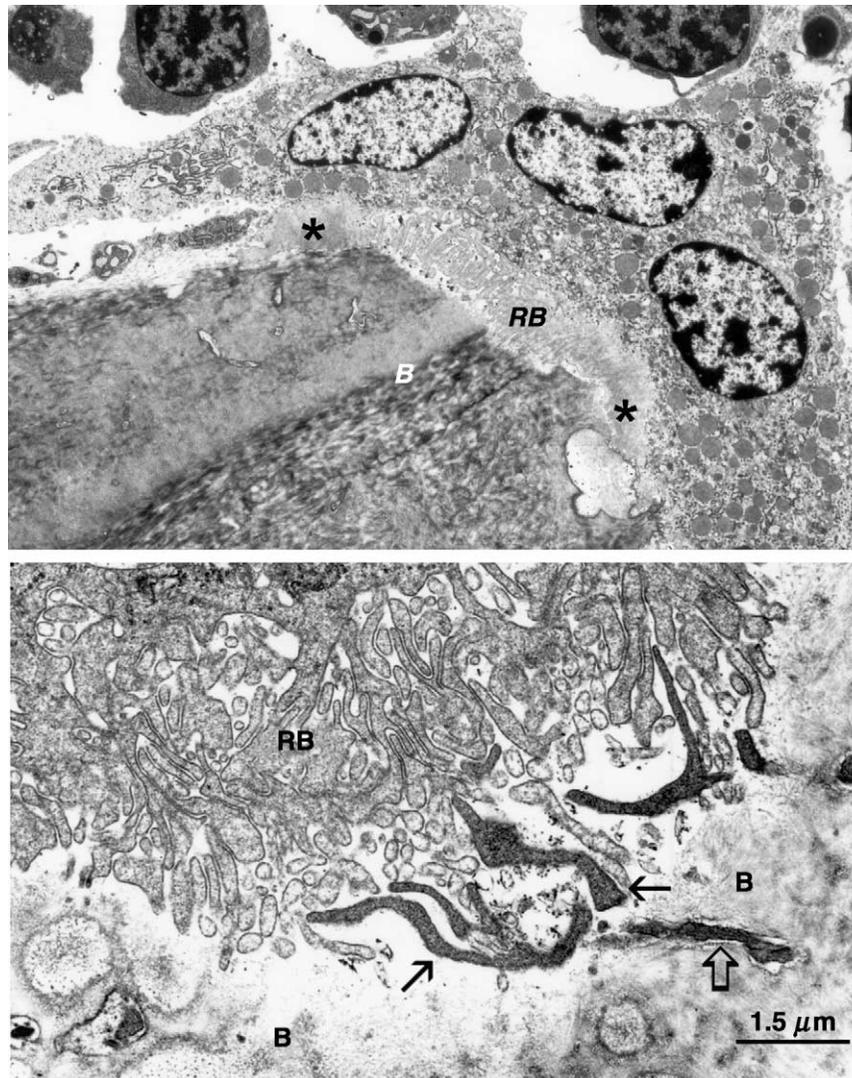
from condensing mesenchyme without an intermediary cartilaginous stage. They also participate in two unique types of joints not found in the appendicular or articular skeleton. At their edges, the flat bones of the skull and face form nonmotile joints called sutures (Fig. 9), in which they interdigitate with adjoining bones to form a tightly interlocked, continuous layer of bone that protects the brain and forms the facial structures. The mandible and maxilla contribute the bony half of the tooth socket, which is composed of a unique type of bone called alveolar bone (Fig. 10). Interactions between tooth and alveolar bone are required to develop and maintain the tooth sockets of the jaws. Without the tooth, all the alveolar bone is resorbed down to the end of the root space.

## BONE MATRIX

The extracellular matrix of bone consists of crystalline precipitates of mineral (the hydroxyapatite form of calcium phosphate) that permeate a dense protein substrate. Type I collagen fibrils constitute 90–95% of the proteinaceous material in bone and provide very high tensile strength, which, gram for gram, is greater than that of steel. Also present are dozens of other glycoproteins, proteoglycans, integrin-binding proteins, other low-abundance collagens, and growth factors. Bone is first laid down as immature, “woven” bone, in which collagen fibrils are oriented randomly (Fig. 11). This bone is later resorbed by osteoclasts



**Figure 7** Illustration showing Haversian bone (also called compact, or cortical, bone). Haversian and Volkmann's canals penetrate the bone, providing contact with the osteocytes that inhabit the bone. Cytoplasmic processes of the osteocytes occupy the canaliculi and permit responses to environmental stimuli, such as mechanical loading and circulating levels of hormones, minerals, and other factors. At the inner (not shown) and outer surfaces, bone forms circumferential lamellae, in which the cells are sufficiently near the surface so as not to require Haversian organization to retain metabolic contact.

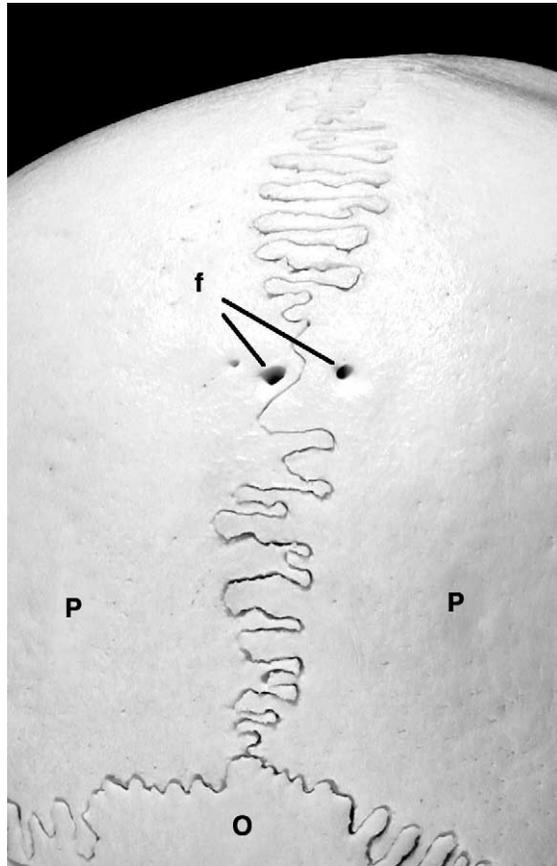


**Figure 8** The osteoclast (transmission electron micrographs of rat tibia). (Top) Three nuclei are visible in this bone-resorbing osteoclast. It is firmly attached to the bone (B) by the so-called clear zone (asterisks), a dense ring of actin filaments that seals off the compartment under the cell where bone resorption occurs. The ruffled border (RB) is a specialized region of highly invaginated plasma membrane where active secretion and resorption take place. Proton pumps in the ruffled border membrane acidify the extracellular compartment to dissolve the mineral, and proteolytic enzymes degrade the proteinaceous matrix. The degraded bone is engulfed and transported through the osteoclast and released from another specialized domain on the free surface. Several marrow cells are visible along the top of the image. (Bottom) Higher magnification of a ruffled border reveals details of the elaborate ruffling that occurs over the area of bone (B) being resorbed. Several cytoplasmic processes of osteocytes can be seen in the process of being broken down (arrows), with one (open arrow) still inside its surrounding canaliculus.

and replaced by osteoblasts as mature, lamellar bone, in which the collagen fibrils are ordered. Lamellar bone contains highly oriented type I collagen fibrils that are layered in alternating directions, much like the grain in layers constituting plywood, supplying maximal strength with minimal mass (Fig. 12).

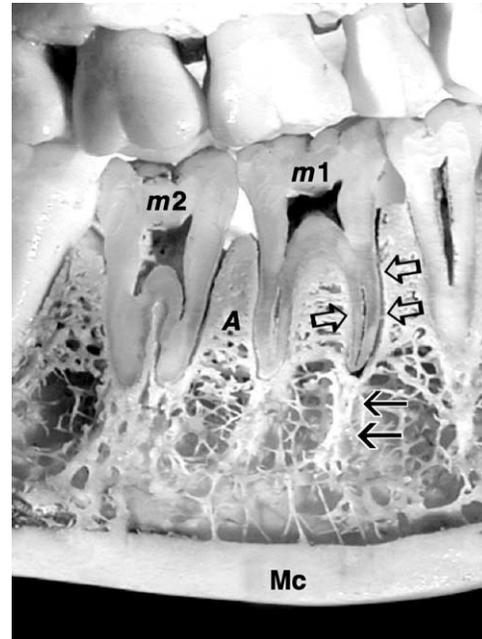
The posttranslational processing of the collagen needed to assemble this ordered extracellular matrix is complex, beginning within the osteoblast and

continuing after secretion. Collagen molecules are assembled in the endoplasmic reticulum of osteoblasts into trimers of highly posttranslationally modified monomers, designated  $\alpha$  chains. Type I collagen consists of two  $\alpha 1(I)$  chains and one  $\alpha 2(I)$  chain, each 1050 amino acids long and encoded by different genes, with a middle region rich in glycine, lysine, and proline. At both the amino and carboxy termini, there are registration peptides that are not a part of



**Figure 9** Sutures of the skull, as seen in a human skull with all cellular material removed. Where they interface with one another, the flat bones of the skull and face form immobile joints called sutures. Lacking a cartilage growth stage, these intramembranous bones can grow only along their edges and on their surfaces. The deep interdigitations provide a rigid, flat joint with high resistance to compressive injury. The figure shows the three-way suture formed at the rear of the skull by the occipital (O) and parietal bones (P), with the cranial ridge at the top. Two foramina, passages for nerves and blood vessels, are also shown (f).

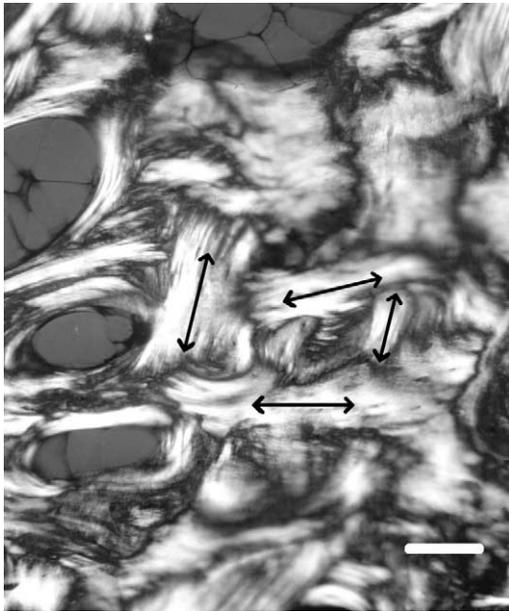
the triple-helix, but are required to align the monomers in proper registration. Within the endoplasmic reticulum, signal peptides are cleaved off and hydroxyl groups are added to the side chains of proline and lysine residues in hydroxyltransferase-mediated reactions that are vitamin C-dependent. The failure of this pathway in the absence of vitamin C is a major contributor to the pathogenesis of scurvy. Disulfide bonds form between the carboxy-terminal registration peptides of adjacent chains, aligning the monomers. Thus modified and registered, three monomers readily form right-handed, triple-helical bundles 300 nm long and 1.5 nm in diameter wound in a C-to-N direction. O-linked sugars are added to some hydroxylysines and N-linked sugars are added to



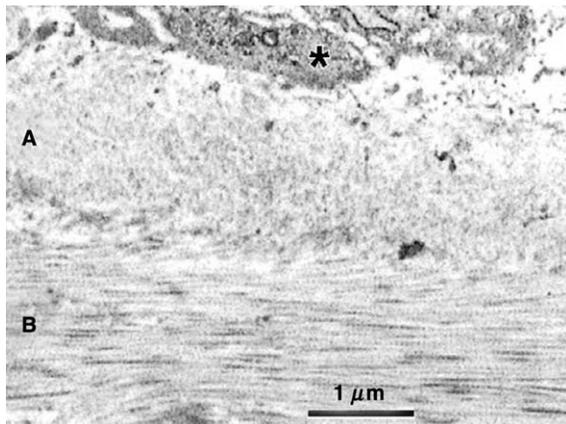
**Figure 10** Bone/tooth interface in the jaw (cutaway section of human mandible with all cellular material removed). The dense bone of the mandibular cortex (Mc) gives way to the spongy alveolar bone of the interior, in which trabeculae are present in proportion to the mechanical support required by the teeth. First and second molars are labeled m1 and m2. Note the very close apposition (open arrows) of dense alveolar bone (A) to the tooth roots. In the absence of teeth, the entire supporting structure of alveolar bone is resorbed by osteoclasts down to the point where the roots end. Note that the trabeculae also form a bony passage that protects the blood vessels and nerve that supply the tooth through each root (arrows).

the terminal regions. The assembled trimers, or “procollagen” molecules, move to the Golgi apparatus, where they are packaged into secretory vesicles and secreted.

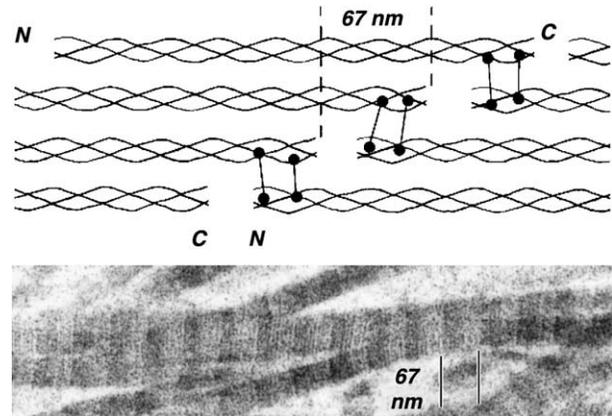
Once outside the cell, the terminal peptides are cleaved off and the resulting triple-helical molecules are called “tropocollagen.” Tropocollagen then polymerizes in aligned bundles, with covalent bonds forming between the C and N termini of adjacent trimers to assemble highly ordered collagen fibrils. The periodicity of the overlapping collagen molecules within the fibrils is seen at very high magnification as a banding pattern 67 nm in length (Fig. 13). The alternating orientation of the fibrils in successive layers of bone collagen fibrils that is crucial to the mechanical resilience and strength of bone can be seen in electron micrographs (Fig. 12). The surface of bone is generally covered by a “capsule” of collagen fibers, called the circumferential lamellae, arranged parallel to the surface. The exception to this occurs at the points of attachment of tendons and ligaments, whose collagen



**Figure 11** Woven bone in polarized light (dog mandible). Collagen fibril orientation is visible in light micrographs with polarized illumination. Double-headed arrows indicate the orientations of several domains. This “first-generation” bone matrix is brittle and weak and needs to be resorbed by osteoclasts and replaced by osteoblasts as mature, lamellar bone to attain full strength and resilience. Bar, 75  $\mu\text{m}$ .



**Figure 12** Collagen fibrils in lamellar bone (transmission electron micrograph of rat tibia). Mature lamellar bone is so named because of the layered, highly oriented collagen fibrils that constitute it, as seen in this electron micrograph of newly synthesized bone matrix. A portion of an osteoblast can be seen at the top (asterisk). In the upper layer of matrix (A), the cartilage fibrils are seen end-on, i.e., projecting out of the plane of the page. In the lower layer (B), they are cut longitudinally and run along the surface of the page. As a result of this alternating orientation of the collagen fibrils, lamellar bone has much greater strength and resilience than woven bone.



**Figure 13** Type I collagen fibrils. The ends of adjacent tropocollagen trimers are staggered by 67 nm, as diagrammed in the top illustration. The N-terminal region of one trimer forms covalent bonds (represented as ball-and-stick figures) with the C-terminal region of adjoining trimers. The spacing between the overlapped ends is seen as a banding pattern in high-magnification electron micrographs, as in the bottom panel. The cross-linked trimers form collagen fibrils that constitute 90–95% of the proteinaceous component of bone matrix.

fibers penetrate the bone surface at an angle and become continuous with the bone collagen fibers. These are called Sharpey's fibers.

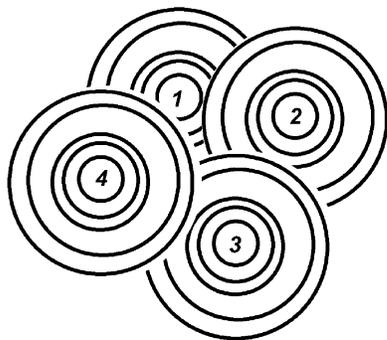
In addition to its high collagen content, the proteinaceous bone matrix secreted by osteoblasts (called “osteoid”) contains many noncollagenous proteins, approximately 5–10% of protein weight. Osteoid provides the microenvironment where mineral precipitates in a controlled way and it is thought to mediate a wide range of other biological processes. Although no definitive *in vivo* functions for the dozens of noncollagenous proteins are known, there are among them many growth factors, cytokines, and integrin-binding proteins as well as a host of other bone-specific proteins. With such an array of cell regulatory molecules thus embedded within the bone matrix, it is clear that bone provides an information-rich environment that helps direct and integrate diverse cellular activities, such as cell attachment, bone synthesis, collagen assembly, and bone resorption, as well as responses to environmental cues, such as mechanical loading or dietary status. Presumably some of these noncollagenous bone matrix proteins are also essential for the remarkable ability of osteocytes to survive while inhabiting lacunae deep within the matrix, interconnected via cytoplasmic processes that inhabit the profuse networks of canaliculi, and for the formation and maintenance of the capillary network that thoroughly permeates the entire Haversian system. Identifying specific biological functions for

the many noncollagenous bone matrix proteins is an important area of ongoing research.

Bone mineral is composed mainly of calcium and phosphate in the form of hydroxyapatite,  $\text{Ca}_4(\text{PO}_4)_6(\text{OH})_3$ , which is secreted in matrix vesicles by osteoblasts. Thus, even though mineral deposition takes place outside the cells, it is a cell-regulated process. Matrix vesicles are seen near osteoblasts that supply mineral to the matrix and are approximately 100 nm in diameter. They contain alkaline phosphatase and other enzymes needed to produce high concentrations of calcium and phosphorus. The vesicles rupture, releasing their contents and permitting the precipitation of hydroxyapatite crystals that eventually permeate the osteoid completely to form the composite material called bone. The mineral adds great rigidity and resistance to compressive and bending forces to the skeletal elements (Fig. 1).

### THE HAVERSIAN SYSTEM: METABOLIC LIFELINE

In large mammals such as humans, compact bone, such as that found in the shafts of the limb bones, is generally too thick to permit adequate nutrient, signal, and waste exchange with the embedded osteocytes. The evolutionary solution to this problem is the Haversian system, in which concentric cylinders of bone surround a nutrient artery and vein located within a central Haversian canal oriented parallel to the long axis of the bone (Figs. 6 and 7). These concentric rings are arranged much like growth rings in trees. One Haversian canal plus its accompanying concentric lamellae is called an “osteon.” Adjacent Haversian systems are interconnected at frequent intervals by blood vessels that travel in transverse



**Figure 14** Remodeling of osteons over time. This diagram depicts the chronology of remodeling of a region of bone over time. The osteons are numbered in order of their formation. As newer osteons form, they invade and erode the preexisting osteons.

passages called Volkmann’s canals. Together, the Haversian and Volkmann’s canals penetrate the entire bone substance and maintain communication among all the osteocytes embedded within the cortex. Note that small mammals, such as rats and mice, do not possess Haversian systems, because their bones are small enough to permit survival of osteocytes without intra-osseous circulatory elements.

Osteons, like other bony structures, are subject to remodeling. Successive waves of resorption by osteoclasts and formation by osteoblasts leave telltale structural evidence. The osteons most recently formed are generally the least densely mineralized and, more noticeably, have the most complete sets of rings. As new osteons form, they invade the territory of older osteons and erode their outer edges. This phenomenon is shown diagrammatically in Fig. 14. This remodeling is essential to maintain the bones and to repair microdamage that accumulates over time. Metabolic studies have revealed that in the adult human, approximately 10% of the skeleton is remodeled per year. This indicates that a complete skeletal replacement occurs approximately once a decade.

### Acknowledgments

This work was supported in part by Grants DE13961 (P. R. O.) and DE07444 (S. C. M.) from the National Institute of Dental and Craniofacial Research (NIDCR). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIDCR or the National Institutes of Health.

### See Also the Following Articles

Body Proportions • Bone Mass Measurement • Bone Turnover Markers • Collagen Metabolism • Collagen Metabolism Disorders • Osteoporosis, Overview • Skeletal Development

### Further Reading

- Kessel, R. G., and Kardon, R. H. (1979). “Tissues and Organs: A Text Atlas of Scanning Electron Microscopy.” W. H. Freeman, San Francisco, CA.
- Marks, S. C., Jr. (1997). The structural basis for bone cell biology. *Acta Med. Dent. Helv.* **107**, 141–157.
- Marks, S. C., Jr., and Odgren, P. R. (2002). The structure and development of the skeleton. In “Principles of Bone Biology” (J. P. Bilezikian, L. G. Raisz, and G. A. Rodan, eds.), 2nd ed. pp. 3–15. Academic Press, San Diego, CA.
- Robey, P. G. (2002). Bone matrix proteoglycans and glycoproteins. In “Principles of Bone Biology” (J. P. Bilezikian, L. G. Raisz, and G. A. Rodan, eds.), 2nd ed. pp. 225–237. Academic Press, San Diego, CA.
- Ross, M. H., Romrell, L. J., and Kaye, G. I. (1995). “Histology: A Text and Atlas,” 3rd ed. Williams & Wilkins, Baltimore, MD.



# Bone Turnover Markers

Patrick Garnero

*INSERM Research Unit 403, Hôpital E. Herriot, and SYNARC, Lyon, France*

Pierre D. Delmas

*INSERM Research Unit 403, Hôpital E. Herriot, Lyon, France*

## Glossary

**bone formation marker** Enzyme or secretory molecule of the osteoblastic cells released during the process of bone synthesis.

**bone resorption marker** Enzyme of the osteoclastic cells or component of the bone matrix released during the process of bone degradation.

**osteoporosis** A systemic skeletal disease characterized by a low bone mass and microarchitectural deterioration of bone tissue, with a consequent increase in bone fragility and susceptibility to fracture risk.

**Paget's disease** A localized disorder characterized by a marked increase in bone turnover, leading to overproduction of bone of poor quality, responsible for hypertrophy, osteosclerosis, and bone fragility.

**B**one turnover is characterized by two opposite activities, the formation of new bone by osteoblasts and the degradation (resorption) of old bone by osteoclasts. Both are tightly coupled in time and space in a sequence of events that define the same remodeling unit.

## INTRODUCTION

Bone mass depends on the balance between resorption and formation within a remodeling unit and on the number of remodeling units that are activated within a given period of time in a defined area of bone. Bone loss characterizing metabolic bone diseases including osteoporosis depends on both an imbalance between resorption and formation within the remodeling unit and an increased activation frequency (i.e., bone turnover) that can be assessed noninvasively

by systemic biochemical markers of bone formation and bone resorption.

The rate of formation or degradation of bone matrix can be assessed either by measuring the enzymatic activity of the bone-forming or -resorbing cells, such as alkaline and acid phosphatase, or by measuring bone matrix components released into the circulation during formation or resorption (Table I). These have been classified as markers of formation and resorption, but it should be kept in mind that in disease states in which both events are coupled and change in the same way, such as in osteoporosis, any marker will reflect the overall rate of bone turnover. Current bone markers cannot discriminate between turnover changes in a specific skeletal envelope (i.e., trabecular vs cortical) but reflect whole body net changes. Circulating levels of these markers can be influenced by factors other than bone turnover, such as the metabolic clearance (liver uptake, renal excretion, and/or trapping on bone hydroxyapatite), and by the pattern of immunoreactive moieties recognized by a given antibody. Thus, each marker needs to be validated carefully in a specific clinical situation before conclusions about its clinical utility can be drawn.

Although bone markers may be useful in the management of a variety of metabolic bone diseases, most studies have focused on their potential use in postmenopausal osteoporosis. Bone turnover markers have been suggested to predict the occurrence of osteoporotic fractures in combination with other risk factors, including bone mineral density (BMD), and to monitor the efficacy of treatment, especially anti-resorptive therapies (hormone replacement therapy, bisphosphonates, and calcitonin) and also, recently, anabolic treatment such as parathyroid hormone. In this article, we review bone marker technology and discuss the use of these markers for the management of osteoporosis and Paget's disease.

**Table 1** Biochemical Markers for Bone Remodeling<sup>a</sup>

Formation
Serum
<b>Osteocalcin</b>
Total and <b>bone alkaline phosphatase</b>
Procollagen type I C propeptide and <b>procollagen type I N propeptide</b>
Resorption
Plasma/serum
Tartrate-resistant acid phosphatase
free pyridinoline and deoxypyridinoline
C-terminal cross-linking telopeptide of type I collagen generated by MMPs
<b>N-terminal and C-terminal cross-linking telopeptide of type I collagen</b>
Urine
<b>Free pyridinoline and deoxypyridinoline</b>
<b>N-terminal and C-terminal cross-linking telopeptide of type I collagen</b>
Calcium
Hydroxyproline
Galactosylhydroxylysine
Type I collagen helicoidal peptide 620–633

<sup>a</sup>The markers with the best performance characteristics in osteoporosis are in bold type.

## BIOCHEMICAL MARKERS OF BONE FORMATION

### Serum Alkaline Phosphatase

The skeletal alkaline phosphatase (ALP) is an enzyme localized in the membrane of the osteoblasts that is released into the circulation by an unknown mechanism. Among the several tissues containing ALP, the liver and bone isoenzymes are the major contributors. Until recently, serum total ALP activity was the most commonly used marker of bone formation, but for the previously mentioned reasons it lacks sensitivity and specificity. In an attempt to improve the specificity and sensitivity of serum ALP measurement, techniques have been developed to differentiate the bone and the liver isoenzymes, which differ only by post-translational modifications because they are coded by a single gene. In general, these assays have slightly enhanced the sensitivity of this marker, but most are indirect and/or technically cumbersome. A real improvement was achieved by using monoclonal antibodies that preferentially recognize the bone isoenzyme. These direct immunoassays have been shown to exhibit a low cross-reactivity with the circulating liver isoenzyme (15–20%) and to be more sensitive than total ALP activity to detect the increase in bone

turnover following menopause. Also, an automated immunoassay for serum bone ALP with improved precision compared to manual immunoassay and high throughput has been developed.

### Serum Osteocalcin

Osteocalcin (OC) is a small, noncollagenous protein that is specific for bone tissue and dentin, but its precise function is unknown. OC is predominantly synthesized by osteoblasts and is incorporated into the extracellular bone matrix, but a fraction of newly synthesized osteocalcin is released into the circulation, where it can be measured by radioimmunoassay. Circulating OC has a short half-life and is rapidly cleared by the kidney. Serum OC concentrations correlate with skeletal growth at the time of puberty and are increased in a variety of conditions characterized by increased bone turnover, such as primary and secondary hyperparathyroidism, hyperthyroidism, Paget's disease, and acromegaly. Conversely, they are decreased in hypothyroidism, hypoparathyroidism, glucocorticoid-treated patients, and in some patients with multiple myeloma and malignant hypercalcemia. Comparisons of serum OC levels with iliac crest histomorphometry and calcium kinetic data have shown that under most conditions, serum OC is a valid marker of bone turnover when resorption and formation are coupled and is a specific marker of bone formation when formation and resorption are uncoupled.

Circulating OC consists of different molecular entities. We found that the intact molecule represents approximately one-third of the immunoreactivity in the adult serum (or plasma), one-third is represented by several small fragments, and another third by a large N-terminal, midmolecule fragment (N-mid fragment). Some *in vitro* studies using purified rat osteoclasts cultured on bovine bone slices suggest that both intact and fragmented OC may be released from bone matrix during bone resorption. Thus, some specific urinary OC fragments may actually reflect bone resorption. However, the relative contribution of osteoclast-versus osteoblast-generated OC to the total circulating level is unknown, and we showed that short-term treatment with a potent bisphosphonate did not decrease either intact or the major circulating N-mid fragment despite a dramatic decrease in bone resorption, suggesting that *in vivo*, serum OC mainly reflects bone formation. After a few hours of incubation of serum samples at room temperature, a significant amount of plasma intact OC is rapidly converted

into the large N-mid fragment, resulting in a significant loss of immunoreactivity with intact OC assay and with most polyclonal antibodies because they recognize the C-terminal end of the molecule. From a practical standpoint, it is recommended to use assays that measure both the intact molecule and the N-mid fragment, providing a more robust and sensitive assay.

### Procollagen Type I Propeptides

During the extracellular processing of type I collagen, there is cleavage of the amino-terminal (PINP) and carboxy-terminal (PICP) extension peptides prior to fibril formation. These peptides circulate in blood, in which they may represent useful markers of bone formation since collagen is by far the most abundant organic component of bone matrix.

In contrast to serum PICP, which is a single glomerular protein the same size as the authentic propeptide, immunoreactive PINP circulates in different forms, including the intact PINP corresponding to the authentic trimeric *in vivo* cleaved propeptide, intact monomeric  $\alpha$  chains, and smaller fragments. The trimeric structure is unstable at 37°C and can be transformed into the stable monomeric forms. The first developed assays for serum PINP, which use immunogen synthetic peptides from the  $\alpha_1$  chain of type I collagen propeptide, recognize both trimeric and monomeric forms and PINP fragments but, as has serum PICP, have been disappointing with regard to osteoporosis. In contrast, assays recognizing the intact trimeric form of PINP have been shown to be more sensitive than PICP and to be as valuable as OC and bone ALP for detecting the increase in bone turnover following menopause and for monitoring the response to antiresorptive therapy. Serum PINP appears to be particularly sensitive for assessing the anabolic effects of PTH in postmenopausal women with osteoporosis.

## BIOCHEMICAL MARKERS OF BONE RESORPTION

### Fasting Urinary Calcium, Hydroxyproline, and Hydroxylysine Glycosides

Fasting urinary calcium measured in a morning sample and corrected for creatinine excretion is the cheapest assay of bone resorption. It is useful in detecting a marked increase in bone resorption but lacks sensitivity. Fasting urinary calcium reflects the amount of calcium released during resorption and also

the renal handling of calcium that is influenced by calcium-regulating hormones and by estrogens. Hydroxyproline (Hyp) is found mainly in collagen and represents approximately 13% of the amino acid content of the molecule. Ninety percent of the Hyp liberated during the degradation of bone collagen is primarily metabolized in the liver. Approximately 10% of the Hyp released by the breakdown of collagen circulates in a peptide-bound form, and these peptides are filtered and excreted in urine without any further metabolism. Urinary Hyp is usually considered an index of bone resorption. However, it should be noted that significant amounts of urinary Hyp are derived from the degradation of newly synthesized collagens, Hyp can be found in other tissues such as the skin, and Hyp can also be liberated from the metabolism of elastin and C1q. Urinary Hyp is therefore considered an unspecific index of collagen turnover and, consequently, has been largely replaced by more specific markers.

Hydroxylysine is another amino acid unique to collagen and proteins containing collagen-like sequences. Like hydroxyproline, hydroxylysine is not reused for collagen biosynthesis, and although it is much less abundant than hydroxyproline, it is a potential marker of collagen degradation. Hydroxylysine is present in part as galactosyl hydroxylysine and in part as glucosyl-galactosyl hydroxylysine. The relative proportion and total content of galactosyl hydroxylysine and glucosyl-galactosyl hydroxylysine vary in bone and soft tissues, with a higher content of galactosyl hydroxylysine in bone, suggesting that its urinary excretion might be a more sensitive marker of bone resorption than urinary hydroxyproline.

### Plasma Tartrate-Resistant Acid Phosphatase

Acid phosphatase is a lysosomal enzyme present primarily in bone, prostate, platelets, erythrocytes, and spleen. Bone acid phosphatase is resistant to L-(+)-tartrate, whereas the prostatic isoenzyme is inhibited.

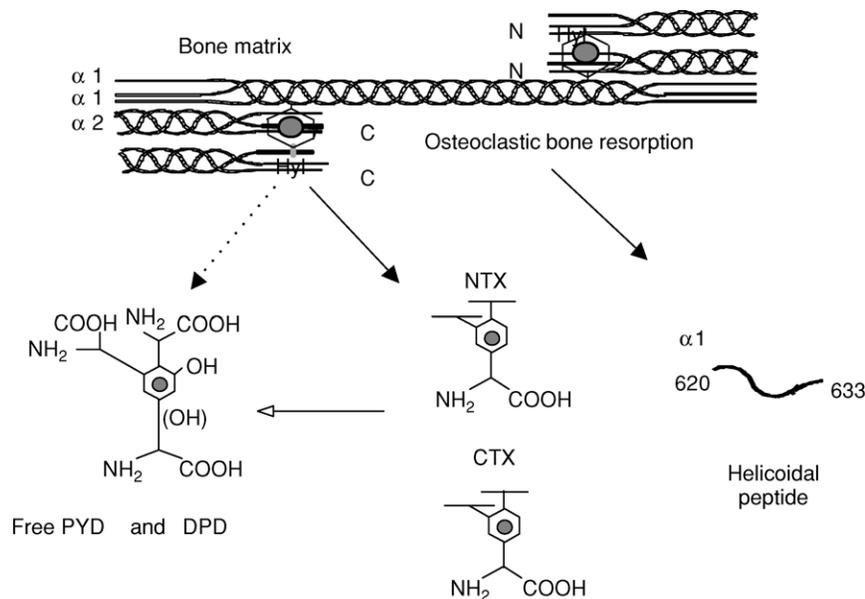
Acid phosphatase circulates in blood and shows higher activity in serum than in plasma because of the release of platelet phosphatase activity during the clotting process. In normal plasma, tartrate-resistant acid phosphatase (TRACP) corresponds to plasma isoenzyme 5, which originates partly from bone because osteoclasts contain TRACP that is released into the circulation. Isoenzyme 5 is represented by two subforms, 5a and 5b, with TRACP 5b being more specific for osteoclasts. Total plasma TRACP activity

can be measured by colorimetric assays. However, the lack of specificity of plasma TRACP activity for the osteoclast, its instability in frozen samples, and the presence of enzyme inhibitors in serum are potential drawbacks that limit the development of clinically useful enzymatic TRACP assays for osteoporosis. To overcome these limitations, antibodies have been developed against TRACP isolated from different sources, including recombinant human TRACP and purified bone TRAP, to develop immunoassays for TRACP 5b.

Plasma TRACP activity measured by conventional colorimetric assays is increased in a variety of metabolic bone disorders with increased bone turnover, including osteoporosis, but it is not clear whether this marker is more sensitive than urinary hydroxyproline. When measured using specific immunoassays, serum TRACP 5b isoenzyme has been shown to be more sensitive than total TRACP for detecting increased bone turnover after menopause and decreased turnover induced by hormone replacement therapy in postmenopausal women, suggesting that this marker may be a useful indicator of the effect of antiresorptive therapy on bone resorption.

### Collagen Pyridinium Cross-Links and Associated Type I Collagen Peptides

Pyridinoline (PYD) and deoxypyridinoline (DPD), also called hydroxylslypyridinoline and lysylpyridinoline, respectively, are nonreducible pyridinium cross-links present in the mature form of collagen. The posttranslational covalent cross-linking generated from lysine and hydroxylysine residues is unique to collagen and elastin molecules. It creates interchain bonds that stabilize the molecule within the extracellular matrix. The highest concentration of PYD (expressed in mol/mol of collagen) is found in articular cartilage, whereas DPD is present in minute amounts in this tissue. PYD and DPD are present in tendon and aorta but absent from the skin, an abundant source of type I collagen. PYD and DPD are released into the circulation after resorption of bone matrix and then excreted in urine, in which they are found as free and peptide bound forms (Fig. 1). Peptide-bound but not free cross-links can be generated directly at the bone resorption site. The total amount of pyridinoline cross-links can be measured by fluorimetry after reversed-phase high-performance liquid

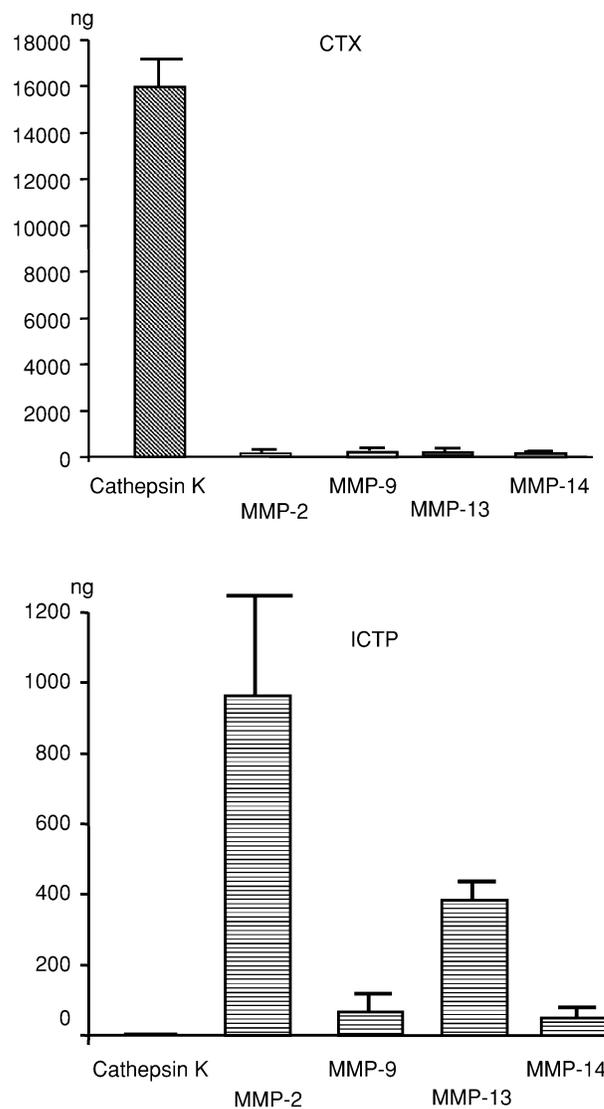


**Figure 1** Type I collagen breakdown products as markers of bone resorption. Type I collagen molecules consist of two  $\alpha_1$  chains and one  $\alpha_2$  chain associated in triple helix except at the ends (telopeptides). In the extracellular matrix, type I collagen molecules are cross-linked by pyridinoline (PYD) and deoxypyridinoline (DPD) in the region of N and C telopeptides. PYD differs from DPD by the presence of a hydroxyl residue shown in *italic*. During osteoclastic bone resorption, fragments originating from the telopeptides and the triple helix region (e.g., helicoidal peptide 620–633) are released into the circulation and excreted in the urine. PYD and DPD are released into the circulation mainly as peptide-bound cross-links [i.e., attached to fragments of C-terminal (CTX) or N-terminal (NTX) telopeptides]. Some of the peptide-bound cross-links are further degraded in the kidney in free cross-links. Immunoassays for detecting free PYD, free DPD, CTX and NTX, and helicoidal peptide 620–633 in serum or urine are available.

chromatography (HPLC) of a cellulose-bound extract of hydrolyzed urine.

Immunoassays for the PYD cross-links and related telopeptide fragments that advantageously substitute for the HPLC measurement of the total excretion are now available and represent the best indices of bone resorption. These comprise measurements of urinary free PYD (free PYD) and DPD (free DPD) and related peptides in urine and, recently, in serum. These peptides include the C-terminal cross-linking telopeptide of type I collagen generated by matrix metalloproteases (MMPs) (S-CTX MMP; also called ICTP), the C-terminal cross-linking telopeptide of type I collagen in serum (S-CTX) and in urine (U-CTX), and the N-terminal cross-linking telopeptide of type I collagen in serum (S-NTX) and in urine (U-NTX). Most of these type I collagen-related resorption markers can be measured either manually or by an automated system with improved precision and high throughput. In contrast to its poor sensitivity in osteoporosis, CTX MMP appears to be a useful bone resorption marker in patients with malignant bone disease. The different sensitivity of this marker in these two clinical situations may be related to differences in the pattern of type I collagen degradation. The epitope of CTX MMP is destroyed by cathepsin K activity, an osteoclastic-specific cysteine protease that is the key enzyme responsible for bone collagen degradation in normal physiological conditions, whereas it is generated by MMPs, whose activity has been suggested to play an important role in collagen degradation associated with cancer. In contrast, CTX and NTX epitopes are highly efficiently generated by cathepsin K and not by MMPS (Fig. 2). Available markers of type I collagen degradation are based on the measurement of either pyridinoline cross-links or associated cross-linked C or N telopeptides, both originating from the telopeptide region. Recently, a type I collagen-specific peptide corresponding to residues 620–633 of the helical region of the  $\alpha_1$  chain was isolated from urine of patients with Paget's disease. We found that the urinary excretion of this helical peptide increased markedly after menopause and was as sensitive as urinary CTX for assessing the antiresorptive effects of bisphosphonate and estrogens, suggesting that it may be useful for the investigation of patients with osteoporosis. However, additional studies are required to fully evaluate its clinical utility.

Urinary and serum markers of bone resorption have significant circadian rhythms. Urinary excretion of PYD and DPD peaks between 2:00 and 8:00 AM and reaches a nadir between 2:00 and 11:00 PM. In healthy premenopausal women, the magnitude of the



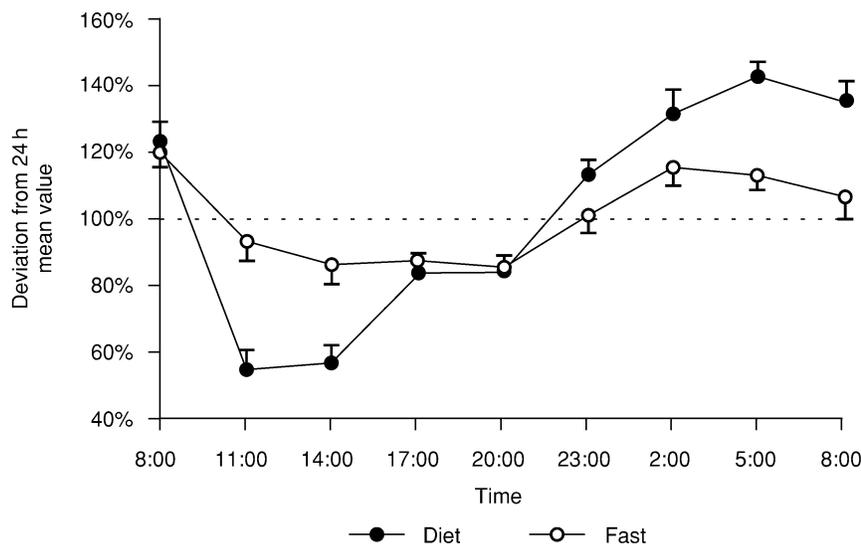
**Figure 2** The type I collagen fragments CTX and ICTP (also called CTX-MMP) are released from human bone collagen by distinct enzymatic pathways. Demineralized insoluble human bone collagen was incubated with active human recombinant cathepsin K, matrix metalloprotease (MMP)-2, MMP-9, MMP-13, and MMP-14 for 24 h at 37°C. CTX and ICTP fragments were measured in the supernatant. The bars represent the mean and SEM of nine (cathepsin K) or three (MMPs) different incubations per experimental condition. As shown in the figure, CTX is released by cathepsin K but not MMPs, whereas ICTP is released by MMPs but not cathepsin K. Reproduced from Garnero, P., Ferreras, M., Karsdal, M. A., NicAmhlaoibh, R., Risteli, J., Borel, O., Qvist, P., Delmas, P. D., Foged, N. T., and Delaissé, J. M. (2003). The type I collagen fragments ICTP and CTX reveal distinct enzymatic pathways of bone collagen degradation. *J. Bone Miner. Res.* **18**, 859–867. Reproduced with permission of the American Society for Bone and Mineral Research.

rhythm may be as much as 100% of the 24-h mean, with a decrease of 25–35% between 8:00 and 11:00 AM. Urinary cross-linked telopeptides have similar rhythms as total cross-link excretion. Serum CTX peaks between 1:30 and 4:30 AM at levels that are more than twice those at the nadir between 11:00 AM and 3:00 PM. The amplitude of the rhythm of CTX MMP is only 15–20% of the 24-h mean, suggesting that the different type I collagen peptides may have different bone specificity and/or reflect different aspects of bone resorption. Fasting results in a diminution in the amplitude of the circadian rhythm of urinary and, more markedly, serum CTX (Fig. 3). Among the factors that may influence the circadian pattern of serum CTX, food intake appears to be the most important determinant; its effect can be mediated by gastrointestinal hormones. In summary, although the amplitude of the circadian rhythm differs according to the markers, its effect is substantial for most of them and indicates the importance of standardizing the time of sampling and the fasting condition of the patient.

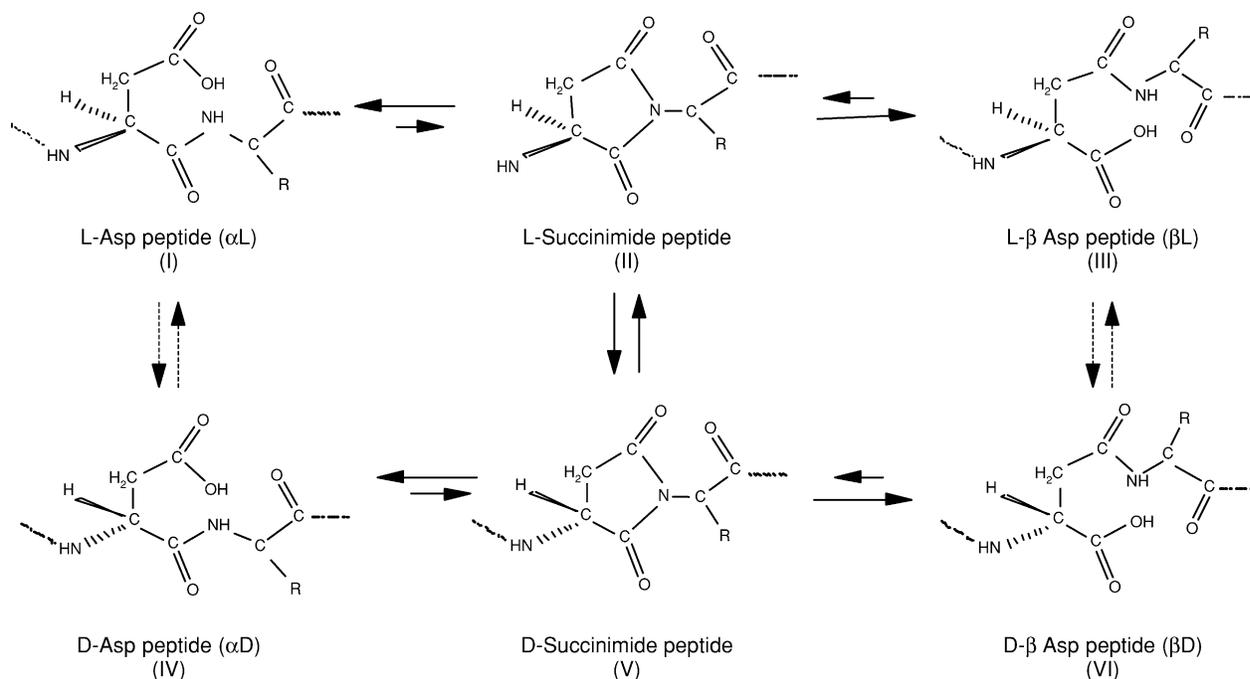
### Posttranslational Modifications of Collagen Molecules

Type I collagen molecules in bone tissue undergo enzymatic and nonenzymatic posttranslational modifications,

and some may be of clinical relevance for the investigation of metabolic bone diseases, including osteoporosis. Among the enzymatic transformations, biochemical studies performed on bone specimens have shown that the extent of hydroxylation of lysine residues and glycosylation of hydroxylysine, the content of nonreducible cross-links, and the PYD:DPD ratio may be associated with bone strength. Nonenzymatic modifications of collagen such as pentosidine, an index of nonenzymatic advanced glycation end products, may also play a role as a determinant of bone strength. Racemization and  $\beta$ -isomerization of the Asp–Gly sequence within the  $^{1209}\text{AHDGGR}^{1214}$  sequence (CTX) of the C telopeptide of type I collagen are other nonenzymatic posttranslational modifications (Fig. 4). Histological studies have shown a decreased degree of isomerization/racemization within the woven pagetic bone (characterized by increased fragility) that can be detected *in vivo* by the differential measurement of native ( $\alpha\text{L}$ ) CTX and isomerized ( $\beta\text{L}$ ) CTX in urine. In a prospective study of postmenopausal women, we found that an increased urinary ratio between native and age-related forms of type I collagen CTX was associated with increased fracture risk, independent of the level of BMD of the hip and the bone turnover rate (Table II). This suggests that a decreased degree of type I collagen isomerization/racemization may be associated with alterations of



**Figure 3** Circadian variation of serum CTX levels: effect of fasting status. Eleven premenopausal women were sampled every 3 h over two consecutive 24-h periods in a crossover design. In one period, women fasted throughout the study (○); in the other period, they were allowed to eat and drink (●). For fasting subjects, the average variation was 13.6%, whereas the variability under nonfasting conditions was 34%. Reproduced with permission from Qvist, P., Christgau, S., Pedersen, B. J., Schlemmer, A., and Christiansen, C. (2002). Circadian variation in the serum C-terminal telopeptide of type I collagen (serum CTx): Effects of gender, age, menopausal status, posture, daylight, serum cortisol and fasting. *Bone* 31, 57–61.



**Figure 4** Racemization and isomerization of type I collagen C-telopeptides. An attack by a peptide backbone nitrogen on the side chain carbonyl group of an adjacent aspartyl residue can result in the formation of a succinimide ring (I → II). The succinimide ring is prone to hydrolysis and racemization producing peptides and β-aspartyl peptides in both the D and L configurations. Racemization is thought to proceed primarily through the succinimide pathway (II → V), but other pathways, such as direct proton abstraction (I ↔ IV and III ↔ VI), may also contribute to the formation of D-aspartyl. The peptide backbone is shown as a bold line. The four types of C telopeptides are present in bone matrix; the native form (αL) and three age-related forms—an isomerized form (βL), a racemized form (αD), and an isomerized/racemized form (βD). With increasing age of type I collagen molecules, the proportion of β isomerized and D racemized forms within bone matrix increases. Degradation products of these four CTX forms of type I collagen can be measured in urine independently by immunoassays using specific conformational monoclonal antibodies.

bone strength properties, a hypothesis that requires confirmation by studies correlating the degree of type I collagen racemization/isomerization with mechanical properties of bone specimens. Clearly, these findings open new perspectives for the clinical use of bone markers, not only to measure quantitative changes of bone turnover but also to assess changes of bone quality.

## CLINICAL USES OF BONE TURNOVER MARKERS IN OSTEOPOROSIS

### Bone Markers and Fracture Risk Assessment

With the emergence of effective-but expensive treatments, it is essential to detect those women at higher risk of fracture. Several prospective studies have shown that a standard deviation decrease of BMD measured by dual X-ray absorptiometry (DXA) or

heel ultrasound is associated with a two- to fourfold increase in the relative fracture risk, including the hip, spine, and forearm. In this context, the question arises as to what extent bone markers could add to bone mass measurement to improve the assessment of fracture risk.

Prospective studies relating levels of bone formation markers to the risk of fracture have yielded conflicting results. Indeed, a decrease, no difference, or an increase in bone formation markers have been reported to be associated with increased fracture risk. The different results among studies may be related to the type of fracture, the population studied, and the duration of follow-up. Thus, whether bone formation marker levels are related to fracture risk remains unclear. In contrast to bone formation markers, data on the relationship between bone resorption markers and fracture risk are more consistent. Results obtained from three large prospective studies (EPIDOS, Rotterdam, and OFELY) indicate

**Table II Degree of Isomerization and Racemization of Type I Collagen and Risk of Fracture in Postmenopausal Women<sup>a</sup>**

Urinary CTX ratio	Relative risk (95% CI) of fracture for values in the upper quartile	
	All fractures	Nonvertebral fractures only
$\alpha$ L/ $\beta$ L		
Unadjusted	2.0 (1.2–3.5)	2.5 (1.3–4.6)
Adjusted for bone ALP	1.8 (1.1–3.2)	2.2 (1.1–4.2)
Adjusted for femoral neck BMD	1.8 (1.03–3.1)	2.2 (1.2–4.0)
Adjusted for bone ALP + femoral neck BMD	1.7 (0.95–2.9)	2.0 (1.04–3.8)
$\alpha$ L/ $\alpha$ D		
Unadjusted	1.8 (1.02–3.2)	1.9 (1.01–3.7)
Adjusted for bone ALP	1.6 (0.92–2.9)	1.7 (0.87–3.3)
Adjusted for femoral neck BMD	1.7 (0.97–2.9)	1.8 (0.95–3.5)
Adjusted for bone ALP + femoral neck BMD	1.6 (0.89–2.8)	1.7 (0.86–3.2)

<sup>a</sup>In this study, 408 healthy untreated postmenopausal women aged 50–89 (mean; 64 years) who were part of the OFELY cohort were followed for a median of 6.8 years. During follow-up, 16 incident vertebral fractures and 55 peripheral fractures were recorded in 65 women. The baseline levels of the four CTX isoforms ( $\alpha$ L CTX,  $\beta$ L CTX,  $\alpha$ D CTX, and  $\beta$ D CTX; see Fig. 4) in women who subsequently had a fracture were compared to those of the 343 women who did not have a fracture. As shown in the table, women with baseline levels of the ratio  $\alpha$ L/ $\beta$ L CTX in the highest quartile had a twofold higher risk of fracture than the other women, independent of bone turnover rate assessed by serum levels of bone alkaline phosphates (bone ALP) and/or femoral neck bone mineral density (BMD). From Garnero, P., Cloos, P., Sornay-Rendu, E., Qvist, P., and Delmas, P. D. (2002). Type I collagen racemization and isomerization and the risk of fracture in postmenopausal women: The OFELY prospective study. *J. Bone Miner. Res.* **17**, 826–833. Reproduced with permission of the American Society for Bone and Mineral Research.

that increased levels of bone resorption markers are associated with increased risk of hip, vertebral, non-hip, and nonvertebral fractures over follow-up periods ranging from 1.8 to 5 years. This predictive value is consistently on the order of a twofold increase in the risk of fracture for levels above the upper limit of the premenopausal range. Both increased levels of serum CTX and urinary CTX and of free DPD have been shown to be associated with a higher risk of hip, vertebral, and nonvertebral fractures. Increased bone resorption is associated with increased risk of fracture

only for values that exceed a threshold, suggesting that bone resorption becomes deleterious for bone strength only when it exceeds the normal physiological range. Because bone resorption rate predicts fracture independently of BMD, these data suggest that increased bone resorption can lead to increased skeletal fragility due to two factors. First, because high levels of bone turnover markers have been shown to be associated with increased bone loss in subsequent years, a prolonged increase in bone turnover will lead to a lower BMD, which is a major determinant of reduced bone strength. Second, increased bone resorption above the upper limit of the normal range may induce microarchitectural deterioration of bone tissue such as perforation of trabeculae, a major component of bone strength.

OC contains three residues of  $\gamma$ -carboxyglutamic acid, a vitamin K-dependent amino acid. It was postulated that impaired  $\gamma$ -carboxylation of osteocalcin could be an index of both vitamin D and vitamin K deficiency in elderly populations. In two prospective studies performed in a cohort of elderly institutionalized women and in a population of healthy elderly women (the EPIDOS study), levels of undercarboxylated osteocalcin (ucOC) over the premenopausal range were associated with a two- or threefold increase in the risk of hip fracture, although total osteocalcin was not predictive. Like markers of bone resorption, the prediction remained significant after adjusting for hip BMD.

Because increased levels of bone resorption markers and of ucOC have been shown to predict the risk of fracture independently of the level of BMD and of clinical risk factors including history of previous fragility fractures, the combination of different risk factors could be useful to improve the identification of women at high risk for fracture. Using the database of the EPIDOS study, we showed that the combination of a bone resorption marker (or ucOC) and hip BMD measurement can detect elderly women at very high risk of hip fracture. Indeed, women with both low hip BMD (according to the World Health Organization definition of osteoporosis) and high bone resorption (or high ucOC) had a four- or fivefold higher risk compared to the general population. However, the use of an odds ratio is not ideal for clinical decision making since the risk may decrease or remain stable with age, whereas absolute risk increases. Thus, calculating absolute risk such as 10-year probabilities, which depend on knowledge of the fracture and death hazards, is probably more appropriate. Based on the probability of hip fracture in the Swedish population and on the marker data of the

EPIDOS study, Johell *et al.* found that combining urinary CTX with BMD or history of previous fracture resulted in a 10-year probability of hip fracture that was approximately 70–100% higher than that associated with low BMD alone (Table III). Thus, clearly the use of multiple risk factors, such as BMD, biochemical markers, and personal history of previous fracture, likely results in better performance than the use of BMD alone.

### Bone Markers for Monitoring Treatment of Osteoporosis with Antiresorptive Therapy

As for most chronic diseases, monitoring the efficacy of treatment of osteoporosis is a challenge. The goal of treatment is to reduce the occurrence of fragility fractures, but their incidence is low and the absence of events during the first year(s) of therapy does not necessarily indicate that treatment was effective. Thus, the use of surrogate markers with a more rapid response is clearly needed for an efficient monitoring of treatment of osteoporosis. A surrogate marker is a laboratory measurement or a physical sign used as a substitute for a clinical meaningful endpoint that directly measures how a patient feels, functions, or survives. Changes induced by a therapy on a surrogate endpoint are expected to reflect changes in a clinical meaningful endpoint.

Measurement of BMD by DXA is a surrogate marker of treatment efficacy that has been widely used in clinical trials. Its use in the monitoring of treatment efficacy in the individual patient, however, has not been validated. Given a short-term precision error of 1–1.5% of BMD measurement at the spine

and hip, the individual change must be >3–5% to be considered significant. With bisphosphonates such as alendronate, repeating BMD 2 years after the initial therapy will allow the determination of whether a patient is responding to therapy (i.e., shows a significant increase in BMD, at least at the lumbar spine, which is the most responsive site). With treatments such as raloxifene or nasal calcitonin that induce much smaller increases in BMD, DXA is not appropriate for monitoring therapy. With any treatment, DXA does not allow the identification of all responders within the first year of therapy. Failure to respond may be due to noncompliance (probably the most important factor), poor intestinal absorption (i.e., oral bisphosphonates), other factors contributing to bone loss, or other unidentified factors. Monitoring may improve compliance, although this needs to be proven for osteoporosis treatment.

Several randomized, placebo-controlled studies found that resorption-inhibiting therapy was associated with a prompt decrease in bone resorption markers that was seen as early as 1 month, with a plateau reached within 3–6 months (Fig. 5). The decrease in bone formation markers is delayed, reflecting the physiological coupling of formation to resorption, and a plateau is usually achieved within 6–12 months (Fig. 5). Several placebo-controlled studies have demonstrated that the magnitude of the decrease in bone turnover markers under antiresorptive therapy including HRT and bisphosphonates, usually expressed as a percentage of the initial value, is strongly correlated with the increase in BMD after 2 or 3 years.

For the clinician, the primary concern is the identification of nonresponders (i.e., patients who fail to demonstrate a significant increase in BMD after

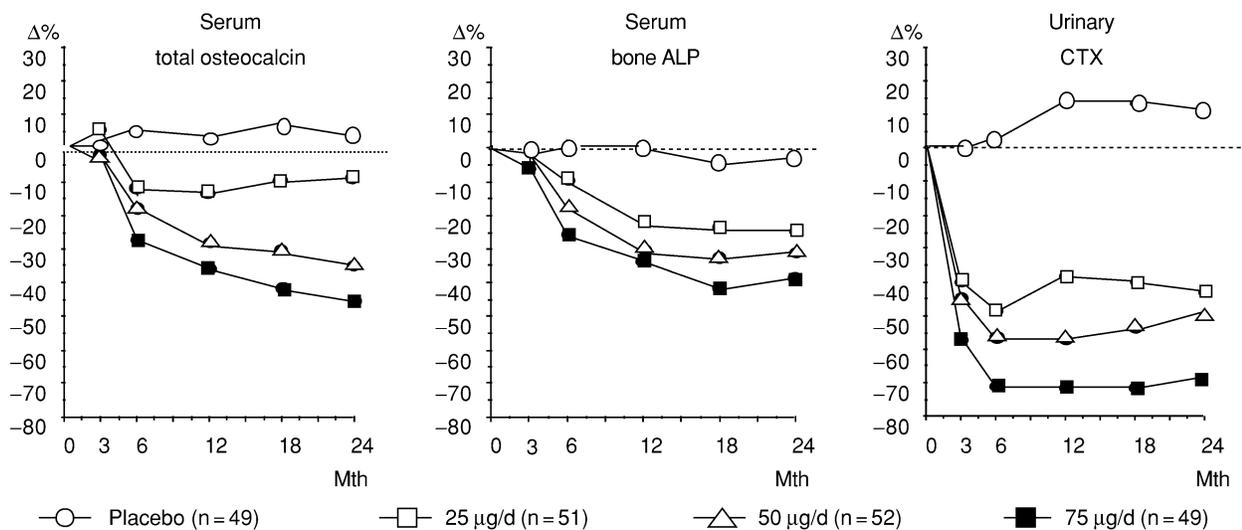
Table III Effect of Risk Factors Alone or in Combination on the Relative Risk of Hip Fracture in Women Age 80<sup>a</sup>

Risk factor	Threshold value	Prevalence (%)	Odds ratio	Relative risk <sup>b</sup>	Ten-year probability (%) <sup>c</sup>
Average	—	100		1.0	18.0
Low BMD	<i>t</i> score < -2.5	56	2.8	1.40	23.6
Prior fracture	Yes	39	3.5	1.77	28.8
High CTX	Above premenopausal values	23	2.4	1.82	29.5
Low BMD + prior fracture	Above premenopausal values	23	4.1	2.39	36.3
Low BMD + high CTX	Above premenopausal values	16	4.1	2.74	40.1
Prior fracture + high CTX	Above premenopausal values	12	5.3	3.50	47.3
All of the above	Above premenopausal values	7	5.8	4.43	54.5

<sup>a</sup>Reproduced with permission from Johnell, O., Oden, A., De Laet, C., Garnero, P., Delmas, P. D., and Kanis, J. A. (2002). Biochemical indices of bone turnover and the assessment of fracture probability. *Osteoporos. Int.* 13, 523–526.

<sup>b</sup>Adjusted for the prevalence of the risk factor in the population.

<sup>c</sup>Probability of hip fracture within the next 10 years.



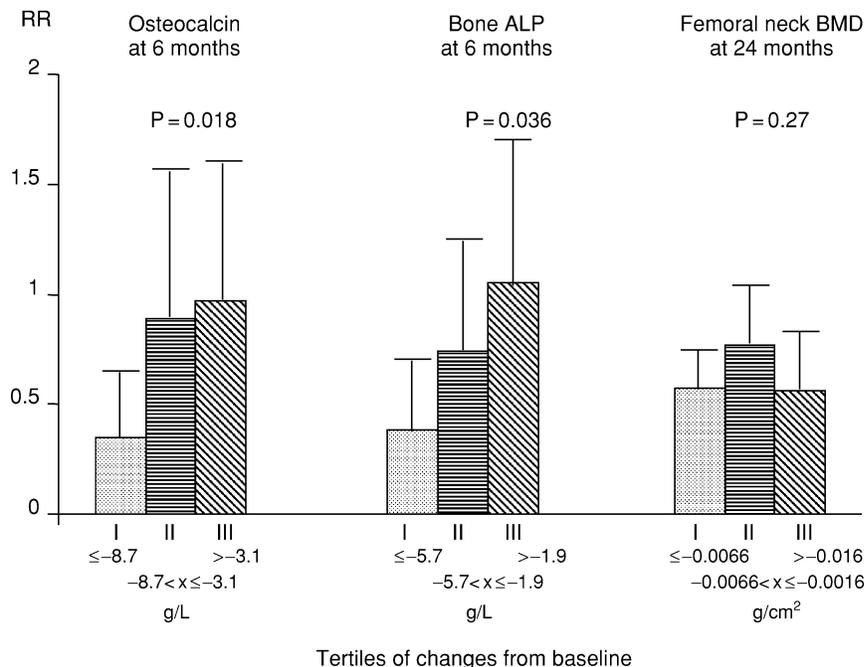
**Figure 5** Effect of different doses of transdermal estradiol on markers of bone turnover. In this study, 201 postmenopausal women aged 40–60 years who had a time since menopause of less than 6 years were given either a placebo or a new matrix delivery transdermal  $17\beta$ -estradiol in a dosage of 25, 50, or 75  $\mu\text{g}$  twice a week for 28 days. The decrease of bone formation markers (osteocalcin and bALP) is delayed compared to that of bone resorption (urinary CTX), reflecting the coupling mechanism. This delayed decreased bone formation is amplified by the transdermal route of administration of estradiol. Reproduced with permission from Cooper, C., Stakkestad, J. A., Radowicki, S., Hardy, P., Pilate, C., Dain, M. P., and Delmas, P. D. (1999). Matrix delivery transdermal  $17\beta$  estradiol for the prevention of bone loss in postmenopausal women. *Osteoporos. Int.* 9, 358–366.

2 years of treatment). Although bone markers do not allow an accurate prediction of bone gain in individual patients, recent studies indicate that the measurement of a marker for resorption and/or formation can provide the same information on therapeutic efficacy as the measurement of BMD. Retrospective analyses of several clinical trials using HRT or alendronate suggest that, for a given marker of resorption or formation, a cutoff value of change from baseline after 3–6 months of treatment, which is dependent on the type of therapy (HRT vs bisphosphonate) and the marker, can be defined that provides adequate predictive value of the subsequent 2- or 3-year BMD response in a single patient. The value of BMD changes to predict the risk of fracture with treatment is debated, especially because treatments such as raloxifene can induce a 30–50% reduction in vertebral fracture rate despite a 2 or 3% increase in BMD at all skeletal sites. Recent reanalyses of placebo-controlled studies suggest that BMD changes under treatment account for only a small part of the efficacy of antiresorptive therapy on fracture risk. Thus, BMD changes may not be an adequate surrogate endpoint to analyze the ability of bone markers to predict fracture risk. There have been few attempts to correlate bone marker changes with fracture risk. In the MORE study, it was found that the short-term changes of serum OC and bone ALP with raloxifene treatment were associated with

the subsequent risk of vertebral fractures in a large subgroup of osteoporotic women, whereas changes in hip BMD were not predictive (Fig. 6). In postmenopausal women with osteoporosis treated with oral risedronate, it was shown that changes in urinary CTX and NTX after 3–6 months predicted the risk of subsequent incident vertebral fractures after both 1 and 3 years (Fig. 7). These changes accounted for 50–70% of the effect of risedronate on fracture risk. Interestingly, in the MORE study, it was shown that the relationship between vertebral fracture risk and changes from baseline in CTX and NTX during 3–6 months of treatment was not linear (Fig. 7) and that there may be a level of bone resorption reduction below which there is no further fracture benefit. A significant association between changes of bone ALP and vertebral, hip, and nonspine fracture was also found in women treated with alendronate who participated in the FIT trial. Ultimately, recommended cutoff values of bone marker changes with treatment should be based on prospective studies with incident fractures as an endpoint.

## BONE MARKERS IN PAGET'S DISEASE

Paget's disease is a localized disorder characterized by a marked increase in bone turnover, leading to

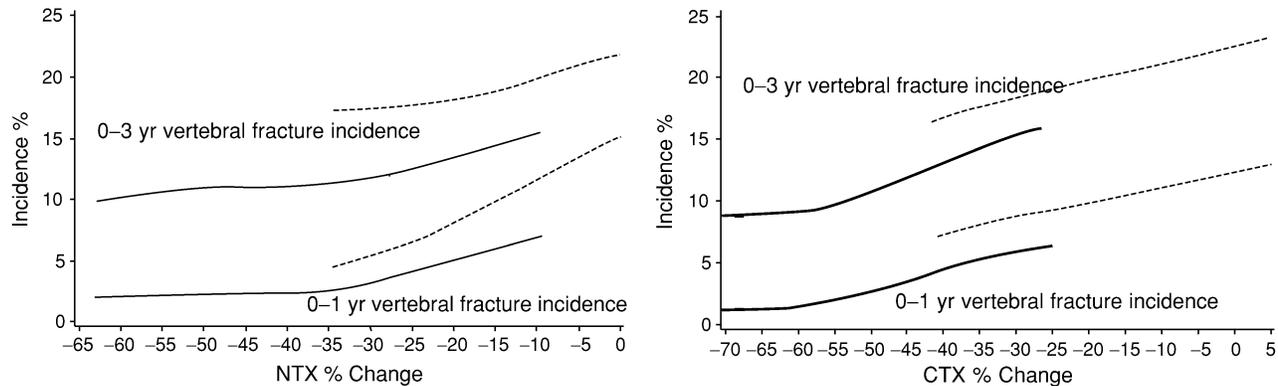


**Figure 6** The relative risk of new vertebral fracture at 3 years (raloxifene vs placebo) by tertile of change in serum osteocalcin, serum bone alkaline phosphatase (bone ALP) after 6 months, and femoral neck BMD after 24 months. Postmenopausal women with osteoporosis participating in the MORE study were treated with raloxifene (60 or 120 mg/day) or placebo for 3 years. Note that there was a significant relationship between the magnitude of changes in bone markers at 6 months and the relative risk of new vertebral fracture at 3 years, whereas changes in femoral neck BMD at 24 months were not predictive. The *p* values are for interaction and indicate the presence of a differential antifracture efficacy across tertile of changes for a model including tertile of change, therapy, and tertile therapy. *n* = 2413, 2403, and 6745 for osteocalcin, bone ALP, and femoral neck BMD, respectively. Reproduced with permission from Bjarnason, N. H., Christiansen, C., Sarkar, S., Mitlak, B., Knickerbocker, R., Delmas, P., Cummings, S., for the MORE Study Group (2001). 6 months changes in biochemical markers predict 3-year response in vertebral fracture rate in postmenopausal, osteoporotic women: Results from the MORE study. *Osteoporos. Int.* **12**, 922–930.

overproduction of bone of poor quality, responsible for hypertrophy, osteosclerosis, and bone fragility. The architecture of the lamellar texture of the pagetic bone matrix is disorganized, with a predominance of woven bone that is characterized by an irregular and patchy arrangement of collagen fibers. Biochemical markers of bone turnover are routinely used for assessing the disease activity and for monitoring the efficacy of bisphosphonate therapy. Because of a marked increase in bone turnover, serum total ALP activity is the most commonly used bone marker for Paget's disease for both applications. Although serum total ALP is adequate to monitor most patients with active disease, this marker may lack sensitivity in patients with monostotic disease affecting a small bone and in patients with purely osteolytic lesions.

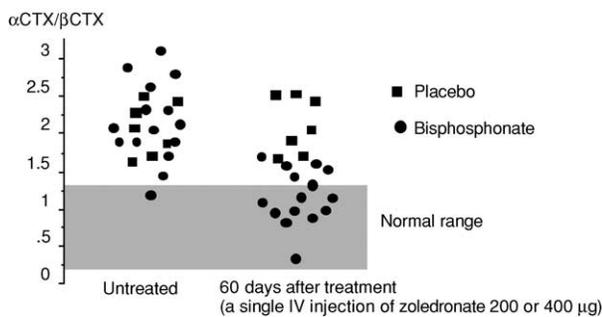
To overcome these limitations, the new markers described previously and validated for osteoporosis have been suggested to be useful for patients with Paget's disease. However, because a given marker may not perform equally in different metabolic bone diseases, they must be tested independently in pagetic

patients. Among formation markers, bone ALP and PINP appear to be the most sensitive markers for assessing disease activity. These markers may be particularly interesting for patients with monostotic disease. Although osteocalcin is one of the most sensitive markers in osteoporosis, it lacks sensitivity in Paget's disease probably because the fraction of newly synthesized osteocalcin that is incorporated into bone matrix may be increased because of the high mineral content of the woven bone. For bone resorption, the type I collagen-related markers have advantageously replaced urinary hydroxyproline. Among these, urinary NTX and urinary nonisomerized  $\alpha$ CTX appear to be the most sensitive bone resorption markers both for assessing disease activity and for monitoring efficacy of bisphosphonate therapy. As discussed previously, the pagetic bone matrix is characterized by an impaired degree of  $\beta$ -isomerization of type I collagen molecules. We found that in patients with active Paget's disease, the urinary excretion of nonisomerized  $\alpha$ CTX was markedly increased (13.5-fold vs controls) compared to  $\beta$ CTX (3.5-fold vs controls),



**Figure 7** Relationship between percentage change in bone resorption markers (urinary NTX and CTX) and the incidence of new vertebral fractures. A total of 693 women with at least one vertebral deformity (mean age, 69 years; SD, 7 years) who received placebo or 5 mg oral risedronate daily for 3 years were studied. The relationships between vertebral fracture risk after 1 and 3 years of treatment and changes from baseline in NTX and CTX were not linear ( $p < 0.05$ ). There was little improvement in fracture benefit after a decrease of 35–40% for NTX and 55–60% for CTX. The placebo group is represented by the dashed line and the risedronate group by the solid line. All patients received calcium supplementation (1000 mg/day) and vitamin D (if levels were low). From Eastell, T., Barton, I., Hannon, R. A., Chines, A., Garnero, P., and Delmas, P. D. (2003). Relationship of early changes in bone resorption to the reduction in fracture risk with risedronate. *J. Bone Miner.* **18**, 1051–1056. Reproduced with permission of the American Society of Bone and Mineral Research.

resulting in a urinary  $\alpha$ CTX: $\beta$ CTX ratio that was 3-fold higher than that in controls. Bisphosphonates induce a marked decrease in bone turnover rate in pagetic patients, and histological studies have shown



**Figure 8** Effect of bisphosphonate treatment on the urinary ratio of nonisomerized to  $\alpha$ -isomerized C-telopeptide breakdown products ( $\alpha$ CTX/ $\beta$ CTX) in Paget's disease. Twenty-eight patients with active Paget's disease were studied before and 2 months after a single injection of the bisphosphonate zoledronate (200 or 400  $\mu$ g) or placebo. Before treatment, most patients (93%) were characterized by an elevated urinary  $\alpha$ CTX: $\beta$ CTX ratio with a mean 2.7-fold increase compared to controls (2.1 vs 0.8). Two months after treatment, the  $\alpha$ CTX: $\beta$ CTX ratio was decreased by 50% in zoledronate-treated patients and returned to the normal range (gray area: mean  $\pm$  2 SD) in 65% of patients. No significant change was observed in the placebo group. Reproduced from Garnero, P., Gineyts, E., Shaffer, A. V., Seaman, S., and Delmas, P. D. (1998). Measurement of urinary excretion of nonisomerized and  $\alpha$ -isomerized forms of type I collagen breakdown products to monitor the effects of the bisphosphonate zoledronate in Paget's disease. *Arthritis Rheum.* **41**, 354–360., with permission of the American College of Rheumatology.

that a few months of treatment induce the formation of normal lamellar bone, suggesting that treatment restores bone quality. In patients with active Paget's disease, we found that a single injection of 200–400  $\mu$ g of zoledronate induced a significant decrease in the urinary  $\alpha$ CTX: $\beta$ CTX ratio, which was 3-fold higher than in controls before treatment, and that it returned to the reference range in most patients after 2 months of treatment (Fig. 8). This observation suggests that bisphosphonate treatment induced progressive replacement of woven bone by lamellar bone with a higher degree of  $\beta$ -isomerization of type I collagen, although this remains to be confirmed by direct histological evidence. Thus, the urinary  $\alpha$ CTX: $\beta$ CTX ratio may be a novel index of bone quality that might be useful for monitoring pagetic patients.

## See Also the Following Articles

Bisphosphonates • Bone Mass Measurement • Bone Remodeling, Dynamics of • Bone Structure • Collagen Metabolism • Osteoporosis in Older Men • Osteoporosis in Older Women • Osteoporosis, Overview • Paget's Disease of Bone

## Further Reading

Cloos, P. A., and Christgau, S. (2002). Non-enzymatic covalent modifications of proteins: Mechanisms, physiological consequences and clinical applications. *Matrix Biol.* **21**, 39–52.  
Cummings, S. R., Karpf, D. B., Harris, F., Genant, H. K., Ensrud, K., LaCroix, A. Z., and Black, D. M. (2002). Improvement in

- spine density and reduction in the risk of vertebral fractures during treatment with antiresorptive drugs. *Am. J. Med.* **112**, 281–289.
- Delmas, P. D. (2000). Markers of bone turnover for monitoring treatment of osteoporosis with antiresorptive drugs. *Osteoporos. Int.* **11**(Suppl. 6), S66–S76.
- Delmas, P. D., and Meunier, P. J. (1997). The management of Paget's disease of bone. *N. Engl. J. Med.* **336**, 558–566.
- Delmas, P. D., Eastell, R., Garnero, P., Seibel, M. J., and Stepan, J. (2000). The use of biochemical markers of bone turnover in osteoporosis. *Osteoporos. Int.* **11**(Suppl. 6), S2–S17.
- Garnero, P. (2000). Markers of bone turnover for the prediction of fracture risk. *Osteoporos. Int.* **11**(Suppl. 6), S55–S65.
- Garnero, P., and Delmas, P. D. (1999). *Laboratory Assessment of Postmenopausal Osteoporosis. Dynamics of Bone & Cartilage Metabolism*, pp. 1–13. Academic Press, San Diego.
- Garnero, P., Hausher, E., Chapuy, M. C., *et al.* (1996). Markers of bone resorption predict hip fracture in elderly women: The Epidos prospective study. *J. Bone Miner. Res.* **11**, 1531–1538.
- Hannon, R., and Eastell, R. (2000). Preanalytical variability of biochemical markers of bone turnover. *Osteoporos. Int.* **11**(Suppl. 6), S30–S44.
- Johnell, O., Oden, A., De Laet, C., Garnero, P., Delmas, P. D., and Kanis, J. A. (2002). Biochemical indices of bone turnover and the assessment of fracture probability. *Osteoporos. Int.* **13**, 523–526.
- Moss, D. W. (1982). Alkaline phosphatase isoenzymes. *Clin. Chem.* **28**, 2007–2016.
- Price, P. A. (1987). Vitamin K-dependent bone proteins. In *Calcium Regulation and Bone Metabolism: Basic and Clinical Aspects* (D. V. Cohn, T. J. Martin, and P. J. Meunier, eds.), Vol. 9, pp. 419–426. Elsevier, Amsterdam.
- Risteli, J., and Risteli, L. (1999). Products of bone collagen metabolism. In *Dynamics of Bone and Cartilage Metabolism* (M. J. Seibel, S. P. Robins, and J. P. Bilezikian, eds.), pp. 275–287. Academic Press, San Diego.
- Stepan, J. J. (2000). Prediction of bone loss in postmenopausal women. *Osteoporos. Int.* **11**(Suppl. 6), S45–S54.



# Brain, Effects of Steroid Hormones

Russell D. Romeo and Bruce S. McEwen  
*Rockefeller University, New York, New York, United States*

## Glossary

**dendrite** Branching neurites that emanate from the neuronal cell body.

**dendritic spine** A small protrusion of the cell membrane of the dendrite that receives excitatory input.

**glucocorticoids** Steroid hormones released from the cortex of the adrenal gland such as cortisol and corticosterone.

**neurogenesis** The production of new neurons.

**sex steroids** Steroid hormones released from the gonads such as androgens, estrogens, and progestins.

**steroid hormones** A family of hormones derived from cholesterol and characterized by their three six-carbon rings and one five-carbon ring.

Steroid hormones have profound effects on the structure and function of the nervous system. Steroids are responsible for organizing neural circuits early in development and for activating the appropriate physiology and behavior during adulthood. Steroids also continue to shape the structure and function of the brain throughout the life span of the organism. This article focuses on these developmental, structural, and functional influences of steroids on the brain and their behavioral and physiological consequences.

## INTRODUCTION

The brain receives a rich supply of blood and is constantly in contact with circulating hormones in the bloodstream. Thus, it is not surprising that the brain is influenced and modulated by hormones. One major class of hormones is the steroid hormones, which gain access to the brain because of their hydrophobic chemical structure that allows them to pass across the blood–brain barrier. Steroid hormones have a wide range of effects on the brain. For instance, steroids help to shape the brain during perinatal development and continue to modulate the structure and

function of the central nervous system (CNS) into adulthood and throughout the life span. Specifically, steroids can influence factors such as neuronal survival, neurogenesis, neurite outgrowth, synaptogenesis, receptor expression, RNA synthesis, and neuronal excitability. The effects of the steroid hormones on the development, structure, and function of the brain are the focus of this article.

## THE MECHANISMS OF STEROID HORMONE ACTION IN THE BRAIN

In mammals, the precursor to all steroid hormones is cholesterol. The sex steroids (e.g., androgens, estrogens, progestins) and glucocorticoids (e.g., corticosterone) are synthesized from cholesterol via enzymatic cleavage. For example, testosterone is produced by a series of enzymatic reactions that convert cholesterol to progesterone, then to testosterone acetate, and finally to testosterone. Testosterone is converted to estrogen by the aromatase enzyme in a reaction called aromatization. Although testosterone is considered the “male sex steroid” and estrogen is considered the “female sex steroid,” it is important to note that testosterone and estrogen are found in both males and females. In fact, estrogen formed locally in the brain of males by the aromatization of testosterone is responsible for masculinization of the male brain and plays a major role in activating male sexual behavior during adulthood. Corticosterone (or cortisol in primates) is produced from a series of enzymatic reactions that convert progesterone to 11-deoxycorticosterone and then to corticosterone. Corticosterone is involved in carbohydrate metabolism and is released in response to stressful stimuli. These steroids bind to specific receptors located throughout the brain. However, the spatial distribution and concentration of these receptors are different for each hormone.

There are three mechanisms through which steroid hormones exert their actions on neurons: genomic, indirect genomic, and nongenomic. The genomic

mechanism of steroid hormone action begins by the hormone diffusing through the plasma membrane of the neuron and binding to its intracellular receptor located in either the cytoplasm or the nucleus. This hormone–receptor complex then undergoes a conformational change allowing the hormone–receptor complex to migrate, dimerize, and bind to particular hormone response elements (HREs) on the DNA. Once the hormone–receptor complex binds to its HRE, transcription or repression of a particular gene is initiated, ultimately allowing the production of proteins that may alter the functioning of the neuron.

Recent advances in our understanding of steroid hormone action in neurons have uncovered several indirect genomic and nongenomic actions of steroids. For instance, steroids have been shown to bind to receptors located in the plasma membrane of neurons, which activate second messenger cascades that initiate gene transcription. These indirect genomic effects of steroids are not mediated through HREs but instead are mediated by other DNA regulatory sites such as the cyclic AMP (cAMP) response element (CRE) and activator protein-1 (AP-1). In addition to these indirect genomic effects, steroids can act through nongenomic mechanisms such as directly modulating the permeability of ion channels and acting as antioxidants in neurons. Steroids may also act at specific local sites in the neurons that do not involve interactions with the DNA. Extranuclear estrogen receptors were found in axon terminals and spines of hippocampal CA1 pyramidal cells of the female rat. Various mRNAs for synaptic proteins and translation machinery have been localized in spines, suggesting that translation of such proteins can occur locally in the dendrite. Thus, estrogens may act directly on the synaptic apparatus and mRNAs to affect synaptic function and synaptogenesis. These various mechanisms of steroid hormone action are presented in [Fig. 1](#).

## ORGANIZATION VERSUS ACTIVATION

Steroid hormones not only allow for the activation of certain behavioral and physiological events during adulthood but also shape the nervous system during development. During perinatal development, the brains of males and females are exposed to different hormonal milieus that ultimately lead to many of the sex differences observed in the brain and behavior of males and females during adulthood. For instance,

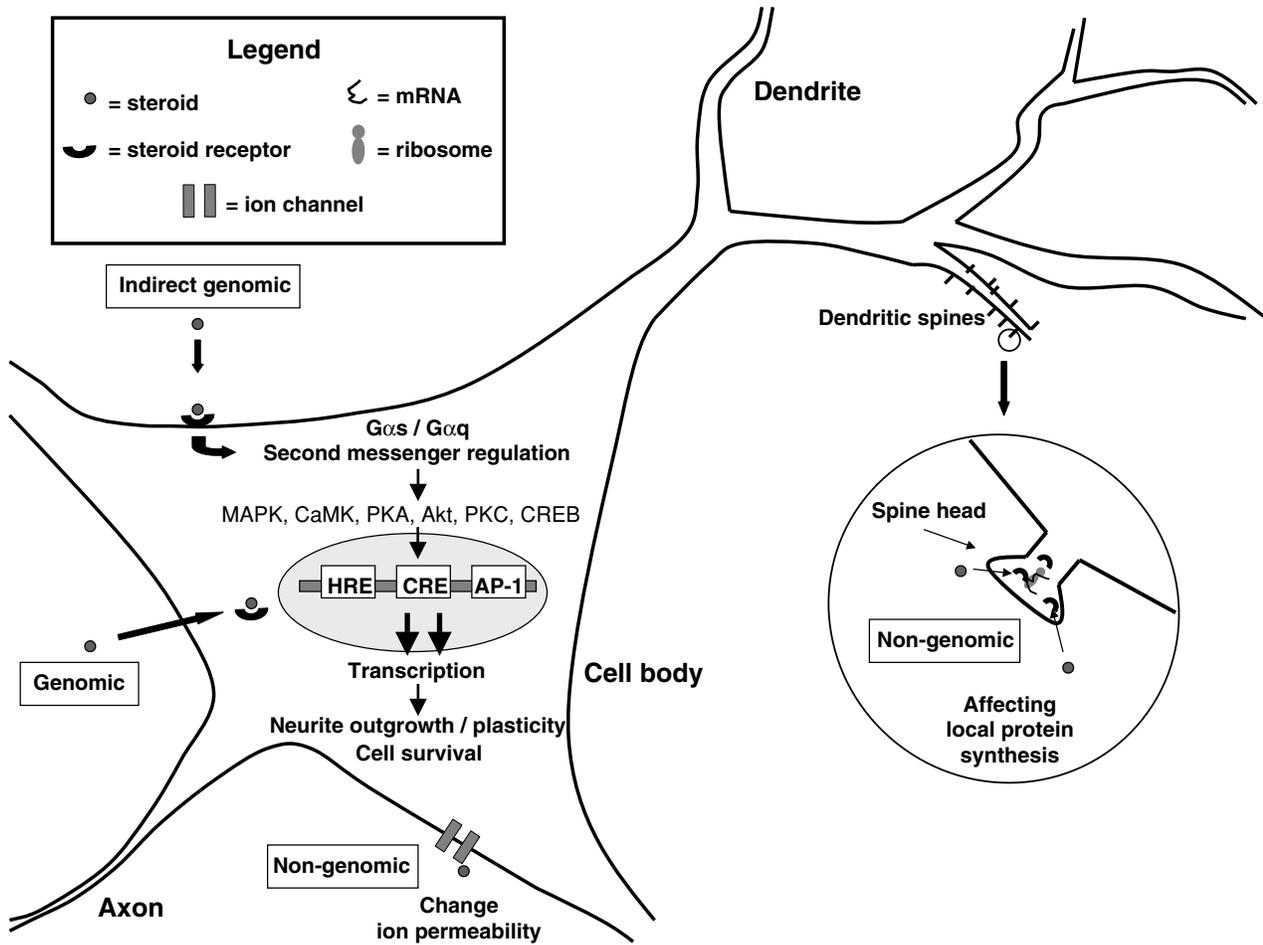
males experience increases in testosterone production and secretion during both prenatal and early neonatal development. This, on conversion in the brain to estradiol via the aromatase enzyme, organizes the brain in a masculine fashion. This early masculinization of the brain then allows the hormonal stimulation received during adulthood to act on these organized neural pathways to activate the appropriate male behaviors. In the perinatal female, the estrogen produced by the mother and the prenatal ovary does not masculinize her brain because this estrogen is bound by alpha fetoprotein and is unable to cross the blood–brain barrier. Indeed, the feminized brain appears to be the default pattern of development in that a brain that does not receive androgenic and estrogenic stimulation during development results in a feminized brain.

The organizing influences of steroid hormones on the perinatal brain and the resulting morphological and functional differences between the sexes can be profound. Indeed, the organizing effects of steroids result in certain sexually dimorphic nuclei in the nervous system that may be as much as five times larger in one sex than in the other. The effects of steroids on the adult brain are much more subtle than these organizing effects during perinatal development. However, steroids are still capable of shaping the adult brain in important ways.

## STEROID-INDUCED NEURONAL PLASTICITY DURING ADULTHOOD

Sex steroids have been shown to affect the density and morphology of dendritic spines. Spines are membranous protrusions emanating from dendrites and are sites of synaptic contact between neurons. Neurons receive excitatory input via these dendritic spines, which contain various types of glutamate receptors, scaffolding proteins, and signaling molecules. Spines come in various sizes and shapes that may be indicative of their maturity. Different steroids can affect the growth and development of spines, thereby modulating excitatory neuronal signaling.

The hypothalamus was the first brain area in which sex steroids were shown to influence synapse formation and spine maturation. Ovariectomized females have significantly fewer synaptic contacts in their ventromedial nucleus (VMN) than do ovariectomized females treated with estrogen. Furthermore, females in the proestrous stage of their estrous cycle (i.e., when estrogen levels are high) have a greater spine density than do females in diestrus (i.e., when



**Figure 1** The genomic, indirect genomic, and nongenomic mechanisms of steroid hormone action on neurons. Indirect genomic mechanisms include the activation of steroid receptors linked to second messenger cascades that activate gene transcription through DNA-binding domains such as cAMP response element (CRE) and activator protein-1 (AP-1), ultimately altering neuronal function. In the genomic mechanism, the steroid binds to the intracellular form of the receptor, permitting the steroid-receptor complex to translocate to the nucleus. This steroid-receptor complex then binds to the hormone response element (HRE), activating transcription of the genome and altering neuronal function. Nongenomic mechanisms of steroid action include the ability of steroids to influence ion permeability and possibly to affect local protein synthesis in the spine head. This later mechanism presumably would allow for the rapid regulation of spine-specific mRNAs and proteins. Akt, protein kinase B; CaMK, Ca<sup>2+</sup>/calmodulin-dependent protein kinase; CREB, cAMP response element-binding protein; G $\alpha$ s and G $\alpha$ q, guanosine triphosphate (GTP)-binding proteins; MAPK, mitogen-activated protein kinase; PKA, protein kinase A; PKC, protein kinase C.

estrogen levels are relatively low). The VMN is a fundamental component of the neural circuit that mediates the expression of female reproductive behavior. Thus, it is not surprising that estrogenic stimulation that activates female mating behavior would also promote spine density and, hence, increase excitatory input to the VMN. In males, VMH spine density increases after castration, whereas estrogen-treated males exhibit a relatively low number of spines, similar to the density observed in intact males. Thus, the effect of estrogen on VMN spine density in males is

opposite that in females. Interestingly, when males are treated with an aromatase inhibitor on the first day of life to block the normal masculine organization of the brain, estrogen treatment during adulthood results in an increase in spine density similar to that observed in females.

The arcuate nucleus (ARC) of the hypothalamus plays an important role in mediating the release of the gonadotropins from the anterior pituitary. The dendrites of these neurons exhibit a sex difference such that females have twice as many spines as do males.

Similar to the VMN, androgenic and estrogenic stimulation received early in development appears to organize these neurons. Specifically, females that receive testosterone neonatally have fewer spines than do untreated females during adulthood. Conversely, adult males that are castrated neonatally show a greater number of spines than do control males. The greater number of spines, and presumably the greater excitatory input to the female ARC, may help to modulate the luteinizing hormone surge that causes ovulation in females.

The effects of sex steroids on brain areas outside of the hypothalamus and their role in biological events other than reproductive physiology and behavior have been explored recently. The sex steroid receptors are located in many extrahypothalamic regions of the brain, such as the hippocampus and the cortex, and steroids have been implicated in mediating diverse phenomena from immunity to cognition. An accumulating body of evidence shows that estrogen can improve an organism's ability to learn and remember. How estrogen mediates these effects on learning and memory is unclear. However, a very interesting and emerging story points to the ability of estrogen to affect the structure and function of the hippocampus, an area of the brain implicated in learning and memory.

Female rats that have been ovariectomized have significantly fewer dendritic spines on the dendrites of the CA1 pyramidal cells of the hippocampus than do estrogen-treated females. Moreover, female rats experiencing the normal fluctuations of estrogen during their 4- to 5-day estrous cycle show changes in spine density. Specifically, females in proestrus, when estrogen levels are at their highest, had a significantly greater number of dendritic spines on their CA1 cells than did females that were sacrificed on the day of estrus, when estrogen levels are relatively low. Additional studies have shown that this increase in spine density is also accompanied by an increase in synaptic input to the apical dendrites of the CA1 pyramidal cells. These structural alterations are paralleled by physiological and behavioral changes as well. Specifically, adult female rats experiencing high levels of estrogen show enhanced hippocampal long-term potentiation (LTP), a putative electrophysiological correlate of learning and memory, and improved memory retention on a hippocampal-dependent spatial memory task.

A study that examined estrogen-treated female nonhuman primates revealed that estrogen increases hippocampal spine density in higher order mammals as well. Whether estrogen increases hippocampal

spine formation in humans is unknown, but estrogen replacement therapy has been shown to promote cognition in postmenopausal women and possibly to aid in the prevention or amelioration of neurodegenerative diseases such as Alzheimer's disease.

Males do not show the same increase in spine density in response to estrogen. However, testosterone has been shown to increase the number of spines on the CA1 pyramidal cells of males. It is unknown whether testosterone enhances LTP or improves memory retention in males, although one study conducted on elderly men suggests that testosterone can increase performance on certain memory tasks.

In addition to synaptogenic effects of estrogen and testosterone on the hippocampus, one study has shown that estrogen can induce greater spine density in the prefrontal cortex of nonhuman primates. Given that the prefrontal cortex of primates is involved in many aspects of cognitive function, it is possible that the memory-enhancing effects of estrogen are mediated, at least in part, by its actions on both the hippocampus and the cortex.

Not all steroid hormones increase spine density and synaptogenesis. Studies investigating the effects of progesterone on synaptogenesis have shown that progesterone in the presence of estrogen actually decreases spine density in the female CA1 hippocampal region. The effects of progesterone alone on hippocampal spine density have not been fully explored. However, recent experiments suggest that progesterone in the absence of estrogen may actually promote spine growth and synaptogenesis.

In contrast to the spine-promoting effects of estrogen and testosterone on the male and female hippocampus, high levels of corticosterone induced by repeated stressors have been shown to decrease the branching of hippocampal CA3 pyramidal cells and to cause dendritic atrophy. In severe cases, extensive exposure to high levels of stress and the resulting chronic elevation of corticosterone can actually kill hippocampal neurons. Animals experiencing chronic stress also tend to perform more poorly on learning and memory tasks than do nonstressed animals. It is important to note that low to intermediate levels of "stress hormones" enhance performance on learning and memory tasks. Studies investigating the effects of corticosterone on hippocampal LTP parallel these behavioral results such that high levels of corticosterone suppress LTP, whereas intermediate levels enhance plasticity. It is interesting to speculate that individuals experiencing severe and prolonged chronic stress will have compromised hippocampal

and cognitive function. Although some human research points to this possibility (e.g., in the case of depressive illness), more research is needed to establish a causal link.

## NEURONAL PROTECTION AND NEUROGENESIS

As discussed previously, steroids can have profound influences on the structure and function of neurons early in development and into adulthood. In addition to these influences, steroids can alter the survival and production of new neurons, even in the adult brain. For instance, at high concentrations, estrogen can protect neurons by working as an antioxidant. Estrogen has also been shown to diminish the neural damage induced by stroke and reduced blood flow to the cortex. These neurotropic effects after stroke may be through the antioxidant properties and/or through the influence of estrogen on growth factor synthesis and receptor expression.

Steroid hormones have also been shown to affect neurogenesis. Estrogen has been demonstrated to increase the production of new neurons in the dentate gyrus of the female hippocampus. Specifically, ovariectomized females treated acutely with estrogen have a greater number of proliferating cells than do untreated ovariectomized females. Moreover, normally cycling females experiencing the relatively high levels of estrogen during proestrus have more cellular proliferation in the dentate gyrus than do females experiencing the relatively low levels of estrogen associated with estrus and diestrus.

Conversely, animals experiencing high levels of corticosterone show less cellular proliferation in the dentate gyrus than do animals experiencing low levels of corticosterone. This corticosterone-induced decrease in neurogenesis suggests that high levels of corticosterone not only promote atrophy (and even loss) of neurons in the hippocampus but also slow the production of new neurons that may be needed to replace the lost cells.

## CONCLUSION

Steroid hormones secreted from the gonads and adrenals act on the brain and are known to influence diverse functions such as reproduction and learning and memory. These hormones can act on neurons through a genomic, an indirect genomic, or a nongenomic mechanism of action. The sex steroids have been shown to organize the male and female perinatal brain so that the proper physiology and behavior can be activated by these hormones during adulthood. Furthermore, steroids continue to shape the adult brain by affecting factors such as synaptogenesis, spine maturation and growth, neuronal branching, neuroprotection, and neurogenesis. Therefore, steroids not only contribute to the development of the nervous system but also continue to influence neuronal structure and function throughout the life span.

## See Also the Following Articles

Alzheimer's Disease and Hormones • Androgens, Gender and Brain Differentiation • Cardiovascular Disease: Impact of Sex Steroid Replacement • Endocrine Disrupters and Male Sexual Differentiation • Estrogen and the Male • GABA (Gamma-Aminobutyric Acid) • Glucocorticoids in Aging: Relevance to Cognition • Steroid Receptors, Evolution of

## Further Reading

- Arnold, A. P., and Breedlove, S. M. (1985). Organizational and activational effects of sex steroids on brain behavior: A reanalysis. *Horm. Behav.* **19**, 469–498.
- Lee, S. J., and McEwen, B. S. (2001). Neurotrophic and neuroprotective actions of estrogens and their therapeutic implications. *Annu. Rev. Pharmacol. Toxicol.* **41**, 569–591.
- McEwen, B. S., and Alves, S. E. (1999). Estrogen actions in the central nervous system. *Endocrine Rev.* **20**, 279–307.
- Sapolsky, R. M. (1999). Glucocorticoids, stress, and their adverse neurological effects: Relevance to aging. *Exp. Gerontol.* **34**, 721–732.
- Simerly, R. B. (2002). Wired for reproduction: Organization and development of sexually dimorphic circuits in the mammalian forebrain. *Annu. Rev. Neurosci.* **25**, 507–536.



# Breast Disease: Impact of Sex Steroid Replacement

Michael E. Carney

University of Hawaii, Honolulu, Hawaii, United States

## Glossary

**hormone replacement therapy (HRT)** Steroidal hormones containing an estrogen with or without a progesterone taken for various indications, including (most commonly) estrogen withdrawal symptoms that occur on cessation of endogenous ovarian sex steroid production; these medications are usually delivered as either oral or transdermal preparations.

**oral contraceptive pill (OCP)** Steroidal hormones usually containing both an estrogen and a progesterone that act in several ways, including ovulation inhibition and implantation hindrance, to effectively and reversibly prevent pregnancy.

**sex steroid hormones** Substances generally derived from a carbon backbone that affect sexual growth and development; in the body, these are secreted by the adrenal gland, ovary, and testis.

An understanding of the relationship between sex steroidal hormones and breast disease begins with an understanding of the connection between the sex steroids and the breast itself.

## INTRODUCTION

Growth and development of the normal healthy breast is dependent on the sex steroid hormones, specifically estrogen and progesterone. This relationship is initially most evident in a woman's life during puberty, when undeveloped and quiescent breast tissue is exposed to increasing amounts of estrogen and progesterone, ultimately resulting in the development of the adult breast. Simply put, exposure to these hormones causes growth and development of the breast tissue. Throughout premenopausal life, during a normal menstrual cycling, physiological changes occur within the breast tissue in response to varying levels of the steroid hormones. Perhaps the most dramatic evidence for this change is during pregnancy, when

the breast responds remarkably to the changing hormone milieu with breast enlargement, engorgement, lactation, and involution after breast-feeding has stopped. Finally, withdrawal of estrogen at menopause, either naturally or surgically, results in breast changes that are clinically measurable.

As medicine has progressed over the past several decades, so has the development of hormonal medications. Tremendous advancements have been achieved using these medications in the areas of pregnancy prevention and relief of vasomotor symptoms related to estrogen withdrawal. Hormonal contraception and hormone replacement are the two most common medications prescribed today in the United States.

Unfortunately, the most common cancer in U.S. women, affecting 1 in 8 women, is breast cancer. The incidence of breast cancer has risen over the past two decades, and there is mounting concern over the connection between exogenous hormone exposure and the development of breast cancer. This concern is not unfounded given that it stands to reason that the exposure of breast tissue to exogenous steroid hormones might affect breast physiology, including growth and development. This article initially examines breast anatomy, growth, and development and the relationship of the breast to sex steroid hormones. Later, it reviews the interaction between exogenous steroid hormones and breast tissue. Finally, it reviews the current literature examining hormone replacement therapy (HRT), including testosterone, oral contraceptives, and the development of breast disease, specifically breast cancer.

## THE BREAST, GROWTH AND DEVELOPMENT, AND SEX STEROIDS

Mammals are defined by the presence of breast tissue. In humans, the development of the breasts begins *in utero* at about the 6th week of fetal life along the "milk ridge" running from the groin to the axilla, similar to other mammals. However, unlike many other mammals, all but the chest portion of the human milk

ridge regresses by the 9th week of fetal life. At approximately the 12th to 16th weeks, mesenchymal cells develop into the smooth muscle of the nipple and areola. Epithelial buds form at 16 weeks and become 8 to 20 strips of epithelium, representing the future breast ducts and secreting alveoli.

In the male fetus, androgens interact with the early epithelial buds to induce suppression and even destruction. In the female at this time, early sweat glands begin to form the breast parenchyma. This early development is genetically determined and appears to be relatively independent of hormonal direction. Later, fetal breast development occurs in the rich hormonal environment of the mother, where estrogen and progesterone are at very high levels being produced by the placenta and maternal ovary. At the 20th week, the branched epithelial tissue is canalized and coalesces with sebaceous glands near the epidermis. Parenchymal differentiation occurs near term with the formation of a lobular–alveolar unit containing colostrom.

The breasts of both the male and female fetus respond to the continuous progesterone and estrogen hormonal milieu. Breast tissue is often palpable in the breasts of both genders at birth, and 85% of newborns will have nipple discharge on the second or third day of life. With the sudden drop of hormones at parturition in the newborn, the discharge and breast swelling quickly resolve. The response of breast tissue to estrogen and progesterone is undeniable, even at this early age. Several investigators, including Nicholas Petrakis, have speculated that the fetal hormonal environment and variable exposure to the estrogen and progesterone might contribute to an individual's risk for the future development of breast cancer.

The time between birth and puberty is relatively quiet in terms of breast tissue growth. Between 10 and 12 years of age in most girls, the hypothalamus begins releasing gonadotropin-releasing hormone (GnRH). When GnRH is released appropriately and rhythmically, it stimulates the pituitary to release luteinizing hormone (LH) and follicle-stimulating hormone (FSH), thereby promoting the primordial follicles in the ovary to produce estrogen. Ovulation does not occur for the first 1 or 2 years; thus, the predominant hormone released at this time is estrogen. The pubertal breast is exposed to estrogen, causing longitudinal ductal epithelial growth and the formation of more terminal ductile epithelial buds. Growth of the surrounding connective tissue, including fat, elastin, and vascularity, is similarly stimulated to grow. Interestingly, progesterone alone cannot cause this breast ductile epithelial growth.

With the onset of ovulation and, hence, exposure to circulating progesterone, full ductular–lobular–alveolar growth is achieved. At maturity, the functional breast is composed of 5 to 10 major nipple ducts that branch into several smaller collecting ducts to form a particular lobe. Each lobe has up to 40 lobules, with each lobule containing up to 100 alveoli that connect to a terminal duct. It is important to understand the role that hormones play in the development and growth of the breast given that virtually all breast cancers arise from the ductular–lobular–alveolar unit, with more than 90% arising from the ductile epithelium itself.

## HORMONAL MICROENVIRONMENT OF THE BREAST

The immature breast ducts and alveoli are lined by two epithelial layers. On exposure to estrogen, this epithelium grows and becomes multilayered with continuous terminal branching and alveolar development. The growth occurs during the proliferative phase of the menstrual cycle, when the ductile cells demonstrate increased mitosis, RNA synthesis, and ribosomal and Golgi apparatus activity. With the surge of progesterone during the mid-luteal phase, the ducts dilate and terminal alveolar cells differentiate into secretory cells, where lipid droplets form and are secreted into the lumen. Premenstrual breast fullness is directly attributable to the growth and secretion that occur as a result of estrogen and progesterone. Breast volume is at its minimum 5 days following menstruation due to the drop in hormone levels.

The hormonal effect on breast growth is even more dramatic during pregnancy. Estrogen causes marked ductile growth, branching, and lobule formation. Progesterone causes significant lobular growth and further formation during the second half of pregnancy. At parturition, the ductular–lobular–alveolar unit stops growing. The breast enlargement that occurs is related more to the secretion and storage of milk and fluid. An interesting cellular event occurs with the termination of lactation at weaning. On withdrawal of the hormones and growth factors of lactation, the terminally differentiated epithelial cells of the lactating gland enter into apoptosis. At menopause, with the decline of circulating estrogen and progesterone, the epithelial structures and stroma regress, leaving an intact ductile system but shrunken and collapsed lobules.

Breast duct growth occurs in response to hormonal exposure due to the presence of estrogen and progesterone receptors within the duct epithelial cells. A

complex symphony of molecular interactions involving the hormone–receptor complex occurs, ultimately resulting in cellular growth and differentiation. This mechanism is not completely understood but does involve many different transmembrane and intracellular growth factor receptors that respond not only to the growth message from the steroid hormone but also to the messages from other hormones, including prolactin, oxytocin, and growth hormone. Certainly, several of the cellular growth factors involved in normal breast growth and development have been implicated in the development of cancer. HER-2/neu, an epidermal growth factor, is just one of those factors known to be overexpressed in some aggressive breast cancers. It is unknown whether over- or underexposure to estrogen or progesterone, either endogenous or exogenous, directly contributes to malfunctions of the growth regulatory system; however, epidemiological and experimental evidence suggests that there may be a connection.

Breast disease, specifically breast cancer, is the most common form of cancer among women. The incidence of this cancer has been increasing recently and simultaneously with the increase in the general use of common hormonal medications, including estrogen and progesterone as either hormone replacement or contraception. Given the clear association between breast growth and the steroid hormones as well as the stimulatory molecular changes that occur at the cellular level, concern has arisen over the likelihood of malignant transformation in the breast as a result of exogenous hormone exposure. Several epidemiological studies have been conducted examining the relationship between hormone use and the development of breast cancer, often with conflicting results. Despite the numerous studies and basic science investigations, there still is no clear direct relationship between exogenous hormone exposure and the development of breast cancer.

## HRT AND BREAST CANCER

By far the largest, most expensive, and most important study ever conducted examining the relationship between hormone replacement and breast cancer is an ongoing study by the National Institutes of Health called the Women's Health Initiative (WHI). This study enrolled 161,809 healthy postmenopausal women (50–79 years of age) and examined the health benefits/risks of a number of interventions, including hormone replacement, in a randomized, blinded prospective fashion. After 5.2 years of follow-up, the portion of that trial following 16,608 enrolled women

with uteri was prematurely stopped and the results were reported. In that subtrial, half of the women were given daily conjugated equine estrogens (0.625 mg) plus medroxyprogesterone (2.5 mg), and half were given placebo. The trial was stopped prematurely as a direct result of the observed increased incidence of breast cancer seen in the women taking the hormone replacement. The risk for invasive breast cancer was 26% higher for the group taking the estrogen and progesterone. In other words, 38 of 10,000 women were observed to develop breast cancer rather than an expected 30 of 10,000. This increase was seen after 4 years of use and appeared to increase further with longer use of hormone replacement.

However, even with the increase in breast cancer, there was no observed difference in mortality between the two groups. Also interesting, the portion of the WHI trial that included women who had undergone hysterectomies and who were randomized to either estrogen without progesterone or placebo was not stopped, suggesting no increased risk of breast cancer in that group. This has led many to speculate that the contribution made to the increased incidence of breast cancer seen in this trial may be a consequence of the progesterone and not of the estrogen. There is epidemiological evidence that estrogen and progesterone affect the breast differently than does estrogen alone.

We can see the effect of progesterone on mammography. In one study, the Postmenopausal Estrogen Progestin Intervention (PEPI) trial evaluating cardiovascular effects of various HRT regimens, the researchers found that the three study groups of women taking different estrogen and progesterone preparations had significantly denser breasts on mammography than did the estrogen only and placebo groups. Several studies have reported that dense breasts on mammography are an independent risk factor for the future development of breast cancer. More abnormalities are found on mammography in patients on hormone replacement.

The WHI represents the only randomized controlled trial to examine and confirm that combined estrogen and progesterone increases the risk of invasive breast cancer in healthy postmenopausal women. However, the study is not the first to suggest a relationship between breast cancer and hormone replacement.

A 1997 study published by the Collaborative Group on Hormonal Factors in Breast Cancer analyzed 51 epidemiological studies reviewing hormone replacement information on 160,000 women, one-third of whom had breast cancer. In that meta-analysis, the research group found a 15% increased risk for breast cancer in women using estrogen plus progesterone for

shorter than 5 years and a 53% increase in those using the combination for longer than 5 years.

The Heart and Estrogen/Progestin Replacement Study (HERS) was conducted to rigorously examine the potential cardiovascular benefits seen in previous studies of women on hormone replacement. In a 2002 publication in *Lancet*, that group reported a 26% (not statistically significant) increase in the risk of breast cancer in women taking hormone replacement at 6.8 years follow-up.

Two separate Southern California studies, a University of California, Los Angeles/Los Angeles County case-control study and a retrospective breast cancer detection project, identified progesterone use as part of hormone replacement to be an independent risk factor for the development of breast cancer. In both studies, the relative risk for the development of breast cancer was 1.4 when progesterone was used with estrogen in HRT. Estrogen alone was found to contribute only a minimal increase in breast cancer risk. These findings seem to be consistent with the WHI's findings that progesterone, when used with estrogen, increases the risk of subsequent development of breast cancer.

More recently, androgen replacement in conjunction with estrogen has become increasingly popular for libido enhancement in postmenopausal women. The impact of exogenous testosterone use on the risk of breast cancer has not been critically evaluated. However, a few articles have reviewed serum sex steroid levels and risk of breast cancer. In a 2002 British study reviewing all of the previous studies examining serum sex steroid hormones and breast cancer risk, the researchers found that elevated levels of all the endogenous sex steroid hormones, including estrogens and androgens, were strongly associated with risk of breast cancer in postmenopausal women. This is in contrast to clinical evidence that androgens inhibit breast growth and proliferation. The physiological interplay among serum sex steroid hormone levels, binding globulins, and receptor status at the cellular level is complex and is the subject of tremendous investigation. A 2002 National Institutes of Health study suggested that the risk of breast cancer in postmenopausal women on hormone replacement might be explained by the suppression of androgens as a result of exogenous estrogen.

## HORMONAL CONTRACEPTION AND BREAST CANCER

Tens of millions of premenopausal women in the United States use hormonal contraception, most

commonly oral contraceptive pills (OCPs) that usually contain both estrogen and progesterone, to prevent unwanted pregnancy. "The pill" represents the most common form of birth control and the most effective for women since its introduction some 40 years ago. Since that time, given the known hormonal responsiveness of breast tissue, there have been concerns about the possible association between OCPs and risk of breast cancer. Over the years, the pill has undergone a progressive and continual modification of its formula. In general, the doses of estrogen and progesterone contained within the pill have been continually lowered so that women are now able to have the lowest effective dose. In addition, new formulations of estrogens and progestones have been introduced such that several different ones are available to be used in the modern OCP.

Unlike the WHI study, no prospective, randomized, placebo-controlled study has ever been conducted to answer the question of association between OCP use and risk of breast cancer. However, many studies have analyzed OCP use and breast cancer in other ways and found varying results. Because OCP formulations and strengths have changed over the years, a study published just 5 or 10 years ago analyzing patients who took the OCP 15 to 20 years ago possibly does not appropriately reflect current risk. The risk of an immediately measurable event, such as stroke or deep vein thrombosis (DVT), related to OCP use has dropped precipitously since the early years of the pill, so it is possible that a more long-term measurable event, such as risk of breast cancer, has also changed.

One of the more recent reports evaluating OCP use and risk of breast cancer was a population-based, case-control study published by the Centers for Disease Control and Prevention (CDC) enrolling more than 9000 women. In that study, the researchers found no increase in the risk of breast cancer in women who were current or former OCP users regardless of duration of use, doses of estrogen, duration of use, or age at initiation of use.

Another CDC study, published in 1983, also found no evidence of a connection between OCP use and risk of breast cancer in the Surveillance, Epidemiology, and End Results (SEER) case-control investigation of 2000 breast cancer cases. In that study, duration of use, age at initial use, family history, and personal history of benign breast disease did not increase the risk of breast cancer. A follow-up CDC study in 1986 found that the relative risk of breast cancer remained similar to that in women not taking OCPs despite different progesterone formulations.

However, not all studies dismiss the risk of breast cancer in OCP users. A Harvard University study examined the relationship between OCP use and breast cancer in 1799 women who developed breast cancer in a prospective cohort of 118,273 nurses. In that study, past users of OCPs (defined as more than 2 years of use prior to a breast cancer diagnosis) realized no increased risk for the development of breast cancer as a result of their OCP use. However, current users realized a relative risk of 1.5.

A 2000 study from the Netherlands following 12,000 women over 7.5 years found 309 cancer cases. When risks were analyzed for the women who developed breast cancer, those women who were over 55 years of age at the time of their cancer diagnosis and who had used OCPs for more than 10 years were twice as likely to have developed breast cancer.

Hormonal contraception can also be achieved by progesterone injection (depot-medroxyprogesterone acetate [DMPA]) or by subcutaneous implantation of progesterone pellets. In several reported studies examining thousands of women, no connection between this type of hormonal contraception and breast cancer could be established.

## CONCLUSION

Sex steroid hormones play a critical role in breast growth and development early in life as well as in breast physiology throughout the remainder of reproductive life. At the cellular level, sex steroid hormones directly affect the function of ductular–alveolar system of the breast. Multiple factors not reviewed in this article, including serum hormone levels, hormone-binding proteins, interaction between hormones, hormone receptor type, hormone receptor expression, and affected intracellular growth mechanisms, all act independently to influence the risk of breast cancer.

The concern regarding the impact of exogenous steroid hormones on breast disease is valid and is the subject of hundreds of studies, some of which were presented in this article. Mounting evidence suggests that hormone replacement increases the risk of breast cancer in women and that the increased risk may be due to the progesterone component of the replacement regimen. One significant limitation of these studies is that relatively few preparations of estrogen or progesterone have actually been examined. Mortality and type of breast cancer may also influence interpretations of these studies. There are also questions about serum levels, other hormone interactions, and specific cellular events that may contribute to that risk, further

clouding the interpretation of large epidemiological studies. It is difficult to draw conclusions about all hormone replacement given these limitations. Even so, given the current studies, there does appear to be an elevated risk for the development of breast cancer in the general population of women taking progesterone and estrogen as hormone replacement.

Similar to hormone replacement, hundreds of studies have been published examining the risk of breast cancer given exposure to oral contraceptives, and a definite risk has not been firmly established. Although most studies find no increased risk, there are some that do in fact point to an increased risk of breast cancer. However, on balance, and given our current understanding, OCPs appear to be safe.

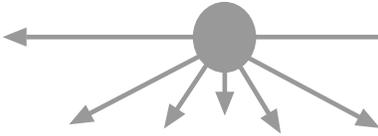
Our understanding of the impact of sex steroid hormones and the risk of breast disease continues to evolve. It is critical to weigh all of the studies, risks, and benefits and to be aware of the limitations of our current understanding of the complex interactions between these hormones and breast tissue.

## See Also the Following Articles

Cardiovascular Disease: Impact of Sex Steroid Replacement • Estrogen Replacement, Oral • Estrogen Replacement, Vaginal • Hormone Replacement, Transdermal • Hot Flash: Impact of Sex Steroid Replacement • Osteoporosis in Reproductive Age Women: Impact of Sex Steroid Replacement • Polycystic Ovary Syndrome: Implications for Cardiovascular, Endometrial, and Breast Disease

## Further Reading

- CASH Study Group. (1983). Long-term oral contraceptive use and the risk of breast cancer: The Centers for Disease Control Cancer and Steroid Hormone Study. *J. Am. Med. Assoc.* **249**, 1591–1595.
- Endogenous Hormones and Breast Cancer Collaborative Group. (2002). Endogenous sex hormones and breast cancer in postmenopausal women: Reanalysis of nine prospective studies. *J. Natl. Cancer Inst.* **94**, 606–616.
- Marchbanks, P. A., McDonald, J. A., Wilson, H. G., *et al.* (2002). Oral contraceptives and the risk of breast cancer. *N. Engl. J. Med.* **346**, 2025–2032.
- Romieu, I., Willett, W. C., Colditz, G. A., *et al.* (1989). Prospective study of oral contraceptive use and risk of breast cancer in women. *J. Natl. Cancer Inst.* **81**, 1313–1321.
- Van Houten, C., Burger, H., Peeters, P. H., *et al.* (2000). Long-term oral contraceptive use increases breast cancer risk in women over 55 years of age: The DOM cohort. *Intl. J. Cancer* **87**, 591–594.
- Writing Group for the Women's Health Initiative Investigators (2002). Risks and benefits of estrogen plus progesterone in healthy postmenopausal women: Principal results from the Women's Health Initiative randomized controlled trial. *J. Am. Med. Assoc.* **288**, 321–333.



# Calcitonin Gene-Related Peptide (CGRP)

María J. Moreno and Danica B. Stanimirovic  
National Research Council of Canada, Ottawa, Ontario, Canada

Edith Hamel  
Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada

## Glossary

### **calcitonin receptor-like receptor (CL receptor)**

G protein-coupled receptor (GPCR); pharmacology of this receptor is defined by its interaction with a group of accessory proteins called receptor activity-modifying proteins (RAMPs). CL receptor functions as a CGRP1 receptor when dimerized with RAMP1 and as an adrenomedullin (AM) receptor when dimerized with RAMP2 or RAMP3.

**CGRP receptor component protein (RCP)** Intracellular protein that facilitates signal transduction through the CGRP1 receptor.

**migraine** Complex neurological disorder characterized by severe head pain, nausea, and exacerbated sensitivity to external stimuli such as light and sound. It presents in one of two forms, migraine without aura or migraine with aura, with the latter characterized by visual disturbances such as zigzag lines and blind spots.

### **receptor activity-modifying proteins (RAMPs)**

A family of three single type I transmembrane proteins, RAMP1, RAMP2, and RAMP3, required for the transport, glycosylation, and pharmacological classification of the calcitonin receptor-like receptor (CL) as CGRP1 or adrenomedullin (AM) receptor.

**C**alcitonin gene-related peptide (CGRP) is a neuropeptide expressed in a variety of cells in the central and peripheral nervous systems. It exerts a broad range of biological effects such as vasodilation, increasing rate and force of cardiac contraction, regulation of calcium and glucose metabolism, reduction of gastrointestinal motility, and inflammation.

## INTRODUCTION

Calcitonin gene-related peptide (CGRP) is a product of alternative splicing of the calcitonin (CT) gene. Its dual occurrence in nerves and endocrine cells suggests

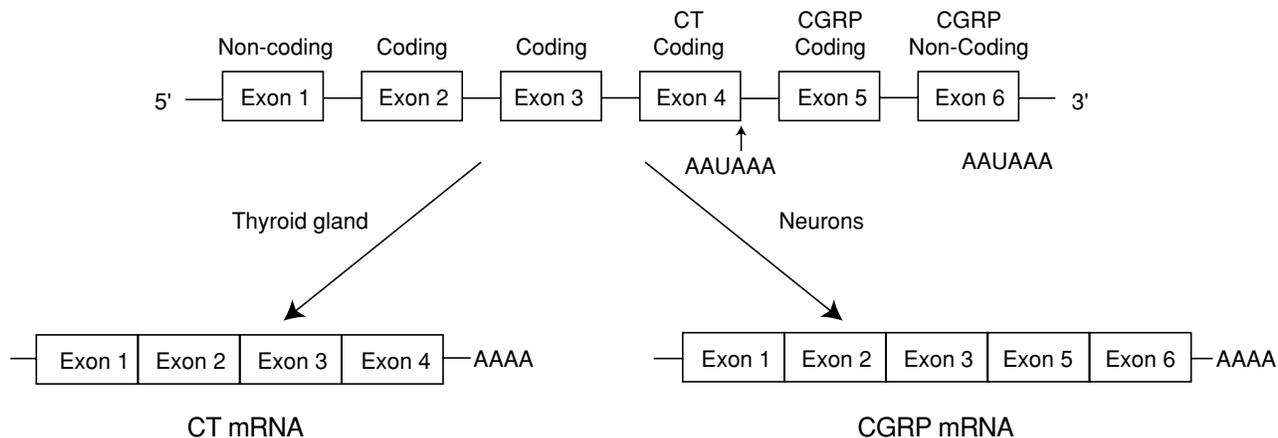
potential functions as a neuromodulator and an endocrine/paracrine hormone. CGRP biological activity is mediated by type II G protein-coupled receptors (GPCRs). Although pharmacological studies suggest the existence of multiple CGRP receptors, only one receptor has been cloned and named CL receptor. In the past few years, the classical concept of GPCRs has been challenged by the discovery of a group of accessory proteins, receptor activity-modifying proteins (RAMPs) and receptor component protein (RCP), that interact with CL receptor and modulate both its ligand-binding selectivity and its signal transduction. Structural and functional analysis of these multiple associations will provide new insights into molecular determinants of multiple CGRP functions and will help to design new drugs to regulate CGRP activity under pathological conditions.

## CT/CGRP GENE, ALTERNATIVE SPLICING, AND TRANSCRIPTIONAL REGULATION

CGRP is a 37-amino acid neuropeptide discovered by Rosenfeld and co-workers in 1983. After Rosenfeld's group observed that serially transplanted rat medullary thyroid carcinoma cells generated tumors in which CT biosynthesis was decreased more than 10-fold, attempted cloning of the CT gene in these tumors revealed a new, longer cytoplasmic mRNA, later identified as  $\alpha$ -CGRP mRNA.

The human genome contains at least three CT/CGRP genes, all of which localize to chromosome 11p. Alternative splicing of the gene and posterior processing of the prepro-peptides generate two different mature peptides: CT and CGRP.

The CT/ $\alpha$ -CGRP gene contains six exons and is processed in two different ways, depending on the tissue (Fig. 1). In parafollicular cells of the thyroid, 95% of the CT/CGRP pre-RNA is processed to include exon 4 followed by polyadenylation, which



**Figure 1** Alternative mRNA splicing of the CT/α-CGRP gene yielding both CT and α-CGRP in a tissue-specific manner. Boxes are exons; lines in between are introns.

leads to the production of the calcium-regulating hormone CT. In neurons, 99% of the pre-CT/CGRP includes exons, 1, 2, 3, 5, and 6 with a poly(A) tail at exon 6. This is translated into the neuropeptide α-CGRP. The key regulatory event in the processing choice is the tissue-specific inclusion or exclusion of the alternative 3'-terminal exon 4. This is modulated by factors that bind to an intronic enhancer element that contains both 3' and 5' splice site consensus sequence elements. Although factors involved in the tissue-specific alternative splicing of the CT/CGRP pre-RNA have not been identified, members of the serine- and arginine-rich RNA-binding protein family seem to participate in multiple interactions that coordinate individual recognition events during the early steps of RNA processing.

A second CGRP gene (β-CGRP) was predicted from analysis of cDNA in rat and human. The β-CGRP gene encodes a molecule closely related to α-CGRP. β-CGRP differs from α-CGRP in one amino acid in rat and in three amino acids in human. The lack of a poly A signal within exon 4 of the β-CGRP prevents alternative splicing; consequently, transcripts from this gene produce only β-CGRP and not CT.

The third CT/CGRP-like gene has also been identified on chromosome 11 and contains nucleotide sequences corresponding to exons 2 and 3 of CT/α-CGRP and β-CGRP genes. However, this gene does not transcribe and has been considered as a pseudogene.

Specific stimuli that modulate the expression of CT/CGRP genes in various cell types and some sequences in the gene promoter responsible for transcriptional regulation have been identified. For instance, thyroid C cells treated with either glucocorticoid hormone or retinoic acid show reduced

expression of the CT/CGRP gene, an effect associated with decreased activity of an enhancer element that normally increases promoter activity in thyroid C cells. In neurons, the CT/CGRP gene is activated by cyclic AMP (cAMP) and nerve growth factor (NGF). Both stimuli induce the phosphorylation of the cAMP response element (CRE)-binding protein (CREB) transcription factor on Ser-133. In the trigeminal ganglia, the regulation of CGRP gene expression has also been associated with MAPK signal transduction cascades. This signaling pathway may be of particular relevance to migraine because many inflammatory mediators implicated in this pathology are known activators of MAPK pathways.

Both α-CGRP and β-CGRP are synthesized as precursor prepro-peptides, consisting of 128 and 127 amino acids, respectively. The mature, C-terminally amidated α-CGRP and β-CGRP are produced by further cleavage of the leader sequence. Although α- and β-CGRP coexist in a large number of neurons, β-CGRP is predominant in the enteric nervous system and in the human pituitary gland, whereas α-CGRP is preferentially found in the trigeminal ganglia, the main source of cerebrovascular CGRP. Moreover, these two peptides activate different CGRP receptor subtypes and mediate distinct biological activities. For example, β-CGRP causes a selective suppression of gastric acid secretion, whereas α-CGRP stimulates renin and aldosterone secretion.

## CGRP RECEPTORS

### Pharmacology of CGRP Receptors

CGRP is a member of a peptide family that also includes CT, amylin (AMY), and adrenomedullin

(AM). Although these peptides have only limited sequence identity, they share a number of structural features, including a six- or seven-amino acid ring formed by a disulphide bond at the N termini (implicated in triggering signal transduction), an amphipathic  $\alpha$ -helix, and a C-terminal amide (responsible for the interaction of the molecule with the receptor). The structural convergence among these peptides results in a significant degree of cross-reactivity between receptors. This, in combination with the lack of fully selective agonists/antagonists, has hindered the accurate classification of CGRP receptors.

CGRP acts via type II GPCRs that primarily stimulate adenylate cyclase through the  $G_s$  subunit. The involvement of other G subunits in the activation of alternative pathways has also been described. Pharmacologically, the existence of at least two classes of CGRP receptors, named CGRP1 and CGRP2, has been proposed based on the differential potencies of CGRP peptide analogues in a variety of *in vitro* and *in vivo* bioassays.

#### **CGRP Receptor Antagonists**

Structure-activity characterization of various fragments of human  $\alpha$ -CGRP ( $h\alpha$ -CGRP) demonstrated the potent antagonist properties of C-terminal fragments such as CGRP<sub>8-37</sub>, CGRP<sub>9-37</sub>, and CGRP<sub>12-37</sub>. Shorter C-terminal fragments, including CGRP<sub>19-37</sub> and CGRP<sub>23-37</sub>, also displayed competitive antagonist properties but with much lower potency when compared with CGRP<sub>8-37</sub>. The ability of the peptide fragment CGRP<sub>8-37</sub> to antagonize  $\alpha$ -CGRP effects with higher ( $pA_2 \geq 6.9$ ) or lower ( $pA_2 \leq 6.6$ ) potency has been used as the main criterion for classifying CGRP receptors in CGRP1 or CGRP2 type, respectively. However, characterization of CGRP receptors based on the antagonist properties of CGRP<sub>8-37</sub> as the only parameter has proven to be limited. Indeed, CGRP<sub>8-37</sub> affinity at CGRP1 receptors varies considerably ( $pA_2$  values 6.9–9.3) depending on species, tissues, and regions within the same organ. CGRP<sub>8-37</sub> also shows higher affinity in radioligand-binding studies than in functional bioassays. It has been difficult to reconcile such a wide range of  $pA_2$  values accounting for a single CGRP receptor subtype. The differences in CGRP affinities have been attributed to various factors such as temperature modulation of the CGRP<sub>8-37</sub> affinity, differences in the distribution of proteolytic enzyme(s) that may affect CGRP<sub>8-37</sub> metabolism, and (most recently) the potential interaction of CGRP with different CGRP/non-CGRP receptor subtypes.

During the past few years, a significant effort has been invested in the identification of the CGRP primary structure and the significance of each single amino acid and its side chain in receptor interaction. These studies have led to a better understanding of the agonist and antagonist properties of multiple truncated CGRP analogues and to the development of smaller molecules with potent antagonist effect such as [Asp31, Pro34, Phe35]CGRP<sub>27-37</sub> and [Pro35, Phe35]CGRP<sub>27-37</sub>. One of the most relevant contributions in this field has been the recent development of the first highly potent and selective nonpeptide CGRP receptor antagonist, BIBN4096BS. This compound is more potent than CGRP<sub>8-37</sub>, exhibits more than 200-fold higher affinity at human CGRP receptors (neuroblastoma SK-N-MC cell line  $K_i = 14.4 \pm 6.3$  pM) than at rat CGRP receptors (spleen  $K_i = 3.4 \pm 0.5$  nM) and does not have significant affinity for other related receptors, including those for CT, AMY, and AM. BIBN4096BS has also become a useful pharmacological tool in identifying other potential CGRP receptors in various animal species such as rat, pig, and human.

#### **CGRP Receptor Agonists**

The use of agonists alone to define CGRP receptor subtypes has been of limited value unless they are combined with CGRP receptor antagonists. Three main CGRP forms, the natural peptide  $\beta$ -CGRP and the linear synthetic  $\alpha$ -CGRP analogues Cys [(acetaminomethyl)<sup>2,7</sup>]human- $\alpha$ CGRP ([Cys(ACM)<sup>2,7</sup>]h $\alpha$ CGRP) and Cys [(ethylamide)<sup>2,7</sup>]human- $\alpha$ CGRP ([Cys(Et)<sup>2,7</sup>]h $\alpha$ CGRP), have been identified as putative CGRP2 receptor agonists. Although  $\alpha$ -CGRP and  $\beta$ -CGRP induce responses with about equal potency in most tissues, the dilation induced by  $\beta$ -CGRP in guinea pig basilar artery, rat aorta, and pig coronary artery is less potently antagonized by h- $\alpha$ CGRP<sub>8-37</sub> than that induced by  $\alpha$ -CGRP. This observation suggests the presence of two CGRP receptors, an  $\alpha$ -CGRP-sensitive receptor and a  $\beta$ -CGRP-sensitive receptor, with the former being more potently blocked by CGRP<sub>8-37</sub>. On the other hand, [Cys(ACM)<sup>2,7</sup>]h $\alpha$ CGRP and [Cys(Et)<sup>2,7</sup>]h $\alpha$ CGRP have also been identified as selective CGRP2 receptor agonists based on their potent agonist effect ( $EC_{50}$  values  $\sim 76.0$  and  $\sim 3.4$  nM, respectively) in the rat vas deferens as compared with the weak inotropic effect ( $EC_{50}$  values  $> 700$  nM and 1  $\mu$ M, respectively) in the guinea pig atria. However, the CGRP2 receptor selectivity of [Cys(ACM)<sup>2,7</sup>]h $\alpha$ CGRP and [Cys(Et)<sup>2,7</sup>]h $\alpha$ CGRP has been challenged by more recent studies indicating

that both peptides are partial agonists with higher or lower activity largely dependent on the experimental conditions. Interestingly, [Cys (Et)<sup>2,7</sup>]hαCGRP seems to stimulate another unidentified class of receptors in rat vas deferens, human fetal astrocytes, and pial vessels.

The molecular identity of CGRP2 receptors has not been revealed to this point. Because of the disparity of responses and affinity values obtained in various cell types and experimental conditions, it is still not clear whether CGRP2 receptors are a homogenous group or whether they represent more than one pharmacologically/molecularly distinct receptor population.

## Molecular Identity of CGRP1 Receptors

### **RDC-1 versus CRLR**

The complexity of the CGRP receptor pharmacology appeared even more obscure after the cloning of two genes, RDC-1 and CRLR, reportedly encoding the same pharmacologically defined CGRP<sub>1</sub>-type receptor. Interestingly, RDC-1 and CRLR share only 10% of sequence identity. With the use of a polymerase chain reaction (PCR)-based approach, several members of the GPCR family were cloned from the thyroid gland. Among them, the canine orphan receptor gene RDC-1 was shown to encode a protein with specific responsiveness to rat α-CGRP in transfected COS-7 cells. However, the failure to reconstitute CGRP receptors by transfecting RDC-1 in other cell lines, as well as the mismatch in distribution between specific RDC-1 *in situ* hybridization signals and CGRP receptor-binding sites in the brain, have raised doubts concerning the real nature of this receptor. RDC-1 seems to be up-regulated during hypoxic conditions, in parallel with an increase in AM mRNA and protein levels. However, there is no conclusive evidence that RDC-1 represents an AM receptor. For this reason, RDC-1 is considered as an “orphan” receptor.

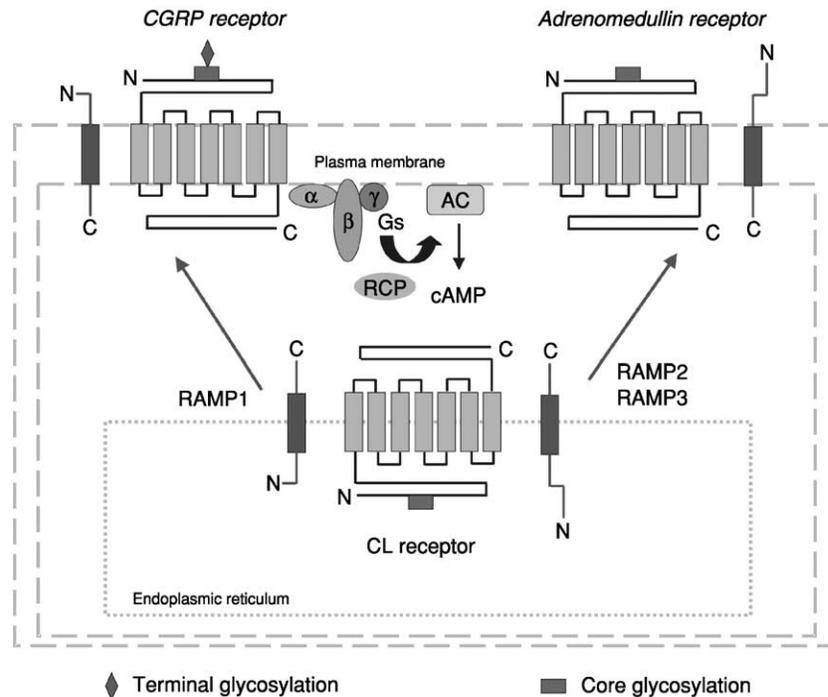
Only a few years after the cloning of RDC-1, another GPCR clone, the calcitonin receptor-like receptor (CRLR), was isolated from rat pulmonary blood vessels. In this article, CRLR is referred to as CL receptor, according to the International Union of Pharmacology (IUPHAR) guidelines. Although initial transfection of human or rat CL receptor into mammalian cells failed to elicit CGRP- or AM-induced responses, a stable expression of this gene in human embryonic kidney 239 (HEK 239) cells resulted in increased density of high-affinity <sup>125</sup>I-CGRP binding sites and stimulation of intracellular cAMP levels

by CGRP. However, the most conclusive proof that CL receptor can behave as a CGRP1 receptor was provided by MacLatchie and colleagues in 1998. The group discovered three accessory proteins called receptor activity-modifying proteins.

### **Receptor Activity-Modifying Proteins**

RAMPs are a family of single transmembrane domain proteins that are required for the transport, glycosylation, and ligand specificity of the CL receptor (Fig. 2). They have a relatively long extracellular amino terminus, a single membrane-spanning domain, and a short intracellular C terminus. Three members, RAMP1, RAMP2, and RAMP3, have been identified in this family by homology searches. They are 148, 175, and 148 amino acids long, respectively, with about 30% of sequence identity. RAMP1, although unable to bind CGRP on its own, can combine with CL receptor to form a CGRP1 fully glycosylated receptor complex of 66 kDa. Interestingly, the combination of CL receptor with RAMP2 or RAMP3 seemingly yields an AM receptor of 58 kDa (core glycosylated) that can cross-react with CGRP only at high concentrations.

The capacity of CGRP family members to interact with various receptors has raised the question as to whether AM-induced responses are mediated by CGRP or by specific AM receptors. Many effects of AM are blocked by α-CGRP<sub>8-37</sub> and not by AM<sub>22-52</sub>, a more selective AM receptor antagonist. In some cases, both α-CGRP<sub>8-37</sub> and AM<sub>22-52</sub> can partially inhibit CGRP- and AM-induced cAMP production. The first report documenting the interaction between CL receptor and RAMP1 in SK-N-MC cells showed that this complex selectively responds to CGRP but not to AM, suggesting that CL receptor/RAMP1 heterodimers define a very selective CGRP1 receptor. Later, it was shown that, although α-CGRP has far lower affinity than does AM for CL receptor/RAMP2 and CL receptor/RAMP3 complexes, the affinity of AM for CL receptor/RAMP1 is comparable to that of αCGRP in some biological systems. Furthermore, in cells expressing higher levels of CL receptor/RAMP1 than of CL receptor/RAMP2 or-RAMP3, the effects of AM were blocked by αCGRP<sub>8-37</sub> but not by AM<sub>22-52</sub>. Combined, these observations suggest that CL receptor/RAMP1 heterodimer defines a functional AM-sensitive receptor potentially antagonized by αCGRP<sub>8-37</sub> and distinct from CL receptor/RAMP2, which is more sensitive to AM<sub>22-52</sub>. This dual interaction of the CL receptor/RAMP1 complex with both αCGRP and AM may explain why many actions of AM are potentially antagonized by αCGRP<sub>8-37</sub>.



**Figure 2** Schematic representation of the CGRP<sub>1</sub>/AM receptor showing the interaction between a ligand-binding protein (CL receptor), accessory proteins (RAMP1, RAMP2, and RAMP3) for trafficking, glycosylation and pharmacology, and a receptor component protein (RCP) for coupling to cellular signal transduction pathways. AC, adenylyl cyclase. Figure modified from McLatchie *et al.* (1998) and Evans *et al.* (2000).

RAMPs act as a chaperon for CL receptor trafficking to the cell surface. When expressed individually, RAMPs and CL receptor are retained intracellularly, with CL receptor sequestered in the endoplasmic reticulum and RAMP1 sequestered in both the endoplasmic reticulum and the Golgi system, predominantly as a disulfide-linked homodimer. When coexpressed, both proteins can be detected on the cell surface, suggesting that the interaction between CL receptor and RAMP1 facilitates their trafficking to the plasma membrane.

RAMP1 is also required for the terminal glycosylation of the CL receptor. After cotransfection of HEK cells with RAMP1 and CL receptor, the molecular weight of CL receptor increases and it becomes resistant to endoglycosidase H. In contrast, RAMP2 does not change the molecular weight of the core glycosylated AM receptor. Differential glycosylation of CL receptor in the presence of RAMP1 or RAMP2 does not define receptor specificity for CGRP or AM. Moreover, core glycosylation is not sufficient for plasma membrane incorporation of CL receptor in the absence of RAMPs. Human CL receptor presents three asparagine (Asn) consensus glycosylation sites at positions 60, 112, and 117 of the amino acid sequence.

Selective mutations of these sites have demonstrated that Asn117 is important for direct interaction of CGRP or AM with CL receptor and RAMP1 or RAMP2 and, therefore, for the functionality of the receptor complex. However, Asn117 substitution does not affect N-glycosylation, association with RAMP1, and/or cell surface expression of receptor/RAMP1 heterodimers.

The extracellular N terminus of RAMP1 plays an important role in determining ligand specificity through association with the amino terminus of CL receptor that results in formation of a CGRP-binding pocket. The observation that CGRP receptor antagonist BIBN4096BS exhibits up to 100-fold higher affinity for the human CL receptor than for the rat CL receptor, despite more than 90% sequence homology of CL receptors between the two species, suggested that this selectivity is likely conferred by the interaction of BIBN4096BS with RAMP1 (homology ~ 71%). In support of this idea, Malle and colleagues subsequently demonstrated that tryptophan 74 in RAMP1 is responsible for the selective high-affinity binding of BIBN4096BS to the human CGRP receptor. Future structural analyses of CL receptor/RAMP complexes are needed to map residues important for

ligand recognition and to provide a basis for design of therapeutics with high affinity and selectivity for these receptors.

RAMPs can also compete in the regulation of CGRP/AM receptor function. For instance, coexpression of RAMP3 and RAMP1 results in reduced RAMP1-dependent CGRP receptor activity in a rabbit aortic endothelial cell line. On the other hand, the expression of RAMPs is also selectively regulated by different factors such as hypoxia and corticosterone. Therefore, balance of expression of various RAMPs, as well as their modulation by physiological or pathological conditions, may affect the affinity, function, and “molecular switch” between CGRP and AM receptors.

In summary, evidence gathered during recent years demonstrates the important contribution of RAMPs to various aspects of CGRP receptor biology: (1) trafficking of CL receptor to the cell surface, (2) glycosylation of CL receptor, (3) ligand specificity, and (4) receptor activity. Furthermore, the importance of RAMPs in receptor pharmacology also extends to other members of the CT/CGRP family. It has been shown that multiple AMY receptors arise from RAMP interaction with the CT receptor gene product. This indicates that the cellular profile of the CT/CGRP family of receptors, and potentially other class II GPCRs, can be dynamically regulated by the level and combination of these sets of proteins.

#### **CGRP Receptor Component Protein**

In addition to RAMPs, the CGRP receptor complex seems to require another protein for optimal function (Fig. 2). A protein cloned from the guinea pig organ of Corti and the cerebellum, known as the receptor component protein (RCP), has been found to confer CGRP responsiveness to oocytes. RCP is expressed in CGRP-responsive tissues, and its expression correlates with CGRP efficacy *in vivo*. RCP is a 146-amino acid protein that does not belong to any known class of proteins, including RAMPs. RCP does not contain any obvious protein motifs that may help to predict its function. RCP is a peripheral membrane protein attached to the membrane by ionic interactions. It seems to facilitate signal transduction by coupling the CGRP receptor to downstream effectors. CGRP binding and receptor density are not affected in RCP antisense-treated NIH3T3 cells, despite reduction in CGRP-mediated signal transduction, suggesting that RCP is not a chaperon protein for CL receptor. The nature of the interaction between RCP and CL receptor is unclear. It has been suggested that RCP may facilitate CL receptor activation, couple the receptor

to G proteins or other effector molecules, or coordinate the receptor–effector complex in the plasma membrane.

The requirement for the expression of at least three proteins (CL receptor, RCP, and RAMPs) to form a functional CGRP receptor may explain the difficulty in determining the molecular identity of CGRP receptors. Structural analysis and antisense studies will allow determination of sites of interaction and assignment of functions to the emerging subset of accessory proteins that bind to and regulate the activity of GPCRs.

#### **SIGNAL TRANSDUCTION**

The binding of CGRP to its receptors activates multiple signaling pathways. The main signaling event triggered by CGRP is stimulation of adenylyl cyclase. However, CGRP has also been reported to activate guanylate cyclase and phospholipase C (PLC) and to increase intracellular calcium concentrations (Fig. 3).

Activation of CGRP receptors stimulates adenylyl cyclase via  $G_{\alpha_s}$ . This induces an increase in intracellular cAMP in a concentration-dependent and saturable manner that, in some tissues, can be more pronounced (two- to threefold) than that produced by forskolin. An increase in cAMP activates protein kinase-A (PKA), which then triggers multiple changes in the downstream signaling pathways. In some cases, the CGRP-induced increase in cAMP is followed by a membrane hyperpolarization associated with the activation of an adenosine triphosphate (ATP)-sensitive potassium channel. In blood vessels, CGRP can increase nitric oxide (NO) production via two different pathways: (1) stimulation of adenylyl cyclase followed by activation of PKA and subsequent phosphorylation of eNOS by PKB through a phosphatidylinositol 3-kinase (PI3K)-mediated mechanism or (2) induction of calcium influx through voltage-gated calcium channels, an effect mediated by a cAMP-dependent PKA. Increased intracellular calcium may stimulate eNOS and nNOS, leading to the generation of NO.

In bone cells, activation of CGRP receptors does not stimulate adenylyl cyclase but instead activates PLC- $\beta_1$  via pertussis toxin (PTX)-insensitive  $G_{\alpha_q/11}$  protein. This induces the formation of diacylglycerol, a direct activator of PKC, and inositol 1,4,5-triphosphate (IP3), which binds to receptors on the endoplasmic reticulum and causes a transient release of calcium.

There is also solid evidence that CGRP, at high concentrations, can signal through  $G_i$  coupled to either calcium or potassium channels in various tissues.



nociceptive, and acoustic information during aversive conditioning. Moreover, CGRP evokes fear-like behavior in rats not trained with aversive stimuli.

A potential role for CGRP in neurological disorders, such as dementia and depression, has been supported by anatomical, biochemical, and clinical data. CGRP receptors are often localized in dopamine (DA)-containing neurons. When administered directly into the brain, CGRP markedly affects DA release and metabolism in selected brain regions as well as learning and memory. On the other hand, both psychomimetics and antipsychotic drugs influence regional brain concentrations and release of CGRP *in vivo*. Taken together, these findings support involvement of CGRP in DA-related disorders. In demented patients, cerebrospinal fluid (CSF) levels of both CGRP-LI and CT-LI are lower, but their concentration ratio is similar to that of age-matched healthy controls. This may be explained by a general neuronal loss or, alternatively, by a down-regulation of the CT/ $\alpha$ -CGRP gene expression. In depressed patients, only CT-LI appears to be diminished, resulting in an increase of the  $\alpha$ -CGRP/CT ratio. This most likely reflects an altered splicing mechanism favoring the formation of  $\alpha$ -CGRP mRNA. Although more studies are needed to draw definitive conclusions, combined measurements of CGRP-LI and CT-LI in CSF may be of diagnostic and perhaps prognostic value in dementia and affective CNS disorders, including depression.

### CGRP in the Cardiovascular System

CGRP and its receptors are expressed in the cardiovascular system. CGRP participates in various functions, including modulation of vascular tone, microvascular permeability, inotropy, chronotropy, cell proliferation, and inflammation.

In healthy volunteers, CGRP plasma levels vary between 2 and 35 pmol/L in a gender-independent manner. Neuronal tissues, especially peripheral nerve terminals, and the thyroid gland contribute to the influx of CGRP into the circulation. CGRP is one of the most potent vasodilating substances known and also exhibits positive chronotropic and inotropic effects. However, the physiological significance of circulating CGRP remains uncertain. Various studies have supported the role of CGRP in the regulation of peripheral vascular tone and regional organ blood flow in physiological conditions through either endothelium-dependent or endothelium-independent mechanisms. Moreover, CGRP is likely involved in

blood pressure regulation given that CT/ $\alpha$ CGRP gene knockout mice display a significantly higher mean blood pressure than do wild-type mice.

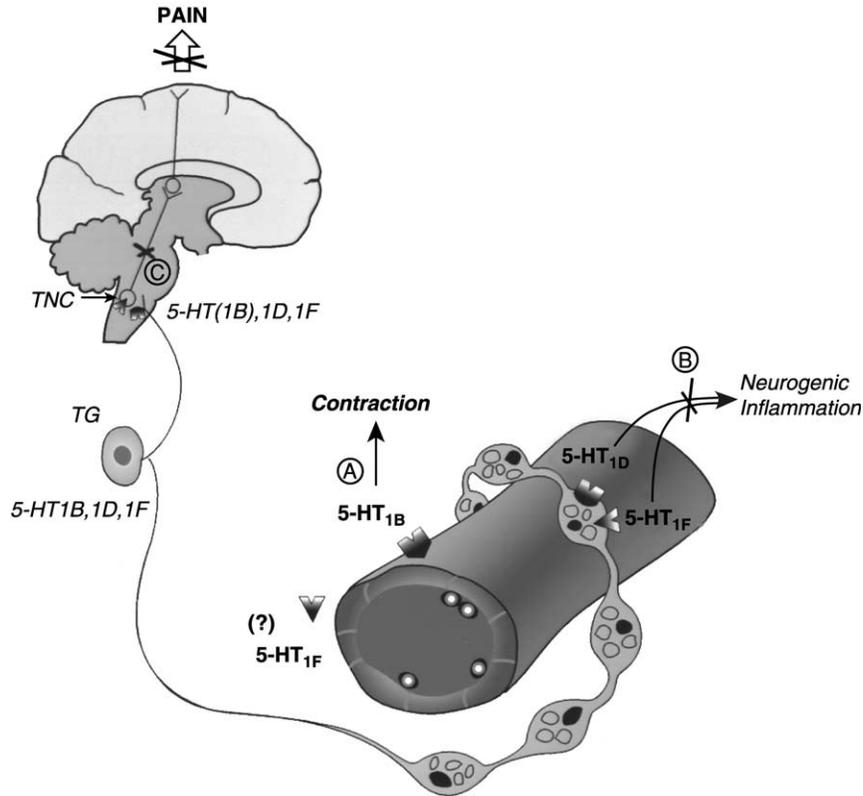
Reduced levels of CGRP have been associated with various vascular pathologies, including hypertension and Reynaud's syndrome. CGRP agonists have shown efficacy in reversing vascular dysfunction in these diseases. Increased release of CGRP is also causally linked to other cardiovascular pathologies, including migraine, inflammation, and cardiogenic shock associated with sepsis. In some cases, increases in CGRP levels are a compensatory mechanism to counteract pathological vasoconstriction or increased plasma volume such as that observed in myocardial ischemia or pregnancy, respectively.

### Role of CGRP in Vascular Disorders Associated with Aging

Wimalawansa and co-workers demonstrated age- and circadian rhythm-related changes of CGRP tissue content. With aging, CGRP levels decrease in neural cells and increase in both the thyroid gland and circulation. The relevance of these changes for deteriorating hemodynamics in the elderly is unclear. CGRP content also decreases in major arteries, including carotid, cerebral, and coronary arteries, a phenomenon likely linked to vasomotor dysfunction and higher risk of ischemic cerebrovascular and cardiovascular events in the elderly. The dorsal root ganglia contain cell bodies of primary afferent neurons that extend CGRP-containing nerves to peripheral blood vessels and the central spinal cord. An age-dependent reduction in CGRP nerve function, including decreases in CGRP mRNA expression in the dorsal root ganglia, neurogenic CGRP release, and CGRPergic nerve-mediated vasodilation, have been observed in spontaneously hypertensive rats. Although various circulating CGRP levels have been reported in essential hypertension in humans, the beneficial effect of calcium intake on lowering the high blood pressure in these patients seems to be associated with a calcium-induced CGRP release into the circulation and the resulting vasodilation induced by the peptide.

### Role of CGRP in Migraine Headache

The implication of CGRP in migraine headache is supported by clinical and experimental evidence. Although the initial triggers of this painful and debilitating condition and the factors involved in susceptibility of migraine sufferers to certain stimuli are not well defined, activation of the trigeminovascular system has been shown to play an important role in



**Figure 4** Schematic representation of the putative sites of action for triptans in the symptomatic treatment of migraine headache. (A) Constriction of distended intracranial blood vessels via activation of 5-HT<sub>1B</sub> receptors in the smooth muscle layer. (B) Inhibition of CGRP release from perivascular trigeminal afferents through the activation of 5-HT<sub>1D</sub> and possibly 5-HT<sub>1F</sub> receptors. (C) Inhibition of nociceptive signaling in trigeminal nucleus caudalis (TNC) via 5-HT<sub>(1B),1D,1F</sub>. TG, trigeminal ganglion. Reprinted, with permission, from Hamel, E. (2000). Serotonin receptors: Relevance to migraine pathogenesis and treatment. In "The Headaches" (J. Olesen, P. Tfelt-Hansen, and K. M. A. Welch, eds.), 2nd ed., pp. 133–139. Lippincott Williams & Wilkins, Philadelphia.

the genesis and maintenance of migraine pain. The trigeminal ganglion consists of bipolar neurons that send fibers to the meninges, to the meningeal (dural) and cerebral (pial) blood vessels involved in the manifestation of head pain, and (centrally) to the caudal brainstem and upper cervical spinal cord. Trigeminal ganglion neurons contain the neuropeptides CGRP, substance P, and neurokinin A. Moskowitz and colleagues documented that abnormally intense neuro-metabolic brain activity is an important activator of trigeminal nerve fibers and trigger of headache in migraine with aura. Activation of the trigeminovascular system causes peripheral release of neuropeptides that induce meningeal blood vessel dilation, blood flow increase, and plasma protein extravasation, in part by central reflex activation of the perivascular parasympathetic system via the superior salivatory nucleus. This inflammatory process is also thought to cause sensitization of nerve fibers and, consequently, lowering of the nociceptive threshold.

CGRP levels increase in the cranial circulation of migraine sufferers during an attack and return to normal values after administration of the antimigraine drug, sumatriptan, concomitant with relief of migraine pain. The efficacy of sumatriptan and other triptans to relieve migraine pain has been attributed to actions at three different levels of the trigeminovascular system (Fig. 4): (a) constriction of distended intracranial blood vessels via activation of 5-hydroxytryptamine (5-HT)<sub>1B</sub> receptors in the smooth muscle layer, (b) inhibition of CGRP release from perivascular trigeminal nerves through the activation of 5-HT<sub>1B,1D</sub> and possibly 5-HT<sub>1F</sub> receptors, and (c) inhibition of nociceptive signaling in trigeminal nucleus caudalis via 5-HT<sub>1D,1F</sub> receptors.

The use of CGRP receptor antagonists has emerged as a potential strategy to regulate CGRP activity during migraine attacks. The nonpeptide CGRP receptor antagonist BIBN4096BS was found to inhibit neurogenic vasodilation induced by

trigeminal nerve stimulation in marmoset monkeys without affecting basal blood pressure or heart rate. This represents an interesting advantage over the application of sumatriptan, which is contraindicated in patients with cardiovascular symptoms. BIBN4096BS has proven to be efficacious in clinical trials and is undergoing more advanced clinical evaluation as a putative therapeutic for treatment of acute migraine headache.

## CGRP IN THE GASTROINTESTINAL SYSTEM

CGRP-containing neuronal afferents form a major component of the sensory innervation in the gastrointestinal tract, including the esophagus, stomach, hepatobiliary tract, pancreas, vasculature, and a portion of nonvascular fibers distributed to the intestinal wall. CGRP nerve fibers arise mainly from dorsal root ganglia, where both  $\alpha$ -CGRP and  $\beta$ -CGRP mRNAs are coexpressed. However, capsaicin treatment was able to completely deplete the rat intestine of  $\alpha$ -CGRP but not of  $\beta$ -CGRP. This observation suggested that distinct neuronal populations impinging on the enteric system express these two peptides in a preferential manner,  $\alpha$ -CGRP by capsaicin-sensitive sensory fibers and  $\beta$ -CGRP by enteric autonomic neurons.

In various animal species, peripheral or central administration of CGRP inhibits gastric acid secretion and motility and increases gastric mucosal blood flow. The central inhibition of acid secretion is due primarily to suppression of the vagal efferent activity, whereas the antisecretory effect of peripherally administered CGRP involves release of gastric somatostatin and inhibition of cholinergic transmission. The central inhibitory effect of CGRP on gastric motor functions is due to increased sympathetic outflow given that adrenalectomy or celiac ganglionectomy abolishes this effect. Peripheral CGRP injections reduce gastric emptying and motility by increasing intracellular cAMP via smooth muscle receptors. CGRP also raises gastric blood flow, an effect that correlates with the capacity of CGRP to prevent ethanol-induced lesions of the gastric mucosa. Therefore, CGRP may have an important protective action against gastric ulcer.

## CONCLUDING REMARKS

There is sufficient evidence to highlight the potential therapeutic benefit of drugs modulating abnormal

CGRP functions in pathological conditions. However, CGRP itself is not suitable for long-term clinical application because it is rapidly metabolized in plasma. Developing synthetic, nonpeptide CGRP receptor agonists and antagonists will open new perspectives for clinical manipulation of CGRP receptor functions. The novel compound BIBN4096BS exemplifies the feasibility of this approach; this highly potent and selective nonpeptide CGRP antagonist is in an advanced stage of clinical trials for the treatment of migraine. In the past few years, considerable progress has been made in understanding the complexity of CGRP receptor structure and pharmacology. The biggest challenge in the future will be to map interactions among different sets of constitutive proteins involved in the formation of functional CGRP receptors. This information will be of great value for a rational drug design aimed at producing specific and potent modulators of CGRP activity.

## See Also the Following Articles

Calcitonin, Overview

## Further Reading

- Aiyar, N., Rand, K., Elshourbagy, N. A., Zeng, Z., Adamou, J. E., Bergsma, D., and Li, Y. (1996). A cDNA encoding the calcitonin gene-related peptide type 1 receptor. *J. Biol. Chem.* **271**, 11325–11329.
- Bolay, H., Reuter, U., Dunn, A. K., Huang, Z., Boas, D. A., and Moskowitz, M. A. (2002). Intrinsic brain activity triggers trigeminal meningeal afferents in a migraine model. *Nat. Med.* **8**(2), 136–142.
- Drissi, H., Lasmole, F., LeMellay, V., Marie, P. J., and Lieberherr, M. (1998). Activation of phospholipase C-beta1 via G $\alpha$ q/11 during calcium mobilization by calcitonin gene-related peptide. *J. Biol. Chem.* **273**, 20168–20174.
- Durham, P., and Russo, A. (2002). New insights into the molecular actions of serotonergic antimigraine drugs. *Pharmacol. Ther.* **94**(1–2), 77.
- Evans, B. N., Rosenblatt, M. I., Mnayer, L. O., Oliver, K. R., and Dickerson, I. M. (2000). CGRP-RCP, a novel protein required for signal transduction at calcitonin gene-related peptide and adrenomedullin receptors. *J. Biol. Chem.* **275**, 31438–31443.
- Freeland, K., Liu, Y. Z., and Latchman, D. S. (2000). Distinct signalling pathways mediate the cAMP response element (CRE)-dependent activation of the calcitonin gene-related peptide gene promoter by cAMP and nerve growth factor. *Biochem. J.* **345**, 233–238.
- Gangula, P. R., Zhao, H., Supowit, S. C., Wimalawansa, S. J., Dipette, D. J., Westlund, K. N., Gagel, R. F., and Yallampalli, C. (2000). Increased blood pressure in alpha calcitonin gene-related peptide/calcitonin gene knockout mice. *Hypertension* **35**, 470–475.
- Gujer, R., Aldecoa, A., Bühlmann, N., Leuthäuser, K., Muff, R., Fischer, J. A., and Born, W. (2001). Mutations of the

- asparagine<sup>117</sup> residue of a receptor activity-modifying protein 1 (RAMP1)-dependent human calcitonin gene-related peptide (CGRP) receptor result in selective loss of function. *Biochemistry* **40**, 5392–5398.
- Hamel, E. (2000). Serotonin receptors: Relevance to migraine pathogenesis and treatment. In "The Headaches" (J. Olesen, P. Tfelt-Hansen, and K. M. A. Welch, eds.), 2nd ed., pp. 133–139. Lippincott Williams & Wilkins, Philadelphia.
- Juaneda, C., Dumont, Y., and Quirion, R. (2000). The molecular pharmacology of CGRP and related peptide receptor subtypes. *Trends Pharmacol. Sci.* **21**, 432–438.
- Kapas, S., and Clark, J. L. (1995). Identification of an orphan receptor gene as a type I calcitonin gene-related peptide receptor. *Biochem. Biophys. Res. Commun.* **217**, 832–838.
- Luebke, A. E., Dahl, G. P., Roos, B. A., and Dickerson, I. M. (1996). Identification of a protein that confers calcitonin gene-related peptide responsiveness to oocytes by using a cystic fibrosis transmembrane conductance regulator assay. *Proc. Natl. Acad. Sci. USA* **93**, 3455–3460.
- Mallee, J. J., Salvatore, C. A., LeBourdelle, B., Oliver, K. R., Longmore, J., Koblan, K. S., and Kane, S. A. (2002). Receptor activity-modifying protein 1 determines the species selectivity of non-peptide CGRP receptor antagonists. *J. Biol. Chem.* **277**, 14294–14298.
- McLatchie, L. M., Fraser, N. J., Main, M. J., Wise, A., Brown, J., Thompson, N., Solari, R., Lee, M. G., and Foord, S. M. (1998). RAMPs regulate the transport and ligand specificity of the calcitonin receptor-like receptor. *Nature* **393**, 333–339.
- Moreno, M. J., Abounader, R., Hebert, E., Doods, H., and Hamel, E. (2002). Efficacy of the non-peptide CGRP receptor antagonist BIBN4096BS in blocking CGRP-induced dilations in human and bovine cerebral arteries: Potential implications in acute migraine treatment. *Neuropharmacology* **42**, 568–576.
- Moreno, M. J., Terron, J. A., Stanimirovic, D. B., Doods, H., and Hamel, E. (2002). Characterization of calcitonin gene-related peptide (CGRP) receptors and their receptor-activity-modifying proteins (RAMPs) in human brain microvascular and astroglial cells in culture. *Neuropharmacology* **42**, 270–280.
- Nagoshi, Y., Kuwasako, K., Ito, K., Uemura, T., Kato, J., Kitamura, K., and Eto, T. (2002). The calcitonin receptor-like receptor/receptor activity-modifying protein 1 heterodimer can function as a calcitonin gene-related peptide-(8-37)-sensitive adrenomedullin receptor. *Eur. J. Pharmacol.* **450**, 237–243.
- Poyner, D. R., Sexton, P. M., Marshall, I., Smith, D. M., Quirion, R., Born, W., Muff, R., Fischer, J. A., and Foord, S. M. (2002). International Union of Pharmacology: XXXII. The mammalian calcitonin gene-related peptides, adrenomedullin, amylin, and calcitonin receptors. *Pharmacol. Rev.* **54**, 233–246.
- Rosenfeld, M. G., Mermod, J.-J., Amara, S. G., Swanson, L. W., Sawchenko, P. E., Rivier, J., Vale, W. V., and Evans, R. M. (1983). Production of a novel neuropeptide encoded by the calcitonin gene via tissue-specific RNA processing. *Nature* **304**, 129–135.
- Tache, Y., Raybould, H., and Wei, J. Y. (1991). Central and peripheral actions of calcitonin gene-related peptide on gastric secretory and motor function. *Adv. Exp. Med. Biol.* **298**, 183–198.
- van Rossum, D., Hanisch, U. K., and Quirion, R. (1997). Neuro-anatomical localization, pharmacological characterization, and functions of CGRP, related peptides, and their receptors. *Neurosci. Biobehav. Rev.* **21**, 649–678.
- Wimalawansa, S. J. (1996). Calcitonin gene-related peptide and its receptors: Molecular genetics, physiology, pathophysiology, and therapeutic potentials. *Endocrinol. Rev.* **17**, 533–585.
- Zhang, X., and Hintze, T. H. (2001). cAMP signal transduction cascade: A novel pathway for the regulation of endothelial nitric oxide production in coronary blood vessels. *Arterioscler. Thromb. Vasc. Biol.* **21**, 797–803.



# Calcitonin, Overview

Sunil J. Wimalawansa

Robert Wood Johnson Medical School, New Brunswick, New Jersey, United States

## Glossary

**bone mineral density (BMD)** A reliable estimate for assessment of risk factor for future fractures; BMD is measured using dual-energy X-ray absorptiometry (DXA) and is expressed as the amount of mineralized tissue in the area scanned.

**calcitonin (CT)** A 32-amino acid polypeptide hormone produced in C cells of the thyroid gland; its main action is to inhibit osteoclast-induced bone resorption.

**calcitonin gene-related peptide (CGRP)** The most potent endogenous vasodilator.

**osteoporosis** A chronic progressive disease characterized by low bone mass and microarchitectural deterioration of bone tissue, leading to bone fragility and increase in fracture risk.

Calcitonin (CT) is a 32-amino acid peptide hormone whose existence was postulated in 1962 by Copp and colleagues. CT is a hormone produced by the C cells in the thyroid gland. Together with parathyroid hormone, it helps to regulate serum calcium and calcium in bone. Synthetic preparations of CT are available for treatment of various metabolic bone disorders (e.g., osteoporosis) and localized bone diseases (e.g., Paget's disease of bone). CT may be administered via nasal spray or injection. Nasal administration is preferred due to ease of use and because the injections tend to be associated with more nausea and flushing. CT is a natural hormone, and there is no evidence of carcinogenesis in animals or humans.

## INTRODUCTION

Calcitonin (CT) is synthesized by parafollicular cells (C cells) in the thyroid in mammals and by C cells associated with the ultimobranchial gland in lower vertebrate animals. C cells derive from the neural crest and migrate forward to become the ultimobranchial body in lower vertebrates and the parafollicular cells in humans. Therefore, during migration, C cells may concentrate in regions other than the thyroid and ultimobranchial body.

## STRUCTURE OF CT AND ITS MEASUREMENTS

CT has been relatively conserved during evolution. CT from nine different species has been identified, generating 11 known sequences (Table I). Six of the invariant amino acid residues are clustered at the amino terminal, and two are at the carboxy-terminal end of the molecule. Furthermore, all CT has a 1–7 disulfide bridge and prolin amide at the C terminus. All 32 amino acids are required for its hypocalcemic activity and its inhibitory effects on osteoclasts. Substitution of some of these amino acids, as with some of the synthetic CT preparations (e.g., synthetic eel CT), enhance its half-life and thereby its biological activity but may enhance allergenicity.

CT is produced as a precursor molecule; a number of posttranslational modifications, including cleavage and C-terminal amidation, will occur prior to secretion of the mature form of biologically active CT<sub>(1–32)</sub> into circulation. The release of CT from C cells is stimulated by cations Ca<sup>2+</sup> and Mg<sup>2+</sup> and also by glucagons, dibutyryl cyclic AMP (cAMP), theophyllin, gastrin, and cholecystokinin. CT causes a dose-dependent elevation of the cAMP level, and the effects of CT on osteoclasts can be mimicked by dibutyryl cAMP. Measurement of plasma immunoreactive calcitonin (i-CT) levels (for hypersecretion) has been used as a screening and diagnostic test for premalignant and malignant C-cell disease (i.e., families of patients with medullary thyroid carcinoma [MTC]). There are a number of diagnostic stimulation tests available for detecting MTC, including infusion of calcium and partagastrin. Table II illustrates some causes of elevated and suppressed serum i-CT levels.

## COMMERCIALY AVAILABLE CT PREPARATIONS

Both parental and nasally administered forms of CT are available. Miacalcin nasal spray (Novartis Co.) is

Table I Amino Acid Sequences of the Nine Fully Characterized and Two \*Predicted Calcitonins

Man	Rat	S-1	S-2	S-3	Eel	*Chick	Bov	Porc	Ovi	*Man
H										
1. Cys	—	—	—	—	—	—	—	—	—	Tyr
2. Gly	—	Ser	Ser	Ser	Ser	Ala	Ser	Ser	Ser	Ser
3. Asn	—	—	—	—	—	Ser	—	—	—	—
4. Leu	—	—	—	—	—	—	—	—	—	—
5. Ser	—	—	—	—	—	—	—	—	—	—
6. Thr	—	—	—	—	—	—	—	—	—	—
7. Cys	—	—	—	—	—	—	—	—	—	—
8. Met	—	Val	—	Val	Val	Val	Val	Val	Val	Leu
9. Leu	—	—	—	—	—	—	—	—	—	Gln
10. Gly	—	—	—	—	—	—	Ser	Ser	Ser	—
11. Thr	—	Lys	Lys	Lys	Lys	Lys	Ala	Ala	Ala	—
12. Tyr	—	Leu	Leu	Leu	Leu	Leu	—	—	—	—
13. Thr	—	Ser	Ser	Ser	Ser	Ser	Trp	Trp	Trp	Leu
14. Gln	—	—	—	—	—	—	Lys	Arg	Lys	—
15. Asp	—	Glu	—	—	Glu	Glu	—	Asn	—	Tyr
16. Phe	Leu	Leu	Leu	Leu	Leu	Leu	Leu	Leu	Leu	Leu
17. Asn	—	His	His	His	His	His	—	—	—	Lys
18. Lys	—	—	—	—	—	—	Asn	Asn	Asn	Asn
19. Phe	—	Leu	Leu	Leu	Leu	Leu	Tyr	—	Tyr	—
20. His	—	Gln	Gln	Gln	Gln	Gln	—	—	—	—
21. Thr	—	—	—	—	—	—	Arg	Arg	Arg	Met
22. Phe	—	Tyr	—	—	Tyr	Tyr	—	—	Tyr	—
23. Pro	—	—	—	—	—	—	Ser	Ser	Ser	—
24. Gln	—	Arg	Arg	Arg	Arg	Arg	Gly	Gly	Gly	Gly
25. Thr	—	—	—	—	—	—	Met	Met	Met	Ile
26. Ala	Ser	Asn	Asn	Asn	Asp	Asp	Gly	Gly	Gly	Asn
27. Ile	—	Thr	Thr	Thr	Val	Val	Phe	Phe	Phe	Phe
28. Gly	—	—	—	—	—	—	—	—	—	—
29. Val	—	Ser	Ala	Ala	Ala	Ala	Pro	Pro	Pro	Pro
30. Gly	—	—	—	—	—	Glu	Glu	Glu	Glu	Glu
31. Ala	—	Thr	Val	Val	Thr	Thr	Thr	Thr	Thr	Ile
32. Pro	—	—	—	—	—	—	—	—	—	—
NH <sub>2</sub>										

Note. Variations of amino acid sequence from human CT are indicated using three-letter amino acid symbol. The invariant residues are clustered at the two ends of the molecule. S, salmon; Bov, bovine; Porc, porcine; Ovi, ovine; Man, predicted.

a synthetic salmon CT that is 10- to 15-fold more potent than human CT. Its major action is the inhibition of bone resorption by a direct action on osteoclasts. In females more than 5 years postmenopause, 200 IU of miacalcin nasal spray once daily is the recommended dose for the treatment of osteoporosis. However, CT should be reserved for those patients who refuse or cannot tolerate other anti-osteoporosis agents or for patients with acute bone pain. Use of CT

nasal spray is recommended in conjunction with a total daily intake of at least 1200 mg of calcium and 400 IU of vitamin D to retard the progressive loss of bone mass.

In some countries, the only available form of CT still is the injectable preparation. Using injectable CT formulations, a dose of 40 to 50 IU three or four times per week is recommended for prevention or treatment of osteoporosis. Small doses of injectable CT such as

Table II Conditions Associated with Calcitonin Underproduction and Overproduction

Conditions associated with calcitonin underproduction	Conditions associated with calcitonin overproduction
<b>Osteoporosis</b>	<b>Calcitonin-secreting tumor</b>
Postmenopausal	Medullary carcinoma of the thyroid
Pregnancy-induced osteoporosis	C-cell hyperplasia
Senile	Other hormone-secreting tumors
Secondary to endocrine disorders	Hypercalcemia
	Neonatal hypocalcemia
	Pseudohypoparathyroidism
	Renal disorders
	Pancreatitis
	Heroin addict
	Graves' disease
	Atrophic gastritis
	Acute gastritis
	Pernicious anemia
	Peptic ulcer
	Gastrointestinal bleeding
	Stress, trauma
	Thyroid surgery
	Hepatic surgery
	Toxic shock
	Myocardial infarction
	Lithium intoxication

10–40 IU three times per week may have the same beneficial effect on bone, may delay onset of receptor-mediated resistance with far fewer side effects, and are more economical. Salmon CT has fewer side effects, but the main drawback with CT is its high cost and a decrease in efficacy secondary to antibody formation.

From a therapeutic point of view, CT can be administered by intravenous, intramuscular, subcutaneous, or an intranasal route. Side effects are greatest with intravenous administration, less with intramuscular and subcutaneous injections, and least with intranasal administration. The recommended dose of intranasal CT is 200 IU per day, and its biological effects are similar to those observed after 50 IU by subcutaneous injection. All forms of CT should be stored at 4°C and protected from light. Shelf life is approximately 2 years. Nonhuman CTs can be allergenic, with roughly 50% of patients developing antibodies over time. However, this leads to actual clinical resistance (e.g., neutralizing antibodies) only in less than a third of patients.

## ADVERSE REACTIONS

### Potentially Life-Threatening Effects

There are no reports of any deaths or long-term side effects attributable to CT use over 30 years of clinical experience. CT is a safe drug with few side effects. The main drawback of CT therapy is its cost, nasal irritation, and only a modest effect on the bone mineral density (BMD) and fracture reduction. In addition, the absorption of nasally administered CT is variable and poor, and a consequent, specific bioavailability is not guaranteed. Therefore, in some patients, the dose may need to be increased by two- to threefold the recommended dose to achieve the desired effects. Alternative routes of administration are under investigation (e.g., pulmonary, rectal, buccal [liposomes], depo-preparations, and oral), but no breakthrough has been yet reported.

### Acute Overdose

Overdose has not been reported. Transient nausea and vomiting is the only potentially hazardous symptom.

### Severe or Irreversible Adverse Effects

Anaphylactic reactions have been reported only rarely in association with CT (nonhuman CT) therapy.

### Symptomatic Adverse Effects

CT is generally well tolerated. With the injectable CT, the most commonly encountered adverse effect is cutaneous flushing, particularly of the face. This is noticed by at least a third of patients and probably occurs to some extent in the majority. The onset may be within minutes of the injection and may last for up to 1 h. Nausea is common but is usually mild. If it is troublesome, the suitable timing of an injection (e.g., on retiring at night) or the prior administration of an antiemetic may often alleviate the problem. Increased urinary frequency occurs in up to 5 to 10% of patients, but diarrhea and vomiting are rare. Fewer than 1 in 10 patients discontinues therapy due to adverse effects even at higher doses. No significant side effects have been observed with long-term treatment with CT. The most common adverse effect of intranasal CT is nasal irritation and runny nose. Rarely it may associate with nasal bleeding and exacerbation of asthma.

### Interference with Clinical Pathology Test

No technical interferences of CT with pathological tests have been reported.

### Drug Interaction

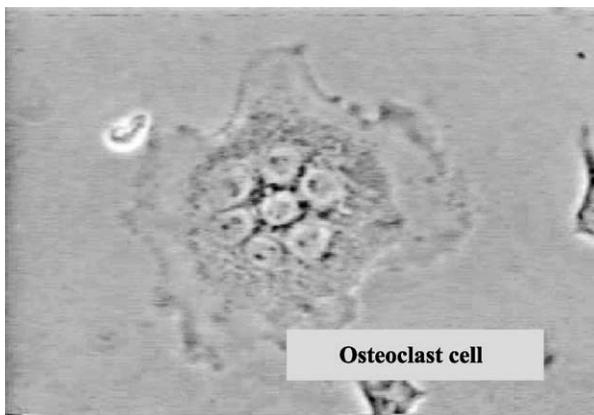
There are no known specific interactions between CT and any other drugs.

## PHYSIOLOGICAL ROLE OF CT

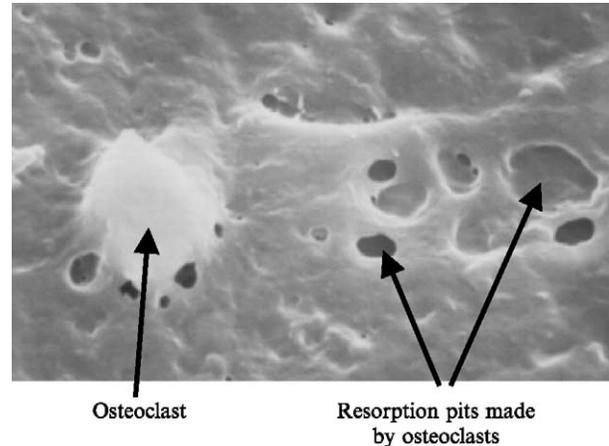
The physiological role of CT is to maintain skeletal mass during periods of calcium “stress” such as growth, pregnancy, and lactation (i.e., preserve the skeleton). During pregnancy and lactation, retention of calcium by the fetus and secretion into milk may occur at the expense of the maternal skeleton (i.e., stimulation via parathyroid hormone-related protein [PTHrP]). In these situations, increased secretion of CT may exert a protective effect to preserve skeletal mineral content and control excessive bone resorption. For example, pregnancy-associated osteoporosis is possibly due to a relative CT deficiency. Testosterone and estrogen both are stimulants for the synthesis of CT by C cells. Physiological concentrations of CT are likely to have a tonic effect to restrict osteoclastic bone resorption.

## CLINICAL PHARMACOLOGY

The main target cells for CT are the osteoclasts. CT suppresses the bone resorption activity of osteoclasts and thereby decreases bone resorption. Therefore, CT is a useful agent in patients with conditions associated with high bone turnover. Inhibition of bone resorption due to CT follows a direct inhibitory effect of it on the osteoclasts. Osteoclasts are multinucleated giant cells (Fig. 1) and create resorption cavities on the bone surface (Fig. 2); CT suppresses this bone-resorbing



**Figure 1** An osteoclast cell, one of the largest cells in the body. Using its pseudopodia, osteoclasts move over the surface of bone, making resorption pits en route.



**Figure 2** Scanning electron micrograph of a bone slice illustrating an osteoclast and its path and showing the number of excavations it has made. These cells make a tight sealing over the bone surface using its cell membrane and secrete acid and proteases to digest the contents underneath. Once a pit is formed, it moves on to a different site. CT will control this bone resorbing activity.

activity. CT also causes a reduction in osteoclast numbers when given over several months. It is not clear whether this decrease in the number of osteoclasts is a consequence of an acute effect of CT on osteoclasts or is an independent effect on its precursor cells. A plasma calcium-lowering effect of CT is not seen in normal adults due to the slow rate of bone turnover. Therefore, the administration of CT has little or no effect on plasma calcium in normal adults. However, when bone turnover is high, as in children or in disease states (e.g., Paget’s disease), administration of CT may be followed by decreases in plasma calcium.

Several pharmacological actions of CT in the gastrointestinal system have been documented. CT enhances the intestinal secretion of sodium, potassium, chloride, and water, and it inhibits both gastric emptying and acid secretion. It also inhibits the secretion of several gastrointestinal regulatory peptides, including gastrin, insulin, pancreatic glucagons, motilin, pancreatic polypeptide, and (perhaps) gastric inhibitory peptide. Inhibitory effects on the secretion of pituitary hormones, including growth hormone, thyroid-stimulating hormone, and luteinizing hormone, have also been reported.

## PHARMACOKINETICS

The most convenient analytical method for CT is radioimmunoassay or enzyme-linked immunosorbent assay (ELISA). The usefulness of the former is flawed

by the fact that it may also detect inactive fragments and precursors. Picogram quantities of CT can be detected by such assays, and the biological activity of similar amounts can now be detected using isolated osteoclast systems. CT is rapidly degraded by gastric contents and therefore needs to be given parenterally. Following subcutaneous injection, the peak plasma concentrations are seen between 15 and 45 min. The initial plasma half-life is short: approximately 4 min for porcine and human CT but approximately 15 min for salmon and eel CT. Animal studies have shown that tissue uptake is greatest in bone, liver, and kidney. The volume of distribution varies widely. Significant protein binding probably does not occur.

## THERAPEUTIC USES

### Osteoporosis

Osteoporosis is a result of an imbalance between bone formation and resorption, resulting in a net loss of bone (leading to a decrease in bone mass) and a loss of trabecular integrity (leading to fractures). CT has been in use for more than three decades to treat patients with various metabolic bone disorders, especially osteoporosis that is characterized by accelerated bone resorption. CT seems to be beneficial in prevention and treatment of osteoporosis, but its effect in fracture prevention is not convincing. The effect of CT in preventing bone loss and decreasing osteoporosis-associated fractures is much less than that of hormone replacement therapy (HRT) or bisphosphonates. Clinically, a reduction in the rate of bone loss, rather than a significant sustained increase in bone mass, should be considered as the therapeutic goal with any CT preparation.

The hormone CT may be used as an alternative treatment to HRT and bisphosphonate in the prevention and treatment of osteoporosis, especially when these agents either are not appropriate or cannot be tolerated due to adverse effects. It is also an alternative therapy for women who cannot tolerate estrogen therapy due to social or medical reasons and for patients who cannot tolerate oral bisphosphonate. However, the efficacy of CT is approximately 50% that of HRT and 25% that of bisphosphonates.

With reference to osteoporosis, other Food and Drug Administration (FDA)-approved drugs are usually recommended over CT because it is not clear whether CT increases bone density or strength and decreases the fracture rate outside the spine. However, because of its pain-relieving (analgesic) effects, CT may be used as a firstline therapy for those who have

an acute intense onset of back pain due to vertebral crush fractures. For pain relief, higher doses of CT (e.g., 600–800 IU) are required, but only for few days (i.e., total of 1–2 vials). The treatment regimen should typically change when the acute pain subsides or if the pain fails to subside within a week or so.

CT has been used successfully for the prevention and treatment of postmenopausal osteoporosis and for treatment of osteoporosis in men. After prolonged administration of CT, it may lose its beneficial effects on bone due to down-regulation of receptors and/or the development of neutralizing anti-CT antibodies (with nonhuman CT) or increased catabolism of the peptide. Administering small doses of CT less frequently, and switching salmon or eel CT to a synthetic or genetically engineered human CT, can minimize these two problems.

Doses of injectable CT as small as 60 IU per week have been shown to be effective in arresting the postmenopausal bone loss. CT is more widely used in Southern Europe and Japan than it is elsewhere in the world. In a small, controlled, 2-year dose-finding study of intranasal CT in postmenopausal women with osteoporosis, spine BMD increased by 2%. There was no effect on other skeletal sites, and slight decrease incidences of vertebral fracture were observed compared with placebo.

Although the ideal treatment for osteoporosis is with anabolic agents to increase bone mass, inhibitors of bone resorption can arrest further bone loss, increase BMD, and decrease fracture rates. CT has also been shown to stabilize or modestly increase indexes of cortical and trabecular bone mass and total body calcium, when it is administered to patients with established osteoporosis. The increments in bone density seen with CT are modest, appear to be transient, and are likely to be due to reduction in bone resorption with bone formation remaining unaffected until remodeling spaces are filled. However, the reason for the reduction of fracture rate in the light of little increase of BMD is not yet understood. Perhaps anti-resorptive therapies (e.g., CT, bisphosphonates), in addition to their well-known effects on decreasing bone turnover, may also have a positive effect on preservation of osteoblasts and osteocytes, enhancement of osteoclast apoptosis, and preservation of the microarchitecture of bone. This may result in disproportionate reduction of fracture rates in comparison with a slight increase of BMD observed.

The PROOF (Prevent Recurrence of Osteoporotic Fractures) study was a 5-year, multicenter, double-blind randomized study designed to test whether miacalcin nasal spray reduces the risk of new vertebral

fractures in postmenopausal women with established osteoporosis. A total of 1255 women with at least one but not more than five vertebral compression fractures were randomized to receive miacalcin nasal spray (100, 200, or 400 IU or placebo daily). Patients also received daily supplements of 1000 mg of calcium plus 400 IU of vitamin D.

The results demonstrated that miacalcin nasal spray, at the recommended dose of 200 IU daily, reduced the risk of new vertebral fractures as compared with placebo. At 5 years, there was a statistically significant reduction in the risk of new vertebral fractures with CT nasal spray (200 IU daily) as compared with placebo in patients with one to five vertebral compression fractures at baseline, but no such effects were seen with either 100 or 400 IU of the same agent. The study did not show hip fracture (or nonvertebral fracture) reduction. The study also showed somewhat of a reduction in the serum type I collagen cross-linked telopeptide (CTX), a biochemical marker for bone resorption. In the study, the frequency of adverse events in all treatment groups was comparable to placebo except for an increase in rhinitis. Other studies using CT have shown small increments in spinal BMD and marginal antifracture efficacy. Most of these studies have fewer fractures and are underpowered to detect a significant effect of CT on fracture reduction.

Synthetic salmon, human, and eel CTs have been developed as therapeutic agents for osteoporosis, but there are no extensive or multiple clinical trial data to show convincing reduction of fracture rates except for the PROOF study. Unfortunately, in many patients, circulating antibodies will be developed with nonhuman CTs (e.g., salmon CT, eel CT) within 2 to 3 years of therapy, and these antibodies may significantly impair the actions of this hormone.

Like other antiresorptive therapies, CT decreases the remodeling phase, thereby achieving a new lower rate of bone turnover. Consequently, increases in bone densities are generally seen only during the first 2 to 3 years of treatment. It may be possible to minimize this by administering CT (e.g., nasal CT 200 IU/day) in a cyclical fashion. There is some evidence to suggest that CT may prevent glucocorticoid-induced bone loss, but the efficacy is considerably less than that of bisphosphonates. It is unclear how CT and similar antiresorptive agents reduce bone loss given that glucocorticoids predominantly depress bone formation rather than stimulate bone resorption. Recent evidence suggests that this may be due to enhanced osteoclast apoptosis with both CT and amino-containing bisphosphonates.

## Paget's Disease of Bone

CT has been used successfully as a treatment for Paget's disease of bone for several decades. Several types of CT have been used, but salmon CT is the most commonly used. CT is a useful treatment for Paget's disease (which is characterized by very high bone turnover) because it not only inhibits osteoclast activity but also prevents the recruitment of new osteoclasts. Inhibition of osteoclastic activity explains the short-term (acute) effects of CT, whereas long-term responses are presumably related to the decreased number of osteoclasts. CT is effective in alleviating bone pain and in improving clinical, biochemical, radiological, and histological features of the disease. The major indication for CT therapy in Paget's disease is bone pain of Pagetic origin. CT relieves bone pain rapidly and helps to decrease excessive blood flows over the affected bones. It can also alleviate some of the chronic neurological problems associated with the disease such as early spinal cord compression due to Paget's disease of vertebrae. CT can decrease biochemical markers of bone turnover, such as serum alkaline phosphatase and N-telopeptide levels, by approximately 40 to 50%. Use of CT will rarely normalize serum bone-specific alkaline phosphatase levels, which is the key marker for Pagetic activity. However, the effectiveness of CT is dependent on its continued use; discontinuation of the drug will reactivate the Paget's disease process within a short period. There are no serious side effects associated with this treatment, although minor side effects are common and include nausea, diarrhea, and pain and redness at the injection site.

## Hypercalcemia

CT produces an acute reduction in plasma calcium in patients with hypercalcemia due to malignancy and hypercalcemic emergency. A fall in plasma calcium of 0.5 to 1.0 mmol per liter may be observed within 24 h of commencing CT therapy. However, in the absence of additional therapy, the effect might not be maintained beyond 48 to 72 h. CT is most likely to be effective in cases of hypercalcemia where a generalized increase in bone resorption is a prominent feature. The calciuric effect of CT may also play a small role in reducing the raised plasma calcium. In patients with hypercalcemia, CT can be used to obtain an immediate calcium-lowering effect. However, in most patients, it is necessary to start a bisphosphonate (e.g., pamidronate, zoledronate) concomitantly for further reduction and maintenance of serum calcium levels.

These agents also have the beneficial effects of controlling metastatic osteolysis.

## ANALGESIC ACTION OF CT

Miacalcin nasal spray is not FDA approved for use as an analgesic. Most of the literature concerning the analgesic action of CT employed the use of the injectable form. The reduction in pain occurs significantly earlier than any demonstrated improvement in skeletal dynamics, and in some instances no objective improvement in the underlying skeletal condition can be observed.

### Use of CT to Control Bone Pain

Although the efficacy of CT in fracture reduction is controversial, there is convincing evidence for efficacy of CT as an analgesic in patients with acute vertebral fractures. In contrast to HRT and bisphosphonate therapies, CT (at three- to fourfold higher doses than that recommended for osteoporosis) reduces pain in patients with recent osteoporotic fractures (e.g., crush fracture syndrome). Because of its analgesic effect, CT can be useful as an adjunct therapy in the management of osteoporotic patients following osteoporotic fractures (i.e., a high-dose short course such as 800 IU/day for 1 week).

CT can be used effectively in patients with osteoporotic fracture syndrome to relieve pain on a short-term basis. In addition to its use in patients with crush fracture syndrome and bone pain, to be cost-effective, CT could be useful in selected patients with high bone turnover and increased bone remodeling. Painful episodes of osteoporosis are not due to the osteoporotic process itself but rather are associated with fractures. The presentation of pain is usually acute, continuous, exacerbated with movement, and ameliorated by rest. The pain can be severe and will generally resolve with bed rest. However, it is essential to mobilize these patients at the earliest possible time to avoid immobilization-associated complications. High-dose, short-term use of CT could assist in this process.

Intranasal CT has been investigated for its analgesic effect in acute back pain due to osteoporosis. In one double-blind, placebo-controlled study, more than 50% of patients receiving 200 IU of CT nasal spray daily showed a "good" response in mobility and functional capacity, whereas patients receiving placebo nasal spray experienced only a moderate response with respect to the same parameters. In addition, intranasal CT appears to be more rapid in providing analgesia to postmenopausal osteoporotic women with acute pain

than does injectable salmon CT. A double-blind, double-placebo study by Pontiroli and colleagues examined the analgesic effect of intranasal CT (200 IU daily) versus intramuscular CT (100 IU daily) versus placebo in 28 women with painful postmenopausal osteoporosis for a period of 4 weeks. In the 24 patients who completed the trial, treatment with intranasal CT resulted in a statistically significant reduction in pain score, as measured by a visual analogue scale, by the second week of treatment. With intramuscular CT, as well as with placebo, the pain scores decreased significantly by the fourth week of treatment. At week 4, the final pain scores were not different among the three treatment groups.

The mechanism of the pain associated with acute fractures is different from that of the pain associated with fractures of long duration. CT nasal spray seems to demonstrate its greatest analgesic activity when administered shortly after a new fracture has occurred. The analgesic effect of various therapies for osteoporosis can diminish over time, and in some studies it is no different from placebo at 4 weeks. This may be due to either the tolerance or the spontaneous improvement that usually occurs over 4 weeks with most vertebral compression and wedge fractures.

### Metastatic Bone Pain

CT has been used for the treatment of bone pain in patients with metastatic bone disease, where an analgesic effect has been observed in three-quarters of patients. CT has also been used successfully in the treatment of intractable pain from advanced malignancy, especially when injected into the subarachnoid space. It is likely that the main site of this analgesic action is in the central nervous system. Although the mechanism of CT's analgesic action is still not fully understood, it has been postulated to occur through both direct action at the hypothalamic level and indirect action through interference with neurotransmitters such as serotonin and prostaglandin, independent of endorphin, and a peripheral action mediated via the inhibition of the chemical factors involved in inflammation.

There have been no placebo-controlled studies conducted in patients with metastatic bone pain. However, Jelic and colleagues used CT to treat 16 patients with pain due to diffuse osteolytic or osteoblastic-osteolytic metastases who had not obtained satisfactory pain relief with opiate-type analgesics. Most of the patients on CT experienced a decrease in pain with an accompanying decreased intake of opiate-type analgesics. Of the 16 patients, 2 were

able to discontinue opiate intake completely, whereas 3 had no pain relief at all. The authors concluded that CT may be useful in relieving metastatic bone pain in some patients resistant to opiate-type analgesics.

### Migraine

Migraine reportedly responds to CT, but the mechanism of action is unclear. Intramuscular doses of 100 IU per day were significantly more effective than placebo in reducing the frequency, intensity, and duration of migraine. However, administration of CT via injections can lead to nausea and thereby may aggravate some of the migraine-associated problems. Salmon CT has also been administered prophylactically via the nasal route and shown to be effective in preventing attacks of common migraine. However, disease-specific agents such as sumatriptan have now superseded these therapies.

### Arthritis

There have been no placebo-controlled studies conducted using CT in patients with rheumatoid or osteoarthritis. There is corroborative evidence to suggest that pain and inflammation of inflammatory arthropathy may respond to CT at least partially. Dainotto and colleagues conducted an open-label study of 70 patients with osteoarthritis. They noted a decrease in spontaneous pain in patients using CT (100 IU intramuscular) for 40 days.

### OTHER CLINICAL USE OF CT

CT has been used successfully in the treatment of Sudeck's atrophy. Clinical experience suggests that the administration of CT shortens the healing time of fractures, but this has not been confirmed. Another possibility is that CT could be administered intrathecally (to decrease the postoperative requirements of narcotics) in patients undergoing hip replacement under spinal anesthesia and in patients with metastatic bone disease.

### CONCLUSION

The 32-amino acid polypeptide hormone CT is a potent inhibitor of the osteoclast cells. Therapy with CT is particularly effective in controlling osteoclastic

bone resorption in disorders characterized by high bone turnover, including Paget's disease of bone, osteoporosis, Sudeck's atrophy, and hypercalcemic states. However, the efficacy of CT in the treatment of osteoporosis is significantly less than that of bisphosphonates, HRT, and SERMs. Nevertheless, CT can be used effectively to control bone pain (e.g., crush fracture syndrome) in patients using higher doses (e.g., 800 IU intranasal daily for 5–7 days). This will allow for early mobilization of patients (and hence prevent immobilization-associated complications) and decrease the doses, or eliminate the necessity, of potent analgesia.

### See Also the Following Articles

Calcitonin Gene-Related Peptide (CGRP) • Hypercalcemia and Hypercalcemia Treatment • Medullary Thyroid Carcinoma • Osteoporosis in Older Men • Osteoporosis in Older Women • Osteoporosis, Overview • Paget's Disease of Bone

### Further Reading

- Azzia, M. (1989). "The Calcitonin: Physiology and Pharmacology." Karger, Basel, Switzerland.
- Chestnut, C. H., III, Silverman, S., Andriano, K., *et al.* (2000). A randomized trial of nasal spray salmon calcitonin in postmenopausal women with established osteoporosis: The Prevent Recurrence of Osteoporotic Fractures Study. *Am. J. Med.* **109**, 267–276.
- Civitelli, S. R., Gonnelli Zacchei, F., Bigazzi, S., Vattimo, A., Avioli, L. V., and Gennari, C. (1988). Bone turnover in postmenopausal osteoporosis: Effect of CT treatment. *J. Clin. Invest.* **82**, 1268–1274.
- Lyrithis, G. P., Paspati, I., Kaiachalios, T., *et al.* (1997). Pain relief from nasal salmon calcitonin in osteoporotic vertebral crush fractures. *Acta Orthopaedica Scand.* **68**(Suppl. 275), 112–114.
- Pontiroli, A. E., Pajetta, E., Scaglia, L., *et al.* (1994). Analgesic effect of intranasal and intramuscular salmon calcitonin in postmenopausal osteoporosis: A double-blind, double-placebo study. *Aging Clin. Exp. Res.* **6**, 459–463.
- Szucs, J., Horvath, K., Kollin, E., Szathmari, M., and Hollo, A. (1992). Three year CT combination therapy for postmenopausal osteoporosis with crush fractures of the spine. *Calcif. Tissue Int.* **50**, 7–10.
- Wimalawansa, S. J. (1989). Calcitonin: Molecular biology, physiology, pathophysiology, and its therapeutic uses. In "Advances in Bone Regulatory Factors: Morphology, Biochemistry, Physiology, and Pharmacology" (A. Pecile and B. Bernard, eds.), pp. 121–160. Plenum, New York.
- Wimalawansa, S. J. (1993). Long- and short-term side effects and safety of CT in man: A prospective study. *Calcif. Tissue Int.* **52**, 90–94.



# Caloric Restriction, Aging and Oxidative Stress

Holly Van Remmen and Arlan Richardson

University of Texas Health Science Center, San Antonio, Texas, United States

## Glossary

**oxidative stress theory of aging** The oxidative stress theory of aging proposes that a chronic state of oxidative stress exists in cells of aerobic organisms even under normal physiological conditions because of an imbalance of pro-oxidants and antioxidants.

**reactive oxygen species** Reactive oxygen species include free radicals and other nonradical molecules such as peroxides that are continually formed endogenously as a consequence of metabolism and result in oxidative modifications to cellular components.

**C**aloric restriction (also called dietary restriction or food restriction) is a consistently proven method of extending the mean and maximum life span in rats and mice based on reducing the total caloric intake without altering the intake of other essential dietary components.

## INTRODUCTION

In 1935, McCay and colleagues first demonstrated that the life span of rodents could be increased by caloric restriction. Since then, numerous laboratories have confirmed that reducing calorie intake (without malnutrition) significantly increases both the mean and the maximum survival in a variety of strains of mice and rats. Furthermore, caloric restriction has been shown to retard a variety of age-related deteriorations in physiologic systems and to delay or prevent the onset of a variety of diseases. Thus, caloric restriction is believed to increase life span by retarding aging; however, the mechanism whereby caloric restriction retards aging has not been elucidated. Increased resistance to oxidative stress and enhanced protection against oxidative damage appear to be important factors in the underlying mechanism of the life span-extending properties of caloric restriction.

## IMPACT OF STUDIES USING CALORIC RESTRICTION

The classic study by McCay *et al.* in 1935 was the first to show that one could increase the life span of rodents by caloric restriction. In these studies, life extension was achieved by a very severe restriction in total food intake, which also stunted the growth and sexual development of the rats. Later studies demonstrated that the life span of rodents could be increased by caloric restriction less severe than the conditions used by McCay *et al.* In fact, many subsequent studies have confirmed that caloric restriction of 30–50% consistently increases both the mean and the maximum life span of both rats and mice. In addition, caloric restriction has been shown to be effective for both long-lived and short-lived strains of rats and mice.

Many studies have provided convincing evidence that caloric restriction increases life span by retarding the aging process. A number of studies show that caloric restriction can delay or reduce the onset of most age-related diseases and alter most physiological processes that change with age. This is an important point because changes in life span alone do not necessarily mean that the aging process per se has been altered.

## POSSIBLE MECHANISMS FOR THE LIFE SPAN-EXTENDING PROPERTIES OF CALORIC RESTRICTION

Current studies are aimed at elucidating the mechanism(s) of caloric restriction—that is, the biochemical/molecular process(es) responsible for the antiaging action of caloric restriction. Several possible mechanisms have been proposed. Initially, caloric restriction was thought to increase survival by retarding growth and development. However, caloric restriction begun in adult life, well after development and maturation, can increase both the mean and the maximum survival

of mice and rats. Body fat content was also thought to play a role, and this was supported by the fact that rodents fed *ad libitum* have a greater body fat content than calorie-restricted rodents. However, no relationship was found between body fat content and survival in calorie-restricted genetically obese (*ob/ob*) and normal mice of the same inbred strain, suggesting that body fat content per se does not play a major role in the biological mechanism responsible for the increased life span of caloric restriction. Reduced metabolic rate was also thought to play a role. However, neither the metabolic rate per unit of lean body mass nor that per unit of "metabolic mass" are reduced significantly by caloric restriction in rats; therefore, caloric restriction cannot be due to a general reduction in the metabolic rate of rodents. Possible mechanisms for the action of caloric restriction are modulation of glycemia and hyperinsulinemia by caloric restriction, hormesis or increased tolerance to stress, and attenuation of oxidative damage.

## RESISTANCE TO OXIDATIVE STRESS

Evidence indicates that caloric restriction may act at least in part by decreasing oxidative stress and increasing antioxidant defense and repair. The oxidative stress hypothesis of aging proposes that a chronic state of oxidative stress exists in cells of aerobic organisms even under normal physiological conditions because of an imbalance of pro-oxidants and antioxidants. Reactive oxygen species produced during the course of normal aerobic metabolism may cause random deleterious damage that accumulates over time and contributes to the process of aging and various age-associated diseases. Evidence indicates that the antiaging action of caloric restriction might occur through a mechanism that involves increased resistance to oxidative stress. For example, studies of invertebrates (nematodes and *Drosophila*) that show life extension suggest that enhanced resistance to stress plays a key biological role in the increased longevity of these organisms. In addition, the activities of one or more of the antioxidant enzymes either remain the same or increase with caloric restriction in rodent tissues. More important, caloric restriction has consistently been shown to reduce oxidative damage in rodent tissues. Caloric restriction can also enhance repair of oxidatively damaged molecules by retarding the age-related decrease in DNA repair proteins and increasing proteolytic removal of damaged proteins.

## OXIDATIVE DAMAGE TO CELLULAR COMPONENTS

A number of studies have measured the levels of oxidative damage in macromolecules in various tissues from young and old *ad libitum*-fed and caloric-restricted animals. Early studies focused on measuring oxidative damage to lipid; recent studies, however, have also measured oxidative damage in protein and DNA. In 1992, the first data were published showing that the levels of oxo8dG in nuclear DNA from liver of 26-month-old male F344 rats were reduced significantly (36%) by caloric restriction. It has also been reported that caloric restriction reduced oxo8dG levels in nuclear DNA from the livers of 28-month-old male F344 rats by 25% compared to *ad libitum*-fed rats. However, this difference was not observed until the rats were 27 months of age. In mice, oxo8dG in the total DNA isolated from brain, skeletal muscle, heart, and liver of 26-month-old male C57BL/6 mice fed a calorie-restricted diet was reduced compared to that of their *ad libitum*-fed counterparts. Caloric restriction has also been shown to significantly reduce (21%) the level of oxo8dG in liver mitochondrial DNA from 24-month-old male F344 rats. The level of oxo8dG is significantly lower in nuclear DNA in tissues of old mice fed a calorie-restricted diet compared to old mice fed *ad libitum*. However, caloric restriction does not always reduce oxidative damage to nuclear DNA in rodent tissues. Even in those tissues in which oxidative damage to nuclear DNA is reduced, the decrease is often relatively mild. Caloric restriction completely prevented the age-related increase in oxo8dG levels in liver mitochondrial DNA in rats and mice, but it had no effect on the age-related increase in oxo8dG levels in rat liver nuclear DNA and reduced oxo8dG levels only 19% in mouse liver. Thus, data suggest that caloric restriction has a much greater effect on the age-related accumulation of oxidative damage to mitochondrial DNA than that to nuclear DNA.

Caloric restriction also alters oxidative damage to proteins and lipids with age. The age-related decline in protein degradation and the accumulation of oxidatively modified or abnormal proteins with age are both delayed by caloric restriction. For example, the level of oxidized proteins has been shown to increase with age in the brain, and this increase is ameliorated by caloric restriction. Age-related changes in lipid peroxidation and in membrane fatty acid composition are also altered by caloric restriction. Membrane fatty acid composition is affected by caloric restriction, resulting in decreased amounts of polyunsaturated fatty acids, reduced peroxidizability, and maintenance

of membrane fluidity. Thus, caloric restriction acts to protect membrane integrity and function.

## MITOCHONDRIA

Since mitochondria continuously produce endogenous reactive oxygen species as a consequence of the action of the electron transport chain, it is possible that mitochondrial production of reactive oxygen species and oxidative damage to mitochondria could be altered by caloric restriction. In fact, caloric restriction has been shown to reduce mitochondrial reactive oxygen species generation and to reduce oxidative damage to mitochondrial DNA. In particular, caloric restriction decreases reactive oxygen species production at complex I of the respiratory chain. Decreased reactive oxygen species production has also been associated with reduced protein oxidation in rat heart mitochondria.

### See Also the Following Articles

Aging, Animal Models for • Body Weight, Body Composition, and Aging • Diabetes Mellitus, Diagnosis and Treatment in the Elderly • Hunger and Satiety • Obesity Regulation • Oxidative Stress and Aging • Stress, Aging, and Central Nervous System Interactions

## Further Reading

- Barja, G. (2002). Endogenous oxidative stress: Relationship to aging, longevity and caloric restriction. *Aging Res. Rev.* **1**(3), 397–411.
- Masoro, E. J. (1999). Caloric restriction. *Aging Clin. Exp. Res.* **10**(2), 173–174.
- Masoro, E. J. (2000). Caloric restriction and aging: An update. *Exp. Gerontol.* **35**(3), 299–305.
- McCay, C. M., Crowell, M. F., and Maynard, L. A. (1935). The effect of retarded growth upon the length of life span and upon the ultimate body size. *J. Nutr.* **10**, 63–79.
- Roth, G. S., Ingram, D. K., and Lane, M. A. (2001). Caloric restriction in primates and relevance to humans. *Ann. N. Y. Acad. Sci.* **928**, 305–315.
- Sohal, R., and Weindruch, R. (1996). Oxidative stress, caloric restriction, and aging. *Science* **273**, 59–67.
- Sohal, R., Agarwal, S., Candas, M., Forster, M. J., and Lal, H. (1994). Effect of age and caloric restriction on DNA oxidative damage in different tissues of C57BL/6 mice. *Mech. Ageing Dev.* **76**, 215–224.
- Van Remmen, H., Guo, Z., and Richardson, A. (2001). The anti-ageing action of dietary restriction. *Novartis Found. Symp.* **235**, 221–230.
- Weindruch, R. (1996). The retardation of aging by caloric restriction: Studies in rodents and primates. *Toxicol. Pathol.* **24**(6), 742–745.
- Yu, B. P. (1996). Aging and oxidative stress: Modulation by dietary restriction. *Free Radical Biol. Med.* **21**(2), 1–18.
- Yu, B. P., and Chung, H. Y. (2001). Stress resistance by caloric restriction for longevity. *Ann. N. Y. Acad. Sci.* **928**, 39–47.



# Captopril

Florentin Tache

University of Bucharest, Bucharest, Romania

Alexandru Farca

Labormed Pharma S.A., Bucharest, Romania

Andrei Medvedovici and Victor David

University of Bucharest, Bucharest, Romania

## Glossary

**angiotensin I** Inactive decapeptide produced by the action of renin, elaborated by the juxtaglomerular cells of the kidney, on angiotensinogen (an  $\alpha_2$  globulin).

**angiotensin II** Octapeptide resulting by the splitting of two amino acids from angiotensin I as a result of angiotensin-converting enzyme (ACE) activity.

**angiotensin-converting enzyme (ACE) inhibitors** Blood pressure-lowering agents that reduce peripheral vascular resistance via blockage of the activity of the ACE, inhibiting the conversion of angiotensin I to angiotensin II.

**mass spectrometry (MS)** Analytical structural method based on the identification of the particular pattern of fragmentation of the organic molecules resulting in electrically charged species analyzed against their mass/charge ratio.

**nuclear magnetic resonance (NMR)** Analytical structural method based on the interaction of some specific nuclides characterized by an uneven quantum number of nuclear spins ( $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{14}\text{N}$ ,  $^{19}\text{F}$ , and  $^{31}\text{P}$ ) with intense external magnetic fields.

**renin-angiotensin system (RAS)** One of the most important regulatory systems, regulating blood volume, arterial pressure, and cardiac and vascular function.

**type 2 diabetes mellitus** More often known simply as type 2 diabetes, a metabolic disease in which the body cannot get its energy from glucose in the normal way because the body does not produce enough insulin or cannot use the insulin it has (known as insulin resistance).

**C**aptopril, (2S)-1-[(2S)-3-mercapto-2-methylpropanoyl]pyrrolidine-2-carboxylic acid, is an inhibitor of angiotensin-converting enzyme (ACE), acting directly on the adrenal gland to stimulate the release of aldosterone. It prevents high blood

pressure by inhibiting the enzymatic conversion of angiotensin I to angiotensin II. Also, during the metabolic pathway, it converts to a disulfide compound.

## HISTORY

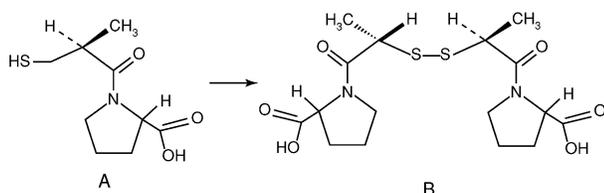
The discovery of Captopril appeared as a result of two basic research directions. The first one emerged from the identification and description of the renin-angiotensin system (RAS); it started in 1934 and was concluded by scientific articles that identified angiotensin I and angiotensin II and were published during the late 1950s. The second direction is based on research in Brazil on the cause of death from snake venom; it identified a natural substance that acts on its victim by fatally lowering blood pressure. It was finalized in 1965 when it was shown that this natural substance blocks the conversion of angiotensin I to angiotensin II. Based on these data, the scientists from Squibb synthesized the first angiotensin-converting enzyme (ACE) inhibitor during the early 1970s. It was subsequently approved for marketing by the U.S. Food and Drug Administration in 1981.

## SYNTHESIS

N-acylation of L-proline with (S)-3-acetylthio-2-methylpropionyl chloride, in the presence of sodium hydroxide, leads to the sodium salt of 3-acetylthio-2-D-methylpropionyl-L-proline, which is treated with hydrochloric acid and hydrolyzed in the presence of ammonia. The resulting captopril ([Fig. 1](#)) is then extracted with ethyl acetate and further purified.

## PHYSICAL-CHEMICAL PROPERTIES

Captopril ( $\text{C}_9\text{H}_{15}\text{NO}_3\text{S}$ ,  $M_w = 217.29$ ) is a white crystalline powder having a characteristic sulfide-like odor



**Figure 1** Captopril (A) and its disulfide derivative (B) chemical structures.

and a melting point between 105 and 108°C. It is freely soluble in water or diluted solutions of alkali hydroxides, in alcohols, in methylene chloride, or in chloroform. The elemental analysis leads to 49.75% carbon, 6.96% hydrogen, 6.45% nitrogen, 22.09% oxygen, and 14.73% sulfur. A solution of 2% captopril in water produces a pH value in the range of 2.0 to 2.6 with a dissociation constant of  $3.7 \pm 0.2$ . It is recommended that captopril be stored in airtight containers, although it is known to be relatively stable up to 50°C.

In aqueous solutions, captopril is oxidized mainly to captopril disulfide (II), which can be favored by higher pH values. There are two chiral carbon atoms in the captopril molecule, so two pairs of diastereoisomers exist. The drug should be characterized by an optical rotation of  $-84 \pm 2^\circ$ . The UV spectrum in methanol shows only a single absorption band situated at 210 nm. The IR spectrum of captopril recorded as KBr dispersion exhibits three characteristic absorption bands: 3450 to 2500  $\text{cm}^{-1}$  (assigned to carboxyl and thiol groups), 1750  $\text{cm}^{-1}$  (specific to the CO group in carboxyl), and 1580  $\text{cm}^{-1}$  (CO from amide group). The  $^1\text{H}$ -nuclear magnetic resonance (NMR) spectrum recorded for a solution in deuterated chloroform exhibits the following signals ( $\delta$  [ppm]): 9.38 (broad), 4.63 (m), 3.61 (m), 2.83 (m), 2.48 (m), 2.31 (m), 2.04 (m), 1.52 (t), and 1.21 (d). The main mass spectrometry (MS) spectral lines correspond to the following mass fragments: 114, 172, 184, and 192 amu.

## METABOLIC FUNCTIONALITY

The thiol group from the captopril structure interacts with zinc ion from the catalytic moiety of ACE inhibitors. Because of this functional group, the tolerance toward nitrates is greatly potentiated. Thus, captopril can enhance the antianginal effect of isosorbide dinitrate.

In general, an excessive hypotension may occur when ACE inhibitors are used concurrently with diuretics, other antihypertensives, or other agents

such as alcohol. An additive hyperkalemic effect is possible in patients receiving ACE inhibitors with potassium-sparing diuretics, potassium supplements (including potassium-containing salt substitutes), or other drugs that can cause hyperkalemia (e.g., ciclosporin, indometacin), and serum potassium concentrations should be monitored. Therefore, potassium-sparing diuretics and potassium supplements should generally be stopped before initiating ACE inhibitors in patients with heart failure. The adverse effects of ACE inhibitors on the kidneys may be enhanced by other drugs such as nonsteroidal anti-inflammatory drugs (NSAIDs), which can affect renal function.

## ANTIDIABETIC ACTIVITY

Hypertension occurs with twice the frequency in the diabetic population compared with the nondiabetic population, and up to 50% of patients with type 2 diabetes mellitus become hypertensive. In addition to being a major risk factor for atherosclerosis in large blood vessels, hypertension in diabetes appears to contribute to small vessel disease and is a risk factor for diabetic nephropathy and possibly for diabetic retinopathy. Recent studies have shown that ACE inhibitors can slow the progression to diabetic nephropathy in patients with type 1 or type 2 diabetes with microalbuminuria or macroalbuminuria.

A major way in which to protect the kidney from the complications of diabetes is to treat high blood pressure aggressively, regardless of the antihypertensive drug types used. In early studies in patients with type 1 diabetes, antihypertensive drugs such as diuretics, beta blockers, and hydralazine were used, indicating that lowering blood pressure reduces proteinuria and slows the decline of renal function. A preponderance of evidence indicates that ACE inhibitors protect the kidney better than do other blood pressure-lowering medications, probably because ACE inhibitors specifically lower the intrarenal pressure.

Lewis and colleagues performed a landmark study in patients with type 1 diabetes, albuminuria, and mildly impaired creatinine clearance, that is, patients who were just beginning to develop renal failure. The ACE inhibitor captopril reduced the risk of a decline in renal function more effectively than did other antihypertensive regimens (not including calcium channel blockers).

Captopril and atenolol are equally effective in reducing blood pressure to means of 144/83 and 143/81 mm Hg, respectively, with similar proportions

of patients (27 and 31%, respectively) requiring three or more antihypertensive treatments. A value lower than 130/85 mm Hg is a reasonable target. It has not been firmly established whether lower blood pressures accrue further benefits.

Captopril and atenolol are also effective in reducing the risk of macrovascular end points. Similar proportions of patients in the two groups showed deterioration in retinopathy by two grades after 9 years (31% in the captopril group and 37% in the atenolol group) and developed clinical-grade albuminuria  $\geq 300$  mg/L (5% and 9%, respectively). The proportions of patients with hypoglycemic attacks were not different between the groups, but mean weight gain in the atenolol group was greater (3.4 vs 1.6 kg).

Blood pressure lowering with captopril and atenolol was similarly effective in reducing the incidence of diabetic complications. Neither drug has any specific beneficial or deleterious effect, suggesting that blood pressure reduction in itself may be more important than the treatment used.

In patients with non-insulin-dependent diabetes mellitus, treatment with captopril or enalapril does not normalize the exaggerated systemic norepinephrine responsiveness.

It was proved that captopril does not produce any significant changes in the blood levels of ionized calcium or phosphorous ions and that it does not alter the serum levels of parathyroid hormone (PTH) and metabolites of vitamin D.

After oral administration of captopril, maximum effect is reached in 1 to 2 h, although the full effect might not develop for several weeks during chronic dosing. The duration of action is dose dependent and may persist for 6 to 12 h.

## PHARMACOKINETICS

After oral administration, 60 to 75% of a captopril dose is absorbed from the gastrointestinal tract and peak plasma concentrations are usually attained within 1 h. Absorption is reduced by 25 to 55% on administration with food, without a significant effect on the antihypertensive activity of captopril.

Pharmacokinetic studies have revealed that about 30% of the captopril amount in human plasma is bounded to proteins. It is eliminated in the urine as unchanged drug (40–50%), disulfide, and other metabolites. The elimination half-time has been reported to be 2 to 3 h, but this is increased in renal impairment (2.5 times for captopril and 4 times for the disulfide metabolite).

## ANALYTICAL ASPECTS OF CAPTOPRIL DETERMINATION IN BIOLOGICAL FLUIDS

The first approaches to captopril determination in biological matrix were made by gas chromatography (GC) and gas chromatography/mass spectrometry (GC-MS).

High-performance liquid chromatography (HPLC) is a challenging method of captopril determination in blood, plasma, and urine. However, the low concentration levels and the relative sensitivity of captopril to oxidation determine that these methods follow three different directions: (1) the use of electrochemical detection, as a more sensitive alternative detection than UV, without any derivatization of the analyte of interest; (2) the UV or fluorescent detection of a derivatization product of captopril, with labeling done prior to or after the chromatographic separation; and (3) the use of preconcentration techniques such as liquid-liquid extraction and multicolumn switching setup.

Concerning the derivatization reagents frequently used for captopril labeling, the following list can be considered: (1) N-(4-benzoylphenyl)maleimide, (2) 1-benzyl-2-chloropyridinium bromide, (3) 2,4'-dibromoacetophenone, (4) N-(4-dimethylamino-3,5-dinitrophenyl)maleimide, (5) N-(4-dimethylaminophenyl) maleimide, (6) N-(1-pyrenyl)maleimide, (7) 7-fluorobenzofurozan-4-sulfonic acid, (8) N-(7-dimethylamino-4-methyl(coumarin 3-yl)maleimide, and (9) 5,5'-dithio-(bis-2-nitrobenzoic acid).

More interest was paid, during the last period, to finding highly specific fluorescent reagents for the thiol function. From this category, monobromobimane was the last one investigated. The derivatization of free captopril with monobromobimane and plasma matrix removal by precipitation with trichloroacetic acid (TCA) should be realized immediately after plasma collection. In such conditions, the plasma samples can be stored for at least 2 months at  $-40^{\circ}\text{C}$ . Collection of blood samples can be achieved either on heparin or citrate. Disulfide can be reduced to captopril with pyrophosphoric acid, allowing total drug determination. The mean plasma concentration profile of this drug, assessed according to the monobromobimane fluorescent labeling method (after a single-dose crossover bioequivalence study), is in very good agreement with previously used methods and is illustrated in Fig. 2.

Some other separation techniques, such as electrophoresis, capillary zone electrophoresis, capillary

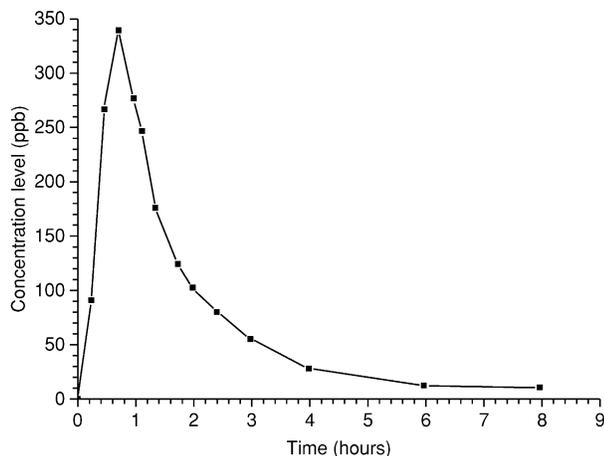


Figure 2 Plasma concentration profile of captopril.

izotachophoresis, and LC–chemiluminescence-based detection, were also used.

### See Also the Following Articles

Diabetes, Type 2 • Hypertension, Renin and • Renin • Tissue Renin-Angiotensin-Aldosterone System

### Further Reading

- Henderson, R. (1994). The evolution of integrative capability: Innovation in cardiovascular drug discovery. *Ind. Corp. Change* **3**, 607–627.
- Lewis, I. F., Hunsicker, L. G., Bain, R. P., *et al.* (1993). The effect of angiotensin-converting enzyme inhibition on diabetic nephropathy. *N. Engl. J. Med.* **329**, 1456–1462.
- Marin-Grez, M. (1982). The involvement of the renin-angiotensin-aldosterone system, the corticotropin-glucocorticoid system, antidiuretic hormone, catecholamines and prostaglandins. *Biochem. Pharmacol.* **31**, 3941–3947.
- Maxwell, R. A., and Shohreh, E. (1990). “Drug Discovery: A Case Book and Analysis.” Humana Press, Clifton, NJ.
- Mogensen, C. E. (1980). Antihypertensive treatment inhibiting the progression of diabetic nephropathy. *Acta Endocrinol.* **238**(Suppl.), 103–108.
- Parving, H. H., Andersen, A. R., *et al.* (1983). Diabetic nephropathy and arterial hypertension: The effect of antihypertensive treatment. *Diabetes* **32**(Suppl. 2), 83–87.
- Sweetman S. C. (ed.) (2002). “Martindale: The Complete Drug Reference,” 33rd ed. Pharmaceutical Press, London.
- Täche, F., Farca, A., David, V., and Medvedovici, A. (2002). Validation of a LC-fluorescence method for determination of free captopril in human plasma, using a pre-column derivation reaction with monobromobimane. *J. Pharm. Biomed. Anal.* **28**, 549–557.
- Townsend, R., Dipette, D. I., Evans, E. E., Dawis, W. R., Green, A., Graham, G. A., Wallace, J. M., and Holland, O. B. (1990). Effects of calcium channel blockade on calcium homeostasis in mild to moderate essential hypertension. *Am. J. Med. Sci.* **300**, 133–137.



# Carbenoxolone

Decio Armanini and Cristina Fiore  
University of Padua, Padua, Italy

## Glossary

**aldosterone** The main mineralocorticoid hormone produced by adrenals.

**bioassay for mineralocorticoid effect** Measurement of the biological effect of mineralocorticoids at the level of target tissues; for example, the potassium/sodium ratio in urine and the rectal potential difference.

**carbenoxolone** A derivative of glycyrrhetic acid; the active principle of licorice root.

**11- $\beta$ -hydroxysteroid dehydrogenase type 1** The enzyme that converts inert cortisone to active cortisol in fat and liver.

**11- $\beta$ -hydroxysteroid dehydrogenase type 2** The enzyme that inactivates cortisol to cortisone at the level of the kidney and other classical target tissues for aldosterone, allowing aldosterone to bind to its receptors.

**glucocorticoid effect** The effect of glucocorticoids at the level of target tissues; for example, the effect on fat deposition and on liver metabolism.

**mineralocorticoid effect** The classical effect of aldosterone in retaining sodium and water and in excreting potassium in urine.

**plasma mineralocorticoid-receptor activity** A method that measures the sum of all the plasma substances that bind to mineralocorticoid receptors in kidney cytosol in relation to their plasma concentration and affinity for the receptor itself.

**pseudohyperaldosteronism** A syndrome characterized by the clinical picture of primary hyperaldosteronism and suppression of renin and aldosterone due to volume expansion related to endogenous or exogenous substances that retain sodium and water.

Carbenoxolone sodium is a hydrogen succinate of glycyrrhetic acid, the active principal of licorice, and has been used for 40 years in the treatment of gastritis and peptic ulcer. Optimal therapeutic effect in gastric ulcer is achieved with a dose of 50 to 100 mg three times daily. The drug can produce endocrine side effects at the level of water and electrolyte

balance and on the metabolism of cortisol, but it also has some potential therapeutic applications in other endocrine clinical situations.

## INTRODUCTION

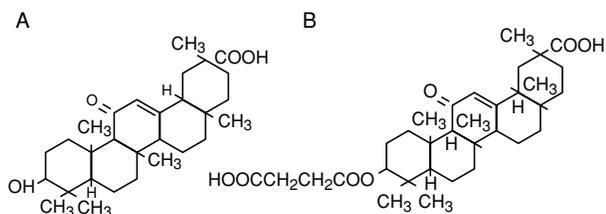
Carbenoxolone sodium is a hydrogen succinate of glycyrrhetic acid, the active principle of licorice (Fig. 1), and has been used for 40 years in the treatment of gastritis and peptic ulcer. Prolonged treatment with carbenoxolone can lead to the appearance of a syndrome of pseudohyperaldosteronism, which is characterized by water and sodium retention, hypertension, hypokalemia, low plasma renin activity, low aldosterone, and impaired glucose tolerance. These side effects are more evident in conditions that predispose to hypokalemia, in old age, and in association with chronic diseases of the liver, kidney, and heart.

## PATHOGENETIC MECHANISMS OF THE MINERALOCORTICOID ACTIVITY

After the introduction of carbenoxolone as a therapeutic agent in 1968, it was demonstrated that the drug shares some aldosterone-like effects in adrenalectomized patients and rats and that it can enhance the action of submaximal amounts of aldosterone, with both effects being abolished by therapy with the aldosterone receptor antagonist spironolactone. Spironolactone can also lessen the therapeutic effect of carbenoxolone. The mineralocorticoid effector mechanism of the drug was investigated by measuring the potassium/sodium ratio in urine or the rectal potential difference.

## Direct Mineralocorticoid Effect and Enhancement of Mineralocorticoid Activity of Aldosterone and of Other Endogenous Steroids

In 1982, Armanini and colleagues found that carbenoxolone has (1) a low but measurable affinity for



**Figure 1** Structure of glycyrrhetic acid (A) and its derivative carbenoxolone (B).

mineralocorticoid receptors in kidney cytosol (1/3000 that of aldosterone), (2) a direct mineralocorticoid effect *in vivo* in adrenalectomized rats at doses consistent with receptor affinity, and (3) a powerful action of amplifying the electrolyte effects of near maximal doses of aldosterone when administered to adrenalectomized rats. Plasma concentration of carbenoxolone after 1 week of oral administration is up to 10 mg/dl, a value consistent with a direct mineralocorticoid effect of the drug, even considering the low receptor affinity.

In another study, Armanini and colleagues evaluated the plasma mineralocorticoid receptor activity, which measures the levels of all plasma factors that bind to mineralocorticoid receptors of rat kidney cytosol, in six volunteers treated with carbenoxolone. The volunteers showed a clear syndrome of pseudohyperaldosteronism, and the plasma mineralocorticoid receptor activity decreased significantly at day 3 of treatment, suggesting a local effect of carbenoxolone in amplifying endogenous steroid action. At day 7, the values were still decreased but were significantly higher than at day 3, suggesting a direct mineralocorticoid effect of the drug as well.

In 1990, Morris and Souness demonstrated that carbenoxolone (1) amplifies the antinatriuretic activity of aldosterone and of desoxycorticosterone, which are not affected by 11-hydroxysteroid dehydrogenase, and (2) confers mineralocorticoid properties to steroids that are not mineralocorticoids such as 11-desoxycortisol and 2 $\alpha$ -methyl cortisone.

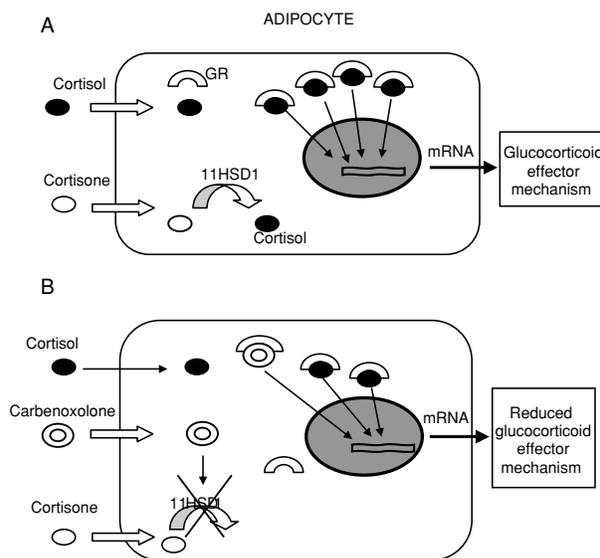
### Inhibitory Effect of 11- $\beta$ -Hydroxysteroid Dehydrogenase

In 1988, Stewart and colleagues emphasized the implication of 11- $\beta$ -hydroxysteroid dehydrogenase in the effector mechanism of aldosterone. This enzyme has two isoforms: (1) 11-hydroxysteroid dehydrogenase type 1 (11HSD1), which is present in the liver and fat and which activates cortisone to cortisol, regulating the physiological functions of these tissues (Fig. 2), and (2) 11-hydroxysteroid dehydrogenase

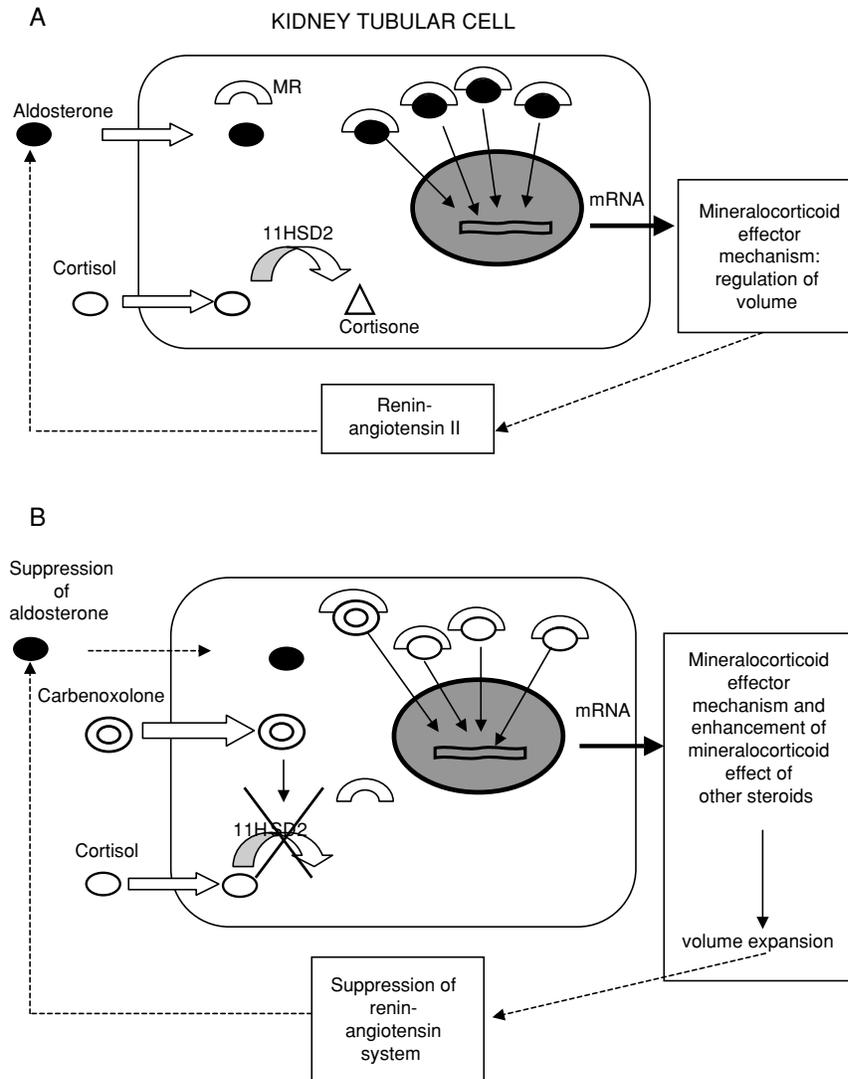
type 2 (11HSD2), which is present in the kidney (Fig. 3) and other classical target tissues for aldosterone (e.g., colon, salivary and sweat glands), where it inactivates cortisol to cortisone, thereby allowing aldosterone to bind to its renal receptors (when the enzyme is not active, e.g., when cortisol becomes a potent mineralocorticoid and produces a severe pseudohyperaldosteronism syndrome due to an inactivating mutation of the 11HSD2 gene). The authors demonstrated that carbenoxolone (1) increases the urinary-free cortisol/free cortisone ratio due to the block of 11HSD2 in the kidney and (2) has no effect on the urinary ratio of tetrahydrocortisol + allo-tetrahydrocortisol/tetrahydrocortisone, which are produced by the liver from cortisol and cortisone and are excreted in urine (thus, their ratio is dependent on liver 11HSD2 and does not increase when this isoform is inactivated).

### Central Effects of Carbenoxolone

In 1992 Gomez-Sanchez and Gomez-Sanchez reported that the infusion of carbenoxolone into the lateral ventricle of the brain produces hypertension without affecting saline appetite, meaning that hypertension was not linked to sodium and volume,



**Figure 2** (A) Mechanism of action of cortisol in fat and liver. The availability of cortisol is increased by the activation of inert cortisone to active cortisol. (B) Mechanism of action of carbenoxolone. The availability of cortisol is decreased due to the inhibition of 11-hydroxysteroid dehydrogenase type 1. This effect reduces the depositing of triglycerides in adipocytes and the differentiation of preadipocytes into adipocytes. GR, glucocorticoid receptor.



**Figure 3** (A) Mechanism of action of aldosterone. Mineralocorticoid receptors are protected from cortisol by the enzyme 11-hydroxysteroid dehydrogenase type 2, which inactivates cortisol to inert cortisone. (B) Mechanisms of mineralocorticoid action of carbenoxolone. The drug can bind directly to mineralocorticoid receptors, can amplify or confer mineralocorticoid activity to other steroids, and can block 11-hydroxysteroid dehydrogenase type 2. All of these effects produce a volume expansion and the suppression of renin and aldosterone. MR, mineralocorticoid receptor.

consistently a central role of mineralocorticoids in blood pressure control.

### Carbenoxolone and Heart Fibrosis

More recent studies have focused on the importance of aldosterone as a profibrotic agent, in particular at the heart level. In 2003, Young and colleagues demonstrated that carbenoxolone can produce a marked mineralocorticoid effect in coronary vascular inflammatory responses, probably mediated by binding of glucocorticoids to mineralocorticoid receptors

due to the block of inactivation of cortisol by 11HSD2. At high concentrations, carbenoxolone can also enhance vasoconstrictor action via selective toxicity to the endothelium and elimination of endothelium-dependent relaxation.

### PATHOGENETIC MECHANISMS OF THE GLUCOCORTICOID ACTIVITY

Carbenoxolone binds to glucocorticoid receptors with low affinity and inhibits 11HSD1, thereby reducing the availability of cortisol at the level of the liver and

adipose tissue (Fig. 2B). The inhibition of 11HSD1 is involved in the physiological regulation of these tissues. An increased activity of the enzyme is probably involved in visceral obesity, which is related to a local action of cortisol.

## CARBENOXOLONE AS A THERAPEUTIC AGENT

### Pseudohypoaldosteronism

Carbenoxolone can suppress the renin–aldosterone system in some patients with pseudohypoaldosteronism type 1. The syndrome is characterized by salt wasting, hyperkalemia, hyperreninemic hyperaldosteronism, and unresponsiveness to mineralocorticoids. In some cases, an inactivating mutation of the mineralocorticoid receptor gene has been demonstrated. The response to carbenoxolone suggests that there is at least a partly functional mineralocorticoid receptor in some patients with renal pseudohypoaldosteronism.

### Obesity

Carbenoxolone in effect has a dramatic inhibitory effect on preadipocyte differentiation into mature adipocytes and reverses the inhibition of insulin release by pancreatic insulae, probably due to the block of 11HSD at this level.

### Hypercholesterolemia

In 2003, Andrews and colleagues found that carbenoxolone reduces total cholesterol in healthy individuals, but not in diabetics, and reduces the release of free fatty acids, the glucose uptake, and the incorporation of glucose into triglycerides in rat epididymal fat pad. All of these studies have pointed out the importance of 11HSD1 in fat tissue distribution and the usefulness of inhibitors of the enzyme in reducing fat mass.

### 21-Hydroxylase Deficiency

This syndrome is characterized by reduced cortisol synthesis. The effect of carbenoxolone on 11HSD1 can explain the ability of the drug to augment the adrenal androgen-suppressing activity of hydrocortisone in patients with 21-hydroxylase deficiency. This combination could improve the treatment of congenital adrenal hyperplasia.

### Gonadic Dysfunction

By blocking 11HSD1, carbenoxolone reduces the availability of cortisol and testosterone in Leydig cells in women with hyperandrogenism.

## CONCLUSION

Carbenoxolone was first introduced as a therapeutic agent. Subsequently, its use was greatly limited by the fear of important side effects. Now, interest in this agent is once again growing due to the finding of a potential therapeutic application in some endocrine diseases.

### See Also the Following Articles

Mineralocorticoids and Mineralocorticoid Excess Syndromes • Tissue Renin-Angiotensin-Aldosterone System

### Further Reading

- Armanini, D., Karbowiak, I., Krozowski, Z., Funder, J. W., and Adam, W. R. (1982). The mechanism of action of carbenoxolone. *Endocrinology* **111**, 1683–1686.
- Armanini, D., Scali, M., Zennaro, M. C., Karbowiak, I., Wallace, C., Lewicka, S., Vecsei, P., and Mantero, F. (1989). The pathogenesis of pseudohyperaldosteronism from carbenoxolone. *J. Endocrinol. Invest.* **12**, 337–341.
- Bader, T., Zoumakis, E., Friedberg, M., Chrousos, G. P., and Hochberg, Z. (2002). Human adipose tissue under *in vitro* inhibition of 11-beta-hydroxysteroid dehydrogenase type 1: Differentiation and metabolism changes. *Horm. Metab. Res.* **34**, 752–757.
- Gomez-Sanchez, E. P., and Gomez-Sanchez, C. E. (1992). Central hypertensinogenic effects of glycyrrhetic acid and carbenoxolone. *Am. J. Physiol.* **263**, E1125–E1130.
- Irony, I., and Culture, G. B. (1999). Effect of carbenoxolone on the plasma renin activity and hypothalamic–pituitary–adrenal axis in congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Clin. Endocrinol. (Oxf.)* **51**, 273–274.
- Lambert, B., Godard, C., and Jacquemin, C. (1978). Effect of carbenoxolone on lipolysis in rat adipose tissue. *J. Pharm. Pharmacol.* **30**, 301–303.
- Morris, D. J., and Souness, G. V. (1990). The 11-β-OHSD inhibitor, carbenoxolone, aldosterone, and 11-desoxycorticosterone. *Am. J. Physiol.* **258**, F756–F759.
- Pinder, R. M., Brogden, R. N., Sawyer, P. R., Sapeight, T. M., Spencer, R., and Avery, G. S. (1976). Carbenoxolone: A review of its pharmacological properties and therapeutic efficacy in peptic ulcer disease. *Drugs* **11**, 245–307.
- Stewart, P. M., Wallace, A. M., and Edwards, C. R. W. (1988). The mineralocorticoid activity of carbenoxolone. In “The Adrenal Hypertension: From Cloning to Clinic,” pp. 48–54. Serono Symposium, Tokyo.
- Young, M. J., Moussa, L., Dilley, R., and Funder, J. W. (2003). Early inflammatory responses in experimental cardiac hypertrophy and fibrosis: Effect of 11-beta-hydroxysteroid dehydrogenase inactivation. *Endocrinology* **144**, 1121–1125.



# Carbohydrate Metabolism and Hormone-Fuel Interrelationships

Keith Tornheim

*Boston University School of Medicine, Boston, Massachusetts, United States*

## Glossary

**carbohydrate** Sugar; chemical compounds of empirical formula  $(\text{CH}_2\text{O})_n$  that are carbon chains with an aldehyde or keto group and two or more hydroxyl groups.

**gluconeogenesis** Metabolic process in which glucose is synthesized from nonsugars.

**glycogen** A branched polymer of glucose; used for storage in mammalian cells.

**glycolysis** Metabolic process in which glucose is broken down to pyruvate or lactate.

**kinase** An enzyme that adds a phosphate to a substrate, usually from ATP.

Carbohydrates [named for their empirical formula,  $(\text{CH}_2\text{O})_n$ ] include simple sugars (monosaccharides), disaccharides consisting of two linked sugars, and polysaccharides such as the glucose storage polymer glycogen.

## INTRODUCTION

By definition, simple sugars are aldehydes or ketones with two or more hydroxyl groups; thus, the smallest sugars are the trioses (three-carbon sugars) glyceraldehyde and dihydroxyacetone. Sugars with three to seven carbons are commonly involved in cellular metabolism. Glucose is a hexose (six-carbon sugar) with an aldehyde group at one end; by convention, this is carbon-1. Fructose is also a hexose, but it has a keto group at carbon-2. Sucrose (common table sugar) is a disaccharide of glucose and fructose.

Glucose is the most important of the biological sugars. It is the essential fuel of the brain and thus control of its blood concentration is critical. It is also

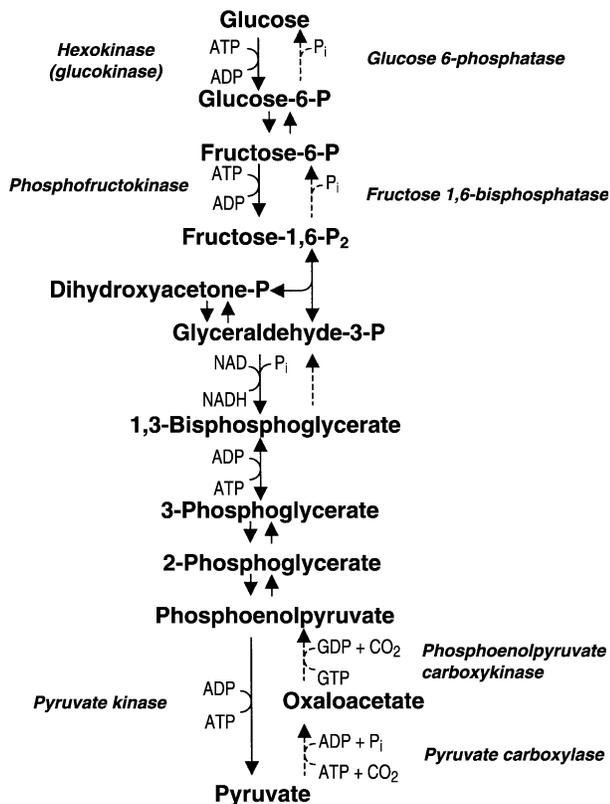
important in powering intense muscular work. A major metabolic function of the liver is to maintain blood glucose through appropriate release of the sugar from the storage polymer glycogen and through resynthesis of glucose from nonsugar precursors (gluconeogenesis). The blood glucose concentration is under hormonal control, with insulin promoting glucose uptake and storage in muscle and liver and the counterregulatory hormones, such as glucagon and epinephrine, promoting glycogen breakdown as well as gluconeogenesis in liver. Lack of or insufficient insulin is the cause of diabetes, and the resulting chronic elevated glucose levels lead to vascular damage and much of the cases of blindness and kidney disease and increased risk of heart disease. This article discusses the metabolism of glucose in the glycolytic pathway and the production of energy; the storage of glucose in glycogen and its breakdown; the resynthesis of glucose through the pathway of gluconeogenesis; the branch to the pentose phosphate pathway that produces ribose 5-phosphate, which is necessary for DNA and RNA synthesis; and how the metabolism of glucose is linked to the secretion of insulin in the  $\beta$  cells of the pancreatic islets.

## GLYCOLYSIS

The chemical logic of the glycolytic pathway is the generation of high-energy phosphate compounds that can be used to phosphorylate adenosine diphosphate (ADP) to adenosine triphosphate (ATP). ATP is the energy currency of the cell. The breakdown of ATP is used to power energy-consuming reactions of the cell, from biochemical syntheses to ion pumping to muscular contraction. The early steps in glycolysis actually involve the consumption of two ATPs, but in the later reactions four ATPs are produced, giving a net production of two ATPs per glucose molecule.

## Reactions

The reactions of glycolysis are shown in Fig. 1. First, glucose is phosphorylated by the enzyme hexokinase (“kinase” means a phosphorylating enzyme) to yield glucose 6-phosphate. Glucose 6-phosphate is isomerized from the aldo sugar to the keto sugar fructose 6-phosphate, which is then phosphorylated at the other end by the key enzyme phosphofructokinase to generate fructose 1,6-bisphosphate. (By current convention, “bis” indicates that the phosphates are on separate carbons, not attached together as in adenosine diphosphate.) Fructose 1,6-bisphosphate is cleaved in the aldolase reaction to two triose phosphates, dihydroxyacetone phosphate and glyceraldehyde 3-phosphate. Since the rest of glycolysis proceeds from the latter, dihydroxyacetone phosphate is converted to glyceraldehyde 3-phosphate by triose phosphate isomerase so



**Figure 1** The glycolytic (glucose breakdown) and gluconeogenic (glucose synthesis) pathways. The purely glycolytic enzymes are listed on the left. All tissues have hexokinase, but liver and pancreatic  $\beta$  cells have the high  $K_m$  isoform glucokinase that makes them sensitive to the glucose concentration. The purely gluconeogenic enzymes are listed on the right and their reactions are shown with dashed arrows. Liver and kidney have glucose 6-phosphatase, which allows them to produce glucose. All other enzyme reactions shown are used in both glycolysis and gluconeogenesis.

that both halves of the glucose molecule can be metabolized. In the glyceraldehyde 3-phosphate dehydrogenase reaction, the aldehyde of glyceraldehyde 3-phosphate is oxidized to a carboxylic acid, but the energy of oxidation is used in part for the uptake of inorganic phosphate ( $P_i$ ) into a phosphate anhydride in the product 1,3-bisphosphoglycerate. In this reaction, the acceptor of the reducing equivalents is the cofactor nicotinamide adenine dinucleotide (NAD), which becomes NADH with the addition of the hydride (hydrogen with both electrons), whereas a second hydrogen is released simply as a proton in solution. The phosphate anhydride generated at carbon-1 is high energy and thus can be used to phosphorylate ADP to ATP in the phosphoglycerate kinase reaction (so named from the reverse reaction). 3-Phosphoglycerate is then converted to the second high-energy intermediate phosphoenolpyruvate in two steps: The phosphate is moved from the 3 to the 2 position by phosphoglycerate mutase and then water is removed in the enolase reaction. Phosphoenolpyruvate is then used to phosphorylate ADP to ATP by pyruvate kinase. It is the highly favored conversion of the initial product enolpyruvate to the normal keto form of pyruvate that pulls the reaction forward and causes phosphoenolpyruvate to be a high-energy phosphate donor.

## Fate of Pyruvate

The fate of pyruvate depends on the oxidative state of the tissue. For glycolysis to proceed, the NADH produced in the glyceraldehyde 3-phosphate dehydrogenase reaction must be reoxidized back to NAD. Under aerobic conditions, the reducing equivalents from NADH may be transferred to the mitochondrial electron transport chain and ultimately to molecular oxygen. However, if there is insufficient oxygen or insufficient activity of the electron transport chain, then the pyruvate must be used to oxidize NADH and regenerate NAD in the lactate dehydrogenase reaction. This is the reason why strong muscular exercise produces lactate (lactic acid). Yeast lack lactate dehydrogenase and so instead under anaerobic conditions the pyruvate is first decarboxylated to acetaldehyde, which is then used in the alcohol dehydrogenase reaction, generating ethanol in the reoxidation of NADH to NAD. This is why fermentation to produce wine or other alcoholic beverages must be done anaerobically in sealed containers; otherwise, the acetaldehyde would be oxidized to acetic acid (vinegar), making a less palatable drink.

If pyruvate is not needed to be converted to lactate in order to reoxidize NADH, then the pyruvate can enter the mitochondria and be converted to acetyl-coenzyme A (CoA) in the pyruvate dehydrogenase reaction. Acetyl-CoA can be completely oxidized to CO<sub>2</sub> in the citric acid cycle. ATP production by oxidative phosphorylation from the two molecules of pyruvate, together with the reducing equivalents transferred from the two NADHs from the glyceraldehyde 3-phosphate dehydrogenase reaction, is on the order of 15 times greater than the two ATPs produced in glycolysis alone. The details of the citric acid cycle and oxidative phosphorylation are not considered here. If there is excess use of glucose beyond what would be necessary for energy production, then acetyl-CoA can also be used for synthesis of fatty acids.

## Regulation of Glycolysis

### **Phosphofructokinase**

The key regulatory enzyme of glycolysis is phosphofructokinase. It is inhibited by ATP and citrate and activated by AMP (and ADP), P<sub>i</sub>, and fructose 2,6-bisphosphate. Although ATP is a substrate of the enzyme, it is also an important product of glycolysis, and the inhibition at a regulatory site is an example of classic feedback inhibition. Since ATP usage produces AMP, ADP, and P<sub>i</sub>, these activators also signal a need for more ATP production, as would a decrease in ATP. Citrate may serve as an indicator of the sufficiency of alternative fuel, particularly fatty acids that are broken down to acetyl-CoA, which may then increase citrate so that glucose may be preserved for the brain. The importance of fructose 2,6-bisphosphate as an activator has principally been established in liver; it is discussed later in the context of reciprocal regulation of gluconeogenesis and glycolysis. (It is made by phosphorylation of fructose 6-phosphate on the 2 position by a phosphofructokinase-2, distinct from the glycolytic phosphofructokinase.)

### **Hexokinase and Glucose Transport**

Hexokinase is nominally the first step in glycolysis, but it is not the major point of regulation because glucose 6-phosphate is a branch point leading to glycogen and to the pentose phosphate pathway. Importantly, glucose 6-phosphate is an inhibitor of hexokinase, so if the other pathways are slow and if phosphofructokinase is inhibited, then glucose 6-phosphate will increase and inhibit hexokinase. Conversely, if phosphofructokinase is activated, such

as by a decrease in ATP and increase in AMP, then fructose 6-phosphate and glucose 6-phosphate will decrease and hexokinase will be disinhibited. Glucose transport into the cell is also stimulated by insulin in muscle and fat cells, which are tissues that have the insulin-regulated transporter Glut4. In contrast, glucose transport in and out of liver is rapid, catalyzed by Glut2.

## Regulation of Pyruvate Dehydrogenase

Regulation of the conversion of pyruvate to acetyl-CoA is of great importance because although pyruvate can be converted back to glucose in the process of gluconeogenesis, there cannot be a net production of glucose from acetyl moieties in animal cells. Pyruvate dehydrogenase is inhibited by its products acetyl-CoA and NADH, and this is countered by the corresponding substrates CoA and NAD; thus, it can be considered that the enzyme is inhibited by high acetyl-CoA/CoA and NADH/NAD ratios. In addition, the enzyme is inactivated by phosphorylation by a specific protein kinase, which is itself activated by high acetyl-CoA/CoA, NADH/NAD, and ATP/ADP ratios. The enzyme can be reactivated by removal of the phosphate by a specific protein phosphatase that is stimulated by high pyruvate levels and by insulin. The pyruvate dehydrogenase reaction is carried out by a multienzyme complex, to which the specific kinase and phosphatase are also bound.

## GLYCOGEN METABOLISM

### Reactions of Glycogen Breakdown

Liver and muscle can contain large amounts of glycogen, a polymer that is a storage form of glucose. It is a branched polymer, with glucose residues largely linked 1–4 but with branches approximately every 10 residues with a 1–6 linkage. The enzyme mainly responsible for breaking down glycogen is phosphorylase (which is short for glycogen phosphorylase because it was the first of its class discovered); it breaks 1–4 linkages by the addition of P<sub>i</sub>, releasing a residue as glucose 1-phosphate. However, it can only come within 4 residues of a branch point. Then a “transferase” transfers the branch, except for the penultimate residue, to the end of the main chain; α-1,6-glycosidase cleaves off the final branched residue as free glucose. With the branch removed, phosphorylase can continue.

Interestingly, the phosphorylase reaction can go in both directions,  $\text{glycogen}(n) + \text{P}_i \leftrightarrow \text{glucose}$

1-phosphate + glycogen( $n - 1$ ), with the equilibrium at  $[P_i]/[\text{glucose 1-phosphate}] = 3.7$ . Carl Cori, the discoverer of phosphorylase, received the Nobel prize for finding the first enzyme that can synthesize a biological polymer since glycogen can be formed in the test tube in the presence of high glucose 1-phosphate. However, *in vivo* the  $[P_i]/[\text{glucose 1-phosphate}]$  ratio is approximately 100, and the enzyme catalyzes glycogen breakdown. This was clearly demonstrated in individuals lacking phosphorylase who have a glycogen storage disease with excess glycogen, not low levels.

Since glucose 1-phosphate is in equilibrium with glucose 6-phosphate through the phosphoglucomutase reaction, the sugar residues released from glycogen can serve as fuel in the glycolytic pathway in muscle. In this context, the fact that glucose 6-phosphate is derived from glycogen phosphorolysis rather than hydrolysis means that there is a net three ATPs produced in glycolysis per glucose residue. In liver, the glucose 6-phosphate produced is generally dephosphorylated to provide blood glucose as needed, along with gluconeogenesis.

## Reactions of Glycogen Synthesis

In the synthesis of glycogen by the enzyme glycogen synthase, the glucosyl residue donor is UDP-glucose. UDP-glucose in turn is synthesized from glucose 1-phosphate and UTP, with the release of pyrophosphate (which can be cleaved by pyrophosphatase to pull the reaction). Glycogen synthase resembles phosphorylase in that it makes only 1–4 linkages. The branches in glycogen are made by branching enzyme, which takes a chain of seven residues from the end of a growing chain and attaches it in 1–6 linkage to a glucosyl residue four positions further on.

## Regulation of Glycogen Metabolism

The controlling enzymes of glycogen breakdown and synthesis are phosphorylase and glycogen synthase, respectively. Their regulation involves allosteric actions by activators and inhibitors as well as phosphorylation and dephosphorylation. The phosphorylation state affects the sensitivity to the allosteric effectors, and the effectors can modulate the phosphorylation state. Hormonal regulation is important to glycogen metabolism. Insulin, a 51-amino acid small protein secreted by the  $\beta$  cells of the pancreatic islets in response to high glucose, promotes glycogen storage. Glucagon, a 29-amino acid peptide secreted by the  $\alpha$  cells of the pancreatic islets in response to

low glucose, promotes glycogen breakdown in liver as well as gluconeogenesis. Glycogen breakdown in both liver and muscle is promoted by epinephrine, a catecholamine secreted by the adrenal medulla in response to nerve impulses, such as when a rabbit sees a fox.

Phosphorylase was originally found to exist in two forms: a nonphosphorylated form (phosphorylase *b*) that required AMP for activity and a phosphorylated form (phosphorylase *a*, with a phosphate on a serine residue of each of its four subunits) that was active in the absence of AMP. Phosphorylase *b* is also inhibited by glucose 6-phosphate and ATP. Glucagon and epinephrine can activate phosphorylase by causing a sequence of reactions (the cascade; Fig. 2) that results in the conversion of phosphorylase *b* to *a*. First, these hormones bind to their specific receptors. (Glucagon is only effective on liver and not on muscle because muscle does not have glucagon receptors.) The binding of the hormone causes activation of adenylyl cyclase, which makes cyclic AMP from ATP. This in turn activates the cyclic AMP-dependent protein kinase (PKA), which can then phosphorylate and activate phosphorylase *b* kinase. The latter protein kinase then phosphorylates phosphorylase *b*, converting it to phosphorylase *a*, which is more active and breaks down glycogen.

In addition to phosphorylation, phosphorylase *b* kinase can be activated by an increase in intracellular free calcium ion. Thus, when muscular contraction is initiated by a release of calcium from the sarcoplasmic reticulum, the increase in calcium can also trigger glycogenolysis to help provide the needed fuel.

Glycogen synthase is also controlled by phosphorylation/dephosphorylation. In this case, the

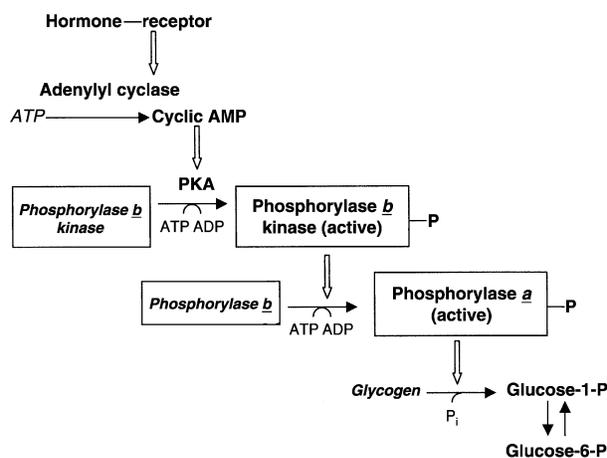


Figure 2 Stimulation of glycogen breakdown by hormones such as epinephrine and glucagon—the cascade.

phosphorylated form is the less active one (originally called glycogen synthase D because it is dependent on glucose 6-phosphate, but now it is called synthase b in correspondence with the less active phosphorylase b; synthase a was originally called synthase I because it is independent of glucose 6-phosphate). The hormonal cascade described previously can cause the phosphorylation of glycogen synthase by PKA, thus inhibiting glycogen synthesis as well as activating glycogen breakdown. However, the situation for glycogen synthase is more complex because there are at least 10 phosphorylation sites that can be phosphorylated by at least nine different protein kinases. The importance or hierarchy of these sites and their kinases remains to be clarified. Insulin's effect in promoting glycogen storage appears to be in part due to inhibition of the action of glycogen synthase kinase 3.

The protein phosphatases that convert phosphorylase a to b and glycogen synthase b to a, particularly the protein phosphatase 1 that has a glycogen-binding subunit, are subject to regulation in several ways. First, binding of allosteric regulators to the protein substrate affects the activity of phosphatases and kinases. Thus, AMP binding to phosphorylase a makes it a poor substrate for the phosphatase because the phosphate is tucked in so that it cannot be hydrolyzed, whereas binding of glucose or glucose 6-phosphate causes the phosphate to become accessible. Binding of glucose 6-phosphate to phosphorylase b also makes it a poorer substrate for the kinase as well as inhibiting phosphorylase activity. Second, inhibitory proteins can bind to the phosphatase, and the action of these proteins is regulated by their phosphorylation and dephosphorylation. Third, the fact that the same phosphatase works on both phosphorylase a and synthase b means that when the phosphatase is bound to phosphorylase a—and cannot hydrolyze it because the phosphate is tucked in—the phosphatase is prevented from working on glycogen synthase.

### Glycogen Storage Diseases

Genetic deficiency of enzymes of the liver related to glycogen breakdown can lead to the engorgement of the liver with glycogen. Thus, patients with Hers disease (type VI) lack liver phosphorylase and must therefore rely on gluconeogenesis to maintain blood glucose. Patients with Von Gierke disease (type I) lack glucose 6-phosphatase, cannot get glucose from either glycogenolysis or gluconeogenesis, and therefore can withstand only limited starvation. Presumably, they have increased glycogen because increased levels of glucose 6-phosphate promote

glycogen synthesis and inhibit glycogen breakdown, as described previously. In McArdle disease (type V), muscle phosphorylase is absent (indicating that there are separate muscle and liver isoforms from separate genes) and exercise capacity is limited.

## GLUCONEOGENESIS

Gluconeogenesis is the synthesis of glucose from non-sugar precursors, such as lactate, pyruvate, and the carbon skeleton of glucogenic amino acids. This is a major metabolic function of the liver since the brain, in particular, is dependent on glucose as a fuel.

### Reactions

In a sense, gluconeogenesis is a reversal of glycolysis, and many of the enzymes are the same for the reactions that are readily reversible. However, specifically gluconeogenic enzymes are needed to reverse three steps in glycolysis that have a large free energy drop. Thus, to reverse the pyruvate kinase reaction, first pyruvate is converted to oxaloacetate by pyruvate carboxylase, a reaction driven by ATP hydrolysis; then the oxaloacetate is decarboxylated and phosphorylated to generate phosphoenolpyruvate by the enzyme phosphoenolpyruvate carboxykinase using GTP (Fig. 1). Note that two high-energy phosphate bonds (from ATP and GTP) are needed to produce the very high-energy intermediate phosphoenolpyruvate. The reactions from there to fructose 1,6-bisphosphate occur via reversal of the glycolytic reactions. Then, to generate fructose 6-phosphate, the 1-phosphate is simply cleaved off by the gluconeogenic enzyme fructose 1,6-bisphosphatase (opposing the glycolytic enzyme phosphofructokinase). Fructose 6-phosphate equilibrates with glucose 6-phosphate in the phosphoglucose isomerase reaction as in glycolysis. Finally, glucose 6-phosphate is hydrolyzed to free glucose by glucose 6-phosphatase (opposing the hexokinase isoform glucokinase). Note that synthesis of a molecule of glucose requires the input of six ATP equivalents: the ATP at pyruvate carboxylase, the GTP at phosphoenolpyruvate carboxykinase, and the ATP at the phosphoglycerate kinase reaction for each of the two halves of the glucose molecule, plus two equivalents of NADH for reversing the glyceraldehyde 3-phosphate dehydrogenase reaction.

### Regulation

Control of net gluconeogenesis involves regulation of the opposing glycolytic enzymes and the correspond-

ing specifically gluconeogenic enzymes. The glucose phosphorylating activity in liver is largely glucokinase, a high  $k_m$  (10 mM) isoform of hexokinase that is responsive to changes in glucose in the physiological range. Glucokinase is not inhibited by glucose 6-phosphate because it lacks the regulatory binding domain present in the other hexokinases. On the other hand, glucokinase has a specific inhibitory protein whose effect is prevented by fructose 1-phosphate and increased by fructose 6-phosphate. (The disinhibition by fructose 1-phosphate explains the stimulation of glucose metabolism by fructose, which in the liver is largely phosphorylated to fructose 1-phosphate by fructokinase.) The liver isoform of pyruvate kinase has several regulatory properties designed to inhibit it so that gluconeogenesis can proceed: (i) strong allosteric inhibition by ATP; (ii) inhibition by alanine, an important gluconeogenic substrate; (iii) dependence on a high level of the activator fructose 1,6-bisphosphate (in order to have the same low  $k_m$  for phosphoenolpyruvate that the muscle isoform has with or without fructose 1,6-bisphosphate); and (iv) inhibition by phosphorylation by PKA in response to glucagon. It was originally thought that phosphofructokinase was similarly inhibited by phosphorylation by PKA, but in fact although phosphorylation of phosphofructokinase occurs, it appears to have little effect on the activity of the mammalian enzymes. Instead, PKA phosphorylation of phosphofructokinase-2/fructose 2,6-bisphosphatase, the bifunctional enzyme that both makes and degrades fructose 2,6-bisphosphate, inhibits the kinase and activates the phosphatase activities, thus reducing the level of fructose 2,6-bisphosphate. Since fructose 2,6-bisphosphate is an activator of the glycolytic phosphofructokinase (sometimes called phosphofructokinase-1 for clarity) as well as an inhibitor of fructose 1,6-bisphosphatase, glucagon can thus reduce glycolytic flux and promote gluconeogenesis at this step as well as at the pyruvate kinase step, in addition to its effect of stimulating glycogen breakdown and inhibiting glycogen synthesis. Fructose 1,6-bisphosphatase is also inhibited by AMP, in contrast to the AMP activation of phosphofructokinase. Pyruvate carboxylase has an absolute requirement for the activator acetyl-CoA; the rationale for this is that gluconeogenesis is quite energy consuming, the required ATP certainly cannot come from glucose metabolism, and the presence of acetyl-CoA would be indicative of sufficient alternative fuel, such as from fatty acids. Finally, these enzymes are adaptive in that the amount of the glycolytic enzymes glucokinase, phosphofructokinase, and pyruvate kinase is increased by a

high-carbohydrate diet, whereas the specifically gluconeogenic enzymes are increased by starvation or a low-carbohydrate diet.

### Substrate Cycling

The opposing gluconeogenic and glycolytic enzymes form energy-consuming substrate cycles—that is, if glucokinase and glucose 6-phosphatase were active simultaneously so that glucose were converted to glucose 6-phosphate and back to glucose, the net reaction would be the hydrolysis of ATP to ADP and  $P_i$ , and similarly for phosphofructokinase and fructose 1,6-bisphosphatase. If phosphoenolpyruvate were converted to pyruvate and back, there would be production of one ATP but utilization of two equivalents, for a net consumption of one ATP. It was originally thought that such substrate cycling would be a waste of energy (hence the original term futile cycle) and that there should be tight regulation to prevent simultaneous operation of the opposing enzymes. However, sophisticated radiolabeling experiments have shown the occurrence of such substrate cycling. For example, if liver cells undergoing gluconeogenesis are provided with glucose with tritium (a radioactive isotope of hydrogen) attached to carbon-5, then after the glucose is metabolized past the phosphofructokinase reaction, the tritium will be released into water on equilibration of glyceraldehyde 3-phosphate with dihydroxyacetone phosphate in the triose phosphate isomerase reaction. Such experiments have shown a substantial glycolytic rate that is largely reduced by glucagon treatment in order to increase the net gluconeogenic rate. Such substrate cycling has the advantage that there is no “dead” range in which neither glycolysis nor gluconeogenesis is operative and thus there is no control. Furthermore, substrate cycling amplifies cellular signals because changes in activators/inhibitors of glycolysis and/or gluconeogenesis lead to much greater percentage changes in the net rate. Finally, as a special case, bumblebees have unregulated fructose 1,6-bisphosphatase in their flight muscles so that substrate cycling with phosphofructokinase generates heat to warm up the muscles on cold days to allow flight, much as an electric heater is used to heat a gasoline engine preparatory to start an automobile in very cold climates.

### Cori Cycle and Alanine Cycle

There are two important interorgan cycles involving glycolysis in muscle and gluconeogenesis in liver. In the Cori cycle, glucose is metabolized to pyruvate and

then to lactate in muscle, the lactate is released into the blood and carried to the liver, where it is reconverted to pyruvate and used for gluconeogenesis, and the resulting glucose is released and travels back to muscle. Lactate is a particularly good gluconeogenic substrate because the reoxidation of lactate to pyruvate in the lactate dehydrogenase reaction also provides the NADH equivalents needed for gluconeogenesis. The alanine cycle is similar to the Cori cycle, except that muscle pyruvate is converted to the amino acid alanine rather than lactate by transamination in the glutamate–pyruvate transamination reaction. A very large proportion of the amino acid put out by muscle is alanine, much more so than its composition in muscle protein, due to the fact that the carbon skeleton derives from glycolytically generated pyruvate. In the liver, alanine is transaminated back to regenerate pyruvate, and the excess amino groups can be disposed of as urea since the urea cycle is also localized to the liver. The importance of alanine as a gluconeogenic substrate in this regard explains why it is a potent allosteric inhibitor of liver pyruvate kinase.

### PENTOSE PHOSPHATE PATHWAY

The pentose phosphate pathway also branches off from glycolysis at glucose 6-phosphate. It has two important products: ribose 5-phosphate, which is needed for synthesis of nucleotides and nucleic acids (DNA and RNA), and NADPH (the reduced form of nicotinamide adenine dinucleotide phosphate), which provides the reducing equivalents for synthetic reactions such as fatty acid biosynthesis. The first reaction of the pathway, catalyzed by glucose 6-phosphate dehydrogenase, generates one molecule of NADPH from NADP, and the sugar aldehyde group is oxidized to a carboxylic acid. However, since the glucose 6-phosphate substrate is in the pyranose (hemiacetal) ring form with the carbon-5 oxygen linked to carbon-1, the initial product is an internal ester, 6-phosphoglucono  $\delta$ -lactone, that is then cleaved by lactonase to yield 6-phosphogluconate. 6-Phosphogluconate dehydrogenase produces a second molecule of NADPH from NADP and also releases the carbon-1 carboxyl group as  $\text{CO}_2$ , leaving the 5-carbon compound ribulose 5-phosphate. The latter (a keto sugar) is converted to ribose 5-phosphate (an aldo sugar) by phosphopentose isomerase. This so-called oxidative branch of the pentose phosphate pathway thus produces two NADPHs and one pentose phosphate. If, however, this gives more pentose phosphate than is needed for nucleotide synthesis, the excess can be converted back

to glycolytic intermediates by the nonoxidative branch. Three pentose phosphates (15 carbons total) are converted to two molecules of fructose 6-phosphate and one glyceraldehyde 3-phosphate as follows: First, transketolase transfers 2 carbons from xylulose 5-phosphate (an epimer from ribulose 5-phosphate made by switching the orientation of the carbon-3 hydroxyl by phosphopentose epimerase) to ribose 5-phosphate, thereby producing the 7-carbon sugar sedoheptulose 7-phosphate plus glyceraldehyde 3-phosphate. Second, transaldolase transfers 3 carbons back from sedoheptulose 7-phosphate to glyceraldehyde 3-phosphate, producing fructose 6-phosphate and the 4-carbon sugar erythrose 4-phosphate. Finally, transketolase transfers 2 carbons to the erythrose 4-phosphate from another molecule of xylulose 5-phosphate, producing fructose 6-phosphate and glyceraldehyde 3-phosphate.

### GLUCOSE METABOLISM AND INSULIN SECRETION

Not only does insulin regulate glucose metabolism but also glucose metabolism in the pancreatic  $\beta$  cells regulates insulin secretion. The  $\beta$  cell is responsive to glucose in the physiological range because it has a high  $k_m$  glucokinase like liver. In the  $\beta$  cell, the increased glucose metabolism causes an increase in the ATP/ADP ratio, which closes ATP-sensitive  $\text{K}^+$  channels in the plasma membrane, causing membrane depolarization and an influx of  $\text{Ca}^{2+}$  through voltage-sensitive calcium channels that triggers secretion. There are also other glucose concentration-dependent enhancements of secretion that may involve lipid metabolism and signaling pathways, such as via protein kinase C. Interestingly, normal insulin secretion is pulsatile, and it has been proposed that this may be due to underlying oscillations in glucose metabolism.

### See Also the Following Articles

Carbohydrate Metabolism: Diabetes Mellitus, Genomic Aberrations • Glucose, Impaired Tolerance • Glucose Physiology, Normal • Glucose Toxicity

### Further Reading

- Berg, J. M., Tynoczko, J. L., and Stryer, L. (2002). "Biochemistry," 5th ed., Freeman, New York.
- Bollen, M., Keppens, S., and Stalmans, W. (1998). Specific features of glycogen metabolism in the liver. *Biochem. J.* **336**, 19–31.
- Lawrence, J. C., Jr., and Roach, P. J. (1997). New insights into the role and mechanism of glycogen synthase activation by insulin. *Diabetes* **46**, 541–547.

- Pilkis, S. J., Claus, T. H., Kurland, I. J., and Lange, A. J. (1995). 6-Phosphofructo-2-kinase/fructose-2,6-bisphosphatase: A metabolic signaling enzyme. *Annu. Rev. Biochem.* **64**, 799–835.
- Prentki, M., Tornheim, K., and Corkey, B. E. (1997). Signal transduction mechanisms in nutrient induced secretion. *Diabetologia* **40**, S32–S41.
- Ruderman, N. B., Tornheim, K., and Goodman, M. N. (2001). Fuel homeostasis and intermediary metabolism of carbohydrate, fat, and protein. In “Principles and Practices of Endocrinology and Metabolism” (K. L. Becker, ed.), 3rd ed., pp. 1257–1271. Lippincott, Philadelphia.
- Van Schaftingen, E., Veiga-da-Cunha, M., and Niculescu, L. (1997). The regulatory protein of glucokinase. *Biochem. Soc. Trans.* **25**, 136–140.



# Carbohydrate Metabolism: Diabetes Mellitus, Genomic Aberrations

Mona M. Sabra

*University of Maryland School of Medicine, Baltimore, Maryland, United States*

Alan R. Shuldiner

*University of Maryland School of Medicine and Baltimore VA Medical Center, Baltimore, Maryland, United States*

## Glossary

**acanthosis nigricans** A scaly and pigmented skin rash, usually on the back of the neck or armpits.

**autoimmune** A condition in which the immune system attacks certain cells, tissues, or organs, resulting in their being damaged or destroyed.

**autosomal dominant** The pattern of inheritance of a disease resulting from the inheritance of a single mutated gene that is passed down from one parent, with each child of an affected parent having a 50% chance of inheriting the condition.

**autosomal recessive** The pattern of inheritance of a disease resulting from the inheritance of two copies of a mutated gene, one from each parent; there is a 25% chance of conceiving an affected child with each pregnancy from unaffected parents each with a single copy of the mutated gene.

**heterozygous** The presence of two different copies of a given gene.

**homozygous** The presence of two of the same copies (normal or mutated) of a given gene.

**hyperandrogenism** Abnormally high levels of male sex hormones, which in woman can result in abnormal facial hair growth, an enlarged clitoris, and/or irregular or absent menses.

**hypertriglyceridemia** Abnormally high levels of triglycerides (fat) in the blood.

**insulin** A hormone secreted by beta-cells of the pancreas that promotes the uptake of glucose from the bloodstream into cells where the glucose is used as energy or is stored for future use.

**insulin resistance** Decreased ability of insulin to promote glucose uptake, use, and storage.

**ketoacidosis** A condition in which toxic ketones and acids accumulate in the bloodstream and organs of the body.

**penetrance** The likelihood that a mutated gene, when inherited, will result in a disease.

**polygenic** The involvement of multiple genes in the development of a disease or condition.

Diabetes mellitus is a chronic disease defined by hyperglycemia (high blood glucose or sugar levels). Varying degrees of insulin resistance and/or dysfunction of the insulin-producing beta-cells in the pancreas cause diabetes. Chronic hyperglycemia is associated with the potential for developing serious complications such as eye disease (retinopathy), kidney disease (nephropathy), nerve disease (neuropathy), and vascular disease (heart attack, stroke, and lower limb amputation). Diabetes mellitus is divided into several different subtypes depending on the underlying genetic and pathophysiological cause. Forms of diabetes due to mutations in single genes tend to be rare, have a high penetrance, and have little influence from environmental factors, whereas the more common forms

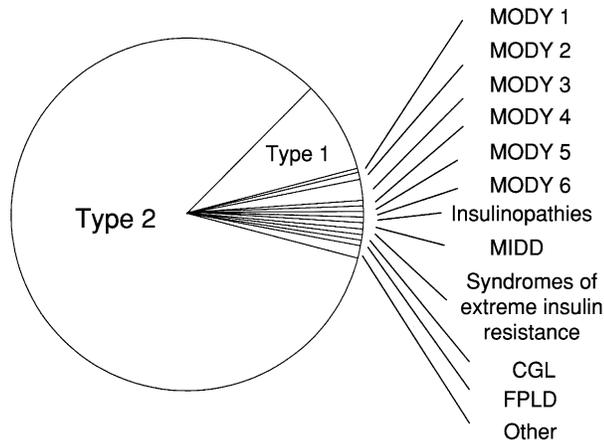
of diabetes (e.g., type 1, type 2) are heterogeneous and caused by multiple genes, with penetrance influenced greatly by the environment (Fig. 1).

## SINGLE-GENE CAUSES OF DIABETES

A list of characteristic features of monogenic diabetes syndromes is provided in Table I.

### Maturity-Onset Diabetes of the Young

Maturity-onset diabetes of the young (MODY) is autosomal dominantly transmitted and is characterized



**Figure 1** Subtypes of diabetes. MODY, maturity onset diabetes of the young; MIDD, maternally inherited diabetes and deafness; CGL, congenital generalized lipotrophic diabetes mellitus; FPLD, familial partial lipotrophic diabetes mellitus.

by onset during childhood or early adulthood, lack of obesity (in most cases), and impaired insulin secretion from the beta-cells of the pancreas without evidence of antibodies against beta cells. MODY is suspected when diabetes is diagnosed during the first to third decades of life and is found in two or more consecutive generations of the same family, with at least one member being diagnosed under 25 years of age. There are at least six forms of MODY differentiated by the specific genetic defects. MODY2 is caused by a mutation in glucokinase, an enzyme produced in pancreatic beta-cells and important for glucose metabolism and sensing and insulin secretion. Hyperglycemia in patients with MODY2 is typically mild and can usually be treated by diet. Chronic diabetic complications are relatively uncommon. MODY1, MODY3, MODY4, MODY5, and MODY6 are caused by mutations in transcription factors that are expressed in the pancreas

**Table 1** Characteristic Features of Monogenic Diabetes Syndromes

Syndrome	Mode of inheritance	Gene	Chromosome	Distinguishing features
Maturity onset diabetes of the young (MODY)	Autosomal dominant	HNF4 $\alpha$ (MODY1) Glucokinase (MODY2) HNF1 $\alpha$ (MODY3) IPF (MODY4) HNF1 $\beta$ (MODY5) Neuro D/Beta2 (MODY6)	20q12–13.1 7p15–13 12q24.2 13q12.1 17cen–q21.3 2q32	Nonobese, onset during young adulthood, deficient insulin secretion
Leprechaunism	Autosomal recessive	Insulin receptor	19p13.2	Extreme insulin resistance, characteristic faces with large ears and small jaw, intrauterine growth retardation, postnatal growth retardation, acanthosis nigricans, sparse fat tissue, low fasting blood glucose, death during first year
Type A syndrome	Autosomal recessive	Insulin receptor	19p13.2	Extreme insulin resistance, acanthosis nigricans, hyperandrogenism
Rabson–Mendenhall syndrome	Autosomal recessive	Insulin receptor	19p13.2	Extreme insulin resistance, abnormal dentition and nails, pineal gland overgrowth, protuberant abdomen, thick rapidly growing hair, early puberty, accelerated growth
Familial partial lipotrophic diabetes (FPLD)	Autosomal dominant	Lamin A/C	1q21.2	Lack of fat usually limited to the lower extremities, insulin resistance, hypertriglyceridemia
Congenital generalized lipotrophic diabetes (CGL)	Autosomal recessive	$\gamma$ 3-like protein (GNG3, seipin) 1-acylglycerol-3-phosphate D-acyltransferase 2 (AGPAT2)	11q13 9q34	Complete lack of subcutaneous fat, Herculean appearance, insulin resistance, hypertriglyceridemia
Maternally inherited diabetes and deafness (MIDD)	Maternal	tRNA <sup>Leu</sup> , deletions	Mitochondrial DNA	Diabetes manifest mostly by decreased insulin secretion and also accompanied by deafness or milder degree of hearing loss

Note: HNF, hepatocyte nuclear factor; IPF, insulin promoter factor; tRNA, transfer ribonucleic acid.

and are thought to play a role in beta-cell development and function. These include hepatocyte nuclear factors 4 $\alpha$  (MODY2), 1 $\alpha$  (MODY3), 1 $\beta$  (MODY4), insulin promoter factor-1 (MODY5), and neuroD/beta2 (MODY6). HNF1 $\alpha$  (MODY3) is the most frequent gene mutated in MODY. Patients with MODY due to mutations in the aforementioned transcription factors typically have a progressive decrease in insulin secretion with increasing endogenous insulin requirements. Long-term diabetes complications occur with similar frequency to more typical forms of diabetes. In addition, renal cystic disease is a common feature of MODY1. Specific diagnostic genetic testing for MODY is currently available only in research settings.

### Genetic Syndromes of Extreme Insulin Resistance

Genetic syndromes of extreme insulin resistance are usually autosomal recessive and due to mutations in the insulin receptor gene. Without properly functioning insulin receptors, insulin cannot effectively stimulate glucose uptake and use. More than 50 such mutations have been identified. Characteristically, glucose levels are normal to elevated and insulin levels are very high (due to compensatory increases in beta-cell insulin secretion). At least three distinct clinical syndromes of extreme insulin resistance exist: type A syndrome, leprechaunism, and Rabson–Mendenhall syndrome. Characteristic features distinguish these syndromes (Table D). The type or location of the mutation does not correlate with the syndrome in any obvious way. These syndromes are very rare, with an estimated incidence of 1 in 4 million live births. Despite very high requirements for insulin (after several thousand units per day), good blood glucose is often difficult to achieve in these patients and chronic diabetic complications are not uncommon.

### Familial Partial Lipomatrophic Diabetes

Familial partial lipomatrophic diabetes (FPLD), also referred to as Dunnigan-type familial partial lipomatrophy or face-sparing lipomatrophy, is an autosomal dominant disorder that results from mutations in the lamin A/C gene located on chromosome 1q21.2. The affected patient has normal fat distribution until puberty, when fat tissue decreases or disappears in the arms, trunk, and lower extremities. The face is typically spared from fat loss, and fat may actually accumulate abnormally in the axillae, back, labia majora, and intra-abdominal region. The patient is insulin

resistant and develops glucose intolerance or overt diabetes during the fourth to fifth decades of life. Hypertriglyceridemia with low levels of high-density lipoprotein (HDL) cholesterol is characteristic and may result in pancreatitis.

### Congenital Generalized Lipomatrophic Diabetes

Congenital generalized lipomatrophic diabetes (CGL), or Berardinelli–Seip syndrome, is a rare, autosomal recessive disorder characterized by the near absence of adipose tissue from birth in association with severe insulin resistance with glucose intolerance or overt diabetes, hyperandrogenism and early puberty, acanthosis nigricans, hypertriglyceridemia, enlarged liver, and prominent muscles. Approximately half of these cases are caused by mutations in the  $\gamma$ 3-linked gene (GNG3; also called the seipin gene) on chromosome 11q13. Other cases of CGL are due to mutations in the 1-acylglycerol-3-phosphate-D-acetyltransferase 2 (AGPAT2) gene on chromosome 9q34.

### Other Single-Gene Diabetes Syndromes

Insulinopathies secondary to mutations in the insulin gene, and maternally inherited diabetes and deafness (MIDD) secondary to mutations in mitochondrial deoxyribonucleic acid (DNA), are also examples of rare single-gene diabetes syndromes. Other genetic diseases in which diabetes is often a feature include Wolfram syndrome, hemochromatosis, Freidreich ataxia, cystic fibrosis, and thiamine-responsive megaloblastic anemia syndrome.

## GENETICS OF COMMON FORMS OF DIABETES

The single-gene forms of diabetes described so far constitute no more than 2 to 5% of cases of diabetes. The more common forms of diabetes are known to have important genetic underpinnings but are not inherited in a predictable pattern. This observation suggests that the common forms of diabetes are likely to be due to several (many) mutant genes that are relatively common in the population. Each gene variant individually is likely to have a modest effect, but together the mutant genes act additively or synergistically. Furthermore, environmental provocations are thought to have important effects on the likelihood that a given gene variant or group of gene variants will express themselves. Thus, the common forms

of diabetes are classified genetically as being heterogeneous and complex.

### Type 1 Diabetes Mellitus

Type 1 diabetes mellitus (T1DM) occurs predominantly in children and adolescents. In the United States, T1DM accounts for approximately 5 to 10% of all diabetic cases diagnosed. The incidence of T1DM varies greatly worldwide, with the highest reported incidence in Northern Europe and the lowest rates in China, Korea, and Mexico. T1DM is characterized by autoimmune-mediated dysfunction and destruction of the beta-cells of the endocrine pancreas, resulting in insulin deficiency and potential to ketoacidosis, coma, and death. The beta-cell loss begins months to years before the actual clinical onset of insulin dependency. Environmental factors such as viral infections and toxins have been implicated in the pathophysiology of this disorder. These environmental provocations may activate cells of the immune system (T lymphocytes, B lymphocytes, and natural killer cells), initiating and perpetuating a chronic inflammatory reaction that ultimately leads to beta-cell loss and insulin deficiency. This autoimmune process is influenced by a cluster of genes on chromosome 6, termed the human leukocyte antigen (HLA) locus, which is known to modulate immune function. The presence of certain HLA haplotypes appears to predispose the individual to T1DM, whereas other HLA haplotypes appear to be protective. Other genes have been implicated in defining the risk of developing T1DM, including variants of the insulin gene (chromosome 11p15.5) and the cytotoxic T-cell lymphocyte associated-4 (CTLA4) gene (chromosome 2q33) as well as several other chromosomal loci for which the specific genes have yet to be determined.

### Type 2 Diabetes Mellitus

Type 2 diabetes mellitus (T2DM) is the most common form of diabetes, accounting for at least 90% of all cases. Its prevalence has reached epidemic proportions, with 16 million people having the disease in the United States alone. Hyperglycemia and overt diabetes result from underlying insulin resistance and impaired insulin secretion secondary to beta-cell dysfunction and failure. Unlike T1DM, the beta cell failure in T2DM is not autoimmune in origin. Both environmental factors (caloric excess and physical inactivity) and genetic predisposition have been implicated in the pathogenesis of T2DM. The strongest

evidence for genetic predisposition to T2DM comes from identical twin studies (with 60–90% concordance rates) and studies that show clustering of the disease in families. The disorder is thought to be heterogeneous and polygenic in nature. Over the past two decades, studies of specific genes have led to the identification of several putative T2DM susceptibility genes. These genes include peroxisome proliferator-activated receptor-gamma (PPARG, chromosome 3p25), beta-3-adrenoceptor (ADRB3, chromosome 8p12–p11.2), fatty acid-binding protein-2 (FABP2, chromosome 4q28–31), plasma cell glycoprotein (PC1, chromosome 6q22–23), insulin receptor substrate-1 (IRS1, chromosome 2q36), regulatory subunit of protein phosphatase 1 (PPP1R3A, chromosome 7q31.1–q31.2), and potassium inwardly rectifying channel (KCNJ9, 11p15.1). Genome-wide searches for T2DM susceptibility genes have resulted in the identification of variants in the calpain 10 gene (CAPN10). Several other chromosomal regions have been implicated, most notably regions on chromosomes 1 and 20.

### CONCLUSION

Diabetes is a heterogeneous group of disorders defined by hyperglycemia. Several forms of diabetes are caused by mutations in single genes. Although these single-gene forms of diabetes are rare (approximately 2–5% of cases), they have provided important insights into glucose regulation and into the underlying disease process. By contrast, T1DM and T2DM are far more common and are due to the interaction of multiple genes with the environment. Identification of the genes responsible for these common forms of diabetes has been slower, but progress is being made. Identification of susceptibility genes for diabetes will lead to new strategies for treatment and prevention.

### See Also the Following Articles

Cardiovascular Disease in Diabetes • Carbohydrate Metabolism and Hormone-Fuel Interrelationships • Diabetes Mellitus, Diagnosis and Treatment in the Elderly • Diabetes, Type 1 • Diabetes, Type 2 • Hypertension and Diabetes • Obesity and Diabetes, Regulation of Food Intake

### Further Reading

Boyle, J. P., Honecutt, A. A., Narayan, K. M., *et al.* (2001). Projection of diabetes burden through 2050: Impact of changing demography and disease prevalence in the U.S. *Diabetes Care* **24**, 1936–1940.

- Fajans, S. S., Bell, G. I., and Polonsky, K. S. (2001). Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young. *N. Engl. J. Med.* **345**, 971–980.
- Hsueh, W.-C., Mitchell, B. D., and Shuldiner, A. R. (2000). Use of genome scans to identify susceptibility genes for type 2 diabetes. *In* “Genetics of Diabetes Mellitus” (W. L. Lowe, Jr., ed.), pp. 231–250. Kluwer, Boston, MA.
- Mercado, M. M., McLenithan, J. C., Silver, K. D., and Shuldiner, A. R. (2001). Genetics of insulin resistance. *Curr. Diab. Rep.* **2**, 83–95.
- Silver, K., and Shuldiner, A. R. (2000). Candidate genes for type 2 diabetes mellitus. *In* “Diabetes Mellitus: A Fundamental and Clinical Text” (D. LeRoith, J. M. Olefsky, and S. I. Taylor, eds.), 2nd ed., pp. 709–719. J. B. Lippincott, Philadelphia, PA.



# Cardiovascular Disease in Diabetes

Helaine E. Resnick and Barbara V. Howard  
MedStar Research Institute, Washington, DC, United States

## Glossary

**cardiovascular disease** Disorders involving both large and small blood vessels; in diabetes, vessels in the eyes, kidneys, legs, head, neck, and heart are commonly affected.

**diabetes mellitus** A heterogeneous group of disorders characterized by high blood glucose levels.

**epidemiology** The study of the distribution and determinants of diseases in populations.

**D**iabetes mellitus (DM) is a serious public health problem that has reached epidemic proportions in developed countries. This epidemic is especially apparent in the United States, where an aging population is accompanied by a trend toward increasing obesity and decreasing physical activity. The most effective way in which to prevent diabetes-associated cardiovascular disease (CVD) is to prevent diabetes. As molecular biology and gene technologies begin to be applied in the clinical setting, new therapies may begin to mitigate the relationship between DM and CVD.

## INTRODUCTION

Diabetes and cardiovascular disease (CVD) both are associated with obesity, a sedentary lifestyle, and an adverse lipid profile. Both conditions increase with age, and their risk can be reduced with simple lifestyle modifications. Diabetes mellitus (DM) is a strong, independent CVD risk factor, and coronary heart disease (CHD) is the most frequent and costly vascular complication of DM.

Cross-sectional epidemiological studies in the United States and community-based studies in U.S. and other populations consistently reveal an association between DM and prevalent CVD. In diabetic individuals as well as in those not yet diagnosed with DM, diabetes is associated with an unfavorable distribution of CVD risk factors. In individuals at high risk

for CVD, prospective epidemiological studies reveal consistent temporal relationships between diabetes and CVD. The association between diabetes and CVD is strong, regardless of geography, age, sex, or racial/ethnic background. Diabetes is associated not only with clinical manifestations of CVD, such as acute myocardial infarction and intermittent claudication, but also with subclinical CVD measures, such as the ankle-brachial index, electrocardiographic abnormalities, conduction velocities of large peripheral nerves, dysfunction of sensory nerves, cardiac autonomic dysfunction, carotid thickening, and retinopathy.

## METABOLIC SYNDROME

Type 2 diabetes has a characteristic physical and metabolic profile that includes obesity and abnormal lipid patterns and concentrations.

## Obesity and Fat Distribution

People with diabetes are generally much heavier than their nondiabetic counterparts regardless of age, although weight tends to vary across racial and gender lines. The high prevalence of diabetes in minority women may well explain why non-White women have higher body mass indexes (BMIs) than do their male counterparts and White women, as demonstrated in epidemiological studies.

An android fat distribution pattern (accumulation of fat in the abdomen) is also typical in individuals with DM. Abdominal visceral fat is thought to be involved in glucose regulation and to potentiate the development of diabetes more than does subcutaneous fat. Although it is not possible to distinguish visceral fat from subcutaneous fat using conventional abdominal girth measures such as waist circumference, circumference measures have been consistently associated with diabetes in several epidemiological studies.

## Dyslipidemia

Diabetic individuals also have a characteristic pattern of dyslipidemia consisting of elevated triglycerides, low levels of high-density lipoprotein (HDL) cholesterol, and small dense low-density lipoprotein (LDL) particles. LDL cholesterol levels might not be higher in diabetic individuals than in nondiabetic individuals. This constellation of physical and metabolic characteristics (often accompanied by hypertension, hyperuricemia, and abnormalities in hemostatic factors) is known as metabolic syndrome, occurs prior to the appearance of frank diabetes, and persists following the diagnosis of diabetes. Metabolic syndrome has been linked to ischemic heart disease, giving further evidence that these metabolic disorders are CVD risk factors.

## BEHAVIORAL RISK FACTORS AND CVD AMONG DIABETIC INDIVIDUALS

Behavioral risk factors, such as smoking and lack of exercise, increase the risk of adverse health events among diabetic individuals, although the effects of these risk factors are not limited to people with diabetes.

Smoking is associated with diabetic complications and increased mortality risk among people with diabetes. The American Diabetes Association recommends smoking prevention and cessation among diabetic individuals.

Diabetic people are less likely to engage in regular physical activity than are their nondiabetic counterparts. Lack of physical activity is associated with mortality in diabetic individuals.

## GENDER EFFECTS

Women have a lower unadjusted risk of CHD than do men. Some, but not all, studies have shown that diabetic women have equal or higher rates of CVD than do men. This excess risk may be explained by several different mechanisms.

## Estrogen

Estrogen is typically associated with an antiatherogenic CVD risk factor profile, including higher HDL, lower LDL and blood pressure, and a peripheral distribution of fat rather than a central one. Therefore, estrogen may protect premenopausal women from CVD. However, cessation of ovarian

function during menopause lowers estrogen and HDL levels and elevates LDL, blood pressure, and abdominal fat deposition. These phenomena also occur with chronological aging and are risk factors for CVD. Thus, after menopause and with aging, the favorable CVD risk factor profile enjoyed by premenopausal women is reduced or eliminated compared with men of similar age. However, changes in CVD risk factors at menopause do not fully explain the association between diabetes and CVD in women.

A period of insulin resistance, which may last for years, often precedes the appearance of frank diabetes. In insulin resistance, peripheral tissues do not respond normally to the biological effects of insulin. Tissue resistance to insulin requires increased pancreatic beta cell activity, ultimately leading to a state of compensatory hyperinsulinemia that helps to maintain euglycemia. Decreased insulin production results in hyperglycemia because the compensatory hyperinsulinemia of insulin resistance is no longer sufficient to maintain euglycemia. Initial diagnosis of diabetes is often accompanied by insulin levels at the high end of the normal range despite the high glucose levels.

Understanding the hyperinsulinemia of insulin resistance and diabetes may be critical to understanding the link between female gender and risk of CVD. Low levels of estrogen and high levels of androgen accompany insulin resistance and diabetes even before menopause, and these changes are associated with an unfavorable distribution of CVD risk factors. Thus, premenopausal women with insulin resistance or diabetes do not benefit from the protective effects of estrogen experienced by women without these conditions. Although CVD itself is relatively uncommon before menopause, CVD risk factors may build up during this time and manifest only after menopause. Thus, the reduction in estrogen during menopause may be an especially powerful CVD risk factor among women who were insulin resistant before menopause. Studies are needed to elucidate the effect of estrogen on CVD risk in both diabetic and nondiabetic women.

## Low-Density Lipoprotein

Additional pathways may link diabetes to higher CVD risk in women than in men. Although levels of LDL (a known CVD risk factor) are similar in diabetic and nondiabetic individuals, LDL composition differs, with diabetic men and women having smaller LDL particle size. After adjustment for other CHD risk factors, including lipids, no differences in LDL size

remained in diabetic men, but unfavorable differences persisted in diabetic women. The presence of small dense LDL may be atherogenic, suggesting that diabetes may increase CVD risk more in women than in men, and indicates that LDL may be an important CVD risk factor even when levels are not elevated.

## MEASUREMENT AND EVALUATION OF NONTRADITIONAL RISK FACTORS

The future of research into the connection between diabetes and CVD will undoubtedly focus on quantifying and assessing other possible CVD risk factors such as formation of advanced glycosylation end products (AGEs) in the tissues of diabetic individuals, chronic inflammation, diabetic autonomic neuropathy, sleep disorders, and genetic susceptibility to diabetes-associated vascular damage.

### Advanced Glycosylation End Products

An important characteristic of diabetic vascular complications may be the creation of AGEs in tissue, a process accelerated by hyperglycemia. Glucose forms early glycosylation products with proteins at a rate proportional to glucose concentrations. However, because the quantity of these products is reversible, depending on glucose concentration, and does not accumulate in stable tissue proteins, the products are not consistently correlated with diabetic complications. Some of the early products change and form bonds with other proteins. AGE levels do not return to normal when hyperglycemia is eliminated; instead, they continue to accrue on large- and small-vessel wall proteins. AGE accumulation in tissue may increase vascular permeability and thicken/stiffen the vessel walls. Unlike early glycosylation products, AGEs are related to diabetic vascular disease. Although many AGEs have been characterized, the relative importance of specific AGEs in diabetes-associated vascular damage is still unknown, as are the potentially differential effects of specific AGEs in different tissues.

As demonstrated *in vitro*, AGEs may accelerate macrovascular disease by linking plasma lipoproteins with matrix proteins, thereby slowing the efflux of lipoproteins from the tissues. AGEs may also, via interaction with their receptors (RAGEs), induce endothelial cell surface adhesion molecules. Blocking the activity of RAGEs may slow the accelerated atherosclerosis characterizing diabetes and may be a target for future therapies.

In addition, binding of AGEs may induce release of inflammatory cytokines. Sustained interaction between the stable AGEs and RAGEs in diabetic tissue may create a long-term proinflammatory environment that increases CVD risk. Study of the potential role of AGEs as a stimulus for the pro-CVD inflammatory process may shed more light on the connection between diabetes and CVD.

Chemical analysis of AGEs reveals that they contain multiple products, each with its own potential link to diabetes and CVD. However, existing assays for polyclonal anti-AGE antibodies do not distinguish among individual AGEs, severely limiting efforts to link specific products with vascular damage. Despite these limitations in epidemiological studies, small clinic-based reports reveal that serum levels of AGEs are elevated in children with type 1 diabetes even before vascular complications appear. The presence of AGEs in cigarette smoke may link smoking with CVD in both diabetic and nondiabetic individuals.

### Chronic Inflammation

Inflammation may potentiate the development of CVD, and diabetes may play a role in this pathway. The constellation of CVD risk factors in insulin resistance syndrome may include markers of inflammation, and these markers may contribute to CVD risk independently of established metabolic abnormalities commonly observed in insulin resistance.

In older adults, markers of inflammation have been shown to predict clinically meaningful increases in fasting glucose levels, suggesting a role for inflammation in the development of glucose abnormalities. Although there is a growing body of evidence linking markers of inflammation to CVD, elevated levels of inflammatory markers are common in several conditions that typically occur in old age, including diabetes. Whether the inflammation is related to CVD or is a common condition of the elderly remains to be elucidated; therefore, data relating these markers to specific outcomes should be interpreted with caution.

### Diabetic Autonomic Neuropathy and Sleeping Disorders

Cardiovascular autonomic neuropathy (CAN) is thought to increase CVD risk among diabetic individuals. Impairment of CAN-associated CVD reflexes, such as heart rate variability, may increase CVD risk in addition to other mechanisms that may link diabetes to CVD via CAN. Among diabetic individuals

with CAN, 25% have been shown to have obstructive sleep apnea, a proportion significantly greater than that of diabetic individuals without CAN. The relatively high prevalence of sleep disturbance in those with diabetic neuropathy implies that the impaired central control of respiration associated with CAN may worsen the effects of sleep disorders on CVD. Increased prevalence of sleep apnea and nocturnal oxygen desaturation in diabetic patients with CAN give further evidence of a relationship between diabetes and sleep-disordered breathing (SDB). In addition to the effects of diabetes on SDB, SDB may actually worsen diabetes by causing dysregulation of glucose metabolism. Reduced glucose tolerance following sleep deprivation has been demonstrated, implying that disrupted sleep may adversely affect some features of normal endocrine function.

### Genetic Susceptibility

Severity of diabetic vascular complications may be determined in part by genetics. Identification of genes that may indicate CVD susceptibility in diabetic individuals may become possible, enabling clinicians to tailor treatment plans for “susceptible” or “protected” patients. Although no such screening is currently possible, recent data suggest that certain genes may confer differential risk of diabetic complications.

Haptoglobin (Hp), a hemoglobin-binding protein that protects against oxidative stress, is one such gene. Oxidative stress is thought to be an important mediator of many pathophysiological processes, including diabetic vascular complications. The two common Hp alleles yield three phenotypes that, due to their different biochemical and biophysical properties, seem to differ in their ability to function as antioxidants. Individuals who are homozygous for the Hp-1 allele (1-1) appear to be protected against the development of diabetic retinopathy, diabetic nephropathy, and coronary restenosis. As the field of genetics continues to grow, new methods of CVD risk stratification that enable clinicians to tailor interventions to the genetic profiles of their diabetic patients may develop.

### See Also the Following Articles

Atherosclerosis • Carbohydrate Metabolism: Diabetes Mellitus, Genomic Aberrations • Cardiovascular Disease: Impact of Sex Steroid Replacement • Diabetes Mellitus, Diagnosis and Treatment in the Elderly • Diabetic Nerve

Disease, Neuropathy • Eye Disease in Diabetes • Foot Disease in Diabetes • Glucose, Impaired Tolerance • Hypertension and Diabetes • Kidney Disease in Diabetes • Obesity and Diabetes, Regulation of Food Intake

### Further Reading

- Brownlee, M., Cerami, A., and Vlassara, H. (1988). Advanced glycosylation endproducts in tissue and the biochemical basis of diabetic complications. *N. Engl. J. Med.* **318**, 1315–1321.
- Cowie, C. C., and Harris, M. I. (1995). Physical and metabolic characteristics of persons with diabetes. In “Diabetes in America” (NIH Publication No. 95-1468), 2nd ed., pp. 117–132. National Institutes of Health, Bethesda, MD.
- Ficker, J. H., Dertinger, S. H., Siegreied, W., Konig, H. J., Pentz, M., Sailer, D., Katalinic, A., and Hahn, E. G. (1998). Obstructive sleep apnoea and diabetes mellitus: The role of cardiovascular autonomic neuropathy. *Eur. Respir. J.* **11**, 14–19.
- Ford, E. S. (1999). Body mass index, diabetes, and C-reactive protein among U.S. adults. *Diab. Care* **22**, 1971–1977.
- Haffner, S. M., Mykkanen, L., Stern, M. P., Paidi, M., and Howard, B. V. (1994). Greater effect of diabetes on LDL size in women than in men. *Diab. Care* **17**, 1164–1171.
- Howard, B. V., Cowan, L. D., Go, O., Welty, T., Robbins, D. C., and Lee, E. T. (1998). Adverse effects of diabetes on multiple cardiovascular disease risk factors. *Diab. Care* **21**, 1258–1265.
- Kannel, W. B., and Wilson, P. W. F. (1995). Risk factors that attenuate the female coronary disease advantage. *Arch. Int. Med.* **155**, 57–61.
- Kiely, J. L., and McNicholas, W. T. (2000). Cardiovascular risk factors in patients with obstructive sleep apnea syndrome. *Eur. Respir. J.* **16**, 128–133.
- Nathan, D. M., Meigs, J., and Singer, D. E. (1997). The epidemiology of cardiovascular disease in type 2 diabetes mellitus: How sweet it is ... or is it? *Lancet* **350**, S4–S9.
- Pickup, J. C., Mattock, M. B., Chusney, G. D., and Burt, D. (1997). NIDDM as a disease of the innate immune system: Association of acute-phase reactants and interleukin-6 with metabolic syndrome X. *Diabetologia* **40**, 1286–1292.
- Resnick, H. E., Harris, M. I., Brock, D. B., and Harris, T. B. (2000). American Diabetes Association diabetes diagnostic criteria, advancing age, and cardiovascular disease risk factor profiles: Results from the Third National Health and Nutrition Examination Survey. *Diab. Care* **23**, 176–180.
- Scheidt-Nave, C., Barrett-Connor, E., and Wingard, D. L. (1990). Resting electrocardiographic abnormalities suggestive of asymptomatic ischemic heart disease associated with non-insulin-dependent diabetes mellitus in a defined population. *Circulation* **81**, 899–906.
- Strohl, K. P. (1996). Diabetes and sleep apnea. *Sleep* **19**, S225–S228.
- Wingard, D. L., and Barrett-Connor, E. (1995). Heart disease and diabetes. In “Diabetes in America” (NIH Publication No. 95-1468), 2nd ed., pp. 429–434. National Institutes of Health, Bethesda, MD.
- Yan, A. D., Stern, D., and Schmidt, A. M. (1997). What's the RAGE? The receptor for advanced glycation endproducts (RAGE) and the dark side of glucose. *Eur. J. Clin. Invest.* **27**, 179–181.



# Cardiovascular Disease: Impact of Sex Steroid Replacement

Jeffrey L. Anderson

University of Utah School of Medicine, Salt Lake City, Utah, United States

## Glossary

**cardiovascular (CV) disease** CV disease includes coronary heart disease and other forms of CV disease. Important among these are hypertension and hypertensive heart disease, stroke (especially ischemic or thromboembolic stroke), and peripheral arterial occlusive disease.

**coronary heart disease (CHD)** CHD represents the clinical consequence of coronary artery atherosclerosis, including angina pectoris, fatal and nonfatal myocardial infarction, sudden death, and heart failure due to ischemic heart disease.

**estrogen replacement therapy (ERT)** Use of unopposed estrogen as sex steroid replacement therapy. ERT is best tolerated in women who have had a hysterectomy; its use in others is less well tolerated (it may be associated with endometrial hyperplasia).

**hormone replacement therapy (HRT)** A synonym often used in the literature for sex steroid replacement.

**selective estrogen receptor modulators (SERMs)** These are engineered, synthetic sex hormones that have different receptor and tissue specificity than natural estrogen. Their potential for CV risk or protection may thus also differ. One SERM, raloxifene, is undergoing testing for its CV and other preventive potential.

**sex steroid replacement (SSR)** Replacement of female sex hormones after menopause, specifically estrogen with or without a progestin.

Hormonal differences between the sexes are believed to play an important role in delaying coronary heart disease (CHD) in women. A prevalent belief has been that sex steroid replacement (SSR) may prevent or reverse adverse metabolic (lipid and glucose) changes, maintain vascular health, and reduce CHD risk in women after menopause. A multitude of observational studies lent support to this hypothesis. However, randomized clinical trials of both secondary prevention (e.g., the Heart and Estrogen/Progestin Replacement Study) and primary prevention (the Women's Health Initiative) did not

demonstrate net benefit or suggested increased cardiovascular risk potential. Thus, SSR can no longer be recommended for primary or secondary CHD prevention. Whether a different outcome applies to estrogen replacement therapy (e.g., in women without a uterus) or to selective estrogen receptor modulators (e.g., with raloxifene) will be determined by ongoing clinical trials.

## BACKGROUND

### Risk of Heart Disease Is High in Elderly Women

Despite widespread belief to the contrary, cardiovascular (CV) disease ranks first as a cause of death and disability in women as well as men. In a poll, women perceived cancer (breast cancer, etc.), Alzheimer's disease, and arthritis/bone loss to be their major health problems. Asked about the leading cause of death in women, 55% responded breast cancer and 22% heart disease. In fact, cancer is the cause of 1 of every 5 deaths and breast cancer is the cause of 1 of every 30 deaths, whereas CV disease causes 1 of every 2.4 deaths in women. Of nearly 530,000 U.S. coronary heart disease (CHD) deaths in 1999, 263,000 were in women. Of all CV deaths, 513,000 occurred in women and 446,000 in men. By comparison, all cancers accounted for 264,000 female and 286,000 male deaths. Thus, heart disease, is an "equal opportunity killer," is the leading cause of death in both men and women, and is equal in magnitude to the next seven leading causes of death combined. Furthermore, whereas the death rate from CV disease has steadily declined in men, it has increased in women. Also, hospitalizations for stroke and heart failure for women now exceed those for men.

### Presentation of CHD Differs in Men and Women

Despite similarities in prevalence, the pattern of CHD differs by gender. Women are 10–15 years older than

men when signs of CHD first appear. The symptoms of CHD often are more subtle or “atypical,” confusing and delaying diagnosis. However, when a heart attack occurs, survival rates are lower for women.

### Risk Is Low before and Increases Rapidly after Menopause

Heart disease risk is low in premenopausal and early postmenopausal women. After age 65, CHD rates increase over time similarly in women and men. By age 75, CHD prevalence in women approximates that in men. Increases in CHD risk after menopause are associated with adverse changes in lipids and lipoproteins: Total cholesterol, low-density lipoprotein cholesterol (LDL-C), and triglycerides increase (10–15%), lipoprotein (a) increases, and high-density lipoprotein cholesterol (HDL-C) decreases (5–10%). LDL composition shifts toward smaller, denser particles (associated with greater CHD risk). Glucose intolerance increases, and endothelial dysfunction becomes more common.

## ESTROGEN AND VASCULAR BIOLOGY

### Might Sex Steroid Replacement Return CHD to Naturally Low, Premenopausal Levels?

The previously discussed observations led to the prevalent belief that hormonal differences between the sexes play an important role in delaying the appearance of CHD in women. A corollary hypothesis naturally developed that sex steroid replacement (SSR) in women might reverse the adverse metabolic changes of menopause, maintain vascular functional health, and reduce CHD risk to premenopausal levels.

### Estrogen and Estrogen Receptors

The biological activity of estrogen is mediated in large part through specific estrogen receptors (ERs). Two major ER types, ER- $\alpha$  and ER- $\beta$ , have been described. Their differing molecular mechanisms and cellular distributions provide the basis for tissue selectivity. Each has ligand- and DNA-binding domains. Each has receptor-associated proteins and coactivators that are cell specific.

### Molecular Mechanisms of Estrogen Activity

After binding to ER, estrogen is transported to the nucleus, where homo- or heterodimers form. DNA binding follows at estrogen response elements and DNA and mRNA expression is induced in distinct profiles. Repressors regulate ER availability and function, and crosstalk occurs with other pathways (e.g., NF- $\kappa$ B). Nonclassic genomic pathways of estrogen/ER action have also been described. Estrogen may also function through nongenomic pathways to promote vascular health (through estrogen-binding proteins, ERs, or other mechanisms). Estrogen induces vascular nitric oxide and exerts antioxidant activity. Estrogen prevents, but does not reverse, vascular inflammation induced by a high-fat diet (atherogenic mouse model).

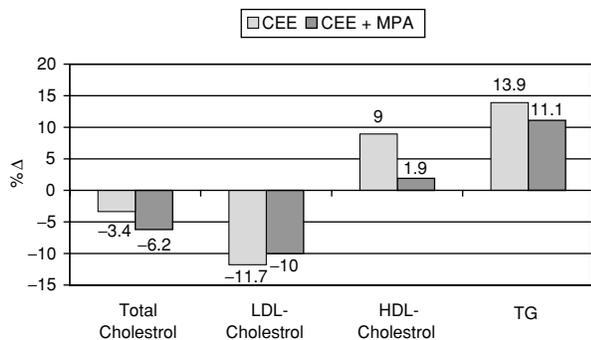
### Estrogen and Lipoprotein Metabolism

Facilitation of reverse cholesterol transport is important for estrogen’s potential preventive role. Estrogen acts to increase apolipoprotein (apo)-A1 and HDL particles, reduce hepatic lipase activity, decrease HDL uptake by hepatic SR-B1 scavenger receptors, and facilitate LDL clearance by hepatic LDL receptors.

With SSR, LDL cholesterol, apoB, and lipoprotein (a) decrease, and HDL2-C, total HDL-C, apoA1, and triglyceride (TG) increase (Figs. 1 and 2). The effects on lipoprotein profiles of estrogen, various estrogen/progestin combinations, and selective estrogen receptor modulators (SERMs) are qualitatively generally similar but differ quantitatively. Estrogen causes the greatest increase in HDL2-C. The response of HDL-C to SSR may be augmented in women with specific ER- $\alpha$  polymorphisms (i.e., IVS1-401 C/C). Differences in prothrombotic factors (fibrinogen, PAI-1, F1.2, and FPA) have also been reported.

### The Postmenopausal Estrogen/Progestin Interventions Trial

The Postmenopausal Estrogen/Progestin Interventions Trial (PEPI) randomized 845 postmenopausal women (45–64 years old) to placebo, conjugated equine estrogen (CEE), CEE plus medroxyprogesterone acetate (MPA), or CEE plus micronized progesterone (Days 1–12) and followed them for 3 years. All SSR regimens increased bone mineral density, and combining CEE with a progestin prevented endometrial hyperplasia associated with CEE alone. For adherent

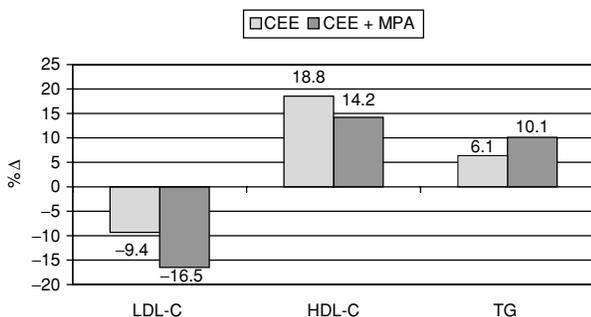


**Figure 1** Effect of SSR on lipoprotein fractions for primary prevention. Data from the PEPI study [*JAMA* (1995), 273, 199-208] of 349 women treated with conjugated equine estrogen (CEE) or CEE + medroxyprogesterone acetate (MPA). HDL, high-density lipoprotein; LDL, low-density lipoprotein; TG, triglyceride.

subjects, LDL-C decreased in each active treatment arm, and TGs and HDL (especially with CEE alone) increased (Fig. 1). Little change in fibrinogen or systolic blood pressure was observed with any regimen, but a greater increase in 2-h glucose level occurred with CEE plus MPA than with CEE (8 or 9% vs 3 or 4%). These observations provided the basis for further testing of estrogen/progestin combinations and, in patients without a uterus, unopposed estrogen.

## OBSERVATIONAL STUDIES AND CARDIOVASCULAR RISK

The beneficial potential of SSR on CHD and other CV disease risk was supported by numerous (>30) case-control and prospective cohort studies. Overviews of these association studies suggest risk



**Figure 2** Effect of SSR on lipoprotein fractions for secondary prevention. Data from the ERA study [*NEJM* (2000), 343, 522-529] of 309 women with CAD. CEE, conjugated equine estrogen; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; MPA, medroxyprogesterone acetate; TG, triglyceride.

reductions of estrogen therapy on CHD of 35–50%. The Nurses' Health Study, the largest prospective cohort study, found that current users of estrogen had approximately half the risk of CHD [relative risk, 0.56 for myocardial infarction (MI) or death], after adjustment for age and other risk factors, as nonusers. Even stronger benefits were found in studies of women with known CHD. Observational studies are limited by the possibility of selection bias. Women who choose to be on SSR may be more health conscious in other ways. Other healthy behaviors, rather than SSR, might explain better CV outcomes. These concerns are best addressed by prospective, randomized trials.

## SECONDARY PREVENTION TRIALS

### The Heart and Estrogen/Progestin Replacement Study

The Heart and Estrogen/Progestin Replacement Study (HERS) is the most important clinical trial of SSR for secondary CV prevention. HERS randomized 2763 postmenopausal women (average age, 67) with documented CHD to 0.625 mg/day CEE plus 2.5 mg/day MPA or placebo and followed them for an average of 4.1 years. At study end, there was no difference between groups in the primary outcome of CHD events (nonfatal MI and CHD-related death). Secondary CV outcomes also did not differ. Post hoc analysis showed a significant difference in outcomes by time: More CHD events occurred with SSR in the first year and fewer events in Years 3–5. The adverse potential of SSR use within the first year of initiation was an unanticipated result that the Women's Health Initiative (WHI) would later replicate. As a consequence, initiation of SSR after a coronary event was proscribed, but CHD patients already taking and tolerating SSR were allowed to continue.

### HERS-II

A major question after HERS was whether the lower event rate in Years 3–5 indicated that benefit would occur with additional years of treatment. This question led to HERS-II, an extended surveillance study of the HERS cohort. Women in the placebo arm were advised not to start SSR, whereas those on therapy could continue. A total of 2321 women in HERS were alive, agreed to enroll in HERS-II, and were followed for an additional 2.7 years. Approximately half continued to take the originally assigned therapy, but no CV benefit of SSR was observed. The unadjusted

hazard ratio (HR) for a primary CHD event was 1.00 [95% confidence interval (CI), 0.77–1.29] in HERS-II and 0.99 (95% CI, 0.84–1.17) for the combined HERS/HERS-II study period. Adjustment for potential confounders, the use of aspirin and statins, and adherence to randomized treatment assignment had little effect on the overall result (HR, 0.96; CI, 0.77–1.19). Similarly, there was no evidence of benefit for the secondary CV outcomes in HERS-II or overall in HERS for coronary revascularization (overall HR, 0.93; 1.02 for bypass surgery and angioplasty), hospitalization for unstable angina (HR, 0.87) or heart failure (HR, 1.05), sudden death (HR, 0.98), stroke or transient ischemic attack (HR, 1.09), or peripheral arterial disease (HR, 0.87). Also, nonfatal ventricular arrhythmia increased (HR, 1.97). Given the failure to confirm a late benefit of SSR in secondary prevention and the documented excess risk of SSR initiation, HERS-II concluded that SSR should not be used to reduce CHD risk in postmenopausal women with CHD.

### Effects of Estrogen Replacement on the Progression of Coronary Artery Atherosclerosis

After HERS, an angiographic trial, Estrogen Replacement and Atherosclerosis (ERA), was undertaken to directly explore the hypothesis that SSR might slow the progression of coronary atherosclerosis. Accordingly, 309 postmenopausal women (average age, 66 years) with angiographically defined coronary artery disease (CAD) were enrolled and randomized to 0.625 mg/day CEE alone, 0.625 mg/day CEE plus 2.5 mg/day MPA, or placebo. After an average of 3.2 years, a follow-up coronary angiogram was obtained and quantitative arteriography performed ( $n = 248$ ). During the study, favorable lipid-lowering effects of SSR were observed for the CEE and CEE plus MPA regimens (with LDL-C reductions of 9 and 17% and HDL-C increases of 19 and 14%, respectively) (Fig. 2). However, no difference in adjusted minimal coronary artery diameter or in diameter change scores were observed for estrogen, estrogen/progestin, and placebo regimens ( $p > 0.2$  vs placebo). Thus, neither estrogen alone nor in combination with progestin beneficially affected the progression of CAD.

### Women's Angiographic Vitamin and Estrogen Trial

The Women's Angiographic Vitamin and Estrogen Trial (WAVE) studied 425 postmenopausal women

with at least one mild to severe coronary stenosis. After 2.8 years of blinded therapy, progression in coronary stenosis scores showed neither a benefit nor a favorable trend in the CEE, CEE plus MPA, or vitamin (E plus C) groups. Indeed, adverse trends for the combination of death, nonfatal MI, and stroke were observed in both the active SSR and the vitamin groups compared with their respective placebo controls.

### Summary of Secondary Prevention Studies

The ERA and WAVE mechanistic studies, together with the HERS/HERS-II outcomes study, are consistent with an absence of meaningful benefit of SSR over 3–7 years in secondary CHD prevention and support the recommendation that SSR should not be used in women with CHD with the expectation of cardiovascular benefit.

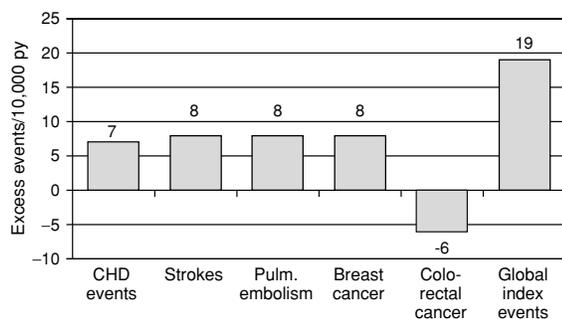
### PRIMARY PREVENTION TRIALS

After HERS, advocates of SSR took refuge in the possibility that SSR might fare better for primary CV prevention: SSRs might have greater potential for preventing the initiation of atherosclerosis in younger women nearer menopause than in older cohorts in which SSR had failed to prevent progression of established disease. Experimental and intermediate endpoint studies also seemed to be supportive.

### Women's Health Initiative

The report of the estrogen plus progestin component of the WHI randomized controlled primary prevention trial dispelled that hope. WHI randomized 16,600 postmenopausal women (average age, 63 years) to CEE, 0.625 mg/day CEE plus 2.5 mg/day MPA, or placebo. The primary efficacy outcome (expected to be beneficial) was CHD (nonfatal MI and CHD death). Invasive breast cancer was the primary safety (adverse) outcome of interest. A global index summarizing the balance of major health benefits and risk was the third major outcome measure of interest. The global index included the two primary outcomes plus stroke, pulmonary embolism (PE), endometrial cancer, colorectal cancer, hip fracture, and death due to other causes.

After 5.2 years, the data and safety monitoring committee recommended stopping the largest estrogen/progestin component of the trial early because the number of cases of invasive breast cancer in the



**Figure 3** Excess events [per 1000 person-years (py)] for SSR in the WHI Primary Prevention Trial.

active treatment group had crossed the boundary of increased risk, and the global index suggested that treatment was causing more harm than good. The hazard ratio for breast cancer ( $n = 290$  cases) prompting study termination was 1.26 (CI, 1.00–1.59). This was not offset by cardiovascular benefit; indeed, the relative hazard of a CHD event ( $n = 286$ ) increased with estrogen/progestin therapy, with a HR of 1.29 (CI, 1.02–1.63), the excess risk being due primarily to nonfatal MI. Other components of the global index that increased were stroke (HR, 1.4; with persisting risk at 5 years) and PE (HR, 2.1), whereas colorectal cancer (HR, 0.63) and hip fracture (HR, 0.66) were reduced. Endometrial cancer and total mortality rates were not adversely affected. Overall, the global index HR was 1.15 (CI, 1.03–1.28), a modest but significantly adverse result.

When applying clinical trials to medical practice, results should be translated into absolute risk. From this perspective, the risk to an individual woman is small. Among 10,000 women taking estrogen/progestin for 1 year, WHI suggests that there will be 7 more CHD events, 8 more invasive breast cancers, 8 more strokes, and 8 more pulmonary emboli but 6 fewer colorectal cancers and 5 fewer hip fractures (Fig. 3). On balance, 19 more global index events would be observed annually (approximately 100 excess events over 5 years, or 1 per 100 women). Despite the small individual risk to a patient, WHI's lack of CV benefit (indeed, net risk) argues strongly that estrogen/progestin therapy should not be prescribed for long-term primary CHD prevention. As Fletcher and Colditz noted, *primum non nocere* should especially apply to preventive health care.

### WHI, Part II, and the Estrogen in Prevention of Atherosclerosis Trial

The WHI trial comparing estrogen alone with placebo among women without a uterus is ongoing.

While awaiting WHI ERT, the Estrogen in Prevention of Atherosclerosis Trial (EPAT) is interesting to ponder. EPAT assessed the effect of unopposed ERT on progression of subclinical atherosclerosis, represented by carotid intimal medial thickness (IMT) measured by high-resolution ultrasound. A total of 222 postmenopausal women (average age, 62 years) without clinical CV disease but with elevated LDL-C were enrolled. Therapy was unopposed micronized  $17\beta$ -estradiol (1 mg/day) or placebo. Lipid-lowering medication (primarily a statin) was prescribed for LDL  $>160$  mg/dl. Carotid IMT was measured every 6 months for 2 years. The rate of IMT progression was lower with ERT [by 0.0053 mm/year; a small but significant difference ( $p = 0.046$ )]. The benefit of ERT was prominent in women who did not take lipid-lowering medication (0.0147 mm/year,  $p = 0.002$ ), whereas progression did not differ in those taking a lipid-lowering drug. The EPAT investigators suggested that unopposed ERT might serve as an alternative to lipid lowering for primary prevention and that natural estrogen as  $17\beta$ -estradiol might have advantages over the complex mixture of SSR represented by CEE. Given the excellent safety record and clearly demonstrated CV benefits of statins, I believe that statins, rather than unopposed  $17\beta$ -estradiol, should be preferred for primary prevention, where appropriate, pending WHI part II.

### INCOMPATIBILITY OF OBSERVATIONAL STUDIES AND CLINICAL TRIALS

#### Wrong Chemistry?

Given the promise of SSR, it is natural to wonder what went wrong. Was there a fundamental flaw in the SSR prevention strategy, or could changes in the formula rescue the chemistry? Was the dose wrong? Was the wrong estrogen or progestin used? Would unopposed ERT fair better? Was the wrong population targeted? Is it possible to fashion a beneficial CV preventive strategy of SSR? There are no clear answers. CEE contains a mixture of molecules with differing properties: Would estradiol fair better? Similarly, might an adverse impact of MPA be overcome by using another progestin (or no progestin)? Would tapering to a lower dose avoid serious adverse effects while contributing to vascular health? Would the transdermal route for SSR delivery, avoiding first-pass hepatic effects, provide safer, more reliable effects? All of these possibilities deserve consideration

given the important public health impact of SSR-related issues. However, the previously mentioned negative trials will likely make sponsors leery of major new initiatives with traditional SSRs.

### Not Made for Marriage?

Of course, the trials also could be signaling a basic incompatibility of SSRs with CV protection after menopause. Animal models might have been too simplistic, inappropriately weighting the potential benefits and underestimating adverse risks. The surprising results of the controlled clinical trials call for careful reassessment of the original hypothesis. Indeed, an exhaustive meta-analysis of prior observational studies found a more modest reduction in relative CHD risk (HR, 0.80; CI, 0.68–0.95) of current SSR use than generally reported. Also, there was no benefit for patients who had ever used SSR compared with those who had never used SSR. Furthermore, when only studies that controlled for socioeconomic status (SES) were considered, no benefit for even current SSR use was seen (HR, 0.97; CI, 0.82–1.16).

Consistent with the hypothesis that SES might explain differences in CHD rates in observational studies, we and others reported prominent differences in incident CHD event rates by SES status, and SES affected women more than men. Thus, the “healthy woman” hypothesis may largely explain the disparity between earlier observational and current randomized trial results. A large component of CV disease is preventable by compliance with healthy behaviors, and effective prevention and treatment are further assisted by better health care access, especially for women.

### Proinflammation to Blame?

Atherosclerosis is an established inflammatory disease. Signs of inflammation are present in plaque (macrophages, T lymphocytes, cytokines, and chemokines) and systemically. C-reactive protein (CRP), a circulating marker of inflammation, is a CV risk factor. In the Nurses’ Health Study, CRP was a stronger predictor of incident coronary events than even LDL-C. Risk was dramatically increased with elevations of both CRP and LDL-C. SSR may increase CRP; hence, concern for a proinflammatory effect has been raised. However, other circulating inflammatory markers, such as interleukin-6, are not increased by SSR, and the risk relation with CRP is blunted in women on SSR. It is possible that augmented hepatic synthesis of CRP, rather than

promotion of vascular inflammation, explains the increased CRP with SSR.

### Is Prothrombosis the Cause?

An increase in venous and arterial thrombotic and thromboembolic events has been observed consistently in oral contraceptive and SSR studies. Thromboembolic events affect only a small percentage of women and may be explained in part by predisposing genetic factors. For example, in one study, oral contraceptive use alone conferred a 6-fold increased risk, but use together with a variant allele in genes for either prothrombin or factor V increased risk 16- to 20-fold.

### Patient Selection by Pharmacogenetics?

If thrombotic risk is excessively increased only in a small percentage of women harboring predisposing genes, then genetic screening could potentially eliminate this risk. Similarly, interacting environmental factors (e.g., smoking or other drugs) and diseases (e.g., diabetes) could elevate SSR-related risk excessively and, if precluded, might reduce CV risk.

## RECOMMENDATIONS FOR SEX STEROID USE

The following CV health recommendations for SSR derive from the previous discussion and consensus statements:

### Secondary CV Prevention

1. SSR should not be initiated for the secondary prevention of CV disease.
  - a. The decision to continue or stop SSR in women with CV disease undergoing long-term SSR therapy should be based on established noncoronary benefits, a benefit/risk assessment (including other therapeutic options), and patient preference. Due to the HERS-II findings, other established therapeutic options (e.g., bisphosphonates for prevention of osteoporosis) may be preferable.
  - b. If an acute CV event occurs during SSR therapy, it is prudent to consider discontinuation of SSR. Similarly, if immobilization occurs during SSR therapy for other reasons, discontinuation of SSR should be considered and/or prophylactic therapy for venous thromboembolism should be

provided. Resumption of SSR after resumed mobilization should be based on noncoronary benefits and risks, alternative therapies, relative CV risks, and patient preference.

### Primary CV Prevention

1. SSR with estrogen/progestin should not be prescribed for long-term primary CV prevention
2. Firm recommendations about the use of unopposed ERT in women without a uterus await the results of ongoing study (WHI-ERT). However, based on negative results for other SSR regimens and populations, a net benefit of ERT should not be assumed. Hence, unopposed ERT cannot be recommended for long-term primary CV prevention for women with or without an intact uterus.
3. Short-term use of SSR for severe menopausal symptoms can probably continue to be justified as sufficiently safe. However, emphasis should be on the lowest dose of medication for the shortest period of time.
4. Long-term use of SSR for other purposes (e.g., prevention of osteoporosis and skin and mucous membrane health) should be reconsidered in light of health risks (WHI), and alternative therapies should be considered.

### Alternatives to SSRs for CV Risk Reduction

1. Lipid lowering: Lipid lowering, especially with statins, is much better established for CV risk reduction than SSRs. An overview of four major randomized trials of secondary and primary prevention has shown that risk reduction is equivalent in women and men (29 vs 31% CHD event reduction). Thus, where appropriate, statin therapy is highly effective and well tolerated for CV risk reduction.
2. Other preventive therapies: Other established CV preventive therapies also apply to women as well as men. Aspirin (75–325 mg/day) is recommended for all patients with CHD or equivalent (e.g., stroke and peripheral arterial disease) and for those at high (>1%/year) primary risk (e.g., with two or more risk factors, such as age  $\geq$ 65 years, smoking, diabetes, dyslipidemia, and hypertension). A heart-healthy diet and regular exercise also form an important component of CV risk reduction. Smoking cessation is a highly effective intervention for CHD prevention. Diabetes counters the advantage that women have over men in regard to CV risk; an aggressive approach to prevention and treatment of dysglycemia with diet, oral antidiabetic

drugs, and insulin is appropriate. Hypertension is an important risk factor for stroke and CHD and can be effectively managed. Following MI, beta-blockers and angiotensin-converting enzyme inhibitors are usually indicated to reduce secondary CV risk. A role for nutritional and herbal remedies, such as soy or other phytoestrogen sources, for postmenopausal CV risk reduction has not been well established.

3. Balance of benefit and risk: When SSR is considered for relief of menopausal symptoms or other non-CV indications (e.g., osteoporosis), an evaluation of likely benefit versus potential risk (CV and non-CV) should be made. The result of this benefit/risk assessment, which will vary for individual women, forms the basis for a decision regarding SSR therapy.

### UNANSWERED QUESTIONS AND FUTURE RESEARCH

Technically, the results of HERS and WHI apply only to the specific SSR combination of CEE and MPA. Several other combinations and unopposed SSR products and doses are available. However, it is likely that sponsors will be wary of investing in further large outcomes trials using traditional SSRs. However, interest in SERMs remains.

#### WHI-ERT

The role of unopposed ERT in women without a uterus is undergoing testing, and recommendations in this setting must await its outcome.

#### WISDOM

A large European prevention study, WISDOM, was under way when WHI was stopped. After much debate, WISDOM was also stopped. Findings of WISDOM, not yet reported, will be of interest.

#### SERMS

SERMS are designer drugs developed to have differential effects on estrogen receptors to accomplish specific therapeutic objectives. Tamoxifen has been useful in breast cancer prevention and treatment. Raloxifene has been tested for osteoporosis and is undergoing evaluation for CV prevention. Indeed, the 10,000-patient Raloxifene Use for The Heart (RUTH) trial is under way with no health warnings from the data and safety monitoring committee. Thus, interest and hope remain for SERMS in CV prevention.

## Parting Thought

There is no assurance of an early reconciliation between SSRs and CV disease prevention. However, our understanding of the complex role of sex hormones in normal and disease-related physiology is embryonic and hope remains that improved understanding of sex steroid-related biology will allow the development of SSR or SERM-based preventive strategies, perhaps customized to individual patients, that will be safe and effective.

## See Also the Following Articles

Atherosclerosis • Brain, Effects of Steroid Hormones • Breast Disease: Impact of Sex Steroid Replacement • Cardiovascular Disease in Diabetes • Hot Flash: Impact of Sex Steroid Replacement • Osteoporosis in Reproductive Age Women: Impact of Sex Steroid Replacement • Polycystic Ovary Syndrome: Implications for Cardiovascular, Endometrial, and Breast Disease

## Further Reading

American Heart Association (2002). "Heart and Stroke Statistical Update." American Heart Association, Dallas, TX.

- Bittner, V. (2002). Lipoprotein abnormalities related to women's health. *Am. J. Cardiol.* **90**(Suppl.), I77–I84.
- Douglas, P. S. (2001). Coronary artery disease in women. Hormones: Risks and benefits of estrogen. In "Heart Disease. A Textbook of Cardiovascular Medicine" (E. Braunwald, D. P. Zipes, and P. Lippy, eds.), 6th ed., pp. 2040–2043. Saunders, Philadelphia.
- Fletcher, S. W., and Colditz, G. A. (2002). Failure of estrogen plus progestin therapy for prevention. *J. Am. Med. Assoc.* **288**, 366–368.
- Grady, D., Herrington, D., Bittner, V., *et al.* (2002). Cardiovascular disease outcomes during 6.8 years of hormone therapy. Heart and Estrogen/Progestin Replacement Study follow-up (HERS II). *J. Am. Med. Assoc.* **288**, 49–57.
- Mosca, L., Collins, P., Herrington, D. M., *et al.* (2001). Hormone replacement therapy and cardiovascular disease. A statement for healthcare professionals from the American Heart Association. *Circulation* **104**, 499–503.
- Stampfer, M. J., Colditz, G. A., Willett, W. C., *et al.* (1991). Postmenopausal estrogen therapy and cardiovascular disease. Ten-year follow-up from the Nurses' Health Study. *N. Engl. J. Med.* **325**, 756–762.
- U.S. Preventive Services Task Force. (2002). Postmenopausal hormone replacement therapy for primary prevention of chronic conditions: Recommendations and rationale. *Ann. Intern. Med.* **137**, 840–849.
- Writing Group for the Women's Health Initiative Investigators. (2002). Risks and benefits of estrogen plus progestin in healthy postmenopausal women. *J. Am. Med. Assoc.* **288**, 321–333.



# Catecholamines

Brian B. Hoffman

VA Boston Health Care System and Harvard University School of Medicine, Boston, Massachusetts, United States

## Glossary

**adrenal medulla** Part of the adrenal gland, surrounded by the adrenal cortex. The major function of the adrenal medulla is to synthesize epinephrine, which is released into the blood; epinephrine acts on distant cells as a hormone.

**adrenergic receptors** Endogenous catecholamines and various sympathomimetic drugs activate cells by interacting with one or more of the three classes of adrenergic receptors. These receptors are characteristic members of the very large family of G protein-coupled receptors.

**catecholamines** Endogenous catecholamines include dopamine, norepinephrine (noradrenaline), and epinephrine (adrenaline). These substances have in common a 3,4-dihydroxybenzene moiety (catechol) and a primary amino group.

**monoamine oxidase** An enzyme that plays an important role in the metabolism (inactivation) of the endogenous catecholamines and certain drugs.

**pheochromocytoma** Tumors that secrete an excess concentration of catecholamines; this may lead to severe hypertension. Therapy involves surgery in most cases and the use of adrenergic antagonist drugs.

**sympathetic nervous system (SNS)** A major component of the autonomic nervous system. Via its primary transmitter norepinephrine, the SNS modulates the function of many cells and is especially important in the regulation of stress responses and blood pressure.

The endogenous catecholamines in humans are dopamine, norepinephrine, and epinephrine. Norepinephrine is the major transmitter in the terminals of the sympathetic nervous system; norepinephrine acts on postsynaptic cells. On the other hand, adrenaline is the primary hormone released by the adrenal medulla into the bloodstream. Catecholamines activate three major adrenergic receptor families; these receptors are part of the larger family of G protein-coupled receptors. Utilizing several signal transduction mechanisms, catecholamines have the capacity to regulate the activity of multiple different cell

types in organs throughout the body. A major role of the sympathetic nervous system and the adrenal medulla is to integrate responses triggered by stress; i.e., the “fight or flight” response conceptualized by Walter Cannon in the 1930s. Abnormalities in catecholamine function contribute to a wide variety of diseases; for example, pheochromocytoma is a tumor that typically secretes excess catecholamines, leading to potentially severe hypertension.

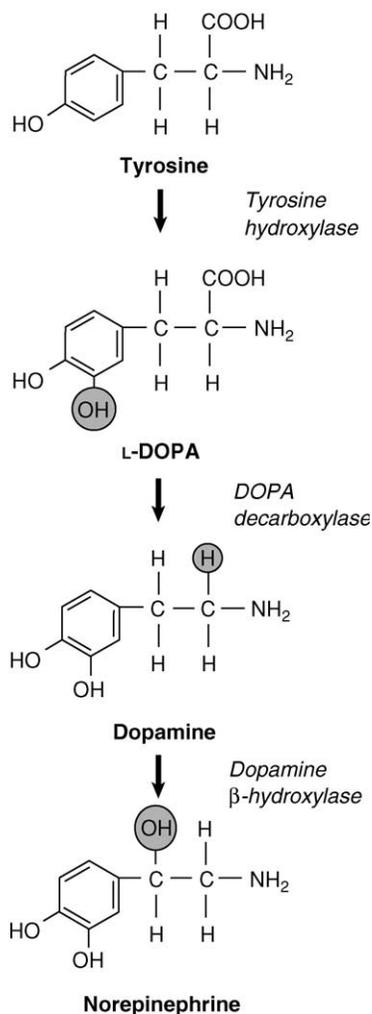
## DISCOVERY OF CATECHOLAMINES

Understanding of the physiological role of the adrenal medulla was unavailable until the demonstration in 1894 by Schafer and Oliver that extracts of the medulla caused marked rises in blood pressure. This discovery immediately stimulated considerable interest in isolating the active substance in the medulla that caused vasoconstriction. Between 1897 and 1901, a number of groups working independently isolated epinephrine (adrenaline) from these preparations; the priority in this discovery is somewhat controversial. In 1904, epinephrine became the first hormone to be isolated and crystallized, to have its structure determined, and then to be successfully chemically synthesized in a laboratory. Epinephrine was rather quickly developed as a drug for a surprisingly broad multiplicity of indications; early on it was found to decrease bleeding in a range of surgical procedures.

## SYNTHESIS, RELEASE, AND TERMINATION OF ACTION OF CATECHOLAMINES

### Synthesis of Catecholamines

Norepinephrine is primarily synthesized in sympathetic nerves, starting from the amino acid tyrosine; several enzymes expressed in these cells are involved in this process (Fig. 1). The rate-limiting step in catecholamine synthesis is that in which the enzyme tyrosine hydroxylase is used, a step that is regulated both translationally and posttranslationally. Up to the point at which dopamine is synthesized, synthesis takes place



**Figure 1** Norepinephrine synthesis in sympathetic nerve endings. Norepinephrine, the major transmitter in postganglionic sympathetic nerve endings, is synthesized from the amino acid tyrosine. The conversion of tyrosine to L-DOPA is the rate-limiting step in the pathway. In the adrenal medulla, epinephrine is synthesized from norepinephrine by methylation.

in the cytoplasm of the cells expressing these enzymes. Dopamine is transported into storage granules, where it is converted into norepinephrine by the enzyme dopamine  $\beta$ -hydroxylase. Epinephrine is primarily synthesized in the adrenal medulla, which expresses the enzyme phenylethanolamine-*N*-methyltransferase, which methylates norepinephrine that has diffused out of the storage granules. Epinephrine is then itself taken up into storage granules, where it is stored until it is released. Reserpine is an antihypertensive drug that inhibits the uptake of norepinephrine into these storage granules, ultimately leading to the depletion of catecholamines from storage granules.

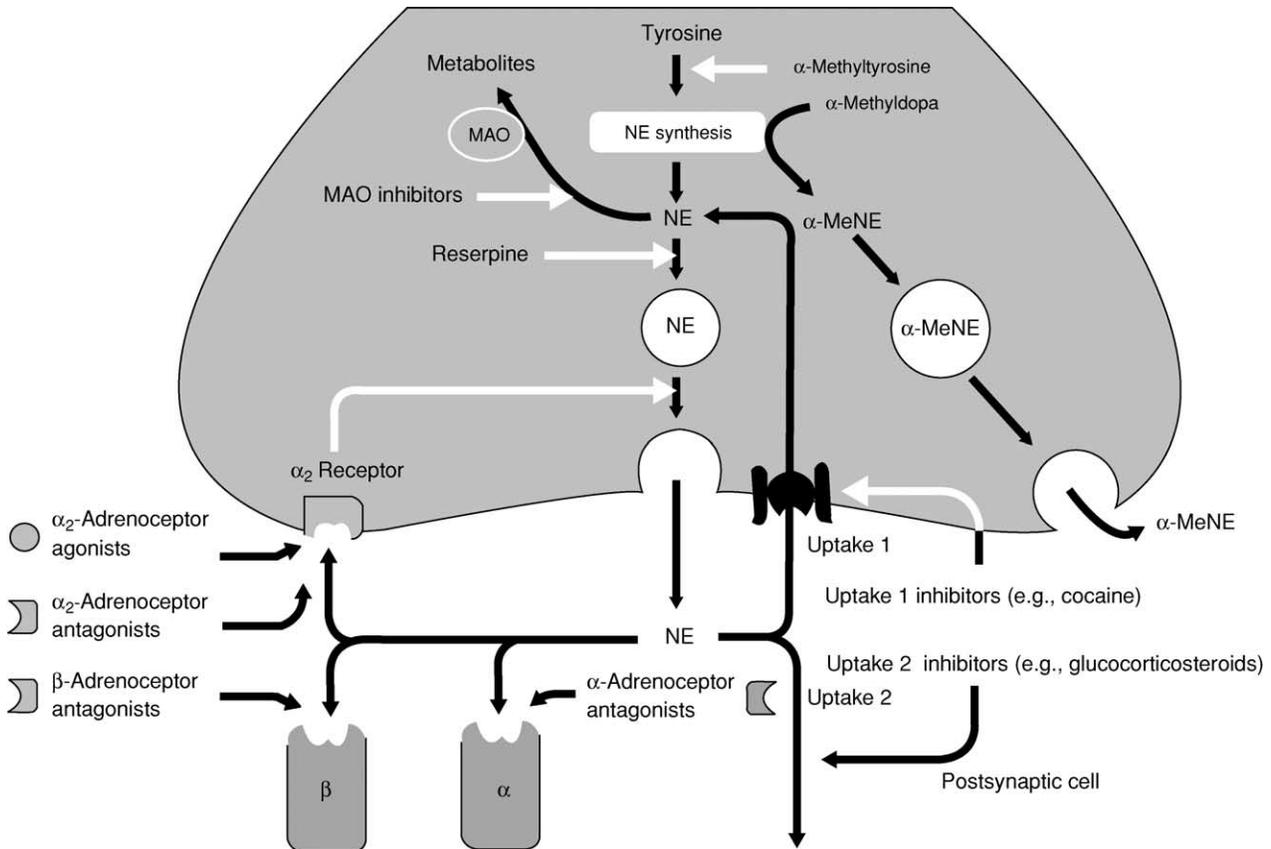
## Release of Catecholamines

Norepinephrine is released into the synaptic cleft when the nerve containing the storage granules depolarizes (Fig. 2). Although the full sequence of events is not yet clear, the mechanism critically involves the rise in intracellular calcium concentrations that occurs with depolarization. As might be expected, this process is tightly regulated, including regulation by a variety of hormone and neurotransmitter receptors expressed on sympathetic neurons. Indeed, these nerve endings typically contain a receptor that is activated by norepinephrine itself, which serves to autoinhibit norepinephrine release with subsequent neuronal firings (the  $\alpha_2$ -adrenergic receptor). The drug amphetamine has a major action of directly promoting the release of norepinephrine from these nerve endings independently of the firing of the sympathetic nerve.

## Termination of the Action of Catecholamines

After interacting with adrenergic receptors on postsynaptic cells, the major mechanism of terminating the action of norepinephrine is reuptake into the sympathetic nerve ending. There is a specific transport pump for norepinephrine expressed on these neurons. Interestingly, the activity of this transporter is blocked by cocaine. Consequently, whereas the major psychotropic actions of cocaine involve the inhibition of activity of dopamine transporters in brain centers, some of the cardiovascular actions of cocaine may involve this effect on norepinephrine uptake in the periphery.

There are three other mechanisms leading to termination of the action of norepinephrine, namely, diffusion out of the synaptic cleft into the bloodstream, uptake into other tissues, and transformation to inactive metabolites. Monoamine oxidase (MAO) plays a major role in inactivating catecholamines. MAO A and MAO B are subtypes of this enzyme. MAO is expressed mainly in nerve terminals and liver but also in several other tissues. The mitochondrial-associated MAO in nerve endings can inactivate norepinephrine not found in secretion granules. MAO in liver prevents substances such as tyramine, which is found in wine and certain foods, from reaching the systemic circulation in an intact form. In other words, blood draining the intestines passes to the liver, inactivating the tyramine before it can reach the systemic circulation. Since tyramine is very efficacious in causing the direct release of norepinephrine from nerve endings, there is a major risk of adverse reactions, such as marked increases in blood pressure, due



**Figure 2** Norepinephrine (NE) release and action in sympathetic synapses with postsynaptic target cells. NE is stored in granules that are released into the synaptic cleft when the nerve fires. NE then interacts with postsynaptic adrenergic receptors on target cells, for example, in the heart, blood vessels, and lungs. The action of NE is terminated largely by reuptake into the nerve ending (uptake 1); other mechanisms include uptake into other tissues (uptake 2), metabolism by monoamine oxidase (MAO) or catechol-*O*-methyltransferase (not shown), or diffusion out of the synapse. Many of these specific steps can be inhibited by important drugs that are used clinically or illicitly.

to ingestion of certain foods or wine for patients taking a MAO inhibitor that inhibits liver enzyme activity. Another important enzyme in the metabolism of catecholamines is catechol-*O*-methyltransferase (COMT). Catecholamines that reach the systemic circulation are extensively methylated by COMT.

### ADRENERGIC RECEPTORS AND SIGNAL TRANSDUCTION MECHANISMS

Catecholamines modulate cellular function by activating adrenergic receptors (also termed adrenoceptors). There are three major classes of adrenergic receptors, termed  $\alpha_1$ ,  $\alpha_2$ , and  $\beta$ . These receptor classes have been delineated by a combination of classical pharmacological experiments as well as by molecular cloning. It turns out that each of these classes of receptor has

additional subtypes:  $\alpha_1A$ ,  $\alpha_1B$ , and  $\alpha_1D$ ;  $\alpha_2A$ ,  $\alpha_2B$ , and  $\alpha_2C$ ; and  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$ . This demarcation has been considerably influenced by comparisons of predicted protein sequences of the various receptors. These receptors are widely distributed with varying levels of expression of particular subtypes on cells throughout the body. There has been considerable interest in designing drugs that are highly specific for one or another class of receptor as well as drugs that can be distinguished by the various subtypes of each of these receptors. Extensive work has been carried out to identify a key receptor subtype that might mediate the major actions of catecholamines that might have relevance for a particular disease. For example, it has been known for quite some time that  $\alpha_1$ -adrenergic antagonist drugs are efficacious in the relief of bladder obstruction in men who have enlargement of the prostate gland. Much has been invested in determining which of the various

$\alpha$ 1 receptor subtypes might be the most important in causing contraction of smooth muscle in prostate, a response that contributes to obstruction of the bladder neck. There is a drug available that has a modestly higher affinity for  $\alpha$ 1A receptors that play a major role in smooth muscle contraction in the prostate. It is a considerable challenge to predict whether or not one subtype is most important when many subtypes may be expressed on target cells and indeed it can be very difficult to know in advance of considerable research efforts whether blockade of a single subtype will have therapeutic efficacy.

In addition to the heterogeneity of the various adrenergic receptors, there are a multiplicity of signal transduction pathways utilized by these receptors to regulate cellular function, some of which are illustrated in Fig. 3. Each of the  $\beta$ -adrenergic receptors activates the G protein  $G_s$ , which in turn activates the enzyme adenylyl cyclase, leading to increased synthesis of cyclic AMP (cAMP). The rise in cAMP concentration in turn activates cAMP-dependent protein kinase, which phosphorylates a number of target proteins, leading to immediate changes in the activity of these enzymes as well as phosphorylating transcription factors such as CREB (cAMP response element binding protein), which alters the transcription of a number of target genes. One of the signaling pathways of  $\alpha$ 2-adrenergic receptors acts to inhibit adenylyl cyclase activity by signaling via the inhibitory G protein  $G_i$ . In addition, the subunits of  $G_i$  can lead to the activation of a variety of other signal transduction pathways. A main signaling pathway of  $\alpha$ 1-adrenergic receptors involves activation of phospholipase C, leading to the release of inositol 1,4,5-trisphosphate and diacylglycerol, which have important roles in signaling via increasing intracellular concentrations and the activation of protein kinase C, respectively. In addition,  $\alpha$ 1 receptors activate a variety of downstream protein kinases that have considerable overlap with those activated by diverse tyrosine kinase cell surface receptors.

## CATECHOLAMINES AND DISEASE

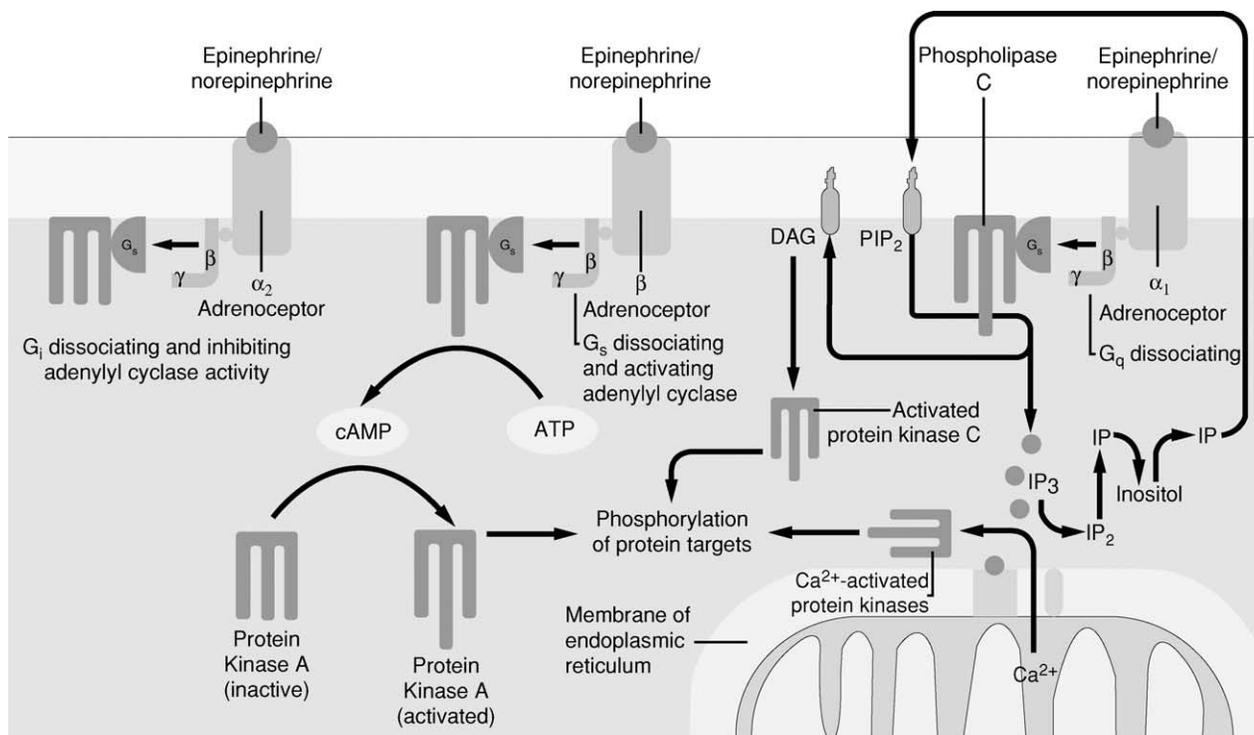
### Pheochromocytoma

Pheochromocytoma is an uncommon tumor that secretes catecholamines to excess and in an unregulated manner. Patients may have, continuously or intermittently, markedly elevated concentrations of catecholamines in the blood. In addition, metabolites of catecholamines may be present at increased concentrations in both blood and urine. Clinical manifestations of pheochromocytoma can be highly varied.

The most common symptoms are headaches, palpitations, and increased sweating. Many patients have hypertension, which may be sustained or intermittent depending at least in part on the pattern of release of active catecholamines into the blood. Some patients may have only subtle symptoms leading to delays in diagnosis, whereas others may have profound "spells," occasionally or infrequently, that are highly characteristic of the disease. Once the diagnosis is suspected, a variety of catecholamine-related assays are available to confirm the diagnosis. Subsequently, the next step is typically to identify the location of the tumor(s) in each patient. Approximately 90% of pheochromocytomas are localized in an adrenal gland. These can usually be identified with computed tomography or magnetic resonance imaging scanning techniques. Approximately 10% of tumors are extra-adrenal (ranging from the neck to the urinary bladder) and require more sophisticated techniques to localize. The primary treatment of pheochromocytoma is surgical excision of the tumor, which is possible in a large majority of cases. A small fraction of tumors are malignant, requiring sustained medical treatment. These tumors are typically slow-growing and the adverse effects of the excess catecholamines can be mitigated by using drugs that antagonize the actions of catecholamines at  $\alpha$ -and/or  $\beta$ -adrenergic receptors or by drugs that inhibit catecholamine synthesis, for example, by inhibiting the enzyme tyrosine hydroxylase, which catalyzes the rate-limiting step in catecholamine synthesis.

### Autonomic Insufficiency

Whereas pheochromocytoma is a disease characterized by catecholamine excess, failure of the sympathetic nervous system is a good example of a disease due to inadequate norepinephrine. The sympathetic nervous system is critically involved in the capacity of humans to attain an upright posture. On standing, blood would tend to pool in dependent parts of the body, for example, in veins in the legs, due to gravity. On standing, activation of the sympathetic nervous activity leads to the stimulation of  $\alpha$ -adrenergic receptors in the veins, leading to smooth muscle contraction and diminished venous pooling so as to maintain the return of blood to the heart and maintain cardiac output. In the absence of an adequate response of the sympathetic neurons with norepinephrine release, there is typically pooling of blood with a consequent fall in cardiac output and a fall in blood pressure, which may lead to inadequate cerebral perfusion and fainting.



**Figure 3** Signal transduction mechanisms of  $\alpha_1$ -,  $\alpha_2$ -, and  $\beta$ -adrenergic receptors. Catecholamines, such as epinephrine and norepinephrine, activate cells by binding to specific cell surface receptors, leading to association of these receptors with specific G proteins, including members of the  $G_s$  ( $\beta$  receptors),  $G_i$  ( $\alpha_2$  receptors), and  $G_q$  ( $\alpha_1$  receptors) families. The  $\beta$  receptor's major mechanism of action involves increasing cyclic AMP (cAMP) synthesis, whereas  $\alpha_2$  receptors inhibit the accumulation of cAMP.  $\alpha_1$  Receptors stimulate the production of inositol 1,4,5-trisphosphate ( $IP_3$ ) and diacylglycerol (DAG), which promote an increase in intracellular  $Ca^{2+}$  concentration and activation of protein kinase C, respectively. Not shown in the figure are the effects of these G protein-coupled receptors on other effectors, such as ion channels, and the role that these receptors play in modulating the activity of other kinase families, including the tyrosine kinases.

There are a number of primary diseases of the nervous system that lead to inadequate sympathetic nervous system function; in addition, patients with diseases such as diabetes mellitus may have autonomic dysfunction due to associated neuropathies. Postural hypotension in these patients may be treated with a range of therapeutic modalities including the use of synthetic drugs to mimic some of the effects of noradrenaline, especially in order to activate  $\alpha$  receptors.

### See Also the Following Articles

Adrenergic Receptors • G Protein-Coupled Receptors • Neurotransmitters, Overview • Norepinephrine Receptors • Norepinephrine Transporter • Normetanephrine and Metanephrine • Pheochromocytoma

### Further Reading

Akduman, B., and Crawford, E. D. (2001). Terazosin, doxazosin, and prazosin: Current clinical experience. *Urology* **58**, 49–54.

- Esler, M., Lambert, G., Brunner-La Rocca, H. P., Vaddadi, G., and Kaye, D. (2003). Sympathetic nerve activity and neurotransmitter release in humans: Translation from pathophysiology into clinical practice. *Acta Physiol. Scand.* **177**, 275–284.
- Lenders, J. W., Pacak, K., and Eisenhofer, G. (2002). New advances in the biochemical diagnosis of pheochromocytoma: Moving beyond catecholamines. *Ann. N. Y. Acad. Sci.* **970**, 29–40.
- Pacak, K., Eisenhofer, G., Carrasquillo, J. A., Chen, C. C., Whately, M., and Goldstein, D. S. (2002). Diagnostic localization of pheochromocytoma: The coming of age of positron emission tomography. *Ann. N. Y. Acad. Sci.* **970**, 170–176.
- Parmer, R. J., and Zinder, O. (2002). Catecholaminergic pathways, chromaffin cells, and human disease. *Ann. N. Y. Acad. Sci.* **971**, 497–505.
- Piascik, M. T., and Perez, D. M. (2001).  $\alpha_1$ -Adrenergic receptors: New insights and directions. *J. Pharmacol. Exp. Ther.* **298**, 403–410.
- Pierce, K. L., Premont, R. T., and Lefkowitz, R. J. (2002). Seven-transmembrane receptors. *Nat. Rev. Mol. Cell Biol.* **3**, 639–650.
- Rockman, H. A., Koch, W. J., and Lefkowitz, R. J. (2002). Seven-transmembrane-spanning receptors and heart function. *Nature* **415**, 206–212.
- Waldeck, B. (2002).  $\beta$ -Adrenoceptor agonists and asthma—100 years of development. *Eur. J. Pharmacol.* **445**, 1–12.



# CCK (Cholecystokinin)

Rodger A. Liddle

Duke University, Durham, North Carolina, United States

## Glossary

**gastrin** A gastrointestinal polypeptide hormone produced by G cells of the gastric antrum that stimulates gastric acid secretion.

**G protein-coupled receptor** A family of cell surface receptor proteins consisting of seven transmembrane regions; the intracellular region interacts with GTP-binding proteins that, on ligand binding, transduce signals within the cell, leading to a cellular response (e.g., secretion, motility, growth).

**microvilli** Finger-like extensions along the apical surface of intestinal mucosal cells; increase the absorptive surface of the cell (on enterocytes); allow potential stimuli greater exposure to their targets (on endocrine cells).

**myenteric plexus** Also known as Auerbach's plexus, a system of nerves and ganglia lying within the longitudinal and circular muscle layers of the intestine; its nerves innervate numerous targets, including the myenteric externa, mucosa, and sympathetic prevertebral ganglia.

**sphincter of Oddi** The muscular region surrounding the distal ends of the common bile duct and pancreatic duct as they enter into the duodenum; when constricted, this sphincter prevents flow of bile and pancreatic juice into the duodenum and restricts reflux of duodenal contents back into the bile and pancreatic ducts.

**submucosal plexus** A network of nerves and small ganglia found in the submucosa of the intestine; comprised of outer and inner layers; transmits secretomotor and vasodilator stimuli to the mucosa; primary sensory nerves contained in this plexus, which also communicates with the myenteric plexus.

Cholecystokinin, a peptide hormone, is produced and secreted by endocrine cells of the upper small intestine following food ingestion. It is the major hormone responsible for stimulating pancreatic enzyme secretion and gallbladder contraction. Cholecystokinin promotes satiety, delays gastric emptying, potentiates insulin secretion, and may regulate bowel

motility. It may also play a role in learning and memory, anxiety, analgesia, and thermoregulation.

## INTRODUCTION

Cholecystokinin (CCK) is a peptide hormone produced by endocrine cells of the upper small intestine. It is secreted on ingestion of a meal and is the major hormone responsible for pancreatic enzyme secretion and gallbladder contraction. CCK was discovered by Ivy and Oldberg in 1928 when they recognized that intestinal extracts could stimulate gallbladder contraction in dogs, and it was named the substance cholecystokinin (“cholecyst” [gallbladder] and “kinin” [to move]). In 1943, Harper and Raper noted that a similar extract, which they named “pancreozymin,” stimulated pancreatic enzyme secretion. It was not until CCK was purified and its amino acid sequence was determined by Mutt in 1968 that it was proven that CCK and pancreozymin were the same hormone that possessed the ability to stimulate both the gallbladder and the pancreas. The original name, cholecystokinin, is the term commonly used.

Over the past three decades, CCK has been found to have many other biological effects. In experimental animals and humans, CCK has been shown to delay gastric emptying, potentiate insulin secretion, and regulate bowel motility. One of the most noteworthy actions of CCK is its ability to induce satiety and reduce food intake. Until the development of reliable assays for measuring blood levels of CCK, the physiological effects of CCK remained controversial. However, it has now been shown in humans that physiological levels of CCK stimulate gallbladder contraction and pancreatic enzyme secretion, inhibit gastric emptying, potentiate insulin secretion, and reduce food intake. CCK may regulate bowel motility and, in certain species, promote pancreatic growth, although that has not yet been proven physiologically. Less well described but fascinating actions of CCK include its effects on learning and memory, anxiety, analgesia, and thermoregulation.

In the small intestine, CCK is produced by discrete endocrine cells within the mucosa. However, CCK is even more abundant in the brain and is found in peripheral nerves innervating the intestine where it functions as a neurotransmitter.

## MOLECULAR FORMS

The original CCK peptide isolated from the intestine was a triacontapeptide (CCK-33). Since then, several larger and smaller molecular forms of CCK have been found in the intestine, brain, and blood of experimental animals and humans. The biologically active region of CCK resides in its carboxyl terminus, and all forms of CCK possess an identical carboxyl 5-amino acid sequence (-Gly-Trp-Asp-Met-Phe-NH<sub>2</sub>) (Fig. 1). This region is common to gastrin, and as a result, gastrin has some CCK-like activity (albeit weak) and CCK shares some weak gastrin-like activity. The amino acid sequence shared by the two hormones has made developing assays for CCK difficult because antibodies directed against the biologically active region of CCK may cross-react with gastrin. This problem is accentuated by the finding that circulating levels of gastrin are 10- to 100-fold greater than those of CCK.

CCK is produced from a single gene that encodes a 115-amino acid preprohormone. By posttranslational processing, molecular forms of CCK ranging in size from 4 to 83 amino acids have been identified in tissues and blood. However, the major molecular forms of CCK are CCK-8, CCK-33, and CCK-58.

In humans, the CCK gene is located in the 3q12–3pter region of chromosome 3. CCK expression is

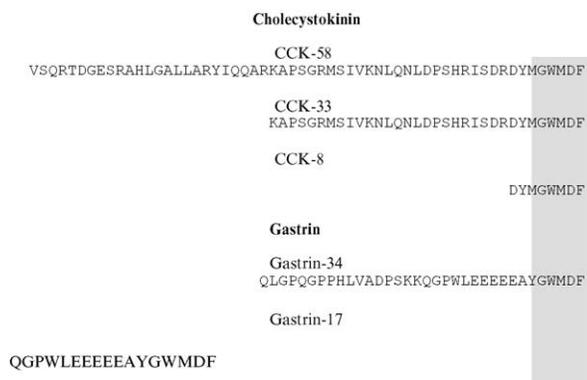
both tissue specific and developmentally regulated. In the intestine, the CCK gene is expressed prenatally and, after birth, is regulated primarily by ingestion of foods that stimulate CCK secretion. In the central nervous system, stimuli regulating neuronal CCK gene transcription include growth factors, second messengers such as cyclic AMP, the neurotransmitter dopamine, and hormones such as estrogen.

## DISTRIBUTION

Cholecystokinin cells are individual flask-shaped cells that are scattered throughout the mucosa of the small intestine. The concentration of CCK cells is greatest in the proximal small intestine and diminishes in a gradient fashion toward the distal jejunum and ileum. CCK cells arise from progenitor cells in the intestinal crypts and, along with enterocytes, migrate up the villus. Residing in the mucosa, the apical surface of CCK cells is open to the lumen of the intestine. Here, cells can actually “sample” luminal contents such as food and releasing factors. These enteroendocrine cells also possess microvilli that increase the exposed surface area, thereby allowing greater exposure to potential stimuli.

Like other gastrointestinal hormones, CCK is a “brain–gut” peptide, meaning that the same transmitter is found in both the central nervous system and the intestine. In the brain, CCK is highly concentrated in the striatum, hippocampus, and cerebral cortex. CCK-containing neurons may also synthesize dopamine. Such nerves have been shown to extend to the limbic forebrain and ventromedial hypothalamus, where they may participate in controlling food intake. CCK has been demonstrated to modulate dopamine release, dopamine-mediated reward, and receptor binding and function. These actions have implications for a role of CCK in drug abuse and neurological and psychiatric disease.

CCK is prevalent in peripheral nerves of the gastrointestinal tract. It is most abundant in nerves innervating the colon, with fewer CCK nerves in the ileum. In these locations, CCK is present in myenteric and submucosal plexi, where it innervates ganglionic bodies. CCK is also abundant in the vagus nerve and celiac plexus. In the intestine, CCK stimulates acetylcholine release and causes smooth muscle contraction. Postganglionic CCK-containing neurons also terminate around the islets of Langerhans, where CCK may stimulate islet hormone secretion such as insulin and glucagon release.



**Figure 1** The amino acid sequences of the common molecular forms of CCK and gastrin. All biologically active forms of CCK and gastrin share an identical carboxyl terminal pentapeptide sequence (shown in gray).

## CCK RECEPTORS

CCK exerts its biological actions by binding to specific receptors on its target tissues. Gastrointestinal CCK receptors reside on tissues of the pancreas, gallbladder, stomach, lower esophageal sphincter, ileum, and colon. In the nervous system, CCK receptors are abundant in brain and on some peripheral nerves. Two types of CCK receptors have been identified. CCK-A (for alimentary) receptors are the primary CCK receptor and mediate most of CCK's effects in the gastrointestinal tract. The CCK-B (for brain) receptor is identical to the gastrin receptor and is the major CCK receptor subtype in the nervous system. It is also abundant in the stomach. Both receptors are G-protein-coupled, seven-membrane-spanning proteins but arise from different genes [Wank et al., 1992; Kopin et al., 1992]. CCK receptor antagonists have been extremely useful in pharmacological and physiological studies in defining the physiological role of CCK. The first CCK receptor antagonists useful for *in vitro* studies were cyclic nucleotide analogues (e.g., dibutyryl cGMP). Subsequently, amino acid derivatives (e.g., CR-1409), carboxyl-terminal CCK analogues, and substituted benzodiazepines (e.g., devazepide) were developed and would be used *in vivo*. Clinical studies have shown CCK receptor antagonists to inhibit CCK- and meal-stimulated gallbladder contraction, accelerate gastric emptying, and induce hunger (i.e., reverse satiety), indicating that CCK has important physiological roles in each of these processes.

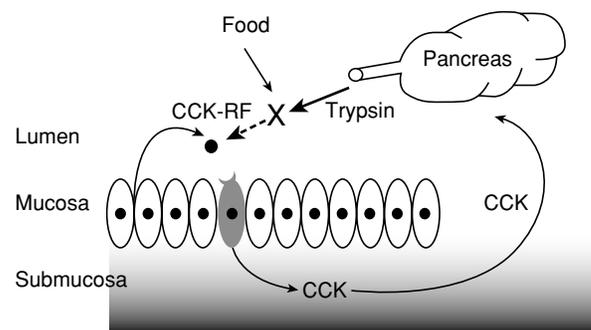
## CCK RELEASE

CCK is secreted from specialized endocrine cells of the mucosa (known as I cells) into the extracellular space where it is taken up into the bloodstream. [Polak et al., 1975b] It is by this mechanism that circulating CCK reaches distant target tissues such as the pancreas and gallbladder. Enteric endocrine cells package CCK in secretory granules that are stored along the basolateral surface of the cell, thereby allowing CCK to be secreted into the interstitium when the cell is stimulated. *In vivo*, ingested proteins and fat are the major dietary stimulants of CCK release. However, in some species (including humans) and in cell preparations *in vitro*, partially digested nutrients such as amino acids, peptides, and fatty acids are potent releasers of CCK, indicating that these components directly interact with the CCK cell.

Circulating blood levels of CCK average approximately 1 pM in the fasting state and increase to

between 5 and 8 pM after eating. Postprandial levels remain elevated for 3 to 5 h as food empties from the stomach into the upper small intestine. Gastric distention does not influence CCK release. Although fat and protein are the primary stimulants of CCK, carbohydrate also has a modest effect on secretion.

It is well recognized that inactivation of protease activity in the lumen of the small intestine of rodents stimulates CCK release and pancreatic exocrine secretion. This phenomenon is now known as negative feedback control of CCK release. Not only has this principle been demonstrated in rodents, but it also applies to other species, including humans. Thus, CCK release is controlled in part by the presence or absence of pancreatic enzymes in the intestine (Fig. 2). This concept indicated that there existed intestinal-releasing factors that are secreted into the intestine and that stimulate CCK secretion [Spannagel et al., 1996; Herzig et al., 1996]. When pancreatic enzymes are present in the intestine, these CCK-releasing factors are inactivated; however, when pancreatic enzymes are inhibited or pancreatic secretion is reduced, these CCK-releasing factors are active and stimulate CCK secretion. Similarly, with ingestion of a meal that temporarily binds trypsin and other digestive enzymes, CCK-releasing factors are also available to stimulate CCK secretion. To date, a human counterpart of a CCK-specific releasing factor has not been identified.



**Figure 2** Model for regulation of CCK release by an intraluminal releasing factor. One or more CCK-releasing factors (CCK-RF) stimulates intestinal CCK secretion. CCK-RF is normally secreted in the intestinal lumen, where it is exposed to pancreatic enzymes. Under basal conditions, CCK-RF is inactivated by even small amounts of enzyme; however, following a meal, food competes for enzyme binding, allowing CCK-RF to interact with CCK cells to stimulate hormone secretion. CCK in the circulation stimulates pancreatic secretion that, following digestion of food, restores CCK-RF and CCK secretion to basal levels. Modified from Liddle, R. A. (1995). Regulation of cholecystokinin secretion by intraluminal releasing factors. *Am. J. Physiol. Gastrointest. Liver Physiol.* **269**, G319–G327.

## BIOLOGICAL ACTIONS OF CCK

CCK is the major hormonal regulator of gallbladder contraction. Coincident with this effect, CCK also relaxes the sphincter of Oddi, which also promotes bile secretion into the intestine. In humans, the predominant CCK receptor type in the pancreas is CCK-B. This differs from most species' CCK, where CCK-A receptors predominate in the pancreas and where CCK is a potent stimulant of pancreatic secretion. Therefore, in humans, although CCK stimulates pancreatic secretion, its role may be limited.

At physiological blood concentrations that occur after a meal, CCK delays gastric emptying, and this may be important for its ability to reduce food intake and induce satiety. Because of its effects on gallbladder contraction, pancreatic secretion, and gastric emptying, CCK coordinates many digestive processes. Thus, CCK plays an important role in the ingestion and digestion of a meal. Although it has been shown that CCK causes relaxation of the lower esophageal sphincter and promotes intestinal motility, it appears as though these effects are neural rather than hormonal. CCK receptors have been found on some gastrointestinal and lung cancers; however, it remains unknown whether CCK plays a role in human cancer growth.

## CLINICAL USES OF CCK

CCK is used along with secretin as a test of pancreatic function. In patients with pancreatic insufficiency, low levels of pancreatic juice are recovered following intravenous injection of these hormones. CCK can also be used clinically to stimulate gallbladder contraction and is helpful in radiographic testing of the gallbladder function. For diagnostic purposes, CCK has facilitated the collection of bile and pancreatic juice for cytological examination.

Therapeutically, CCK injections have been administered to patients who are unable to eat (e.g., parental alimentation) to stimulate gallbladder contraction. This therapy has been effective in reducing gallbladder sludge and in preventing gallstone formation.

Low blood levels of CCK have been reported in patients with celiac disease and bulimia nervosa and in conditions that delay gastric emptying. The defect in celiac disease is likely due to reduced CCK secretion from diseased small intestinal mucosa. The cause of abnormal CCK responses in bulimia is unknown but may be related to alterations in gastric emptying given that normal postprandial CCK release is dependent on delivery of food from the stomach to the

small intestine. It is not known whether CCK deficiency contributes to any specific pathological consequences. There are no known diseases of cholecystokinin excess.

## Acknowledgments

This article was modified from Liddle, R. A. (2003). Cholecystokinin (CCK). In "Encyclopedia of Hormones" (H. L. Henry and A. W. Norman, eds.), pp. 264–268. Elsevier Science (USA) San Diego.

## See Also the Following Articles

Gallbladder and Biliary Secretion • Gastrin • Gastrin-Releasing Peptide • GI Hormones Outside the Gut: Central and Peripheral Nervous System • Hunger and Satiation

## Further Reading

- Calam, J., Ellis, A., and Dockray, G. J. (1982). Identification and measurement of molecular variants of cholecystokinin in duodenal mucosa and plasma. *J. Clin. Invest.* **69**, 218–225.
- Doty, J. E., Pitt, H. A., Porter-Fink, V., and Denbesten, L. (1985). Cholecystokinin prophylaxis of parenteral nutrition-induced gallbladder disease. *Ann. Surg.* **201**, 76–80.
- Geraciotti, T. D., Jr., and Liddle, R. A. (1988). Impaired cholecystokinin secretion in bulimia nervosa. *N. Engl. J. Med.* **319**, 683–688.
- Gibbs, J., Young, R. C., and Smith, G. P. (1973). Cholecystokinin decreases food intake in rats. *J. Comp. Physiol. Psychol.* **84**, 488–495.
- Hansen, T. v. O. (2001). Cholecystokinin gene transcription: Promoter elements, transcription factors, and signaling pathways. *Peptides* **22**, 1201–1211.
- Harper, A. A., and Raper, H. S. (1943). Pancreozymin, a stimulant of secretion of pancreatic enzymes in extracts of the small intestine. *J. Physiol.* **102**, 115–125.
- Herzig, K. H., Schon, I., Tatemoto, K., Ohe, Y., Folsch, U. R., and Owyang, C. (1996). Diazepam binding inhibitor is a potent cholecystokinin-releasing peptide in the intestine. *Proc. Natl. Acad. Sci. USA* **93**, 7927–7932.
- Ivy, A. C., and Oldberg, E. (1928). A hormone mechanism for gallbladder contraction and evacuation. *Am. J. Physiol.* **65**, 599–613.
- Jorpes, E., and Mutt, V. (1966). Cholecystokinin and pancreozymin: One single hormone?. *Acta Physiol. Scand.* **66**, 196–202.
- Konturek, S. J., Kwiecien, N., Obtulowicz, W., Kopp, B., Oleksy, J., and Rovati, L. (1990). Cholecystokinin in the inhibition of gastric secretion and gastric emptying in humans. *Digestion* **45**, 1–8.
- Kopin, A. S., Lee, Y.-M., McBride, E. W., Miller, L. J., Lu, M., Lin, H. Y., Kolakowski, L. F., Jr., and Beinborn, M. (1992). Expression cloning and characterization of the canine parietal cell gastrin receptor. *Proc. Natl. Acad. Sci. USA* **89**, 3605–3609.
- Liddle, R. A. (1994). Cholecystokinin. In "Gut Peptides: Biochemistry and Physiology" (J. H. Walsh, and G. J. Dockray, eds.), pp. 175–216. Raven, New York.
- Liddle, R. A. (1995a). Regulation of cholecystokinin secretion by intraluminal releasing factors. *Am. J. Physiol. Gastrointest. Liver Physiol.* **269**, G319–G327.

- Liddle, R. A. (1995b). Regulation of cholecystokinin secretion by intraluminal releasing factors. *Am. J. Physiol.* **269**, G319–G327.
- Liddle, R. A., Gertz, B. J., Kanayama, S., Beccaria, L., Coker, L. D., Turnbull, T. A., and Morita, E. T. (1989). Effects of a novel cholecystokinin (CCK) receptor antagonist, MK-329, on gallbladder contraction and gastric emptying in humans: Implications for the physiology of CCK. *J. Clin. Invest.* **84**, 1220–1225.
- Liddle, R. A., Goldfine, I. D., Rosen, M. S., Taplitz, R. A., and Williams, J. A. (1985). Cholecystokinin bioactivity in human plasma: Molecular forms, responses to feeding, and relationship to gallbladder contraction. *J. Clin. Invest.* **75**, 1144–1152.
- Lund, T., GeurtsvanKessel, A. H. M., Haun, S., and Dixon, J. E. (1986). The genes for human gastrin and cholecystokinin are located on different chromosomes. *Hum. Genet.* **73**, 77–80.
- Mutt, V. (1988). Secretin and cholecystokinin. In "Gastrointestinal Hormones" (V. Mutt, ed.), pp. 251–320. Academic Press, San Diego.
- Niederau, C., Heintges, T., Rovati, L., and Strohmeyer, G. (1989). Effects of loxiglumide on gallbladder emptying in healthy volunteers. *Gastroenterology* **97**, 1331–1336.
- Owyang, C., Louie, D. S., and Tatum, D. (1986). Feedback regulation of pancreatic enzyme secretion: Suppression of cholecystokinin release by trypsin. *J. Clin. Invest.* **77**, 2042–2047.
- Polak, J., Pearse, A., Bloom, S., Buchan, A., Rayford, P., and Thompson, J. (1975a). Identification of cholecystokinin-secreting cells. *Lancet* **2**, 1016–1021.
- Polak, J. M., Pearse, A. G. E., Bloom, S. R., Buchan, A. M. J., Rayford, P. L., and Thompson, J. C. (1975b). Identification of cholecystokinin-secreting cells. *Lancet* **2**, 1016–1017.
- Rehfeld, J. F. (1989). Cholecystokinin. In "Handbook of Physiology: The Gastrointestinal System, Vol. 2: "Neural and Endocrine Biology" (G. M. Makhlof, ed.), pp. 337–358. Oxford University Press, New York.
- Rozengurt, E., and Walsh, J. H. (2001). Gastrin, CCK, signaling, and cancer. *Annu. Rev. Physiol.* **63**, 49–76.
- Spannagel, A. W., Green, G. M., Guan, D., Liddle, R. A., Faull, K., and Reeve, J. R., Jr. (1996). Purification and characterization of a luminal cholecystokinin-releasing factor from rat intestinal secretion. *Proc. Natl. Acad. Sci. USA* **93**, 4415–4420.
- Szelenyi, A. (2001). Cholecystokinin and thermoregulation: A minireview. *Peptides* **22**, 1245–1250.
- Wank, S. A., Harkins, R., Jensen, R. T., Shapira, H., De Weerth, A., and Slattery, T. (1992). Purification, molecular cloning, and functional expression of the cholecystokinin receptor from rat pancreas. *Proc. Natl. Acad. Sci. USA* **89**, 3125–3129.
- Wank, S. A., Pisegna, J. R., and De Weerth, A. (1994). Cholecystokinin receptor family: Molecular cloning, structure, and functional expression in rat, guinea pig, and human. *Ann. NY Acad. Sci.* **713**, 49–66.
- Weller, A., Smith, G. P., and Gibbs, J. (1990). Endogenous cholecystokinin reduces feeding in young rats. *Science* **247**, 1589–1591.
- Wolkowitz, O. M., Gertz, B., Weingartner, H., Beccaria, L., Thompson, K., and Liddle, R. A. (1990). Hunger in humans induced by MK-329, a specific peripheral-type cholecystokinin receptor antagonist. *Biol. Psychol.* **28**, 169–173.



# Chemokines

Naofumi Mukaida  
Kanazawa University, Kanazawa, Japan

## Glossary

**chemokinesis** Increased random migratory activity of cells.

**chemotaxis** Increased directional migration of cells in response to concentration gradients of certain chemotactic factors.

**inflammation** A series of host reactions to tissue damages such as infections, traumas, tumors, and foreign particles; characterized by increased blood supply, increased vascular permeability, and transendothelial migration of leukocytes.

Chemokines are a large group of cytokines that act on receptors with seven-transmembrane domains and that have a variety of chemotactic and cell-activating properties on selected populations of target cells, particularly leukocytes.

## INTRODUCTION

Since the first description of CXCL8/interleukin-8 (IL-8) as a neutrophil chemotactic cytokine in 1987, a family of structurally related cytokines has been identified. Because most of them exhibit chemotactic activity against a limited spectrum of leukocytes, they are now called chemokines (*chemotactic cytokines*). Chemokines direct the movement of various types of cells, particularly leukocytes, along a concentration gradient by modulating expression and structure of adhesion molecules and cytoskeletal proteins of the target cells. Because of their molecular stability and target specificity, chemokines are presumed to be crucial in leukocyte infiltration and subsequent activation in inflammation. Moreover, chemokines participate in other biological phenomena such as immune reactions, angiogenesis, and organ development.

All chemokines except CX3CL1/fractalkine possess four cysteines at the well-conserved positions.

Chemokines exhibit 3- $\beta$  sheets with  $\alpha$  helix at the carboxyl-terminal portion due to the presence of two disulfide bridges formed between the first and third cysteines as well as between the second and fourth cysteines. Most chemokines are secreted proteins with a molecular weight of approximately 10 kDa. They exhibit a basic nature, and their  $\alpha$  helix structure is responsible for preferential binding to proteoglycans on the vascular endothelial cells and to extracellular matrix proteins. Chemokine receptors comprise a large branch of the rhodopsin family of cell surface G protein-coupled receptors (GPCRs) with seven-transmembrane domains. Functional binding to the target cells and subsequent signaling are mediated by these receptors.

Chemokines are divided into four subgroups: CXC, CC, C, and CX3C. CXC, CC, and CX3C chemokines have four cysteines, whereas C chemokines have only two cysteines, corresponding to the second and fourth cysteines in the other groups. CXC and CX3C chemokines are distinguished by the presence of one and three amino acids, respectively, whereas the first two cysteines are adjacent in CC chemokines. A large number of CXC (>15) and CC chemokines (>25) have been identified in humans and other species, but human lymphotactin  $\alpha$  and  $\beta$  and fractalkine and their equivalents in other species are the only examples of C and CX3C chemokines, respectively.

A novel nomenclature system has been proposed for chemokines and their receptors. Systematic chemokine names, shown in Table I with their common names, are built from cysteine subclass roots, followed by "L" for "ligand." The numbers correspond generally to the same number used in the corresponding gene nomenclature. Because most chemokine receptors are restricted to a single chemokine subclass, the nomenclature system of chemokine receptors is rooted by the chemokine subclass specificity, followed by "R" for "receptor" and the number.

Table 1 Chemokines and Their Receptors

Proposed names	Commonly used names	Main receptor
CXC chemokine		
CXCL1	Gro $\alpha$ , Melanoma growth stimulating activity (MGSA)	CXCR2
CXCL2	Gro $\beta$	CXCR2
CXCL3	Gro $\gamma$	CXCR2
CXCL4	Platelet factor-4 (PF-4)	Unknown
CXCL5	Endothelial-derived neutrophil attractant-78 (ENA-78)	CXCR2
CXCL6	Granulocyte chemotactic protein-2 (GCP-2)	CXCR1, CXCR2
CXCL7	Neutrophil-activating protein-2 (NAP-2)	CXCR2
CXCL8	Interleukin-8 (IL-8)	CXCR1, CXCR2
CXCL9	Monokine induced by interferon $\gamma$ (Mig)	CXCR3
CXCL10	Interferon-inducible protein-10 (IP-10)	CXCR3
CXCL11	Interferon-inducible T-cell $\alpha$ chemoattractant (I-TAC)	CXCR3
CXCL12	Stroma cell-derived factor-1 (SDF-1)	CXCR4
CXCL13	B-lymphocyte chemoattractant (BLC)	CXCR5
CXCL14	Breast and kidney (BRAK)	Unknown
CXCL15	Unidentified in human	
CXCL16		CXCR6
CC chemokine		
CCL1	I-309	CCR8
CCL2	Monocyte chemoattractant protein-1 (MCP-1)	CCR2, CCR11
CCL3	Macrophage inflammatory protein-1 $\alpha$ (MIP1 $\alpha$ )	CCR1, CCR5
CCL4	MIP-1 $\beta$	CCR5, CCR1
CCL5	Regulated on activation, normal T expressed and secreted (RANTES)	CCR5, CCR1, CCR3
CCL6	Unidentified in human	
CCL7	MCP-3	CCR1, CCR2
CCL8	MCP-2	CCR2, CCR5, CCR11
CCL9	Unidentified in human	
CCL10	Unidentified in human	
CCL11	Eotaxin	CCR3, CCR5
CCL12	Unidentified in human	
CCL13	MCP-4	CCR2, CCR3, CCR11
CCL14	Human hemofiltrate CC chemokine-1 (HCC-1)	CCR1
CCL15	HCC-2	CCR1, CCR3
CCL16	HCC-4	CCR1, CCR2, CCR3, CCR8
CCL17	Thymus and activation-regulated chemokine (TARC)	CCR4
CCL18	Pulmonary and activation-regulated chemokine (PARC)	Unknown
CCL19	EBI1-ligand chemokine (ELC), MIP-3 $\beta$	CCR7
CCL20	Liver and activation-regulated chemokine (LARC), MIP-3 $\alpha$	CCR6
CCL21	Secondary lymphoid tissue chemokine (SLC), 6CKine	CCR7
CCL22	Macrophage-derived chemokine (MDC)	CCR4
CCL23	Myeloid progenitor inhibitory factor-1 (MPIF-1)	CCR1
CCL24	Eotaxin-2, MPIF-2	CCR3
CCL25	Thymus-expressed chemokine (TECK)	CCR9
CCL26	Eotaxin-3	CCR3
CCL27	Interleukin-11 receptor $\alpha$ -locus chemokine (ILC)	CCR10
CCL28	Mucosae-associated epithelial chemokine (MEC)	CCR10, CCR3
CX3C chemokine		
CX3CL1	Fractalkine	CX3CR1
C chemokine		
XCL1	Lymphotactin	XCR1

## SIGNAL TRANSDUCTION MECHANISM

All chemokine receptors are membrane-bound molecules composed of seven-transmembrane domains and coupled to G proteins at the carboxyl-terminal portion and possibly the third intracellular loop. Following ligand binding, chemokine receptors are presumed to be internalized and subsequently recycled and reappear on the cell surface rapidly within 60 min, as observed on the human CXCL8 receptor system. In the case of the human CXCL8 receptor system, the inhibition of recycling reduced CXCL8-mediated chemotaxis, suggesting its important role in chemokine-mediated chemotaxis.

The ligand binding activates pertussis toxin-sensitive and receptor-coupled G proteins, particularly G $\alpha$ i proteins. G-proteins, on conversion to the guanosine triphosphate (GTP)-bound form, dissociate into G $\alpha$ - and G $\beta\gamma$ -subunits. Generated G $\beta\gamma$  activates phosphatidylinositol 4-phosphate kinase (PIP-K), phospholipase C (PLC)- $\beta$ , and phosphatidylinositol-3-kinase- $\gamma$  (PI3K $\gamma$ ). PIP-K and PLC- $\beta$  generate inositol 1,4,5-trisphosphate (IP $_3$ ) and diacyl glycerol (DAG). IP $_3$  induces a transient increase in intracellular Ca $^{2+}$  through mobilization of the intracellular Ca $^{2+}$  store, whereas DAG activates protein kinase C. These steps are required for superoxide production and granule release but not for chemotaxis.

The activation of PI3K $\gamma$  leads to the generation of PIP $_3$ . PIP $_3$ , in turn, activates protein kinase B (Akt) and small GTPases, resulting in chemotaxis and adherence. A human hematopoietic cell-specific CDM family protein, DOCK2, is indispensable for Rac activation and subsequent T- and B-lymphocyte migration in response to SDF-1/CXCL12 and BLC/CXCL13. However, the roles of these family proteins in other chemokine receptor-mediated cell migration remain unclear. PIP $_3$  also activates Raf and the mitogen-activated protein kinase (MAPK) pathway, leading to the transcription of various genes.

GPCR-mediated signals can be down-regulated by regulator of G-protein signaling (RGS) proteins. RGS proteins appear to enhance the endogenous GTPase activities and, thus, to decrease the half-life of the active GTP-bound state of both trimeric G proteins and small GTPases. Among RGS proteins, RGS1, RGS3, and RGS4 reduce CXCL8-mediated migration and adherence. Moreover, expression of RGS proteins also reduces CXCL8-induced MAPK activation. Major signal transduction pathways are shown in Fig. 1.

## CHEMOKINES IN THE ENDOCRINE SYSTEM

The number of neutrophils increases markedly in the thecal layer of the leading follicle immediately before the time of ovulation in humans. In *in vitro* fertilization (IVF), the administration of human chorionic gonadotropin (hCG) induced a rapid increase in CXCL8 concentration in follicular fluids. Moreover, during the normal menstrual cycle in humans, CXCL8 concentrations were higher in follicular fluids from dominant follicles of the late follicular/ovulatory phase than in those from the mid-follicular phase. Furthermore, pharmacological doses of CXCL8 could induce follicular maturation and growth in rats and rabbits. Macrophages invade the regressing corpus luteum to evacuate apoptotic luteal cells. Moreover, CCL2/monocyte chemoattractant protein-1 (MCP-1) mRNA and protein expression was enhanced at the site of the regressing corpus luteum, suggesting that CCL2 can regulate macrophage infiltration inside the corpus luteum. On the contrary, because no apparent defect in ovulation was observed in mice deficient in their cognate receptors, the roles of these chemokines in the ovulatory process still remain elusive.

During pregnancy, the amniotic fluid levels of several chemokines, including CXCL8, CXCL1/gro, CCL2, and CCL5/regulated upon activation, normal T expressed and secreted (RANTES), were increased progressively as the gestation stages advanced, reaching a maximum level at labor. Moreover, microbial infection of the amniotic cavity further enhanced chemokine levels in amniotic fluids. Furthermore, at parturition, the human lower uterine segment contains high levels of CXCL8 as well as leukocyte enzymes, matrix metalloproteinase-8 (MMP-8), and MMP-9 that may be responsible for cervical dilatation.

Leydig cells can constitutively produce CCL2 and CXCL10/IP-10. Moreover, Leydig cells can produce various chemokines in addition to CCL2 and CXCL10 by stimulation with lipopolysaccharide and viral infection. Interestingly, CXCL10 can inhibit hCG-induced testosterone formation.

Rat anterior pituitary gland cells express CXCL1, and the incubation of these cells with CXCL1 can increase prolactin secretion but suppress luteinizing hormone (LH) and follicle-stimulating hormone (FSH) secretion. In humans, CXCR2 is expressed by various normal endocrine cells, including pituitary cells, adrenal medulla, pancreatic islet cells, thyroid C cells, and neuroendocrine cells in the bronchi and gastrointestinal tracts. Moreover, CXCR2 is also

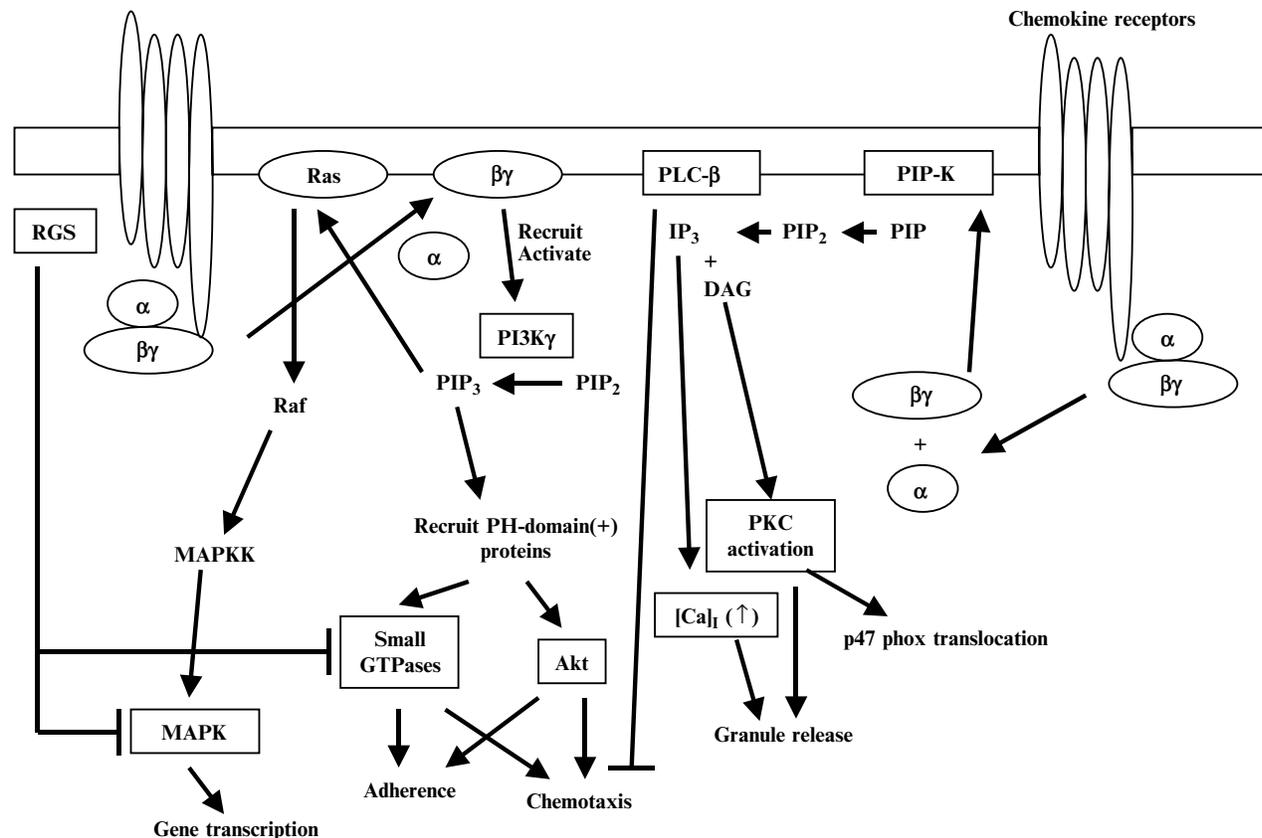


Figure 1 Major signaling pathways of chemokines.

expressed by carcinoids, pituitary adenomas, pheochromocytomas, and medullary carcinomas of the thyroid. Because some pituitary adenoma cells express CXCL8 constitutively, CXCL8 may be involved in malignant transformation of these cells in an autocrine and/or paracrine pathway.

Thyroid follicular cells can produce several chemokines, including CXCL8, CXCL12/stromal cell-derived factor-1 (SDF-1), CCL2, and CCL5, by the incubation with proinflammatory cytokines such as IL-1 and IL-6. In Graves' disease, a large number of CCR5<sup>+</sup> or CXCR3<sup>+</sup> lymphocytes infiltrated into the thyroid gland, and their ligands, CCL5, CXCL9/Mig, and CXCL10, were detected in the gland. Interestingly, CXCL9 and CXCL10 expression was maximal in recent-onset Graves' disease patients and correlated well with interferon- $\gamma$  expression. In autonomous thyroid adenomas, CXCL12 expression was reduced compared with that in normal thyroid tissues.

Parathyroid cells possess specific receptors for CXCL8 that can increase parathyroid hormone (PTH) mRNA expression and subsequent PTH release.

Transgenic CCL2 in pancreatic islets produced monocyte-rich insulinitis without diabetes, suggesting the involvement of an additional factor(s) in the development of diabetes. Several lines of evidence suggest the involvement of Th1-dominant immune response in diabetes development in nonobese diabetic (NOD) mice. In line with this assumption, CCL3/macrophage inflammatory protein-1 $\alpha$  (MIP-1 $\alpha$ ), CXCL9, and CXCL10 expression are enhanced, together with infiltration of CCR5<sup>+</sup> or CXCR3<sup>+</sup> lymphocytes, in the pancreas of NOD mice. Moreover, aberrant expression of these chemokines was documented in human type 1 diabetes patients. Furthermore, one analysis of NOD mice suggests that lymphocytes bearing CCR4 and their ligand, CCL22/macrophage-derived chemokine (MDC), also have crucial roles in the development of diabetes.

### See Also the Following Articles

Adipocytokines • Cytokine Actions, Cellular Mechanism of • Cytokine Receptors • Cytokines, Constitutive Secretion • Cytokines, Evolutionary Aspects and Functions • Cytokines,

Extracellular Transport and Processing • Janus Kinases and Cytokine Receptors

### Further Reading

- Kim, S. H., Cleary, M. M., Fox, H. S., Chantry, D., and Sarvetnick, N. (2003). CCR4-bearing T cells participate in autoimmune diabetes. *J. Clin. Invest.* **110**, 1675–1686.
- Murphy, P. M., Baggiolini, M., Charo, I. F., Herbert, C. A., Horuk, R., Matsushima, K., Miller, L. H., Oppenheim, J. J., and Power, C. A. (2000). International Union of Pharmacology XXII: Nomenclature for chemokine receptors. *Pharmacol. Rev.* **52**, 145–176.
- Olson, T. S., and Ley, K. (2002). Chemokines and chemokine receptors in leukocyte trafficking. *Am. J. Physiol. Regulatory Integrative Comp. Physiol.* **283**, R7–R28.
- Wong, K. H., Negishi, H., and Adashi, E. Y. (2002). Expression, hormonal regulation, and cyclic variation of chemokines in the rat ovary: Key determinants of the intraovarian residence of representatives of the white blood cell series. *Endocrinology* **143**, 784–791.
- Zlotnik, A., and Yoshie, O. (2000). Chemokines: A new classification system and their role in immunity. *Immunity* **12**, 121–127.



# Childhood Cancer, Endocrine Effects of

Bernadette M. D. Brennan

*Royal Manchester Children's Hospital, Manchester, United Kingdom*

Steve M. Shalet

*Christie Hospital NHS Trust, Manchester, United Kingdom*

## Glossary

**body composition** Indicates the relative proportions of various tissues within the body—fat, lean, and bone mass, broadly speaking.

**bone mineral density** A measurement of the mineral content of bone expressed either as a proportion to the surface area or as the volume of bone; reflects the strength of the bone and, hence, the risk of fracture.

**chemotherapy** Treatment of disease by means of chemical substance or drug; usually used in reference to neoplastic disease.

**growth** The increase in size of a living being or any of its parts occurring during the process of development.

**puberty** Sequence of events by which a child becomes a young adult through the development of secondary sexual characteristics and reproductive function.

**radiotherapy** The use of ionizing radiation to cause DNA damage in tumor cells and, hence, cell death.

Late endocrine effects after childhood cancer is the study of the long-term effects on growth and development of a child as a result of previous chemotherapy or radiotherapy for cancer during childhood. Integral to this is the effect on other endocrine organs such as adrenal gland, thyroid gland, and pancreas.

## INTRODUCTION

As a result of modern treatment protocols, the majority of children with cancer will now survive to adulthood, particularly patients with acute lymphoblastic leukemia, who form a significant proportion of this cohort of survivors. Therefore, the long-term complications of their treatment have become increasingly important. Furthermore, because of the ability to treat children with acute lymphoblastic leukemia from the

1970s onward, many members of this cohort are now reaching their third or fourth decade of life.

Children with cancer may have received chemotherapy and/or irradiation; thus, the type of drug used or the site of irradiation will determine whether endocrine late effects occur. For example, deficiency of one or more anterior pituitary hormone frequently occurs if the hypothalamic pituitary axis lies within the field of irradiation; growth hormone secretion is well recognized to be the most vulnerable to irradiation-induced damage. Therefore, it is important to consider the late endocrine effects of cancer therapy in the context of previous treatment, which itself is highly dependent on the tumor type.

## GROWTH AND GROWTH HORMONE STATUS

Children with acute lymphoblastic leukemia who have previously received cranial irradiation show a significant reduction in final height. This includes those children who have received doses as low as 18 gray to the cranium as described by Helena Davies in 1993. Furthermore, a proportion of these children are growth hormone deficient, with the exact prevalence being dependent on the number of fractions, the fractionation size, and the duration of cranial irradiation schedule. Therefore, a number of these children will go on to receive growth hormone therapy. In 1996, Adnan and colleagues assessed growth rate and final height in 43 long-term survivors of acute lymphoblastic leukemia who had received cranial irradiation. The final height was significantly greater in the cohort of those who were growth hormone deficient and received growth hormone therapy than in the cohort of those who were growth hormone deficient but did not receive growth hormone therapy.

Following cranial irradiation for brain tumors, short stature is a common complication, as described

by Stephen Shalet in 1988. Growth hormone deficiency in this cohort of cancer survivors is well reported but might not be the only factor to contribute to the growth disturbance. Other factors may include spinal irradiation, poor nutrition, steroid therapy, the effect of cytotoxic chemotherapy on the growth plate, and early puberty. In 1995, Amanda Olgilvy-Stuart studied 47 survivors of brain tumors with growth hormone deficiency who had received at least 1 year of growth hormone replacement. Growth hormone significantly increased leg length velocity, with a mean of 4.3 versus 2.0 cm per year pre-growth hormone therapy. In all, there was significant loss in mean final height, which was greatest in those who had also received spinal irradiation and/or chemotherapy.

The overall consensus of pediatric endocrinologists has been to offer growth hormone replacement to children with radiation-induced growth hormone deficiency and to continue to final height given that growth hormone deficiency is at least one of the important factors that contribute to poor growth and, hence, reduced final height.

A significant proportion of these children who receive cranial irradiation will survive to adulthood. In particular, those who have had acute lymphoblastic leukemia will have received growth hormone therapy. In 1998, we assessed growth hormone status in 32 young adult survivors of childhood acute lymphoblastic leukemia who had received cranial irradiation. Of these, 9 (28%) had severe growth hormone deficiency and 12 (38%) had growth hormone insufficiency. This level of growth hormone status abnormalities has also been demonstrated in long-term survivors of childhood brain tumors. In studies examining the effect of childhood onset growth hormone deficiency during adulthood, there is an increase in fat mass and total cholesterol; reduced lean body mass, bone mineral density, and cardiac function; and impaired quality of life. Because growth hormone therapy has benefited the childhood onset growth hormone deficiency cohort, these individuals need to be monitored not only during childhood but also as young adults; they may benefit from growth hormone therapy not only for growth during childhood but during adulthood as well.

A proportion of children with malignancies will go on to have high-dose therapy with stem cell rescue. The conditioning regimens prior to stem cell transplant include chemotherapeutic agents such as cyclophosphamide and busulphan, and a large proportion will also receive total body irradiation. The effects of total body irradiation on growth and growth hormone

status have been well described. Loss in final height following fractionated total body irradiation is not as substantial as loss in height following single-fraction total body irradiation; however, in 1999, Jean Sanders observed a greater impact on final height in those children who were under 6 years old at the time of transplant (mean final height  $-3.50 \pm 1.77$  SD below the mean). Although there is plenty of literature demonstrating that a significant proportion of children have growth hormone deficiency following total body irradiation, relatively little information is available on the effect of growth hormone therapy in this cohort, particularly on final height data. The majority of the therapeutic results are short term, usually 1 year, with occasional studies extending to 3 years.

## PUBERTY IN CHILDREN WITH CANCER

The effect of childhood cancer on puberty may be direct through a mass lesion effect on the reproductive axis or indirect through hormones secreted by tumors (e.g., human chorionic gonadotropin), weight loss, or the actual presence of a chronic disease process *per se*. The most frequent pubertal problem faced by children with cancer are due to the impact of treatment on the central nervous system, the hypothalamic-pituitary axis, or the gonad.

### Cranial Irradiation and Puberty

Low-dose cranial irradiation of 18 to 24 gray used in the treatment of acute lymphoblastic leukemia is paradoxically associated with precocious puberty. However, the effect is gender dependent; boys, but not girls, enter puberty at the normal time, but peak height velocity during pubertal growth is attenuated in both genders. In girls with acute lymphoblastic leukemia who enter puberty early, the decision to treat with a combination of gonadotropin-releasing hormone (GnRH) analogue and growth hormone replacement is a difficult one given that the timing of puberty is altered only moderately, with the mean peak height velocity at 10.7 to 11.0 years versus 12.1 years in normal individuals and the mean time of menarche at 12.2 years versus 13.4 years in normal individuals, as demonstrated by the Elizabeth Didcock study in 1995. Therefore, this decision should be made only on an individual basis.

Cranial irradiation at doses exceeding 50 gray to the hypothalamic-pituitary axis may render a child gonadotropin deficient. However, for most children

with brain tumors, the irradiation dose to the hypothalamic–pituitary axis is lower and puberty is early. The onset of puberty occurred at an early age in both genders (mean 8.5 years in girls, 9.2 years in boys, plus 0.29 year for every year of age at the time of irradiation), with a significant linear association between age at time of irradiation and age at onset of puberty. All children were growth hormone deficient in the study reported by Ogilvy-Stuart in 1994. Therefore, the consequence of early puberty relates primarily to its limitation on further growth potential, particularly in those children irradiated before 5 years of age.

It is recognized that the majority of children irradiated for brain tumors will be growth hormone deficient and receiving growth hormone replacement therapy, primarily to improve their growth prognosis. It follows that there is excellent justification, in such children entering puberty early, to halt pubertal progression with the use of a GnRH analogue, with a view to further improving the growth prognosis. Thus, the combined use of GnRH analogue and growth hormone replacement therapy does not usually present a clinical dilemma in such children, despite the fact that there is no well-established evidence base supporting the efficacy of such an approach. There are limited studies observing better final height outcome for patients who receive GnRH analogue along with growth hormone therapy.

### Chemotherapy and Radiation-Induced Gonadal Damage

The effect of combined cytotoxic chemotherapy on the gonad depends on the nature and dose of the drugs used and, to some extent, the gender of the child. Drugs implicated include the following alkylating agents: cyclophosphamide, chlorambucil, nitrosureas, procarbazine, vinblastine, cytosine arabinoside, cisplatin, and (possibly) ifosfamide. Direct irradiation to the gonad is also detrimental, with the adult testis being extremely sensitive to the effects of external irradiation. However, neither the threshold dose of irradiation required to damage the germinal epithelium during childhood nor the dose above which irreversible damage occurs is known. In 1989, Hamish Wallace demonstrated that in irradiated girls, the response of the ovary involves a fixed pool of oocytes that cannot be replaced once they are destroyed. The dose of irradiation that will kill 50% of oocytes, as the human ovary has been estimated not to exceed 4 gray.

### Boys

The combination chemotherapy used for acute lymphoblastic leukemia in the past and currently may cause gonadotoxicity, a complication that depends largely on the drugs used and their total dose. In 1991, Wallace assessed the effect of combination chemotherapy on the testes by assessing adults who had been treated for acute lymphoblastic leukemia during childhood. The majority had normal germinal epithelial function. But in the few who did have testicular damage, it was not clear whether this was related to their previous total dose of cyclophosphamide or cytosine given that some would have received one or both of these drugs at doses greater than 1 g/m<sup>2</sup>. If modern acute lymphoblastic leukemia treatment continues to intensify, particularly with the use of cytosine and/or cyclophosphamide, more testicular damage may be seen in the future.

Boys who receive direct testicular irradiation as part of acute lymphoblastic leukemia treatment, usually with a dose of 24 gray, will show Leydig cell failure and absent sperm production postpuberty with no evidence of recovery. In the presence of Leydig cell dysfunction, with no signs of puberty by 12 or 13 years of age, androgen replacement therapy will be required.

There have been no reports of chemotherapy-induced testicular damage interfering with the onset or normal progression of pubertal development. However, chemotherapy-induced damage to the germinal epithelium, resulting in infertility and failure to acquire a normal-size testis appropriate for the stage of puberty achieved, is well reported. Thus, at some point, discussion regarding future fertility prospects is required. Obviously, the timing of such discussion with a teenager is based very much on individual circumstances, nature, and personality.

The use of lomustine with or without procarbazine chemotherapy in boys with brain tumors leads to a high prevalence of primary gonadal dysfunction. In 1988, Peter Clayton demonstrated that when boys treated with these drugs for brain tumors reached adulthood, they had inappropriately small testicular volumes and a raised basal concentration of follicle-stimulating hormone (FSH). Because most of these boys would have also received spinal irradiation, with scatter to the testes estimated at 0.46 to 1.20 gray, it is likely that this contributed to the observed testicular damage. Other clinical scenarios where there could be scatter irradiation to the testes include children who receive whole abdominal irradiation for Wilm's tumors, particularly in the early treatment protocols. Whole abdominal radiotherapy is rarely used in

modern Wilm's tumor treatment protocols. The cohort study by Stephen Shalet in 1978 demonstrated significant damage to the germinal epithelium, but all members of the cohort progressed through puberty spontaneously.

Combination chemotherapy has greatly improved the prognosis for children with Hodgkin's disease, but at the expense of future fertility in boys. In a large study in the United Kingdom involving 101 postpubertal individuals in 1996, Erica Mackie demonstrated that 89% of males had germ cell epithelium damage and, hence, infertility. The alkylating chemotherapy agents chlorambucil and procarbazine were the likely causative agents. Although current treatment protocols are trying to reduce the total doses of these drugs, particularly that of procarbazine, these drugs are unlikely to be omitted completely; hence, current treatment schedules for Hodgkin's disease are still likely to affect fertility, particularly in boys.

Chemotherapy agents such as cyclophosphamide and, more recently, ifosfamide are integral to many treatment regimens in current childhood cancer treatment, particularly for tumors such as B-cell non-Hodgkin's lymphoma and sarcomas. Doses of cyclophosphamide greater than 3 g/m<sup>2</sup> in boys have resulted in germ cell epithelium damage, but as yet the effect of ifosfamide, other than that provided in anecdotal reports, is unknown. Cisplatinum, particularly when used for bone tumors, has also been implicated as a likely cause of infertility.

An additional problem arising from chemotherapy-induced testicular damage is an increased prevalence of pubertal gynecomastia. For cosmetic reasons, subareolar mastectomy may be required postpuberty to resolve this problem.

### **Girls**

Although abnormalities of ovarian function with elevated FSH levels have been reported following gonadotoxic chemotherapy for brain tumors, bone tumors, acute lymphoblastic leukemia, and Hodgkin's disease, failure to initiate or progress through puberty is rare following chemotherapy alone. Other studies have looked at fertility as well as pregnancy outcome in women treated for acute lymphoblastic leukemia and demonstrated that pregnancy does occur. Another possible risk group for ovarian failure induced by chemotherapy during childhood is those girls treated for Hodgkin's disease. Mackie's study in 1996 demonstrated that 53% of women had raised gonadotropin levels and variable estradiol levels. This study reported a much higher prevalence of ovarian failure than had been thought previously. Serial follow-up of

this cohort will determine whether ovarian function recovers or whether progression to a premature menopause is inevitable.

There have been few studies of ovarian function following ovarian irradiation uncomplicated by the effects of other cytotoxic agents in humans. Ovarian morphology following whole abdominal irradiation (20–30 gray) demonstrates inhibited follicular growth, and in the majority the number of oocytes is reduced markedly. In clinical parallel with histological observations, the same irradiation protocol will induce pubertal failure in more than two-thirds of such girls. However, when whole abdominal irradiation is replaced by flank irradiation, ovarian function is normal in nearly all patients. Clearly, brain tumor patients and long-term surviving leukemia patients who received spinal irradiation are at risk for developing ovarian failure from scatter irradiation to the ovaries.

Some have advocated transplanting the ovary outside the irradiation field, particularly when pelvic irradiation is required. There is little or no evidence base for this, nor are there results of outcome where performed. In clinical practice, this technique is rarely used and may be superseded by ovarian tissue storage in the future.

## **Gonadal Toxicity and Bone Marrow Transplantation**

### **Boys**

The observations derived from dose–response relationships between irradiation dose and Leydig cell function, plus the relatively mild impact of chemotherapy on Leydig cell function, allow us to predict the outcome for boys receiving total body irradiation, with a dose of 13 to 15 gray, and chemotherapy in the preparation for bone marrow transplant. It was predicted that they would progress through puberty spontaneously, and this has been confirmed in several studies. However, the germinal epithelium is universally damaged in boys receiving chemotherapy and total body irradiation for bone marrow transplant, with rare reports of germ cell dysfunction recovering in those boys who possibly received only high-dose chemotherapy. Therefore, it is inevitable that infertility will follow in boys who receive high-dose chemotherapy and irradiation preparative regimens prior to transplantation.

### **Girls**

For girls undergoing total body irradiation preparation for bone marrow transplant, the reproductive

outcome is highly dependent on the age at treatment. In girls receiving single-dose total body irradiation during pubertal and early adult life, the incidence of ovarian failure is probably close to 100%. However, if total body irradiation is used in prepubertal life, normal pubertal development occurs in about 50% with the potential for future fertility. Clearly, the induction or maintenance of puberty with estrogen is carried out in exactly the same way as for other young girls with alternative causes of ovarian failure such as Turner syndrome.

## BODY COMPOSITION

Most studies on body composition in childhood cancer focus on the incidence of obesity following childhood cancer, and particularly studies of survivors of childhood acute lymphoblastic leukemia demonstrate that obesity is present at the end of puberty in approximately 50% of cases. Most patients have received cranial irradiation and, hence, damage to the hypothalamic-pituitary axis, resulting in abnormal growth. Therefore, growth hormone secretion has been implicated as the underlying cause.

In this cohort, another potential effect of cranial irradiation and resultant hypothalamic damage is leptin insensitivity. In 1999, we examined the relationship among leptin, growth hormone status, and lean and fat mass in a cohort of adults who had been treated during childhood for acute lymphoblastic leukemia. In those who were growth hormone deficient, leptin levels were significantly raised even after controlling for differences in fat mass. It is not clear whether this is due to irradiation-induced hypothalamic damage or to an effect of growth hormone deficiency per se. Because leptin is thought to act as a satiety signal, leptin insensitivity could contribute to the obesity in this group of patients.

Other factors involved in the mechanism of obesity include the use of steroids in acute lymphoblastic leukemia treatment. Furthermore, recent studies looking at changes in body mass index in acute lymphoblastic leukemia survivors who did not receive cranial irradiation also demonstrated increasing evidence of obesity. These data, however, have only a short follow-up period and small numbers; therefore, we do not know the significance of these initial findings or whether they are still present when final height is achieved. The amount of exercise and energy expenditure used by these children may be another contributing factor to the increased incidence of obesity. Both have been shown to be reduced and

related to the increase in fat mass demonstrated directly by dual energy X-ray absorptiometry scans.

Obesity does not seem to be a significant problem in children who are treated for other malignancies or who have not received cranial irradiation. However, children treated for brain tumors that includes cranial irradiation who become growth hormone deficient during childhood remain so throughout adult life, and it is well recognized that adults with childhood onset growth hormone deficiency have increased fat mass.

Following bone marrow transplant for childhood acute lymphoblastic leukemia, body composition, and hence the degree of fatness, is less well described. Using body mass index measurements alone, studies have indicated that obesity does not follow bone marrow transplant that includes total body irradiation because the body mass index falls after bone marrow transplant. Using direct measurements of fat mass, such as was done in Karston Nyson's study in 2001, fat mass is actually increased compared with that in normal controls. Furthermore, this study demonstrated that there was a reduction in lean mass. Further interpretation of these data is difficult because more than half of the patients in the study had previously undergone cranial irradiation as part of their acute lymphoblastic leukemia treatment; therefore, because cranial irradiation is associated with increased obesity, any findings have to be interpreted with caution. Larger numbers of patients who have received only total body irradiation and not prior cranial irradiation must be studied.

## THYROID

It has long been recognized that the thyroid gland is very radiosensitive and that the risk of thyroid cancer following neck irradiation during childhood is increased in a dose-related fashion. The latency period between thyroid irradiation and clinical presentation with a thyroid tumor may be many years. The number of reports of thyroid cancer following total body irradiation are few, and the incidence is low. In a recent report of second malignancies following bone marrow transplant for acute leukemia, 5 thyroid carcinomas occurred among 3182 children, with a strong relationship between age at transplant and the occurrence of thyroid cancer. In light of the latter risk, the majority of endocrinologists would treat a patient with an elevated thyroid-stimulating hormone (TSH) level with thyroxine replacement because there are a considerable number of animal studies suggesting that

lowering a previously raised TSH level into the normal range in a child who has received irradiation to the thyroid gland is likely to reduce the risk of thyroid malignancy. However, there are no follow-up data to support this strategy in patients treated with thyroxine for long periods.

The major effect following neck irradiation, however, is thyroid dysfunction, ranging in degree from frank hypothyroidism to compensated hypothyroidism with raised TSH and normal thyroxine levels. Following irradiation doses to the neck of 30 to 50 gray, two-thirds of patients will be expected to have thyroid dysfunction, most occurring within 6 years of irradiation. Other groups of children who would have received irradiation directly or scatter irradiation to the neck during treatment for their childhood cancer include those who have received craniospinal irradiation, particularly children with brain tumors.

Finally, a small group of patients who also receive significant irradiation to the neck are those with neuroblastoma who receive meta-iodobenzyl guanidine therapy. Although the irradiation dose to the thyroid during the treatment is low compared with external irradiation, there still may be a risk of thyroid dysfunction despite prophylaxis with potassium iodide. These patients must be checked for thyroid dysfunction and followed up with regular assessment of their thyroid function.

Central hypothyroidism has rarely been reported following cranial irradiation for children unless there has been involvement of the hypothalamic–pituitary axis by the tumor itself such as craniopharyngioma or germ cell tumor. This was disputed by work published in 1999 by Susan Rose and colleagues, who diagnosed central hypothyroidism in 34% of survivors of childhood cancer who had received varying doses of cranial irradiation. The definition of central hypothyroidism, however, was a blunted TSH surge in response to thyroid-releasing hormone. It is noteworthy that only 5 patients (8%) had free thyroxine levels below the lower limits of normal. This suggests that subtle central hypothyroidism may be a cause of poor growth in these patients despite apparently normal thyroxine levels.

## ADRENAL

The adrenal gland is relatively radio resistant, and impairment of adrenocortical-stimulating hormone secretion is not seen in children receiving 18 to 24 gray cranial irradiation prophylactically for acute lymphoblastic leukemia. However, if a patient receives

glucocorticoid therapy for a significant period of time, such as for graft versus host disease following bone marrow transplantation, suppression of the pituitary–adrenal axis will occur; once glucocorticoid therapy is stopped, persistence of pituitary–adrenal axis suppression is related to the duration and intensity of the previous glucocorticoid exposure. Initially, there was some concern that a significant number of children might require glucocorticoid replacement therapy following total body irradiation and bone marrow transplantation, but this has not been supported by recent studies. Clearly, tumor involvement of the hypothalamic–pituitary axis may directly affect adrenocortical-stimulating hormone secretion and, hence, may cause adrenocortical dysfunction.

## PANCREAS

An initial study by Teintrum and colleagues in 1995 concluded that, as a result of abdominal irradiation, the risk of diabetes mellitus due to pancreatic damage was increased and that assessment of glucose tolerance should be performed in such patients. Subsequently, in a letter to the *Lancet*, Mike Hawkins described his group's data from the long-term study of survivors of childhood cancer within the population of the British National Registry of Childhood Tumors. Hawkins could not support Teintrum and colleagues' conclusions, indicating instead that diabetes is unlikely to be a major long-term problem for survivors of childhood cancer.

The risk following bone marrow transplant that includes total body irradiation may, however, be increased. A Finnish study reported by Mervi Taskinen in 1999 observed a higher than expected incidence of type 2 diabetes in 23 long-term survivors (median age 20 years) of bone marrow transplantation. Abdominal obesity, but not overweight, was a common finding among members of this cohort. Some of these features resembled growth hormone deficiency in adults, although growth hormone status was not formerly assessed (despite the possibility that growth hormone deficiency might explain these findings).

## BONE MINERAL DENSITY AFTER CHILDHOOD CANCER

There have been many cross-sectional studies assessing bone mineral density following treatment for acute lymphoblastic leukemia both during childhood and in adult survivors of childhood acute lymphoblastic leukemia. In general, they have shown a reduction

in bone mineral density, with the majority of studies including patients who had received multi-agent chemotherapy, intermittent steroid therapy, and (usually) cranial irradiation. Because the latter is associated with abnormalities in growth hormone secretion that may affect bone mineralization, it is difficult to determine the major causative factor responsible for the reduced bone mineral density.

In modern treatment for acute lymphoblastic leukemia, the majority of children will not receive cranial irradiation but glucocorticoids continue to be used. This is especially the case with dexamethasone, which has more glucocorticoid potency, and overall chemotherapy, including methotrexate, is more intensified. In addition, not all studies assessing bone mineral density in either children or adults following childhood acute lymphoblastic leukemia treatment have taken into account bone size or height and its effect on bone mineral density. Thus, the method of assessment of bone mineral density and the adjustment of the bone mineral density for the height of the individual are important in considering the significance of any results obtained.

Several studies have suggested that cranial irradiation is an important contributing factor to the reduced bone mineral density and, therefore, have postulated a role for growth hormone deficiency in explaining the reduced bone mineral density found. In our study published in 1999, we concluded that although bone mineral density is reduced, it is not more reduced than in those individuals who are severely growth hormone deficient. Therefore, other factors could be involved in the reduction in bone mineralization, including low physical activity and the previous use of steroids.

Bone mineral density has been examined in large cohorts of acute lymphoblastic leukemia survivors who did not receive cranial irradiation. In 2001, Nina Kadan-Lottick and colleagues found that whole body bone mineral density was normal in 75% of individuals who had previously been treated for acute lymphoblastic leukemia.

It is likely that children treated for other cancers with protocols that do not include cranial irradiation or a significant dose of steroids will have a normal bone mineral density, but only a limited number of studies have been performed to provide any meaningful conclusions to this point.

### See Also the Following Articles

GI Hormones in Cancer • Irradiation, Thyroid and • Pancreatic Cancer • Parathyroid Cancer • Prostate Cancer • Thyroid Carcinoma

### Further Reading

- Brennan, B. M. D., and Shalet, S. M. (2002). Endocrine late effects after bone marrow transplant. *Brit. J. Haematol.* **118**, 58–66.
- Murray, R. D., Brennan, B. M. D., Rahim, A., and Shalet, S. M. (1999). Survivors of childhood cancer: Long-term endocrine metabolic problems dwarf the growth disturbance. *Acta Paediatr. Suppl.* **433**, 5–12.
- Shalet, S. M., and Brennan, B. M. D. (1998). Growth and growth hormone status following treatment for childhood leukaemia. *Horm. Res.* **50**, 1–10.
- Shalet, S. M., and Brennan, B. M. D. (2002). Puberty in children with cancer. *Horm. Res.* **57**, 39–42.
- Wallace, W. H. B. (1997). Growth and endocrine function following the treatment of childhood malignant disease. In "Paediatric Oncology Clinical Practice and Controversies" (C. R. Pinkerton, ed.), pp. 706–731. Chapman & Hall, London.



Cholecystokinin

see CCK



# Chondrodysplasias

Nancy B. Schwartz and Miriam Domowicz  
University of Chicago, Chicago, Illinois, United States

## Glossary

**extracellular matrix** Milieu outside the cell composed of fibrous proteins (collagens), adhesive glycoproteins, and proteoglycans.

**endochondral ossification** Replacement of cartilage with bone in the growth plate of the skeleton.

**growth plate** Cartilaginous region at the ends of vertebrate long bones where bone growth occurs.

The chondrodysplasias are a group of rare, inherited disorders of skeletal development and growth that are genetically and clinically heterogeneous. Patients present with short stature and/or skeletal deformities, often having disproportionately short limbs compared to trunk length. The chondrodysplasias result from mutations in genes that are critical to the process of endochondral ossification, which is responsible for skeletal growth.

## INTRODUCTION

Over 100 distinct chondrodysplasias have been identified in humans. Although they are defined as disorders of the cartilage components of the developing skeleton, other tissues may be affected. Chondrodysplasias occur in all vertebrates and, in fact, studies of naturally occurring and genetically engineered mutant mice and other model organisms have resulted in a better understanding not only of skeletal development but also of the human disease process. The nomenclature and classification of the chondrodysplasias have evolved over the past 50 years, at various times employing clinical features, radiographic findings, inheritance patterns, morphology, pathogenesis, and genetic etiology as the primary basis of classification. A dynamic, gene-based scheme, put forth by the International Working Group on Bone Dysplasias in the late 1990s, directly interfaces with the Online Mendelian Inheritance in Man database, which

provides references for the listed disorders and links to other informational services. Table I presents a list of molecularly defined chondrodysplasias.

Chondrodysplasia-causing mutations act through a variety of mechanisms and involve genes that encode different types of proteins, including transcription factors, growth regulators, cartilage matrix proteins, membrane receptors, modifying enzymes, and ion transporters. The number of genes responsible for the large number of human chondrodysplasias is a fraction of the clinical phenotypes described, indicating that a limited number of genes are critical to skeletal development. The fact that mutations in a limited number of genes give rise to a wide range of chondrodysplastic phenotypes indicates that multiple allelism is common. For instance, more than half of the human disorders involve just three genes, *COL2A1*, *FGFR3*, and *DTDST*, with resulting phenotypes that range from very mild to lethal. Chondrodysplasia-causing mutations disrupt the function of the growth plate where endochondral ossification leading to bone growth occurs. To understand the pathogenesis of the chondrodysplasias, an understanding of the temporal and molecular events involved in endochondral ossification is required.

## GROWTH PLATE DEVELOPMENT

Growth and development of the long bones of the skeleton constitute a complex, multistep, precisely timed, and spatially organized process regulated by several genetic, endocrine, and mechanical programs. During limb formation, mesenchymal cells migrate in a condensation process to establish areas of high cell density; at this stage, the position, number, and shape of the future skeletal elements are determined. Following the establishment of a cartilage template, endochondral bone growth occurs in the epiphyseal growth plate, a specialized tissue found at the ends of long bones. In this process, chondrocytes become arranged in columns parallel to the longitudinal axis of the long bones and differentiate along that axis,

Table 1 Chondrodysplasia Genes

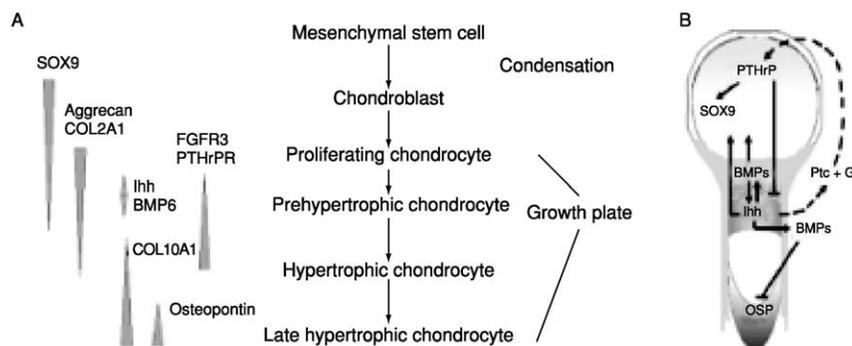
Gene locus	Product	Disease
<b>ECM molecules</b>		
COL2A1	Type II collagen $\alpha$ 1 chain	Achondrogenesis II Hypochondrogenesis Late-onset spondyloepiphyseal dysplasia Spondyloepiphyseal dysplasia congenital Kniest dysplasia Stickler dysplasia
COL9A2	Type IX collagen $\alpha$ 2 chain	Multiple epiphyseal dysplasia
COL9A3	Type IX collagen $\alpha$ 3 chain	Multiple epiphyseal dysplasia
COL10A1	Type X collagen $\alpha$ 1 chain	Schmid metaphyseal chondrodysplasia
COL11A1	Type XI collagen $\alpha$ 1 chain	Stickler-like dysplasia
COL11A2	Type XI collagen $\alpha$ 2 chain	Stickler-like dysplasia
COMP	Cartilage oligomeric matrix protein	Pseudo-chondrodysplasia Multiple epiphyseal dysplasia
HSPG2	Perlecan	Dysegmental dysplasia: Silverman-Handmaker and Schwartz Jampel syndrome
MATN3	Matrilin 3	Multiple epiphyseal dysplasia
<b>Metabolic/modifying molecules</b>		
DTDST	DTD sulfate transporter	Diastrophic dysplasia Achondroplasia 1B Atelosteogenesis type 2 Recessively inherited multiple epiphyseal dysplasia Spondyloepimetaphyseal dysplasia
PAPSS2	3'-phosphoadenosine 5'-phosphosulfate synthase 2	
CTSK	Cathepsin K	Pycnodysostosis
SEDL	Sedlin	X-linked spondyloepimetaphyseal dysplasia tarda
RMRP	RNA component of mitochondrial RNA-processing ndoribonuclease	Cartilage-hair hypoplasia
EIF2AK3	Eukaryotic translation initiation factor 2- $\alpha$ kinase 3	Wolcott-Rallison dysplasia
PEX7	Peroxisomal biogenesis factor 7	Rhizomelic chondrodysplasia punctata 1
GNPAT	Glyceronephosphate <i>O</i> -acyltransferase	Rhizomelic chondrodysplasia punctata 2
AGPS	Alkylglycerone phosphate synthase	Rhizomelic chondrodysplasia punctata 3
ARSE	Arylsulfatase E	X-linked chondrodysplasia punctata 1
<b>Signaling molecules</b>		
FGFR3	FGF receptor 3	Thanatophoric dysplasia I Thanatophoric dysplasia II Achondrodysplasia Achondrodysplasia Hypochondroplasia
PTHR	PTHrP receptor	Jansen metaphyseal chondrodysplasia Blomstrand metaphyseal dysplasia
<b>Transcription factors</b>		
SOX9	SRY box 9	Campomelic dysplasia
SHOX	Short stature homeobox	Leri-Weill dyschondrosteosis Mesomelic dysplasia Langer type

forming histologically distinct zones. Closest to the epiphyseal end of the growth plate are metabolically less active resting chondrocytes, followed by a zone of high proliferation that leads to lengthening of the cartilage anlagen. After several mitoses, chondrocytes begin to mature, becoming encased in a specialized extracellular matrix (ECM) and acquiring a characteristic round morphology (prehypertrophic). These chondrocytes then exit the mitotic cycle, increase in size, and eventually hypertrophy. Mature hypertrophic chondrocytes die by apoptosis and are replaced by mineralized bone and bone marrow via vascular invasion, resorption of the cartilaginous matrix, and recruitment of osteoblasts that deposit a bone-specific matrix. Removal of chondrocytes from their native positions in the surrounding cells or the molecular milieu largely destroys their ability to continue this program, suggesting that the process is not only directed by autonomous gene expression but also that the orderly differentiation of chondrocyte cell types and progressive changes in developmental programs are highly context-dependent.

Concomitant with this morphological series of events, a highly organized gene expression profile is evident, encompassing an array of growth regulators, transcription factors, and signaling molecules that orchestrate the cellular and molecular events underlying the formation, maturation, and eventual destruction of transient cartilage. Two secreted factors, Indian hedgehog (Ihh), produced in the prehypertrophic region, and parathyroid hormone-related peptide (PTHrP), synthesized by periarticular chondrocytes, interact in a negative feedback loop to regulate the onset of hypertrophic differentiation (Fig. 1). The bone morphogenetic proteins (BMPs), produced by prehypertrophic and hypertrophic chondrocytes, are potential interactors of the Ihh–PTHrP feedback loop

regulating bone formation and BMP signaling modulates the expression of Ihh, thereby integrating the regulation of chondrocyte proliferation and the onset of hypertrophy. BMPs may also function independently of Ihh in regulating chondrocyte proliferation and in delaying the maturation of terminally hypertrophic chondrocytes. Other factors have been indicated as being negative regulators of chondrocyte maturation, e.g., transforming growth factor- $\beta$  family members, Wnt-5a, and Notch receptor ligand Delta-1, whereas transcription factors Cbfa1/Runx2 and Dlx5 appear to positively regulate chondrocyte maturation. Observations in Fgfr3 and FGF18 knockout animals suggest that these factors function to inhibit cell proliferation and studies in bone explants *in vitro* argue that fibroblast growth factor (FGF) signaling accelerates terminal differentiation in hypertrophic chondrocytes. Clearly, the rate and progression of hypertrophic chondrocyte maturation during endochondral ossification are regulated by a coordinated balance between negative and positive signals, providing multiple loci where mutations may affect the process (Fig. 1B).

The ECM is of critical importance to the limb's developmental process, since specific matrix components define the milieu of the various differentiating chondrocyte cell types and are synthesized in a program that is developmentally regulated. Expression of ECM-specific genes produces the structural components that both provide a scaffold to support the chondrocytes and, most likely, interact directly with growth factors and signaling molecules and/or provide a milieu or gradient through which these must pass from cell to cell. Chondrocytes organize their pericellular matrix, composed of type II collagen and aggrecan–hyaluronan–link protein aggregates, and facilitate its assembly and retention at the surface of the



**Figure 1** (A) Gene expression correlated with cell differentiation in the chondrocyte lineage. (B) Signaling pathways that regulate the developing growth plate. Ptc, patched; Gli, Zinc finger protein Gli.

cells. In general, the collagens provide the ECM with strength and resilience, whereas the proteoglycans, predominantly polyanionic aggrecan, are multiply hydrated, conferring compressibility and forming a gel-like medium that acts as a selective diffusion barrier. The most definitive evidence that proper matrix composition is essential to growth plate development comes from the study of genetic alterations that specifically affect cartilage formation and endochondral ossification: the inherited skeletal disorders known as the chondrodysplasias. Although the pathogenesis of the chondrodysplasias is complex, with the increasing elucidation of mutations responsible for human disorders and the advent of animal model systems, knowledge is accumulating rapidly with regard to the pathways of control and the spectrum of phenotypes that can arise from defects in a single gene. Representative examples of such phenotype are discussed.

## ECM COMPONENTS

### COL2A1

Type II collagen constitutes the major component of collagen fibrils in hyaline cartilage. Mutations of *COL2A1*, the gene encoding type II collagen, lead to a spectrum of disorders (Table I) from the relatively mild, late onset spondyloepiphyseal dysplasia to the perinatal lethal achondrogenesis type II (Langer-Salidino). Disorders of intermediate severity include hypochondrogenesis, spondyloepiphyseal dysplasia (SED) congenital, SED Strudwick, Kniest dysplasia, and Stickler dysplasia. Collectively, these fall under the designation SED and comprise over 40 mutations associated with SED and SED-like phenotypes. The SEDs encompass clinical phenotypes that primarily manifest radiological abnormalities of the spine and epiphyses, although other organ systems may also be affected; e.g., Stickler syndrome is usually dominated by eye abnormalities. Although there is some histological variation, the growth plates of all SED patients are noticeably affected. The epiphyseal cartilage of the most severe forms of achondrogenesis type II exhibit hypercellularity, scant matrix, and disorganized or absent growth plate. In Kniest dysplasia cartilage, vacuolar degeneration of chondrocytes and matrix is the most distinguishing feature, whereas cytoplasmic inclusions and dilated rough endoplasmic reticulum are characteristic of SED congenital and SED Strudwick, respectively.

The molecular mutations that lead to the multiple SED phenotypes are heterogeneous and mostly map

to the triple-helical domain of the type II collagen  $\alpha 1$  chain, where they are dispersed fairly evenly along the chain. There is a tendency for C-terminal mutations to produce more severe phenotypes. Most *COL2A1* mutations involve amino acid substitutions and most of these are glycine (found in every third position in triple-helical domains) to serine substitutions. The molecular consequences of these mutations include disruption of folding, posttranslational modification, triple-helix formation, and delayed secretion and/or accelerated degradation of type II collagen molecules. The synthesis of abnormal collagen reduces the number of collagen fibrils in the cartilage growth plate matrix and the degree of fibril reduction appears to correlate with phenotype severity, regardless of the specific mutation. As with any of the major matrix components, it is still unclear how the reduction in the number of type II collagen molecules or the production of abnormal molecules leads to the observed phenotypes affecting cartilage development and growth potential. Since collagen fibrils are assumed to provide structural integrity to the extracellular matrix, it is likely that fewer or abnormal fibrils would reduce the tensile strength of the supporting matrix during the growth and remodeling process. It is possible that more direct cell-matrix interactions or diffusibility of nonmatrix components may be altered as well.

Although type II collagen is the major species in hyaline cartilage and a majority of chondrodysplasias are due to defects in the *COL2A1* gene, a significant number of human disorders result from mutations in genes that encode other matrix proteins, including the type IX, X, and XI collagens and cartilage oligomeric matrix protein (COMP). For instance, more than 25 heterogeneous mutations in the *COL10A1* gene have been reported, but are all clustered near the C-terminus. *COL10A1* mutations are associated with Schmid metaphyseal chondrodysplasia. The noncollagen molecule COMP, a member of the thrombospondin gene family, has more than 50 mutations associated with it. Mutations in the COMP gene cause a range of phenotypes from normal to mild short stature, with or without joint problems, to severe disproportionate short stature with joint symptoms. COMP is thought to interact with types I and II collagens through their C-terminal globular domains, which presumably would be disrupted by mutations in the COMP gene product; however, clear genotype/phenotype correlations are not available.

Several mouse strains harboring mutations that result in phenotypes that are similar to the human chondrodysplasias have been identified. The disproportionate micromelia (*Dmm*) mouse contains a

deletion in the portion of the *Col2a1* gene that encodes a highly conserved region in the C-propeptide of type II collagen. Disruption of intrachain disulfide bonding leads to misassembly of type II collagen molecules. Homozygotes for this defect exhibit severe dwarfism and die at birth. *Dmm* has been proposed to be a model for Stickler dysplasia type I. Other mouse models have been engineered to be null for *Col2a1*; these exhibit a severe reduction in collagen abundance and are models for the SED human chondrodysplasias. Clearly, such studies of naturally occurring and genetically engineered mouse mutations have provided significant insights into the etiology of skeletal development as well as improved the understanding of human skeletal disorders.

### Aggrecan

There are multiple loci where mutations may affect proteoglycan-based phenotypes, i.e., genes for the proteoglycan core proteins, glycosaminoglycan-modifying enzymes, or signaling, growth, and transcription factors that regulate the synthesis and secretion of proteoglycans. Despite the abundance of the cartilage-specific proteoglycan aggrecan, there are no known human aggrecan-based chondrodysplasias. However, several chondrodysplasias have been identified in animal models—nanomelia (nm) in chick, cartilage matrix deficiency (cmd and cmd-Bc) in mouse, and brachymorphism (bm) in mouse—that predominantly affect the production or biochemical properties of aggrecan, resulting in altered growth plate development. The cartilage matrix deficiency alleles, *Agc1<sup>cmd</sup>* and *Agc1<sup>cmd-Bc</sup>*, bear autosomal-recessive lethal mutations due to deletions in the mouse aggrecan core protein gene and cause short limbs and snout, enlarged abdomen, protruding tongue and cleft palate in newborns, and perinatal death. Mutant embryonic limb chondrocytes are tightly packed with very little extracellular matrix between the cells. The cartilage fails to form chondrocyte columns and lacks the characteristic demarcation of the growth plate into resting, proliferative, and hypertrophic zones. The fact that two allelic deletion mutations have been identified in the murine aggrecan gene but no mutations have been reported for the human aggrecan gene, even though numerous skeletal disorders have been reported in humans due to defects in other ECM molecules, is puzzling. The lack of known counterparts in humans may reflect differing frequencies of “hot spots” for illegitimate recombination in the human and murine genes. Interestingly,

a variable-number-of-tandem-repeat polymorphism found in chondroitin sulfate (CS) attachment region of the human aggrecan gene is associated with bilateral hand osteoarthritis.

The first mutation identified in the aggrecan core protein gene was in the nm chick, which exhibits a phenotype in homozygotes that is nearly identical to that of the cmd mouse: an extreme form of micromelia with reduced head and trunk size and gross skeletal abnormalities including short, broad, and malformed limbs. The cmd mouse limbs have a rather homogeneous growth plate cell population, devoid of matrix and with loss of growth zone demarcation. Analysis of the nm mutant indicates that the absence of aggrecan in the growth plate ECM is accompanied by discordant expression of certain signaling molecules and transcription factors responsible for regulating growth plate maturation as well as by altered levels of apoptosis and proliferation in growth plate chondrocytes. The profound changes observed in the aggrecan-deficient milieu suggest a more dynamic role for this ECM component than only that of a structural scaffold or polyanionic hydrant. Lack of functional aggrecan in the ECM may account for the reduction in limb length characteristic of these chondrodysplasias because less space is occupied per element due to the bulk matrix loss or because active maintenance of the normal chondrocytic phenotype is absolutely dependent on the expression of aggrecan. These mutants provide superb model systems with which to assess the contribution of this key factor to skeletal growth and development.

## SULFATE TRANSPORT AND MODIFYING PROTEINS

### DTDS Transporter

In addition to chondrodysplasias associated with mutations in the aggrecan core protein gene, growth disorders caused by defects in biosynthetic pathways involved in the modification of proteoglycan glycosaminoglycan (GAG) chains have been identified. Several human disorders associated with defects in the transport of sulfate into the cell, a process necessary for the sulfation of GAG chains, lead to undersulfated proteoglycans and chondrodysplasias (Table I). Four distinct phenotypes with different degrees of severity are associated with mutations in the same transporter gene, diastrophic dysplasia sulfate transporter (DTDST/SLC26A2), which encodes a plasma membrane sulfate/chloride exchanger. Patients with one of the earliest of these mutations to be characterized,

diastrophic dysplasia (DTD), exhibit disproportionately short stature, short trunk and limbs, and generalized joint dysplasia. Although the phenotype is severe and progressive, DTD is usually nonlethal. Atelosteogenesis type 2 (AO2) is a rare recessive perinatally lethal chondrodysplasia (micromelia) that is phenotypically similar to DTD and is also caused by mutations in the *DTDST* gene. Achondrogenesis type 1B (ACG-1B), one of the most severe forms of chondrodysplasia in humans, is characterized by extremely short extremities and short trunk and is invariably lethal, often before birth. Recessively inherited multiple epiphyseal dysplasia has been added to the DTDST chondrodysplasia family. The phenotype includes epiphyseal dysplasia but only mildly short or normal stature. DTDST is predicted to have 12 transmembrane domains and a C-terminal cytoplasmic tail. More than 40 types of mutations have thus far been reported in the *DTDST* gene; mutational analysis of affected individuals suggests that most mutations are in the coding region and cause structural changes to the transporter. There appears to be extensive allelic heterogeneity with only a few common mutations and the high rate of mutation suggests that no additional loci are responsible for DTDST-like chondrodysplasias. With respect to proteoglycan disorders in general, it is curious that only human chondrodystrophies involving this very late posttranslational modification (sulfation) have been identified. The range of severity observed for disorders resulting from mutations in the same sulfate transporter gene suggests that the amount of residual transport activity produced by the affected proteins may be responsible for the modulated expression observed clinically. Nonetheless, limitation in sulfate uptake into the cell, which presumably leads to undersulfated aggrecan, can in some circumstances be lethal.

### PAPS Synthetase

In addition to chondrodystrophies associated with mutations in the sulfate transporter gene, growth disorders caused by defects in the biosynthetic pathway responsible for synthesis of the high-energy sulfate donor phosphoadenosylphosphosulfate (PAPS) have been identified in humans and mice. In particular, studies in a murine model revealed that there are at least two PAPS synthetase genes. Unlike the aggrecan core protein mutations that have no counterpart in humans, the identification of the PAPS synthetase 2 (*Papss2*) isoform mutation in the *bm* mouse was followed by the elucidation of a *Papss2* mutation in

human spondyloepimetaphyseal dysplasia (Table I). This disorder is characterized by short and bowed lower limbs, enlarged knee joints, and early onset of degenerative joint disease in the hands and knees.

The elucidation of the human PAPS synthetase disorder was facilitated by the extensive previous work on the *bm* mouse. The nonlethal growth disorder brachymorphism is characterized by a dome-shaped skull, short thick tail, and shortened but not widened limbs. The phenotype is inherited as an autosomal recessive and *bm* homozygotes breed normally, have normal life spans, and are of average size at birth, but then exhibit a 50% reduction in limb length over the first month of development. Biochemical analysis showed that *bm* cartilage contains normal levels of glycosaminoglycans that are significantly undersulfated, due to a defect in PAPS synthetase activity. Genetic linkage studies localized two murine members of the PAPS synthetase family, *Papss1* and *Papss2*, to chromosomes 3 and 19, respectively; *Papss2* is tightly linked with the marker for the *bm* locus in chromosome 19. Sequence analysis of *Papss2<sup>bm</sup>* cDNA revealed a missense mutation in a highly conserved portion of the APS kinase domain and expressed *bm* *Papss2* failed to catalyze the adenosine 5'-phosphosulfate (APS) kinase reaction or to synthesize PAPS. These human and murine PAPS synthetase defects underscore the importance of proper sulfate metabolism for cartilage development and skeletal growth and the multiplicity of genes that might be affected in chondrodysplasias.

## REGULATORY MOLECULES

### FGFR3

Some of the most common forms of human chondrodysplasias are due to mutations in the fibroblast growth factor receptor 3 (FGFR3). FGF signaling is carried out through the dimerization of four membrane-spanning tyrosine kinase receptors. The formation of dimers is developmentally and tissue-specifically regulated and plays an important role in skeletal differentiation. Mutations in three of these receptors are responsible for a variety of human skeletal dysplasias and craniosynostosis syndromes. Different mutations in FGFR3 are responsible for all four phenotypes in the achondrodysplasia family of skeletal dysplasia: achondrodysplasia, hypochondrodysplasia, thanatophoric dysplasia, and severe achondrodysplasia with developmental delay and acanthosis nigricans (SADDAN). Achondrodysplasia is characterized by relatively short limbs and macrocephaly with frontal bossing and midface hypoplasia. In most of the

reported cases, mutations that result in amino acid substitutions in the transmembrane domain of FGFR3 are responsible for the achondrodysplasia phenotype. Hypochondrodysplasia is very similar to achondrodysplasia but is milder as the rhizomelic dwarfism and craniofacial manifestations are not as pronounced. More than 80% of hypochondrodysplasia cases are due to mutations resulting in substitutions in the intracellular domain of FGFR3. The two forms of the neonatally lethal thanatophoric dysplasias I and II are characterized by very severe rhizomelic dwarfism, strong midface hypoplasia, and extremely small thorax. The type I form has curved femurs with a cloverleaf skull and is associated with a variety of mutations in different domains of FGFR3; in contrast, type II exhibits straight femurs with or without a cloverleaf skull and is always caused by mutations resulting in a specific amino acid substitution in the intracellular tyrosine kinase domain of the FGFR3. The SAADAN disorder results from a different substitution at the same residue that is affected in thanatophoric dysplasia type II. Evidence from mouse models indicates that FGFR3, the expression of which is restricted to proliferating and hypertrophic chondrocytes, is a negative regulator of cell proliferation in the growth plate. Thus, the pathological mechanism through which mutations in FGFR3 are manifested involves gain of function, whereby mutated receptors provide a relatively ligand-independent, constitutively active receptor form. The study of FGFR mutations in human skeletal dysplasia syndromes has provided significant information about these signaling pathways and their involvement in skeletal development.

### PTHrP–Ihh Loop

Further insights into growth regulation are provided by mutations that affect the PTHrP receptor. A constitutively active receptor is observed in patients with Jansen metaphyseal chondrodysplasia, an autosomal-dominant disorder characterized by severe hypercalcemia, hypophosphatemia, and short-limb dwarfism (Table I). Blomstrand chondrodysplasia, an autosomal-recessive lethal disorder caused by mutations that generate inactive receptors, is characterized by extremely accelerated chondrocyte differentiation and thus premature ossification. PTHrP receptors belong to the G protein-coupled receptor family and mediate the action of two hormones, parathyroid hormone and PTHrP. Studies in animal models indicate that PTHrP facilitates the proliferation of chondrocytes in the growth plate and negatively regulates

their differentiation to hypertrophic chondrocytes and ensuing programmed cell death. Also, it has been proposed that the PTHrP receptor indirectly mediates the action of Ihh in delaying hypertrophy in the growth plate through a negative feedback loop. How these two rare genetic disorders characterized by activation and inactivation of PTHrP receptor lead to abnormal growth plate development is not clear, but these findings suggest that bone growth and elongation are regulated through a complicated autocrine/paracrine system involving at least PTHrP, its receptor, and Ihh.

### Transcription Factors

As more transcription factors and growth modulators that control growth plate development are identified, it is expected that mutations in the genes for these regulatory factors will be associated with additional skeletal abnormalities. One example is campomelic dysplasia, a congenital skeletal dysplasia in humans that exhibits shortening and bowing of the long bones and abnormal facial features including macrocephaly, micrognathia, cleft palate, and flat nasal bridge, as well as male-to-female sex reversal (Table I). Affected babies usually die in the neonatal period or in early infancy of respiratory distress. Chromosomal rearrangements that affect transcription regulatory elements of the SOX9 transcription factor gene promoter or loss-of-function mutations in the coding region of the gene are responsible for this disorder. SOX9 is expressed in mesenchymal condensations and during chondrogenesis and has been proposed to be involved in the regulation of the expression of a number of extracellular matrix molecules, such as type II collagen and aggrecan. Thus, it is not surprising that campomelic dysplasia phenotypic characteristics resemble the human and animal disorders that involve extreme defects in these two ECM molecules.

### SUMMARY

As illustrated by the presentation of these representative examples, the chondrodysplasias are a genetically, biochemically, and clinically heterogeneous group of vertebrate disorders. They nonetheless all share the common hallmark of disproportionately short stature owing to abnormal bone growth. The basis for this major phenotypic feature involves disruptions in the normal program of endochondral development due to gene mutations at different levels of the process.

Consequently, the study of diseases that affect skeletal development and growth is providing valuable insights into the role not only of individual genes, but of broader regulatory pathways.

### See Also the Following Articles

Albright's Fibrous Dysplasia • Collagen Metabolism • Collagen Metabolism Disorders • Fibroblast Growth Factor (FGF) • Skeletal Development • Skeletal Development During Childhood and Adolescence

### Further Reading

- Francomano, C. A., and Muenke, M. (2002). Craniosynostosis syndromes and skeletal dysplasias caused by mutations in fibroblast growth factor receptor genes. *In* "Connective Tissue and Its Heritable Disorders: Molecular, Genetic and Medical Aspects" (P. M. Royce and B. Steinmann, eds.), pp. 961–991. Wiley-Liss, New York.
- Horton, W. A., and Hecht, J. (2002). Chondrodysplasias: Disorders of cartilage matrix proteins. *In* "Connective Tissue and Its Heritable Disorders: Molecular, Genetic and Medical Aspects" (P. M. Royce and B. Steinmann, eds.), pp. 909–937. Wiley-Liss, New York.
- Horton, W. A., and Hecht, J. (2002). Chondrodysplasias: General concepts and diagnostic and management considerations. *In* "Connective Tissue and Its Heritable Disorders: Molecular, Genetic and Medical Aspects" (P. M. Royce and B. Steinmann, eds.), pp. 901–908. Wiley-Liss, New York.
- Juppner, H. (2000). Role of parathyroid hormone-related peptide and Indian hedgehog in skeletal development. *Pediatr. Nephrol.* **14**, 606–611.
- Karsenty, G. (2003). The complexities of skeletal biology. *Nature* **423**, 316–318.
- Kronenberg, H. M. (2003). Developmental regulation of the growth plate. *Nature* **423**, 332–336.
- McLean, W., and Olsen, B. R. (2001). Mouse models of abnormal skeletal development and homeostasis. *Trends Genet.* **17**, S38–S43.
- Mundlos, S., and Olsen, B. R. (2002). Defects in skeletal morphogenesis. *In* "Connective Tissue and Its Heritable Disorders: Molecular, Genetic and Medical Aspects" (P. M. Royce and B. Steinmann, eds.), pp. 993–1093. Wiley-Liss, New York.
- Schwartz, N. B., and Domowicz, M. (2002). Chondrodysplasias due to proteoglycan defects. *Glycobiology* **12**, 57R–68R.
- Schwartz, N. B., and Domowicz, M. (1998). Proteoglycan gene mutations and impaired skeletal development. *In* "Skeletal Growth and Development" (J. A. Buckwalter, M. G. Ehrlich, L. J. Sandell, and S. B. Trippel, eds.), pp. 413–433. American Association of Orthopedic Surgeon Publications, Rosemont, IL.
- Superti-Furga, A. (2002). Skeletal dysplasias related to defects in sulfate metabolism. *In* "Connective Tissue and Its Heritable Disorders: Molecular, Genetic and Medical Aspects" (P. M. Royce and B. Steinmann, eds.), pp. 939–960. Wiley-Liss, New York.
- Zelzer, E., and Olsen, B. R. (2003). The genetic basis for skeletal diseases. *Nature* **423**, 343–348.



# Chronic Fatigue Syndrome and Fibromyalgia

D. J. Torpy

University of Queensland, Greenslopes Private Hospital, Brisbane, Queensland, Australia

## Glossary

**chronic fatigue syndrome** Unexplained debilitating fatigue (>6 months) in association with at least four of the following eight features: sore throat, tender cervical or axillary lymph nodes, impaired hearing and concentration, headaches of a new type, joint pains, muscles pains, prolonged or severe postexertional fatigue exacerbation, and unrefreshing sleep.

**fibromyalgia** Widespread chronic (>3 months) pain and specific tender points (>11 of 18) on examination.

**stress** Any external or internal factor that threatens homeostasis, or the stable internal environment.

**stress system** The biological response system to stress comprising the hypothalamic–pituitary–adrenal axis, with its main product, cortisol, and the sympathetic nervous system, which produces norepinephrine at sympathetic nerve endings (sympathoneural) and epinephrine from the adrenal medulla (sympathoadrenal).

Chronic fatigue syndrome and fibromyalgia are relatively common clinical syndromes. There are syndromal diagnostic criteria for each condition, although fatigue and chronic pain are frequent in both conditions, suggesting possible overlap in their etiopathogenesis. The cause of these conditions remains unknown, and specific, highly successful therapies are not available. Many possible causative theories have been offered, mainly involving infectious agents and psychological factors. The frequency and heterogeneity of these conditions suggest that the causes are likely multifactorial. Evidence suggests that these disorders may represent disorders of the stress system, comprising the hypothalamic–pituitary–adrenal (HPA) axis and the sympathetic nervous system (Table I).

## CHRONIC FATIGUE SYNDROME AND FIBROMYALGIA: CLINICAL CONCEPTS

Fatigue may be regarded as pathological when pervasive and unrelated exclusively to exertion. Fatigue, as a sense of tiredness or lack of energy,

may be appreciated centrally in terms of concentration, memory, and motivation or appreciated peripherally, where symptoms are often referred to the muscles.

Unexplained and chronic (>6 months) fatigue has been classified as idiopathic chronic fatigue. The clinical diagnosis and definition of chronic fatigue syndrome (CFS) are based on symptoms such as chronic (>6 months) unexplained fatigue with associated disability and at least four of eight of the following symptoms: impaired short-term memory or concentration; sore throat; tender cervical or axillary lymph nodes; muscle pain; multijoint pain without arthritis; headaches of a new type, pattern, or severity; unrefreshing sleep; and prolonged or severe postexertional fatigue exacerbation. CFS is more common in young adults, with a peak age of onset between 20 and 40 years, and it is more common among women, typically occurring at a ratio of 2 or 3:1. The prevalence of CFS is approximately 0.2–0.7%. Idiopathic chronic fatigue has a prevalence of 2%. Fibromyalgia is characterized by widespread pain and at least 11 of 18 specific tender points on examination. The prevalence of fibromyalgia is thought to be approximately 1–4% of the general population. Diagnosis of CFS or fibromyalgia is based on symptoms and exclusion of other diseases that may present with similar symptoms.

No specific treatment is widely successful; however, a positive diagnosis and reassurance that many patients recover can help affected patients. Cognitive-behavioral therapy may assist and involves the provision of information and counseling to reduce the psychological impediments to recovery as well as encouraging the patient to participate at an appropriate level of social and occupational activity.

On the basis of the tendency for abrupt onset and the observation that many patients experience postinfectious fatigue, it has been proposed that CFS results from an infection. Initial attention focused on the Epstein–Barr virus in CFS, but many CFS patients lack Epstein–Barr virus seropositivity, and many others have antibody titers to other viruses.

Moreover, global increases in humoral immune responses are seen in a number of chronic stress states, and neurohormonal changes may account for these immune aberrations.

CFS is generally not considered to be a psychiatric disorder. Although depression is frequently seen in CFS and fibromyalgia, most patients do not exhibit the characteristic self-reproach or biological features of endogenous depression. Hence, depression may be reactive. Personality attributes such as achievement orientation have also been associated with CFS.

### CHRONIC FATIGUE SYNDROME AND FIBROMYALGIA AS STRESS SYSTEM DISORDERS

Stress is defined in biology as any threat to homeostasis and may involve injury, sepsis, or psychic disturbance. To respond to stressors, the body has two neuro-humoral effectors: the hypothalamic–pituitary–adrenal axis, with cortisol as its principal hormonal product, and the catecholamines norepinephrine and epinephrine, produced by sympathetic nerve endings or the adrenal medulla, respectively. Cortisol secretion by the human adrenocortical zona fasciculata is regulated by pituitary adrenocorticotropin (ACTH) and, in turn, by hypothalamic corticotropin-releasing hormone (CRH) and arginine vasopressin. ACTH secretion is influenced by stress, a light-entrained circadian rhythm, and negative feedback at the hypothalamus and inhibitory brain centers. Basal (unstressed) cortisol secretion acts to prevent arterial hypotension by augmenting the effects of catecholamines and to maintain normoglycemia through insulin counter-regulation. Stress-induced cortisol secretion activates the central nervous system, increases blood pressure, elevates blood glucose, and suppresses the inflammatory/immune response to prevent tissue damage. Many of the circulatory and metabolic effects of catecholamines are synergistic with those of cortisol.

A total of 90 to 95% of circulating cortisol is bound to a 383-amino acid glycosylated transport protein, corticosteroid-binding globulin (CBG). Cortisol diffuses through cell membranes and binds the 777-amino acid glucocorticoid receptor- $\alpha$ , which stimulates glucocorticoid response elements of many genes via central “zinc fingers” and largely inhibitory effects through inhibition of the action of *c-fos*, *c-jun*, NF- $\kappa$ B, and other transcription factors on the activated protein-1 and other DNA binding sites.

Several studies have examined basal and stimulated pituitary–adrenal gland function in CFS. Studies in

the broader CFS patient group have generally detected relative hypocortisolism and altered dynamic responses, providing indirect evidence of a central nervous system understimulation of pituitary–adrenal function. Two different types of heritable disorders of this axis have been described, in which fatigue is the principal symptom: (i) glucocorticoid resistance due to glucocorticoid receptor abnormalities and (ii) mutations of the corticosteroid-binding globulin gene, the main cortisol transport protein. These disorders are probably rare, but they reinforce the notion that primary pituitary–adrenal abnormalities may produce chronic fatigue symptoms. Evidence for stress system dysfunction in CFS and fibromyalgia is summarized in [Table I](#).

Plasma basal morning and evening cortisol measurements, salivary cortisol, and 24-h urine free cortisol levels have been found to be reduced in CFS. Reduced plasma cortisol diurnal variation has been reported, with lower AM cortisol and elevated PM cortisol levels. Although the levels were not significantly different from those of controls, the extent of diurnal variation reached clinical significance. DHEA and its long half-life sulfated metabolite, DHEA-S, are major adrenal gland products in terms of mass. They are important contributors to circulating androgen activity, particularly in women. DHEA and DHEA-S levels were lower in 15 CFS patients compared to 11 controls; furthermore, CFS patients did not display the usual decrease in DHEA:cortisol ratio with ACTH stimulation. A preliminary study of 8

**Table I** Evidence of Stress System Dysfunction in Chronic Fatigue Syndrome and Fibromyalgia

<b>Neuroendocrine</b>
Mild hypocortisolism
Reduced ACTH response to CRH stimulators
CRH
Vasopressin
Reduced cortisol response to ACTH
Symptomatic resolution with low-dose hydrocortisone
Sympathoneural instability: elevated norepinephrine, increased responses to stimulation, autonomic-related complaints
<b>Genetic</b>
Association of chronic fatigue with genetic abnormalities involving the HPA axis
Glucocorticoid receptor
Corticosteroid-binding globulin
<b>Syndromal</b>
Onset after prolonged stress
Symptoms exacerbated by adoption of hypo- or hyperstressful environments

selected CFS patients with a subnormal 1  $\mu\text{g}$  ACTH stimulation test showed a 50% reduction in adrenal gland volume on computed tomography scan.

One study of pituitary–adrenal function in CFS patients is notable for its size and use of several complementary methods to assess different components of the axis. This study of 30 patients and 72 normal volunteers found reduced basal evening cortisol, low 24-h urine free cortisol, reduced maximal cortisol response to ACTH, and reduced integrated ACTH response to CRH. These findings were interpreted as suggesting a mild central adrenal insufficiency, perhaps involving endogenous hypothalamic CRH release, as the main regulator of HPA axis function. A reduced ACTH response to arginine vasopressin in 19 CFS patients compared to controls was interpreted as evidence of reduced hypothalamic CRH tone since CRH and vasopressin act synergistically in ACTH release. Although cerebrospinal fluid CRH levels were normal, it has been suggested that CSF–CRH may not reflect hypothalamic CRH secretion.

Relative hypocortisolism, along with the coexistence of fatigue, low blood pressure, and mood alterations in both Addison's disease and CFS, has led to trials of hydrocortisone therapy in CFS. One positive therapeutic study was conducted on 32 CFS patients with disease duration <100 months and a lack of evidence of major depression or other comorbid psychiatric disorder. Five or 10 mg of hydrocortisone was administered for 28 days in a placebo-controlled design. Approximately 28% of patients experienced a reduction in fatigue scores, such that their scores were comparable to those of controls. Only 9% of CFS patients taking placebo experienced similar symptom resolution. A study of full-replacement hydrocortisone (16 mg/m<sup>2</sup> per day in divided doses, approximately 25–35 mg hydrocortisone daily) in 70 CFS patients for 3 months showed slight improvement on symptom scales, particularly in wellness score, but there was evidence of suppressed adrenocortical responsiveness on the basis of basal and ACTH-stimulated cortisol levels in 12 patients. Patients in this study had a rapid evolution of initial CFS symptoms (<6 weeks).

To further examine the endocrine axes, stimulation testing is a classic endocrine paradigm, in which subtle hypofunction may become more evident through the administration of stimulatory hormones or neuroactive agents. Nevertheless, because central control of endocrine axes cannot be directly assessed due to the lack of accessibility of the hypothalamic–pituitary circulation and brain *in vivo*, interpretation of the findings tends to be presumptive.

Dynamic endocrine testing with human CRH (pituitary stimulus) and D-fenfluramine (central serotonergic stimulus) in CFS patients revealed a trend toward lower cortisol responses, which became statistically significant when ACTH responses were analyzed as a covariate. Reduced ACTH responses to the 5-HT-1a receptor agonist ipsapirone were noted in CFS patients, although there was no difference in cortisol response. On the other hand, reduced ACTH responses to CRH have been reported. Another study showed no difference in cortisol responses to D-fenfluramine. Treadmill exercise produced a blunted ACTH response in CFS.

Insulin hypoglycemia is a profound stimulus of ACTH/cortisol release because it likely induces release of many hypothalamic ACTH secretagogues. Studies in CFS have revealed increased ACTH but normal cortisol responses after insulin hypoglycemia. This finding could be interpreted as indicating low CRH tone, with chronic CRH hyposecretion despite an intact CRH neuron, and secondary adrenal atrophy. Similar findings have been reported for the related disorder, fibromyalgia, after insulin hypoglycemia.

Naloxone is thought to stimulate ACTH and cortisol secretion by blocking tonic opioidergic inhibition of the CRH neuron. Naloxone-mediated activation is blunted in CFS, suggesting that the CRH neuron or pathways inhibitory to this neuron lead to HPA axis hypofunction in CFS, rather than increased opioidergic tone.

Stress system alterations in fibromyalgia include reduced urine free cortisol, altered circadian cortisol rhythmicity, an exaggerated response to CRH, and a paradoxical decline in cortisol levels with exercise. Increased norepinephrine excretion and a hyper-response of norepinephrine to exogenous interleukin-6 have been reported, although muscle sympathetic activity was reduced.

Familial clustering has been noted in CFS and fibromyalgia. Moreover, an epidemiological study of 1004 adult twin pairs from the Australian NH&MRC Twin Registry revealed that fatigue, as measured with the Schedule of Fatigue and Anergia (SOFA) questionnaire, likely has a genetic factor, based on multivariate genetic modeling and a monozygotic:dizygotic concordance ratio of 2.69:1.

Corticosteroid-binding globulin, a key cortisol transport protein, participates in the stress response. CBG levels decline during stress, in part due to interleukin-6-induced inhibition of CBG gene transcription. Two inherited mutations of the CBG gene, leading to lack of synthesis (null mutation) or fourfold reduced CBG:cortisol binding (Lyon mutation), have

been associated with chronic fatigue and mild hypotension. Plasma cortisol levels are low in patients with these CBG mutations, but urine free cortisol is normal.

Glucocorticoid resistance implies reduced tissue effect of circulating cortisol. Fatigue as an isolated symptom has been described in a 55-year-old woman with glucocorticoid resistance. The patient's son, in his 20s, also experienced fatigue but her seven sisters were healthy. Interestingly, fatigue was intermittent in this patient but blood pressure was constantly in the low-normal range, despite hypertension in the patient's mother. The patient did not have postural hypotension. Urinary cortisol was elevated (400–800 nmol/24 h; reference range, <300 nmol/24 h), as were plasma cortisol levels. A thermolabile glucocorticoid receptor was noted, specifically a temperature-induced reduction in [<sup>3</sup>H]dexamethasone, although a specific glucocorticoid receptor mutation was not reported. Fatigue has also been reported in other cases of glucocorticoid resistance. It is likely that these mutations are rare, but it is interesting that the two mutations leading to a disorder similar to CFS both relate to the HPA axis. It is possible that less severe mutations of these and other genes may contribute to the etiology of chronic fatigue syndrome and fibromyalgia. Since these mutations often have opposing effects on cortisol levels, such as high cortisol levels in glucocorticoid resistance and low cortisol levels in CBG gene mutations, they may contribute to the heterogeneity of findings observed in classic neuroendocrine testing.

## CONCLUSION

Chronic fatigue syndrome and fibromyalgia are common idiopathic syndromes. It has been argued that these disorders should be considered together rather than separate on organ-based nomenclature. There is a clinical impression that these disorders can be exacerbated during times of physical or psychic stress. Data indicate relative hypocortisolism, which may explain the fatigue and tendency to low blood pressure, a proposition reinforced by the therapeutic response to supplemental doses of hydrocortisone in a subset of patients. Dynamic studies of the HPA axis suggest that this hypocortisolism is of central origin. Sympathetic nervous system dysregulation may result from the hypocortisolism, particularly elevated catecholamine excretion. An underlying genetic tendency to develop these disorders, suggested by twin studies and familial clustering, may be based on inherited variations in key regulatory peptides of the stress system, as has been shown in rare kindreds

with mutations of the glucocorticoid receptor and corticosteroid-binding globulin genes. Further biochemical and molecular studies of the stress system may integrate these concepts into a coherent model to help explain the variability in neuroendocrine findings, which may be based on patient selection or variable etiology.

## See Also the Following Articles

ACTH (Adrenocorticotrophic Hormone) • Corticotropin-Releasing Hormone (CRH) and Inflammation • Glucocorticoid Resistance Syndromes and States • Stress and Endocrine Physiology

## Further Reading

- Bamberger, C. M., Schulte, H. M., and Chrousos, G. P. (1996). Molecular determinants of glucocorticoid receptor function and tissue sensitivity to glucocorticoids. *Endocr. Rev.* **17**, 245–261.
- Bronnegard, M., Werner, S., and Gustafsson, J.-A. (1988). Primary cortisol resistance associated with a thermolabile glucocorticoid receptor in a patient with fatigue as the only symptom. *J. Clin. Invest.* **78**, 1270–1278.
- Clauw, D. J., and Chrousos, G. P. (1997). Chronic pain and fatigue syndromes: Overlapping clinical and neuroendocrine features and potential pathogenic mechanisms. *Neuroimmunomodulation* **4**, 134–153.
- Clare, A. J., Heap, E., Malhi, G. S., Wessely, S., O'Keane, V., and Miell, J. (1999). Low dose hydrocortisone in chronic fatigue syndrome: A randomised crossover trial. *Lancet* **353**, 455–458.
- Demitrack, M., Dale, J., Straus, S. E., Laue, L., Listwak, S. J., Kruesi, M. J., and Chrousos, G. P. (1991). Evidence for impaired activation of the hypothalamic–pituitary–adrenal axis in patients with chronic fatigue syndrome. *J. Clin. Endocrinol. Metab.* **73**, 1224–1234.
- Emptoz-Bonneton, A., Cousin, P., Seguchi, K., Avvakumov, G. V., Bully, C., and Hammond, G. L. (2000). Novel human corticosteroid-binding globulin variant with low cortisol-binding affinity. *J. Clin. Endocrinol. Metab.* **85**, 361–367.
- Fukuda, K., Straus, S. E., Hickie, I., Sharpe, M. C., Dobbins, J. G., and Komaroff, J. A. (1994). The chronic fatigue syndrome: A comprehensive approach to its definition and study. *Ann. Intern. Med.* **121**, 953–959.
- Griep, E. N., Boersma, J. W., and de Kloet, E. R. (1993). Altered reactivity of the hypothalamic–pituitary–adrenal axis in the primary fibromyalgia syndrome. *J. Rheumatol.* **20**, 469–474.
- Torpy, D. J., Papanicolaou, D. A., Lotsikas, A. J., Wilder, R. L., Chrousos, G. P., and Pillemer, S. R. (2000). Responses of the sympathetic nervous system and the hypothalamic–pituitary–adrenal axis to interleukin-6. A pilot study in fibromyalgia. *Arthritis Rheum.* **43**, 872–880.
- Torpy, D. J., Bachmann, A. W., Grice, J. E., Fitzgerald, S. P., Phillips, P. J., Whitworth, J. A., and Jackson, R. V. (2001). Familial corticosteroid-binding globulin deficiency due to a novel null mutation: Association with fatigue and relative hypotension. *J. Clin. Endocrinol. Metab.* **86**, 3692–3700.
- Torpy, D. J., White, C., Bachmann, A. W., and Young, R. (2001). Chronic fatigue syndrome. In "Adrenal Physiology and Diseases." [www.endotext.org](http://www.endotext.org).



# Circadian Rhythms: Hormonal Facets

Johannes D. Veldhuis

Mayo Clinic, Rochester, Minnesota, United States

Ali Iranmanesh

VA Medical Center, Salem, Virginia, United States

Steven Pincus

Guilford, Connecticut, United States

Daniel Keenan

University of Virginia, Charlottesville, Virginia, United States

## Glossary

**circadian** Recurring approximately daily or every 24 h.

**feedback** Repression of a driving stimulus by a downstream product.

**feedforward** A stimulatory signal upstream of a target cell.

**homeostasis** Maintenance of stable biochemical conditions.

**multivalent** Determined by three or more signaling factors.

**pulsatile** Burst-like or episodic.

Hormonal homeostasis is subject to multivalent control by chemical signals operating remotely via time-delayed feedback and feedforward. The ensemble of interactions mediates short-term (pulsatile) and longer term (circadian) control of the endocrine milieu. Precise orchestration of multihormonal adaptations is critical across the human life span: *in utero*, infancy, puberty, adulthood, and aging. The ensemble concept of coalescent pulsatile and circadian secretory control provides a unifying platform for investigating regulatory pathophysiology.

## INTRODUCTION

Circadian rhythmicity abounds in the natural world but is surprisingly difficult to affirm unambiguously in clinical settings. A primary feature is recurrence of measured epochs at a consistent interval (periodicity) of approximately 24 h (Fig. 1A). Second, circadian activity (e.g., cortisol secretion) exhibits a free-running 24-h periodicity, wherein cyclical activity

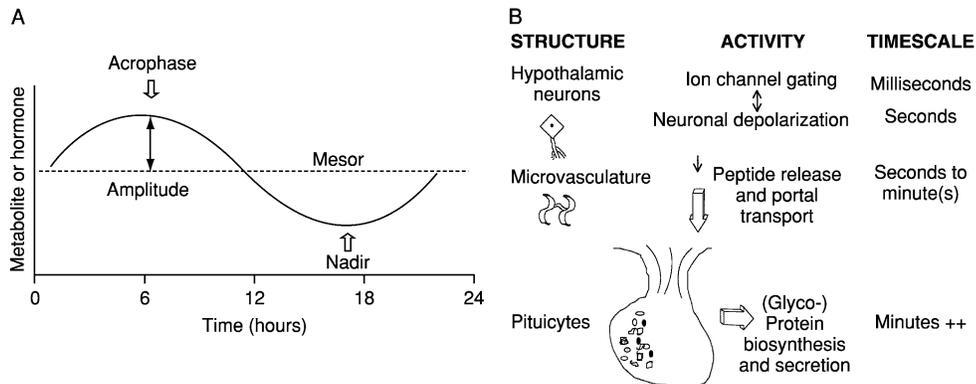
persists in the absence of external cues such as exercise, food intake, and socially directed behavior. This unique facet is explored in clinical isolation experiments such as extended cave dwelling. Third, the 24-h rhythm is temperature compensated, whereby a change in habitat temperature does not abolish periodicity. Fourth, zeitgeber (or time-given) cues operate that harmonize the rhythmic phase (acrophase or timing of the 24-h maximum) to environmental demands.

The suprachiasmatic nucleus (SCN) is a prime locus of circadian rhythm generation and attendant integration of zeitgeber inputs. In many species, including the human, brief exposure to bright light acts via retinohypothalamic pathways to trigger SCN resetting of the phase of sleep, activity, and hormonal rhythms. Unshielded astronauts orbiting the earth are exposed to sesquihoural (90-min) light–dark cycles and, thereby, potential phase resetting.

Few neuroendocrine rhythms in the human, other than cortisol, satisfy the forgoing strict criteria of circadian drive. In this regard, the body temperature rhythm provides a corollary marker of circadian endocrine activity. When the strict circadian definition is not satisfied, terms such as nycthemeral, 24-h cyclic, diurnally varying, and food- or sleep-modulated variations are more apt.

## THE CORTICOTROPIC AXIS

In initial clinical studies, day–night rhythmicity of adrenal secretory control was inferred by a morning elevation and an evening decline in urinary adrenal glucocorticoid metabolites (e.g., 17-hydroxycorticosteroids). Frequent sampling and determination of



**Figure 1** (A) Terms applied in monitoring circadian rhythms. (B) Time scales of hypothalamic neuronal outflow and attendant pituitary hormone secretion.

plasma cortisol and adrenocorticotrophic hormone (ACTH) concentrations by immunoassay corroborated nycthemeral adaptations of this stress-adaptive axis. Further analyses disclosed prominent (three- to fivefold) short-term pulsatile ultradian fluctuations (occurring briefly and repeatedly over 24 h).

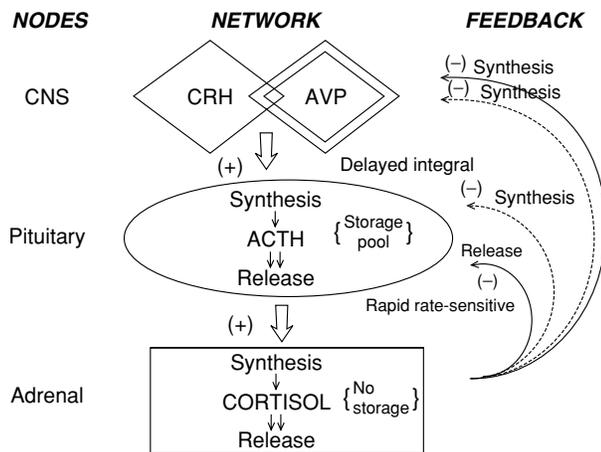
From a physiological perspective, circadian and pulsatile secretion of ACTH and cortisol arise by way of neuroglandular interfaces linking the central nervous system (CNS), anterior pituitary gland, adrenal cortex, and splanchnic plexus. The hypothalamo-pituitary unit is remarkable in transducing regulatory signals that span time scales of milliseconds, seconds, and minutes (Fig. 1B). Release of hypothalamic stress effectors occurs in brief bursts, each lasting just a few minutes. However, a cascade of subsequent events prolongs the effectual stress response; these include the latency of ACTH drive of adrenal steroidogenesis, slow dissociation of secreted cortisol from a cognate high affinity-binding globulin (CBG) in the circulation, and persistence of target tissue transcriptional and translational events initiated by cortisol exposure. For example, in the human, the plasma corticotropin-releasing hormone (CRH) half-life is 2 to 5 min, the ACTH half-life is 14 to 18 min, ACTH stimulates cortisol secretion over 5 to 30 min, the cortisol half-life is 50 to 85 min, and the cellular effects of glucocorticoids unfold over and persist for several hours.

Ensemble control is achieved via time- and concentration-dependent dynamic interactions among CNS sites, the anterior pituitary gland, adrenal cortex, autonomic neurons and peripheral target organs. Fig. 2 illustrates a simplified biomathematical construct that encapsulates this network-like notion. The two primary hypothalamic agonists are CRH and arginine vasopressin (AVP). Both neuropeptides

are released in short (1- to 3-min) secretory bursts. Terminal nerve fields from paraventricular and supraoptic nuclei, respectively, abut fenestrated and valveless portal microvessels in the median eminence. The portal microvasculature allows protein-peptide exchange between the CNS and the anterior and posterior lobes of the pituitary gland. Pulsatile signaling by CRH and AVP drives prominent ACTH secretory bursts, which are conspicuous in hypothalamo-pituitary portal blood in the conscious horse and sheep.

CRH and AVP trigger exocytotic release of ACTH and promote *de novo* biosynthesis and storage of corticotrophic hormone in pituitary cells. Blood-borne ACTH stimulates adrenal biosynthesis of cortisol by activating a complex series of intracellular second messengers and convergent nuclear transcriptional signals. Activation culminates in the transcription, translation, and phosphorylation of pivotal sterol-transport proteins and cytochrome P450-enriched steroidogenic enzymes. Secreted cortisol is transported in blood as free steroid and (predominantly) in association with high-affinity CBG and low-affinity albumin. Important targets of cortisol action include the liver, lymphoid tissue, muscle, fat cells, the pituitary gland, the hypothalamus, and hippocampal neurons. Central glucocorticoid actions mediate repression of CRH, AVP, and ACTH outflow.

The mechanisms that amalgamate pulsatile (ultradian) and 24-h rhythmic activity are not well understood. However, in an integrative context, circadian rhythms in corticotropin and glucocorticoid concentrations arise from three- to fourfold daily variations in the amplitude of underlying ultradian bursts with lesser (if any) change in frequency. The fundamental basis for such amplitude-mediated control is not established. In the experimental animal, hypothalamic



**Figure 2** Fundamental feedforward (positive) and feedback (negative) connections that interlink anatomic loci in the CRH/AVP-ACTH/cortisol stress-adaptive axis. Time-delayed, nonlinear, dose-responsive interactions endow circadian, pulsatile and entropic (pattern-regularity) dynamics. Adapted from Keenan, D. M., *et al.* (2001). *Proc. Natl. Acad. Sci. USA* **98**, 4028–4033.

genes encoding CRH and AVP exhibit 24-h rhythmic variations. There is both synergy and redundancy in CRH and AVP actions. In particular, their combined effect is supra-additive, and the genetically CRH-depleted state still maintains a diurnal pattern of ACTH secretion. Glucocorticoid feedback varies across 24 h. Adrenal steroidogenic responsiveness to ACTH also changes over the day and night. The latter depends on splanchnic neuronal drive in the rodent. Thus, nycthemeral mechanisms arise in each of four loci: the hypothalamus, anterior pituitary gland, adrenal cortex, and peripheral autonomic nervous system.

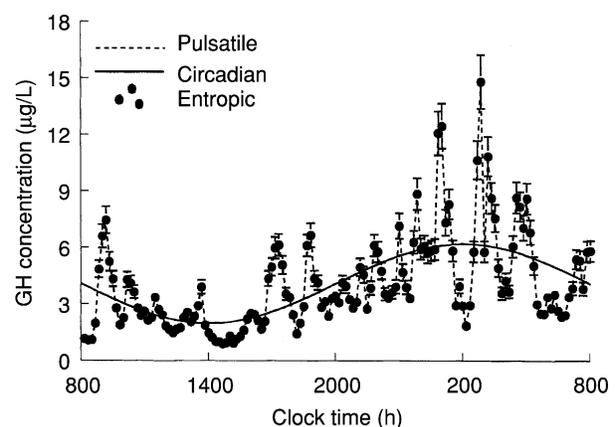
The engineering notion of a servocontrol system (regulated by time-delayed and autoadaptive mechanisms) applies broadly to homeostatic adjustments within the hippocampo–hypothalamo–pituitary–adrenal–splanchnic axis. Feedforward encompasses CRH and AVP’s stimulation of ACTH production and ACTH-specific activation of cortisol secretion, and feedback includes cortisol-mediated inhibition of CRH, AVP, and ACTH output. Ensemble interactions, rather than any single regulatory locus, supervise ultradian and circadian secretion patterns. Greater understanding of such network-level adaptive control should enhance clinical facility in visualizing the mechanistic basis of corticotropic-axis pathophysiology.

Ensemble regulation of complex neuroendocrine systems such as the corticotropic axis is difficult to quantitate precisely. One useful metric is the approximate entropy (ApEn) statistic. This measure

monitors serial subpattern regularity or reproducibility. Changes in this more subtle (subpulsatile) dynamic denote alterations in strength and/or number of signaling connections. Network-like disruption is common in endocrine pathophysiology. For example, ApEn quantitates marked loss of orderly ACTH and cortisol output in patients with corticotropinomas (Cushing’s disease) and identifies significant erosion of hormone pattern regularity in aging. Analogously, growth hormone (GH) and prolactin secretion patterns deteriorate in patients with somatotropinomas (acromegaly) and prolactinomas. Randomness of the hormone release process also increases in aging GH, luteinizing hormone (LH), and insulin.

## THE SOMATOTROPIC AXIS

The somatotropic axis embraces (minimally) CNS agonistic and antagonistic neurotransmitter inputs to hypothalamic peptidergic neurons; stimulatory, inhibitory, and synergistic hypothalamic signals to pituitary somatotrope cells; complex actions of GH on diverse target tissues; and insulin-like growth factor type 1 (IGF-1) production and reception. GH secretion is under the control of nycthemeral, pulsatile, and entropic (feedback-sensitive) mechanisms (Fig. 3). Ultrasensitive assays of this protein hormone unveiled 20-fold lifetime variations and 300-fold daily excursions in serum GH concentrations in healthy individuals. Such extrema reflect nocturnal amplification and daytime suppression of GH secretory-burst amplitude (and, in smaller measure, event frequency).



**Figure 3** Fluctuations in GH concentrations in a healthy young adult. GH release was monitored by chemiluminescence assay. The profile illustrates tripartite patterns; viz., 24-hour rhythmic (putatively partially circadian) [continuous curve], pulsatile [interrupted line], and entropic (feedback-sensitive orderliness of the release process).

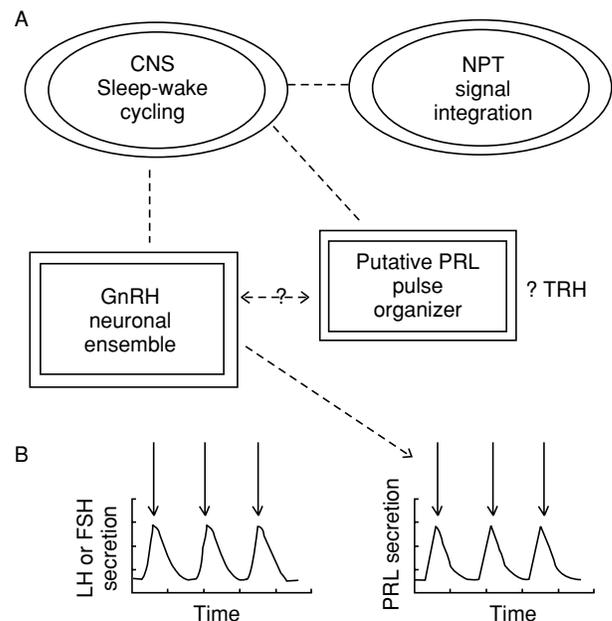
Clinical isolation experiments that establish circadian properties of GH release (e.g., free running, temperature compensated, zeitgeber-dependent control) are not available. However, sleep–wake reversal studies under constant enteral nutrition unmask a significant (35–40%) day–night rhythm in young men. Such assessments are difficult inasmuch as numerous factors govern GH secretion (e.g., fasting, food intake, sleep, exercise, gender, age). Fasting elevates GH concentrations to 10 to 30  $\mu\text{g/L}$ , whereas food ingestion suppresses values to less than 0.07  $\mu\text{g/L}$  in men and less than 0.7  $\mu\text{g/L}$  in women. Entry into deep sleep (stages III and IV) amplifies GH release by three- to sixfold within 2.5 min in young men. Conversely, two keynote GH secretagogues, gonadotropin hormone-releasing hormone (GHRH) and gonadotropin hormone-releasing peptide (GHRP), exert somnifacient effects. However, the physiological role of either effector in triggering natural sleep is not defined. Sleep onset also appears to repress hypothalamic somatostatin outflow. Somatostatin is a potent inhibitory tetradecapeptide that restrains pituitary GH release and hypothalamic GHRH secretion. Muting of somatostatinergic inhibition during deep sleep would predictably heighten somatotrope responsiveness and prompt GHRH secretion. The latter concept of somatostatin withdrawal nocturnally could account for diurnal GH rhythmicity in the face of unvarying infusions of GHRH and/or GHRP-2. Exercise promotes GH release acutely in a gender- and exercise intensity-dependent manner, and physical endurance training elevates the daily mean GH concentration by approximately twofold compared with sedentary controls. Each of the forgoing dominant regulators drive GH secretory pulse amplitude without detectably altering frequency or the timing of nycthemeral rhythmicity.

### NYCTHEMERAL RHYTHMS IN THYROTROPIN, PROLACTIN, AND LUTEINIZING HORMONE SECRETION

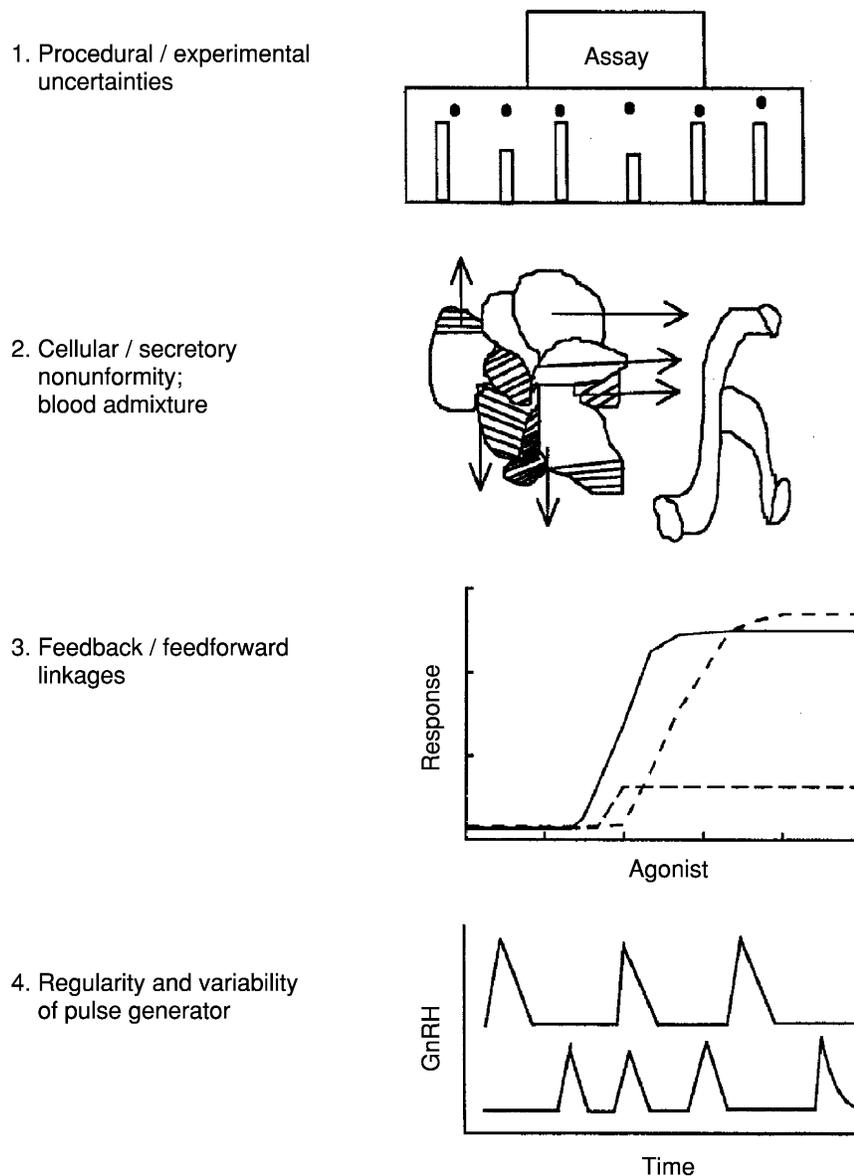
Less is known about the genesis of day–night variations in thyrotropin (TSH), prolactin, LH, and follicle-stimulating hormone (FSH) concentrations in the human. True circadian regulation is not established for these hormones. The amplitude and (less vividly) frequency of TSH and prolactin secretory bursts rise significantly during the late evening (e.g., 2030–2200 h) well before sleep onset. Approximately 35 to 50% of TSH and prolactin secretion may be nonpulsatile (basal), as compared with only 5 to 15% for other anterior and posterior pituitary hormones.

An unresolved physiological issue is whether the nonpulsatile mode of hormone production is regulated over the day and night.

LH concentrations vary by 5 to 15%, and testosterone concentrations vary by 20 to 30%, in healthy young men over 24 h, with maxima occurring at 0430 to 0900 h. The matutinal rise in FSH is minimal and variable. Diurnal rhythmicity of LH and testosterone secretion is evident before adolescence, increases markedly as puberty unfolds, and declines (sometimes to an undetectable level) in older men and women. Continuous EEG recording and frequent (2.5-min) monitoring of reproductive hormone outflow during sleep have unveiled multivalent linkages among sleep stage transitions; LH, FSH, prolactin, and testosterone secretion; and cyclic nocturnal penile tumescence (NPT), which mirrors autonomic neuronal outflow. The forgoing inferred interrelationships are illustrated in Fig. 4A. For example, rapid eye movement (REM) sleep predicts a concomitant surge in NPT activity, whereas non-REM forecasts idiosyncratically time-delayed secretion of LH, testosterone, and prolactin (Fig. 4B).



**Figure 4** (A) Inferred interconnections among sleep-wake activity centers, autonomic neuronal outflow, and hypothalamic neuropeptide signaling in men. Autonomic traffic triggers oscillations in nocturnal penile tumescence (NPT). The latter activity correlates with rapid-eye-movement (REM) sleep. Episodic release of gonadotropin-releasing hormone (GnRH) drives pulsatile secretion of LH during non-REM sleep. Hypothalamic GnRH stimulates, dopamine inhibits, and additional (diverse) signals modulate pulsatile prolactin (PRL) secretion. (B) Nominal time delays among NPT, REM sleep, LH pulses, and testosterone secretion in young men.



**Figure 5** Sources of random (stochastic) variations that impact neurohormone outflow. Adapted from Veldhuis, J. D., *et al.* (1999). *J. Clin. Endocrinol. Metab.* **84**, 3498–3505.

Sleep–wake reversal studies, albeit limited in scope, are consistent with partially sleep-independent modulation of pituitary–gonadal rhythms. The origins and mechanisms of such regulation are unknown. Plausible (but unproven) mechanisms are diurnally cyclic neural signals from the hypothalamus to the pituitary gland and from the spinal cord to the testis.

The frequency of LH pulses declines at night in young men and women (especially during the midfollicular and luteal phases of the menstrual cycle). Nycthemeral frequency adaptations virtually vanish

in older men and postmenopausal women. The latter observation points to a dependency of diurnally varying gonadotropin-releasing hormone (GnRH)/LH pulse generator frequency on gonadal or adrenal sex steroids and/or age. Estrogen and progesterone govern SCN-directed gonadotropin secretion in the female rat.

Epinephrine (adrenalin), insulin, renin, aldosterone, adrenal androgen, leptin, antidiuretic hormone (ADH), oxytocin, and parathyroid hormone concentrations vary over 24 h. Rhythms in these cases arise from multiple (nonexclusive) mechanisms such as

nycthemeral oscillations in autonomic neuronal outflow, diurnal differences in caloric and fluid intake, the sleep–wake cycle, and day–night contrasts in physical activity. For example, aldosterone secretion increases on arising in the morning due to heightened circadian ACTH stimulation, accentuated renin–angiotensin stimulation due to upright posture, increased adrenergic drive associated with nocturnal diuresis, and posturally mediated volumetric adjustments.

## STOCHASTIC ELEMENTS

From a clinical and experimental perspective, facile assessment of hormone release is confounded by several important sources of random (stochastic) variability. System uncertainty arises from technical and biological nonuniformities (Fig. 5). Technical factors include variations among (and within) individuals, sample acquisition, processing and assay, and data reduction (e.g., radioactive particle and photon counting, regression fitting, measurement interpolation). Biological factors originate from inconstant hormone synthesis, secretion, and exocytosis by single cells, clusters of cells, and an entire gland; random molecular admixture within, distribution throughout, and elimination from the circulation; time-varying dose–response coupling; and unpredictable pulse timing.

## CONCLUSION

The SCN supervises circadian activity of the CRH/AVP–ACTH–cortisol axis. Circadian control is less evident in the GH axis, wherein fasting, food intake, sleep, and exercise act dominantly. LH, FSH, TSH, and prolactin rhythms do not satisfy strict circadian criteria. Prominent 24-h variations in pulse amplitude (and, to some degree, frequency) build diurnal rhythms. Kinetic (elimination) processes delay protein removal from plasma, thereby stabilizing effector concentrations. Homeostasis is coordinated via time-delayed, dose-dependent, nonlinear feedback and feedforward interactions within (and among) neuroendocrine axes. Stochastic variability arises jointly from technical and biological sources. The forgoing ensemble of control mechanisms stabilizes hormone secretion within physiological bounds.

## Acknowledgments

The authors thank Jonathan Kuipers and Olivia Veldhuis for graphics and manuscript support. We appreciate stimulating discussions of this theme by Eve van Cauter, Eberhard Haus, Bernard Carroll, and Eugene Yates. This work was supported in part by a National Center for Research Resources grant to the General Clinical Research Center of the Mayo Clinic (M01 RR00585), the National Institutes of Health (RO1 DK 19695, K01 AG19164), a Small Business Innovative Research Award (NIH SBIR R44 DK54104), and a Veterans Affairs Merit Review Award.

## See Also the Following Articles

Brain, Effects of Steroid Hormones • Melatonin • Neuroendocrine System and Aging • Neurotransmitters, Overview • Pineal Gland

## Further Reading

- Evans, W. S., Sollenberger, M. J., Booth, R. A. Jr., Rogol, A. D., Urban, R. J., Carlsen, E. C., Johnson, M. L., and Veldhuis, J. D. (1992). Contemporary aspects of discrete peak detection algorithms: II. The paradigm of the luteinizing hormone pulse signal in women. *Endocrine Rev.* **13**, 81–104.
- Giustina, A., and Veldhuis, J. D. (1998). Pathophysiology of the neuroregulation of GH secretion in experimental animals and the human. *Endocrine Rev.* **19**, 717–797.
- Keenan, D. M., Licinio, J., and Veldhuis, J. D. (2001). A feedback-controlled ensemble model of the stress-responsive hypothalamo–pituitary–adrenal axis. *Proc. Natl. Acad. Sci. USA* **98**, 4028–4033.
- Keenan, D. M., and Veldhuis, J. D. (2001). Hypothesis testing of the aging male gonadal axis via a biomathematical construct. *Am. J. Physiol.* **280**, R1755–R1771.
- Urban, R. J., Evans, W. S., Rogol, A. D., Kaiser, D. L., Johnson, M. L., and Veldhuis, J. D. (1988). Contemporary aspects of discrete peak detection algorithms: I. The paradigm of the luteinizing hormone pulse signal in men. *Endocrine Rev.* **9**, 3–37.
- Veldhuis, J. D. (1999). The neuroendocrine control of ultradian rhythms. In “Neuroendocrinology in Physiology and Medicine” (P. M. Conn and M. Freeman, eds.), pp. 453–472. Humana Press, Totowa, NJ.
- Veldhuis, J. D., Iranmanesh, A., Johnson, M. L., and Lizarralde, G. (1990). Twenty-four hour rhythms in plasma concentrations of adenohipophyseal hormones are generated by distinct amplitude and/or frequency modulation of underlying pituitary secretory bursts. *J. Clin. Endocrinol. Metab.* **71**, 1616–1623.
- Veldhuis, J. D., Straume, M., Iranmanesh, A., Mulligan, T., Jaffe, C. A., Barkan, A., Johnson, M. L., and Pincus, S. M. (2001). Secretory process regularity monitors neuroendocrine feedback and feedforward signaling strength in humans. *Am. J. Physiol.* **280**, R721–R729.



# Collagen Metabolism

Jason P. Mansell and Allen J. Bailey  
*University of Bristol, Bristol, United Kingdom*

**collagen** A family of extracellular supporting proteins comprising 25% of the total body protein.

**cross-links** The enzymic formation of bonds between molecules to stabilize the fiber.

**glycation** The adventitious reaction of glucose with long-lived proteins resulting in a reduced optimal function.

**metabolism** The remodeling of collagen by controlled enzymic degradation and synthesis.

**turnover** The rate at which collagen is replaced.

Collagen is the major supporting protein of the mammalian body—from bone, skin, and tendon to the transparent lens capsule, capillary basement membranes, and renal filtration membranes. It is rapidly remodeled during embryonic development and growth, a process that continues more slowly throughout life. The tissues are also capable of specific repair of the collagen. Therefore, the control mechanisms of collagen metabolism must be extremely specific and consequently very complex. With increasing age, its optimal function is lost due to adverse reactions with body metabolites such as glucose.

## INTRODUCTION

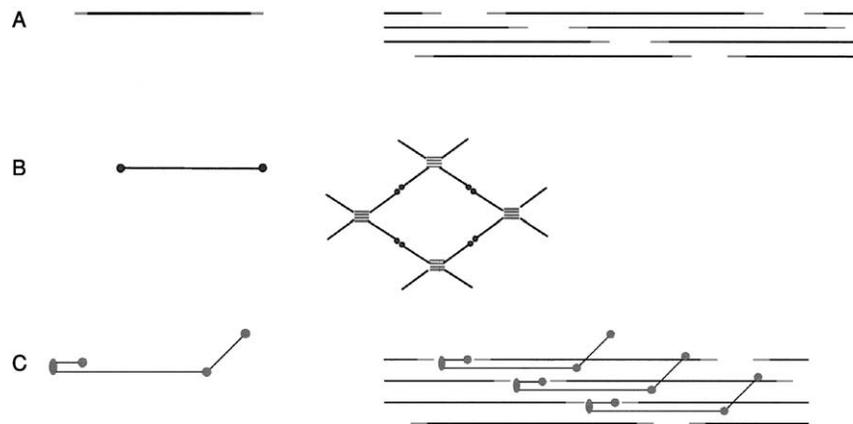
The old perception of collagen as a metabolically inert supporting structure has changed dramatically over the past two decades. Approximately 27 different collagen types have been identified, and their biological properties are extremely diverse in addition to their role as a supporting scaffold for various tissues. Collagen acts as a substrate for cell adhesion, regulates fibril formation, and serves as a linker to other matrix components. In addition, it is involved in cell recognition during development and the transduction of extracellular signals. Collagen is also capable of generating peptide factors responsible for the autocrine regulation of collagen metabolism. For example, the terminal domain of type IV collagen has

been reported to inhibit angiogenesis and tumor growth, and other functions for these domains are likely to be identified in the future. As new members of the collagen protein family are being identified, even more functions are being added. The collagen family clearly plays a major role in embryonic development, remodeling of tissues, repair, and aging. The remodeling of collagenous tissues during growth and development clearly must be a very carefully controlled series of processes.

Remodeling is slower following maturation, and recent studies have demonstrated that collagenous tissues are susceptible to adventitious reaction with glucose, a series of reactions now referred to as non-enzymic glycosylation or, more generally, as glycation. These complex reactions lead to advanced glycation end products (AGEs); some of these modify the amino acid side chains, whereas others cross-link the fibrils, resulting in tissue malfunction and increased resistance to remodeling. The mechanism of synthesis and degradation of collagen during these processes is basically understood, but the regulation that maintains the balance between the two has only recently begun to be unraveled. A detailed understanding of the regulation and specific interactions of matrix collagen is central to our knowledge of growth and development, understanding certain disease etiologies, and generating new therapeutic strategies.

## COLLAGEN STRUCTURE

The collagens are a closely related group of extracellular matrix molecules (ECMs) that adopt very distinct supramolecular organization (Fig. 1). In addition, the collagens exhibit diverse biophysical properties. The characteristic factor of this family of proteins is the unique triple helical domain composed of three polypeptide chains in which glycine occurs every third residue in the repeating sequence Gly–X–Y, where X and Y are often proline and hydroxyproline, respectively. Consequently, these so-called  $\alpha$ -chains are left-handed polyproline helices rather than the



**Figure 1** Diagrammatic representation of the major triple helical collagen molecules and their various modes of extracellular aggregation to form supra-macromolecular supporting structures. (A) Types I, II, III, V, and XI are long, rod-like molecules that aggregate in a parallel fashion to form quarter-staggered end overlap fibers. (B) Type IV is a longer, more flexible molecule that aggregates with the procollagen peptides still intact in an antiparallel fashion to form a network structure that is the basis for basement membranes. (C) Types IX and XIV are FACIT molecules possessing interrupted triple helices that decorate the surface of existing fibers, for example, type IX on type II fibers in articular cartilage.

usual protein  $\alpha$ -helices. Three of these polyproline helices are bound together into a super right-handed triple helix. The triple helix is stabilized by hydrogen bonding, and the glycine side chain (hydrogen) is small enough to be integrated within the triple helix, allowing tight interchain hydrogen bonding between the NH and the CO of the adjacent peptide chains. This tight structure renders the triple helix particularly resistant to proteolysis. More than 20 different collagen molecules, made up of 38 different  $\alpha$ -chains (either as homotrimers or heterotrimers) with two or three different chains, have been identified. The triple helix is not unique to collagen, but other proteins containing the helix are not classified as collagens because they do not aggregate to form ECM assemblies.

### Fibrillar Collagens

Type I, II, III, V, and XI collagens aggregate in a quarter-staggered parallel array to form tightly packed fibrils possessing the characteristic 65-nm axial periodicity when viewed in the electron microscope. Type I is the predominant collagen of skin, tendon, and bone and confers strength to the tissue. Type III exists as narrow fibrils and tends to confer flexibility to tissues comprised of type I fibrils. Type II is the major supporting collagen of articular cartilage. The other fibrillar collagens are present in minor quantities and may play a role in regulating fiber size and the interaction with proteoglycans.

### Nonfibrillar or Network Collagens

Type IV collagen is the major structural scaffold of basement membranes and exists as an open three-dimensional network that is ideally suited for the close incorporation of laminin and heparan sulfates. These membranes act as cell supports and molecular sieves but can also play a role in cell migration and differentiation. Type VII collagen molecules aggregate as dimers in an antiparallel fashion and act as an anchoring fibril of the basement membrane to the underlying matrix. Types VIII and X appear to form hexagonal arrays in Descemet's membrane of the eye and growth plate cartilage, respectively.

### Filamentous Collagens

Type VI collagen aggregates as a loosely packed filamentous structure formed by end-to-end alignment of the tetramers with a repeat period of 100 nm. They occur in many tissues, and it has been suggested that they may align the larger type I fibrils.

### FACIT Collagens

The fibril-associated collagens with interrupted triple helices (FACIT) do not form fibrils but associate with other collagen fibrils. Type IX decorates the surface of type II collagen fibrils and may act to limit the size of the type II fibrils or act as a link between the type II

fibrils to form a stabilizing network within the cartilage. Types XII and XIV are similar minor collagens. Type XII has been shown to interact strongly with dermatan sulfate, suggesting that it may act as a molecular cross-link stabilizing the matrix.

### Other Minor Collagens

The collagens identified recently, largely by cDNA sequencing, have not all been isolated from tissue, nor have their molecular and supramolecular structures been determined.

The aggregated macromolecular assemblies are often heterogeneous. For example, type I may contain small amounts of types III and XII on the surface and type X in the fibril core. Cartilage type II fibrils are decorated with type IX and possess a core of type XI. The six different type IV chains can form different molecules, for example,  $(\alpha 1)2\alpha 2$  in most basement membranes and  $\alpha 3, \alpha 4, \alpha 5$  in glomerula basement membrane.

### Gene Structure

Each of the known 38 procollagen  $\alpha$ -chains is encoded by a separate gene, although some differences in collagens may arise from alternative splicing of primary pre-mRNA transcripts, particularly in the case of type XIII. The genes encoded for these collagen polypeptide chains are widely dispersed among 15 chromosomes and are characterized by high concentrations of guanine and cytosine due to the repeated coding for proline. The type IV genes are located in pairs in head-to-head orientation on chromosome 13 and their promoter regions overlap, but generally the other collagen genes are not closely linked. Most of the exons for the fibrillar collagens have a conserved 54 base-pair repeat but do not appear to be repeated in the nonfibrillar collagens.

## COLLAGEN SYNTHESIS

### Intracellular Events

The fibroblast is the predominant cell type responsible for the production of collagen, although several other cell types (e.g., endothelial cells, osteoblasts, chondrocytes, muscle cells) are also capable of synthesizing collagen. The processes are now well understood and have been well reviewed (Fig 1).

The major steps can be described as follows. Collagen DNA is transcribed in the nucleus, and the primary RNA is spliced to mRNA and translated in

the cytoplasm. The N-terminal signal sequence peptide is cleaved at the rough endoplasmic reticulum to provide a single procollagen chain that interacts with two other procollagen chains to form the triple helical procollagen molecules.

The procollagen  $\alpha$ -chain undergoes a number of posttranslational modifications that include hydroxylation of proline and lysine by prolyl-hydroxylase and lysyl-hydroxylase, respectively, and these reactions involve the cofactors ferrous iron, ascorbate,  $\alpha$ -ketoglutarate, and oxygen (Table I). Hydroxylation occurs on the nascent  $\alpha$ -chain and proceeds until all of the prolines in the Y position are hydroxylated. This process is crucial to the subsequent stability of the triple helix. The hydroxyproline content determines its thermal stability, and underhydroxylated collagen is unstable and denatures at body temperature. Some of the hydroxylysines may then be glycosylated by galactosyl and glucosyl transferases. These modifications take place only on the nascent  $\alpha$ -chains and cease as the triple helix is formed following nucleation of the C propeptides and the subsequent winding together of the polyproline helices into the triple helix. The procollagen molecules are then transported from the rough endoplasmic reticulum to the Golgi apparatus, where they are secreted from the cell.

### Extracellular Processing

The N and C propeptides are released from the procollagen molecules on secretion from the cells. This step results in the formation of fibrils. The C propeptide enters the circulation and, unlike the N terminal, is not degraded until it reaches the liver. Therefore, the quantitation of the C propeptide can be used as an accurate stoichiometric determinant of synthesis in which one intact collagen molecule is produced per procollagen molecule (Fig. 2).

Cleavage of the N and C propeptides by specific endopeptidases occurs during secretion from the cell, but the collagen molecule retains short residual non-triple helical regions at the N (16 residues) and C (20 residues) termini. Each of these telopeptides contains a lysine or hydroxylysine residue that is subsequently involved in stabilizing the newly formed fibers through the formation of intermolecular cross-links within the fibril. This is achieved by the oxidative deamination of these residues by lysyl oxidase to form the corresponding aldehydes that then react with specific lysine and hydroxylysine residues in the triple helix. These initial divalent cross-links are predominant during rapid growth and remodeling of the

**Table I Production of Collagen Fibers: Posttranslational Events**

<b>Intracellular</b>	
Cleavage of signal peptide	Signal peptidase
Hydroxylation (at positions 3 and 4) of proline	Prolyl hydroxylase
Hydroxylation of lysine	Lysyl hydroxylase (four isoforms)
O-galactosylation of hydroxylysine	Hyl-galactosyl transferase
O-glucosylation of gal-hyl	Hyl-gal-glucosyl transferase
Disulfide bond formation	Protein disulfide isomerase
<i>Cis-trans</i> conformation of prolyl bond	Prolyl-peptidyl <i>cis-trans</i> isomerase
<b>Extracellular</b>	
<b>Enzymic</b>	
Cleavage of N propeptides	Procollagen-N-proteinase
Cleavage of C propeptides	Procollagen-C-proteinase
Cross-link formation	Lysyl oxidase
<b>Non-Enzymic</b>	
Glycation	Glucose/Glyoxal/ Malondialdehyde
Isomerization and racemization	L- to D-aspartic acid
Oxidation/Beta elimination	Methionine/Serine

**Table II Potential Regulatory Stages of the Amount of Collagen Synthesised and Deposited as Fibrils**

Transcription	mRNA transcription and splicing mRNA degradation
Translational	mRNA translation Posttranslational modifications
Intracellular	Intracellular degradation of newly synthesized collagen molecules
Extracellular	Extracellular degradation of newly synthesized collagen fibers

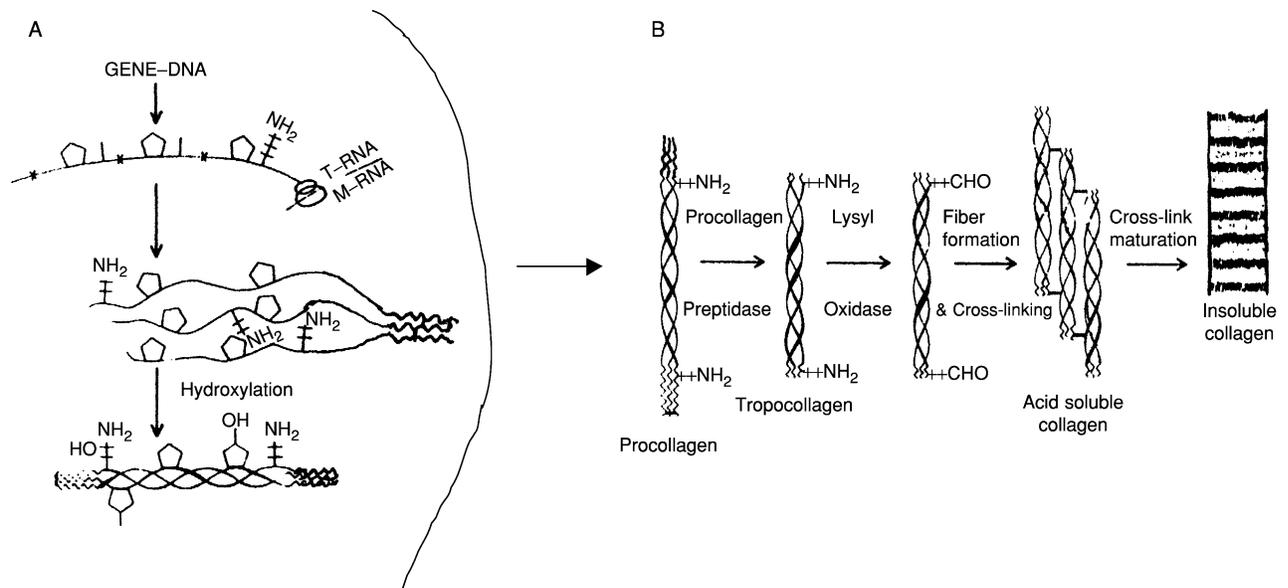
**Regulation of Synthesis**

Regulation of procollagen synthesis occurs at several levels: transcription in the nucleus, translation in the cytosol, degradation within the cell and in the extracellular space, and modulation of the concentration of connective tissue-synthesizing cells (Table II). Clearly, under normal physiological conditions, the overall synthesis of collagen has to be in balance with the overall degradation of the collagen fibrils.

**Transcriptional Regulation**

Synthesis can initially be controlled at the transcriptional level by the activation of particular genes, and recent evidence suggests that the genes must be coordinated. Therefore, the regulation of expression of procollagen genes is clearly a complex process: coordinating genes for the various collagens. The control mechanism is believed to involve both specific DNA sequences (*cis* elements) in the promoter region

collagenous tissues. However, as turnover decreases, they react further to form trivalent stable cross-links. It has been proposed that these trivalent bonds stabilize the mature fibril by forming interfibrillar cross-links.



**Figure 2** Schematic representation of intracellular (A) and extracellular (B) events in the synthesis of collagen and the formation of the collagen fiber.

that are required for efficient transcription and the promoter sequence required for cell and tissue expression; cell-specific (*trans*-acting) factors are involved in the regulation of collagen gene expression. Any modification of these regulatory elements would result in loss of control and, subsequently, in fibrosis or resorption. Several *cis*-acting regulatory elements and *trans*-acting protein factors have now been identified. The absolute quantitation may be modulated, for example, by cytokines mediated at the transcription level, although posttranslational regulation must also be taken into account.

C and N propeptides of procollagen have been reported to be down-regulators of collagen synthesis by a feedback mechanism, and it has been proposed that they may act either by blocking DNA transcription, by decreasing the half-life of procollagen mRNA, or by altering the rate of mRNA translation.

The rate of collagen synthesis will be determined by the equilibrium between the synthesis of mRNA (i.e., the rate of transcription) and the degradation of mRNA. Protein kinase C, cyclic AMP (cAMP), and various cytokines such as interleukin-1 (IL-1) may affect this equilibrium. Growth factors have also been shown to affect collagen production by altering the steady-state levels of mRNA for procollagen; for example, TGF- $\beta$  acts at more than one regulatory level by increasing the rate of transcription of procollagen genes, increasing mRNA stability, and decreasing the degradation of the newly synthesized procollagen. There are several TGF- $\beta$  signaling pathways leading to the activation of collagen genes, one of which is linked directly to the TGF- $\beta$  receptor at the cell surface and involves proteins known as Smads. TGF- $\beta$  also activates the connective tissue growth factor (CTGF).

#### Posttranslational Regulation

The rate of translation of procollagen mRNA is generally not considered to be rate limiting in collagen synthesis. However, the posttranslational modifications, the most important of which is hydroxylation of proline, may be a rate-limiting step.

A second determinant of collagen formation is the rate of the intracellular degradation of newly synthesized procollagen when as much as 30% of the procollagen is degraded. The final step in the deposition of collagen is the proteolytic degradation of the fibrous collagen in a process that is highly regulated, with many cytokines and growth factors affecting the activity of the enzymes involved.

Table III Collagen-Degrading Enzymes: Matrix Metalloproteinases and Cathepsins

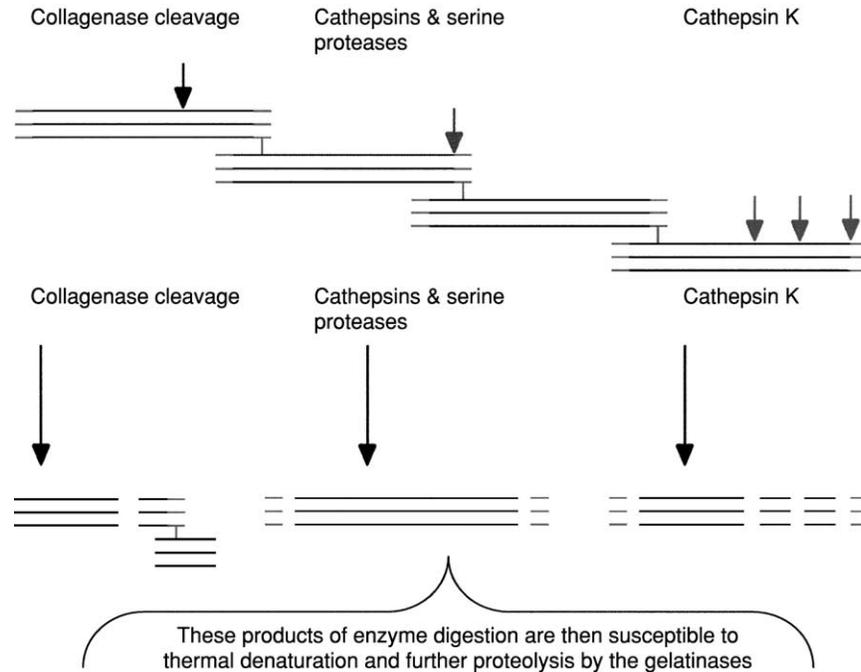
Matrix metalloproteinase	Optimal pH	Major substrate
Interstitial collagenases		
MMP-1, MMP-8, MMP-13	7-8	Triple helix
Gelatinases		
MMP-2, MMP-9	7-8	Gelatin/Propeptides
<b>Membrane types</b>		
MMP-14 (MT-1-MMP)	7-8	Telopeptides/Gelatin
MMP-16 (MT-3-MMP)	7-8	Pro-MMP-2
<b>Cathepsins</b>		
Serine proteinases		
Plasmin, elastase, cathepsin G	7-9	Telopeptides/Gelatin
Cysteine proteinases		
Cathepsin B, L, S, K	3-6	Telopeptides/Gelatin
Aspartic proteinases		
Cathepsin D	3-6	Telopeptides/Gelatin

## RESORPTION

### Degradation

The degradation of collagen is a highly controlled process, as several functionally related proteinases with diverse substrate specificities coordinate to control the degradation and resorption (Fig. 3). The intact triple helical molecules are very resistant to proteolytic degradation, but a specific group of zinc-dependent endopeptidases, known as the metalloproteinases, are capable of degrading collagen (Table III). There are now more than 20 members of this proteinase family. Degradation is initiated by interstitial collagenase (MMP-1) or neutrophil collagenase (MMP-8), both of which cleave the triple helical region into two fragments at a point three-quarters along the length of the molecule at a specific site (Gly-Ileu) near the C-terminal end. The two fragments are thermally unstable and denature at body temperature to random chains of gelatin, which are then susceptible to most proteolytic enzymes, including the gelatinases MMP-2 and MMP-9, which are present in the ECMs.

An alternative route to initial fragmentation/solubilization of the fiber, prior to denaturation and degradation by the gelatinases, is the specific cleavage of the N- and C-terminal peptides on the triple helical side of the cross-links. This can be achieved by the acid cathepsins following phagocytosis by the intracellular lysosomes, but cathepsins can also exist in the extracellular matrix and so can digest the intact fibrils.



**Figure 3** Representation of the various mechanisms involved in the degradation of cross-linked collagen fibers. Mammalian collagenase cleaves the triple helix at a single location to give two fragments three-quarters and one-quarter the length of the molecule. Extracellular cathepsins and serine proteinases cleave the telopeptide region between the triple helix and the cross-links, thereby releasing the molecule into solution. Cathepsin K has been shown to cleave the triple helix at several locations (multiple arrows). The released fragments thermally denature under physiological conditions and are then digested by the gelatinases to small peptides and amino acids. In addition to extracellular degradation of the fiber, damaged fibers are phagocytosed and digested intracellularly within lysosomes by the cathepsins.

Indeed, cathepsin K has recently been reported to be capable of digesting the triple helical region of the type I collagen fibril in bone. The importance of this enzyme in bone metabolism is highlighted in pycnodysostosis, where the congenital deficiency of cathepsin K results in an abnormal skeletal phenotype such as unfused cranial sutures and sclerosis.

Collagenase (MMP-1) is secreted as a proenzyme at 55 kDa or at 57 kDa in the glycosylated form. Collagenase is expressed by most connective tissue cells, and although it has several substrates, its predominant activity is against intact collagen. A novel human collagenase, homologous to rodent collagenase I, has been identified in breast carcinoma and designated MMP-13. Its principal substrates are the interstitial collagens, primarily type II. Guinea pigs possess both MMP-1 and MMP-13; guinea pigs resemble rabbits more closely than they do rodents.

The MMPs share common structural and functional features and can consist of seven defined domains of biological activity; for example, MMP-9 exhibits all seven domains, whereas most others lack at least one of these domains. At the N terminal, the

prepeptide directs the proenzyme to the extracellular matrix and is cleaved during secretion. The second domain contains the "cysteine switch" sequence motif maintaining the proenzyme in its latent state, but this is also lost on activation. The third domain is the catalytic region containing the fibronectin domain to bind gelatinases (MMP-2 and -9) and the zinc-binding domain that is coordinated with three histidine residues at the active site. The fifth short domain contains the type V collagen linker that is proline rich and is probably the recognition site for collagen, followed by the hemopexin domain (hemopexin is a blood glycoprotein with a high affinity for heme) that binds the tissue inhibitors of metalloproteinases (TIMPs). The seventh domain of the membrane MMPs contains a transmembrane domain. A furin-serine proteinase activates the MT-MMPs in the Golgi complex; consequently, these enzymes complex with the plasma membrane in their active form.

## Regulation

The MMPs have the capacity to completely degrade the extracellular matrix; therefore, they need to be

strictly controlled. This may occur at three stages: (1) transcriptional regulation by cytokines and growth factors via cell-signaling pathways, (2) extracellular activation of the proenzyme by plasmin or other MMPs, and (3) inhibition by specific TIMPs that are coordinately expressed with the MMPs.

### **Transcriptional Regulation**

The genes that encode the MMPs (except MMP-2) are inducible by cytokines and growth factors; consequently, they have multiple regions in their promoters AP-1, TPA, and SP-1. Therefore, a complex system exists for regulation at the transcriptional level, allowing a subtle interplay between the members of the MMP family and gene control of remodeling in response to physiological and pathological stimuli.

### **Activation Control**

The MMPs (except MT-MMP) are secreted as proenzymes and remain in their latent state until required. Therefore, activation can occur locally at a required site of action. The proenzyme is activated by cleavage of the 77- to 87-N-terminal amino acid sequence by plasmin or one of the other activated MMPs. This activation, known as the cysteine switch, involves water entering the active site and initiating irreversible autolytic cleavage of the propeptide motif PRCG (V/W) PD (V/L). Plasmin is thought to be the most significant activator of MMPs. (*In vitro* activation can be achieved by aminophenyl-mercuric-acetate (APMA), a compound that disrupts the cysteine-zinc interaction.) Activation of MMP-2 occurs by a different route in that it does not possess a plasmin cleavage site and *in vivo* may be activated via MT-MMP or via a pro-MMP-2/TIMP-2 complex. It can also be activated *in vitro* by APMA.

### **Inhibition Control**

The last line of regulation to control matrix degradation is inhibition of the MMPs by the TIMPs, which bind to MMPs with a 1:1 stoichiometry. They are also inhibited by the general serine proteinase inhibitor  $\alpha$ -2 macroglobulin, but its action is predominantly restricted to the body fluids. The TIMPs can complex with the latent form of MMPs. There are four structurally related TIMPs: TIMP-1, -2, -3, and -4. TIMP-1 and -2 have a molecular weight of approximately 24 kDa and so are readily diffusible in tissue. TIMP-3 is bound to the plasma membrane. The TIMPs are stabilized by disulfide bonds, with one domain being responsible for enzyme recognition and the other being responsible for inhibition.

### **Gelatinases**

The gelatinases MMP-2 and -9 will degrade collagen following its denaturation to gelatin but will also cleave native type IV collagen because of its numerous short disorganized helical regions due to the absence of the strict Gly-X-Y sequence. MMP-9 expression can be induced by cytokines and growth factors and is associated mainly with the inflammation process and particularly following the ingress of neutrophils. MMP-2 is expressed constitutively and is found in all tissues, particularly during remodeling. Pro-MMP-2 is secreted as a 72-kDa zymogen but can exist in a higher molecular weight glycosylated form, particularly in rodents. The gelatinase recognizes gelatin by the affinity for three fibronectin-like repeat sequences and specifically cleaves Gly-Ileu at the sequence Pro-Gln-Gly-Ileu-Ala-Gln-Glu.

## **COLLAGEN TURNOVER RATE**

The amount of collagen finally deposited obviously continues to increase during mammalian growth. However, collagen synthesis rates decrease with age but do not terminate; collagen is continuously synthesized and degraded slowly throughout life. The rate of this turnover differs considerably among the tissues of the body. For example, the collagen of articular cartilage has a biological half-life of more than 100 years, compared with a half-life of 2 days for the periodontal ligament. Unfortunately, accurate determination of the rate of turnover has been fraught with technical difficulties. The early studies involved following the incorporation of  $^{14}\text{C}$ -proline into collagen and determined as  $^{14}\text{C}$ -hydroxyproline indicated a very low turnover of collagen; indeed, it was considered virtually metabolically inert. Further studies revealed the surprising finding that up to 30% of the newly synthesized procollagen was degraded before secretion from the cell. Presumably, this involved the removal of aberrant collagen; for example, underhydroxylated collagen is thermally unstable and would denature and hence be readily degraded. Later studies using a flooding dose of  $^{14}\text{C}$ -proline suggested a similar rapid degradation of the newly formed extracellular fibers. These studies also suggested that the extent of degradation increased with age and that up to 80% of the total synthesis is degraded before formation of stable fibers in adult rats and rabbits. Furthermore, calculation of the rate of turnover of lung collagen was estimated at 10% per day, a surprisingly high figure.

Alternative techniques have been attempted, e.g., the determination of  $^{18}\text{O}$  by mass spectrometry. The

recent identification of isomerization and racemization of L- to D-aspartic acid has been shown to be linearly related to age and hence used to calculate turnover. Based on the rate of D-aspartic acid accumulation in the tissue as determined by high-performance liquid chromatography (HPLC) of the derivitized racemic mixture following hydrolysis, the turnover of human collagen has been estimated at 14 years for skin, 100 years for articular cartilage, and 500 years for dentine. The transformation of D-aspartic acid appears to provide a means of determining the rate of turnover of collagen, and further studies on other collagenous tissues would be valuable. It also needs to be assessed whether the transformation reaches equilibrium at a certain proportion of D-aspartic acid formation.

The glycation cross-link pentosidine has been shown to increase linearly with the age of skin and cartilage and so can be an additional marker of the metabolism of these tissues.

## AGING

### Glycation

The turnover of collagen decreases following maturation, and this leads to long-term exposure to reducing sugars and the consequent nonenzymic glycation of collagen (Fig. 4). Glycation is a posttranslational modification that is of singular importance in long-lived proteins such as collagen. These Maillard-type reactions can result in random cross-linking, thereby changing the physical and chemical properties, alterations that include reduced turnover to such an extent that they are believed to play a fundamental role in the aging process.

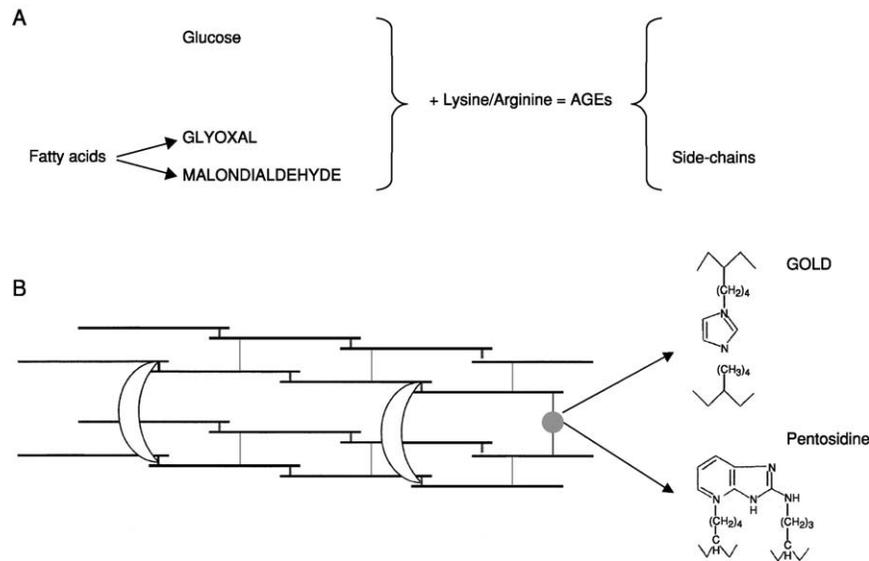
The initial reaction is between the aldehydic group of the open chain form of glucose with either the  $\epsilon$ -amino group of lysine residues or the  $\epsilon$ -guanidino group of arginine to form a Schiff base. The resulting Schiff base is then stabilized by undergoing an Amadori rearrangement. The Amadori product undergoes further oxidative reactions, resulting in the formation of stable AGEs. Other pathways to the formation of AGEs may involve metal catalyzed oxidation products of glucose, reaction with lipid oxidation products such as malondialdehyde, and the reaction with natural active metabolites such as methylglyoxal. The many different pathways depend on the conditions available and lead to a diverse range of AGEs. Some have been identified, but these probably represent only the tip of the iceberg in terms of products yet to be identified.

The AGEs may be stable modifications of the amino acid side chains of lysine and arginine (e.g., carboxymethyl-lysine, argpyrimidine, imidazolones of arginine), and these modifications may affect important cell-collagen interactions. Other AGEs may form intermolecular cross-links at random between collagen molecules, thereby stiffening the fiber and reducing its optimal function and turnover. The formation of these AGE cross-links is believed to be responsible for the stiffening of the aorta, thickening of the glomerula basement membranes, cataract formation, and reduction of renal and pulmonary function. The AGE cross-links have also been implicated in the amyloid plaques in neurodegenerative diseases. These changes generally occur slowly with age but are accelerated in diabetes where the glucose levels are much higher, a process often referred to as "accelerated aging." The involvement of AGEs in these deleterious processes is strongly supported by similar changes occurring following the administration of a mixture of AGEs prepared *in vitro* to normal mice.

There are many cell-mediated effects of AGEs on collagen turnover, possibly through the AGE receptors that, when bound, activate key signaling molecules, such as NF- $\kappa$ B, that alter gene expression. Fibroblast cell lines exposed to glycated albumin can exhibit reduced synthesis of type I collagen and MMP-2 production. In addition, there is a decrease in the adhesion of cells to the matrix, possibly through modifications in the RGD (Arg-Gly-Asp) recognition sequence for the integrins.

The extent of glycation depends on the rate of collagen turnover for a particular tissue; thus, greater glycation has been reported in articular cartilage where turnover is very low, whereas matrix glycation is lower in bone because of the higher metabolic rate.

Glycation may be inhibited by several compounds that compete during the initial stages of the reaction such as aminoguanidine, aspirin, and pyridoxine. Studies have reported a reversal of the effects of glycation in terms of an impressive return of the elasticity of various tissues, including the vascular system by the use of phenyl-dimethyl-thiazolium chloride. Although this type of compound seems to be effective, its proposed mode of action to cleave the glycation cross-links is controversial and may simply involve the removal of specific reaction metabolites. It has also been proposed that the effects of AGEs may be reduced *in vivo* by the presence of AGE-binding scavenger receptors such as those found on a variety of cell types, including macrophages, monocytes, and endothelial cells. Therefore, there does appear to be the



**Figure 4** (A) Schematic diagram of the nonenzymic glycosylation (glycation) of collagen by reaction with glyoxal (derived from glucose or fatty acids) and malondialdehyde (derived from fatty acids) to form AGEs. These products may be glycation cross-links or side chain modifications. (B) The proposed location of inter-fibrillar glycation cross-links (double arrows) and the structure of two glycation cross-links: pentosidine and a lysyl-imidazolium product (GOLD). The proposed locations of the lysyl-oxidase-derived immature cross-links (up and down arrowheads) and mature cross-links (crescent symbols) are also shown.

potential for an in-built control mechanism to reduce the damaging effects of glycation *in vivo*.

The accumulation of glycation products certainly correlates with aging, and the glycation products do affect the optimal functioning of vital tissues, but do they cause aging? Inhibition would slow down the loss of function in the cardiovascular and renal tissues and so would lead to better health during old age, but if it is the cause of aging, it would lead to an extension of maximum life span. It is highly unlikely that glycation controls aging.

### Isomerization and Racemization

Amino acids exist in proteins in the L form, but isomerization and racemization of these residues in long-lived proteins have been known for some time. Aspartic acid has one of the fastest rates of racemization, and determination of the rare D-aspartic acid form in tissue has been shown to be age related and so can be used as a measure of turnover of the particular protein. Type I collagen undergoes  $\beta$ -isomerization of the Asp-Gly bond within the non-helical C-terminal peptide, and the extent of the transformation increases linearly with age. There are two spontaneous transformations: the racemization of the L-enantiomorph form to the D form and the

isomerization of the peptide backbone by transfer from the  $\alpha$ -carboxyl to the  $\beta$ -carboxyl of the peptide-bound aspartic acid. However, the nature of the structural change within the telopeptide, in contrast to glycation, is unlikely to affect the rate of turnover of the modified collagen.

Other chemical modifications accumulate with age, deamidation, oxidation of methionines, *cis-trans* isomerization of prolines, and  $\beta$ -elimination of serine and cysteine, but glycation appears to have the most profound effect on the metabolism of collagen.

### See Also the Following Articles

Bone Turnover Markers • Collagen Metabolism Disorders • Fibroblast Growth Factor (FGF)

### Further Reading

- Bailey, A. J., Paul, R. G., and Knott, L. (1998). Mechanisms of maturation and ageing of collagen. *Mech. Ageing Dev.* **106**, 1–56.
- Bienkowski, R. S. (1989). Intracellular degradation of newly synthesised collagen. In "Current Trends in the Study of Intracellular Protein Degradation" (S. Grisolia and E. Knecht, eds.), pp. 423–443. Springer, New York.
- Kielty, C. M., Hopkinson, I., and Grant, M. E. (1993). Collagen: The collagen family, structure, assembly, and organization in the extracellular matrix. In "Connective Tissue and Its

- Heritable Disorders: Molecular, Genetic, and Medical Aspects” (P. M. Royce and B. Steinmann, eds.), p. 113. Wiley-Liss, New York.
- May, P. K., and Laurent, G. J. (1994). The regulation of collagen and elastin gene expression in normal lung and during pulmonary disease. *In* “Molecular Biology of Lung Disease” (P. J. Barnes and R. A. Stockley, eds.), pp. 216–260. Blackwell Science, Oxford, UK.
- Myllyharju, J., and Kivirikko, K. I. (2001). Collagen and collagen related diseases. *Ann. Med.* **33**, 7–21.
- Parks, W. G., and Mecham, R. P. (1998). “Matrix Metalloproteinases.” Academic Press, San Diego.



# Collagen Metabolism Disorders

Jason P. Mansell and Allen J. Bailey  
*University of Bristol, Bristol, United Kingdom*

**collagen** A family of extracellular supporting proteins comprising 25% of total body protein.

**metabolism** The ratio of synthesis to degradation.

**turnover** The rate at which collagen is replaced.

The health and maintenance of collagenous structures involve the delicate balance between the synthesis of collagen types and their degradation and removal. Disruption in either of these processes can have serious consequences for the performance of the affected tissues.

## INTRODUCTION

The turnover and maintenance of functionally competent collagenous tissues involve a delicate balance between synthesis and degradation. An imbalance would lead to either a loss of collagen (i.e., resorption) or an excess of collagen (i.e., fibrosis). The loss or resorption of collagen is probably best exemplified by the fragility of bone in osteoporosis, but loss does occur in other tissues. In pathological fibrosis, the excess synthesis continues unabated, disrupting the normal tissue architecture and function. Fibrosis can occur in nearly any organ of the body, with the most frequently affected being the lung, liver, kidney, heart, and skin.

The initiating factors resulting in a metabolic imbalance, as well as factors initiating the resorption of collagenous tissues, may well be different in the fibrosis of various tissues. However, there are common cellular and molecular events underlying progression of fibrosis; hence, there is a common basis on which therapeutic strategies could be developed. To provide a basis for any therapy, an understanding of the cellular and molecular events is crucial. The metabolism of collagen in terms of the mechanisms of synthesis and degradation has been discussed previously, and the inhibition or stimulation of any of these stages that

are rate limiting could be used to prevent fibrosis or resorption.

We have classified a number of examples of collagen metabolism diseases involving an imbalance in both synthesis and resorption. The latter is best exemplified by bone collagen, and the former is best exemplified by soft tissue collagen. Therefore, it is worth briefly summarizing the properties of collagen in bone. Approximately 90% of the organic matrix of bone tissue consists of collagen. Bone tissue is a composite material consisting mostly of type I collagen, which provides a template for mineralization by hydroxyapatite crystals. The function of the small amount of type V collagen (~3%) is uncertain, but based on the retention of its amino-terminal propeptide, it may be involved in the regulation of type I collagen fiber size. The collagen is synthesized and mineralized by the osteoblasts and is removed primarily by the bone-resorbing osteoclasts. In addition to its important supportive role, collagen provides a means of cellular communication, most notably to the entombed osteocytes, a main function of which is believed to be the transduction of mechanical stimuli that has important consequences for the regulation of bone mass.

Bone loss, or osteopenia, and the subsequent development of osteoporosis are known to occur as a consequence of changes in the circulating levels of some systemic hormones. The reduction in serum estrogen postmenopause and its impact on bone metabolism have probably gained the greatest global interest, but bone loss in response to raised prolactin, parathyroid hormone (PTH), thyroxine (T4), and glucocorticoids, as well as to reduced insulin, has also been studied.

## METABOLIC DISORDERS DUE TO EXCESSIVE DEGRADATION

### Estrogen Deficiency

The study of bone tissue in response to reduced estrogen has been the subject of intense academic

and clinical interest because osteoporosis is a leading cause of morbidity, mortality, and social costs. Fracture of the femoral neck exerts the greatest impression on the health services. Indeed, of those who present with this type of fracture, nearly a fifth will die within the first 6 months, and of the remainder, approximately 50% will lose their independence. It has been predicted that the number of individuals over 60 years of age will increase to 700 million by 2025, so clearly the financial burden on the health services is a major concern.

When the circulating levels of estrogen fall with the onset of menopause, bone collagen turnover increases (i.e., the indexes of resorption and formation both are elevated) but the rate of collagen catabolism outweighs collagen synthesis. In addition, this rate of turnover is greater in cancellous sites due to their larger surface area. This increased rate of synthesis of collagen results in a change in quality, for example, higher hydroxylation of the lysine residues and modified cross-linking, both of which lead to a more fragile fibril. The change in quality accelerates the probability of bone fracture and almost certainly accounts for the absence of a correlation of bone mineral density (BMD) and bone fracture observed in clinical practice. Clearly, the quality of the collagen needs to be taken into account.

It has also been demonstrated that osteoblasts mobilize a number of very potent bone-resorbing stimuli in response to diminished estrogen; these include prostaglandins, macrophage colony-stimulating factor (MCSF), interleukins-1 and -6, and tumor necrosis factor- $\alpha$  (TNF- $\alpha$ ). It has also been found that the bone tissue levels of transforming growth factor- $\beta$  (TGF- $\beta$ ) are reduced following estrogen deficiency, and this in turn may lead to greater osteoblast losses through programmed cell death and/or reduced collagen synthesis. The sensitivity of bone to PTH has also been reported to be greater with estrogen deficiency.

Estrogen replacement is a well-established treatment for osteoporosis. Estrogen reduces the rapid phase of bone loss, but the mechanism is unknown, although estrogen receptors are known to be present in both osteoblasts and osteoclasts. The inhibition of bone collagen turnover by estrogen allows the collagen to stabilize by normal enzymic cross-link maturation. However, we found that in the long term (>5 years), the estrogen had an anabolic effect, with the synthesis of new collagen being stimulated. These results indicate that long-term high doses of estrogen have a therapeutic role in postmenopausal women with osteoporosis.

Bisphosphonates are also used in osteoporosis and are believed to inhibit osteoclast activity, but the precise mechanism is unknown.

### Cushing's Disease

Since it was first described by Harvey Cushing in 1932, the syndrome of glucocorticoid excess has been recognized as one of the most important causes of bone loss. Indeed, osteoporosis features significantly in the morbidity associated with Cushing's syndrome. What of the mechanism of bone loss in Cushing's syndrome? Osteoblast-like cells and calvarial bone cultures on receipt of glucocorticoids have reduced mRNA levels of  $\alpha$ 1(I) collagen (COL1A1). This reduction in type I collagen synthesis is attributed to both a reduction in COL1A1 transcription and a decrease in the stability of the actual mRNA. It has been proposed that glucocorticoids act at a TGF- $\beta$ -responsive site to block COL1A1 transcription. Because the ability of osteoblasts to produce type I collagen is reduced during periods of raised serum cortisol, there is a reduction in bone collagen synthesis and mineralization. However, the resorption of bone with prolonged glucocorticoid exposure is increased in response to raised serum PTH and decreased estrogen.

### Hyperparathyroidism and Hyperthyroidism

Hyperparathyroidism is a common endocrine disease and is due to an excessive production of PTH by the parathyroid gland. It is characterized clinically by hypercalcemia and osteopenia. Albright and Reifenstein referred to it as a "disease of the bones and stones and abdominal groans occasionally complicated by psychological moans."

It is generally agreed that the actions of PTH on bone are via the osteoblast, although binding of iodinated PTH (residues 1–84) has been described for osteoclasts. PTH results in the catabolism of collagen (by stimulating the synthesis of collagenase) as well as in the inhibition of type I collagen synthesis. It has been demonstrated that PTH, by acting through the cyclic AMP (cAMP)–protein kinase A pathway, inhibits the promoter region for type I collagen, possibly through the generation of a protein that binds to the COL1A1 promoter. Another possible mechanism is thought to include the PTH-induced autocrine feedback by prostaglandin E<sub>2</sub> on osteoblasts, resulting in the suppression of COL1A1 gene transcription.

Hyperparathyroidism results in significant bone losses because of heightened bone matrix resorption coupled with reduced type I collagen synthesis. Primary disease results from either parathyroid gland hyperplasia or a single adenoma, whereas secondary hyperparathyroidism is most commonly a consequence of chronic renal failure. A hormone strikingly similar to PTH, PTH-related protein, is the known causal factor for the hypercalcemia associated with malignancy.

More than a century has passed since von Recklinghausen first described bone loss in thyrotoxicosis. Since this initial observation, there has been compelling evidence of a role for thyroid hormones in bone metabolism. Bone loss as a consequence of hyperthyroidism is the result of an imbalance between bone formation and resorption. Cultured rat and human osteoblasts display heightened proliferation and reduced collagen synthesis when treated with triiodothyronine (T3). The ability of cells to mineralize a collagenous matrix *in vitro* is also reduced when osteoblasts are treated with T3 and T4.

### Prolapse

As an example of an imbalance in a soft connective tissue turnover, genitourinary prolapse is a very common problem, the precise pathophysiology of which is unknown. However, recent studies have demonstrated that the disorder may be due to a change in the metabolism of the uterine collagen in favor of excessive resorption. This results in a decrease in the mechanical strength of both the uterine tissues and the supporting ligaments.

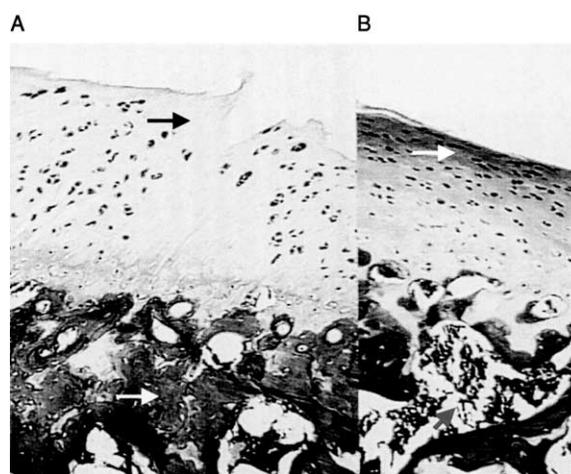
## METABOLIC DISORDERS DUE TO EXCESSIVE COLLAGEN SYNTHESIS

### Osteoarthritis

Osteoarthritis (OA) represents the most common form of joint disease. This cruelly debilitating disease is characterized by local and systemic increases in BMD, bony outgrowths (osteophytes), thickening of bone tissue (sclerosis), bone cysts, loss of articular cartilage, and joint space narrowing. However, it has generally been considered to be a disorder of articular cartilage degradation. Although both Radin and Dequecker proposed several decades ago that OA is a whole "joint disease," it is only during the past few years that researchers have looked beyond the degradation of the articular surface in attempting to understand the etiology of the disease. Besides the characteristic local joint changes, another example of

bone tissue abnormality in OA is the striking systemic increases in BMD for nonsynovial sites such as the lumbar spine. The factor(s) responsible for this rise in BMD has not yet been defined.

It was proposed many years ago that increased bone stiffness, attributable to altered collagen metabolism and/or composition, was responsible for placing the increase in shear stresses at the joint surface, with catastrophic consequences for joint function. Such changes in bone collagen stiffness below the articular cartilage would have a significant bearing on the integrity of cartilage tissue; the replacement of bone tissue with the stiffer methacrylate composite produces demonstrable articular cartilage splitting and losses. Like the human, the guinea pig and macaque also develop spontaneous OA. In addition, both of these species display marked morphological changes in bone tissue prior to cartilage tissue involvement (Fig. 1), findings that support earlier claims of an inherent problem of bone collagen metabolism and/or composition in OA. One of the first direct studies that aimed to address bone collagen metabolism in OA reported an increase in the excretion of the mature bone collagen cross-link, lysyl pyridinoline. Subsequent technetium-labeled bisphosphonate studies in



**Figure 1** Skeletal tissue changes in an animal model of spontaneous osteoarthritis. (A) Articular cartilage losses and disruption in response to thickened subchondral bone tissue in the Dunkin-Hartley guinea pig. Note the reduced staining for proteoglycan in the articular cartilage (black arrow) and the marked increase in subchondral bone tissue (white arrow). (B) Typical appearance of healthy joint structures in an age-matched guinea pig strain that does not develop osteoarthritis. There is considerably more bone marrow occupying the spaces between the bone trabeculae, and the matrix proteoglycan in the cartilage layer is clearly visible on staining with hematoxylin and eosin (black arrow).

**Table I** Metabolic Changes That Have Been Reported to Occur within the Trabecular Bone Compartment of Osteoarthritic Femoral Heads (Means  $\pm$  SEM)

Parameter	Healthy bone tissue	Osteoarthritic bone
Collagen content ( $\mu\text{g}/\text{mg}$ tissue)	79 $\pm$ 7	140 $\pm$ 3
Collagen calcification (mmol calcium/mol collagen)	8.4 $\pm$ 0.14	6.2 $\pm$ 0.25
C-terminal propeptide of type I collagen (ng/mg protein)	11 $\pm$ 6	288 $\pm$ 22
Alkaline phosphatase activity (mUnits/mg protein)	40 $\pm$ 10	370 $\pm$ 30
Change in latent MMP-2		Two-fold increase
Change in active MMP-2		Four-fold increase
TGF- $\beta$ (ng/mg protein)	32.9 $\pm$ 24	123.4 $\pm$ 13

affected patients identified potential changes in mineral metabolism and hence the possibility of altered collagen calcification in osteoarthritic bone. Taken together, these preliminary findings started to point toward bone abnormalities in a disease that was largely thought to center around a defect in articular cartilage biology.

A comprehensive investigation of bone collagen turnover and composition was conducted by Mansell and Bailey (Table I), who addressed the question of whether changes in collagen metabolism could occur at sites other than those juxtaposed to cartilage damage but at regions nearer the existing head-neck boundary of diseased femoral heads. The overall findings supported altered bone collagen composition and metabolism in OA. Although collagen turnover was affected most profoundly nearest the existing articular surface (subchondral bone), there were clear increases in bone collagen metabolism in the cancellous bone at more distant sites. The synthesis of type I collagen, as determined by the tissue levels of the carboxy-terminal propeptide (PICP), was markedly elevated in osteoarthritic bone, and increased degradation was determined by metalloproteinase levels. The levels of alkaline phosphatase were also greatly elevated, supporting a general up-regulation in collagen synthesis and potential for calcification. However, the extent of collagen mineralization was significantly lower in OA specimens than in age-matched controls, indicating a greater proportion of immature matrix. Additional microscopic investigations revealed disorganized collagen fibrils in diseased bone tissue. The latter may be related to a change in the phenotypic expression of collagen. Under these conditions, bone collagen contains an increasing proportion of the new type I homotrimer  $[(\alpha 1)_3]$ , as compared with the normal one  $[(\alpha 1)_2\alpha 2]$ , and an increased level of lysine hydroxylation. Both of these effects actually lead to a weaker collagen fiber in the thickened subchondral bone.

Together with the finding that osteoblasts from OA bone are capable of degrading cartilage, this suggests that a mechanical basis for OA is debatable.

The factor(s) responsible for driving the increase in bone collagen turnover has yet to be identified. Interestingly, two studies have reported raised bone tissue levels of TGF- $\beta$ , and one of these found an increase in this factor at a nonsynovial joint site (iliac crest). The significance of these findings is reflected in the ability of this agent to stimulate type I collagen synthesis by osteoblasts, and because the collagen content in OA bone tissue is significantly elevated, TGF- $\beta$  may represent the candidate molecule responsible for the changes observed.

### Paget's Disease

Another common bone disease of the elderly involving a change in the metabolism of collagen is Paget's disease, characterized by excessive collagen production and disorganized bone remodeling. The rates of collagen turnover can be increased up to 20-fold. The newly synthesized collagen fibers are disorganized in structure, producing both woven and lamellar bone. This lack of alignment is probably the cause of the most common complication, that is, fracture after trivial injury.

The nature of the biochemical changes in the collagen apparently has not been studied, but clearly the increased metabolism could lead to a high lysine hydroxylation, thereby affecting the fiber size and the cross-linking, both of which lead to increased fragility. Whether there is a change in the genetic type of collagen has not been investigated, but it is probable that the type I homotrimer levels are increased.

The disease is inhibited by antiosteoclastic drugs, suggesting that the primary abnormality is in the osteoclast, but the precise mechanism is unknown.

Calcitonin and bisphosphonates appear to be the preferred treatments.

### Cardiac Disease

Cardiac tissue contains collagen types I, III, IV, V, and VI. The most abundant of these is type I collagen, accounting for approximately 80% of the total collagen. Type III collagen forms a further 10%, and the remainder is comprised of the other collagen types. It is the presence of type I collagen that accounts for the diastolic stiffness of the heart.

The accumulation of additional collagen in the heart, or cardiac fibrosis, is known to occur in response to raised levels of the mineralocorticoid aldosterone. Although endothelial cells, myocytes, and fibroblasts have been found to express high-affinity mineralocorticoid receptors, the mechanism of aldosterone-induced cardiac fibrosis is thought to involve angiotensin II. In addition, rodent models indicate that the effect of aldosterone on cardiac fibrosis is critically dependent on salt status; animals infused with aldosterone on a restricted salt diet do not have a fibrosis of the heart. Although mechanical loading has a pivotal role to play in collagen metabolism, the hypertension resulting from raised aldosterone and salt intake is not responsible for the fibrosis. Rather, the reactive interstitial fibrosis that occurs is the result of altered endocrine signaling. Specifically, the myocardium expresses greater levels of the angiotensin receptor (AT1) to which angiotensin II will bind. This ligand, in turn, is known to stimulate the production of TGF- $\beta$ 1 by fibroblasts, and this stimulates collagen synthesis by these cells. Also, raised bradykinin levels in aldosterone-salt excess may also stimulate increased collagen deposition in the heart. Finally, aldosterone will increase the expression of endothelin receptors, and endothelins are potent stimulators of collagen synthesis. The application of the endothelin antagonist bosentan reduces cardiac fibrosis in response to elevated aldosterone-salt intake. The cardiac fibrosis occurring in response to raised aldosterone, as would take place, for example, with adrenal adenoma, involves the complex interplay between systemic factors and local signaling pathways.

### Pulmonary Disease

Pulmonary fibrosis involves the deposition of collagen within the expanded interstitium. Inflammation, fibroblast proliferation, and deposition of collagen occur during the development of pulmonary fibrosis.

No genetic mutations that can be directly associated with pulmonary disease have been established for collagen or elastin. A large number of cytokines could be involved, but the emphasis has been on TGF- $\beta$ , which has been localized at active sites of collagen production.

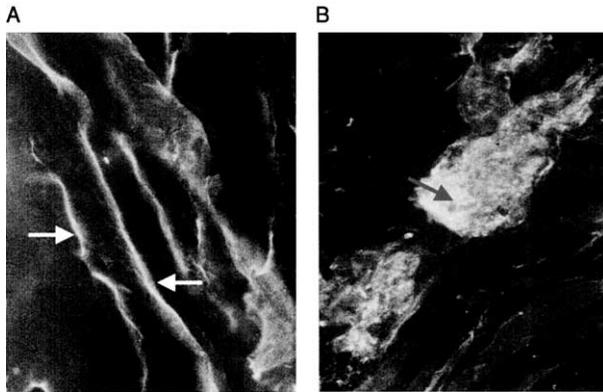
The collagen synthesis rate decreases with age, as in other tissues, but does not terminate. In fact, the turnover rates of lung collagen in rats and rabbits have been estimated at 10% per day using a flooding dose of labeled radioactive proline. However, turnover values are notoriously difficult to determine, and several methods are generally required for confirmation.

### Dupuytren's Disease and Peyronnie's Disease

Dupuytren's disease (DD) involves the progressive and irreversible deposition of excess fibrous collagen, resulting in contraction of the fingers. The apocryphal story of DD is that Count Dupuytren thought that his coachmen developed DD from the action of pulling on the horses' reins and so released the tension by slashing their fingers with his sword. Surgery is still the only treatment. However, there is no evidence of a mechanical initiation of the disease.

The lesion primarily occurs in the palmar aponeurosis, that is, the collagen that separates the flexor tendon from the overlying fibro-fatty layer. The major questions yet to be answered are what stimuli are responsible for excess collagen synthesis, how this can be controlled, and what the mechanism of contraction is. The stimuli that activate these processes are not known, but considerable progress has been made in detailing the metabolism and biochemical changes in DD.

Structurally, the aponeurosis in DD exhibits highly cellular nodules and fibrous cords. Analysis of the aponeurosis reveals striking increases in the extent of lysine hydroxylation; the level rises from approximately 5 residues per 1000 to 13 per 1000 in DD. There is also a corresponding increase in the proportion of lysine glycosylation that maintains the ratio of hydroxylysine to glycosylated hydroxylysines. Type III collagen is more abundant in the nodules of DD (10–15%), as compared with normal individuals (1–3%), and the concentration of this collagen type is even greater in the fibrous cords (30–40%). In addition, there is a higher level of immature collagen cross-links, indicating greater turnover in DD, a feature that is also observed in the granulation tissue of dermal wounds. However, in the latter case, the



**Figure 2** Type III collagen organization in the palmar aponeurosis in both healthy individuals (A) and those with Dupuytren's contracture (B). (A) Fluorescence immunolocalization of type III collagen (white arrows) molecules surrounding the type I collagen fibers in the normal aponeurosis. (B) Diffuse type III collagen distribution (white arrow) in Dupuytren's contracture. This breakdown in collagen organization and loss of the fascicular structure is a typical feature of the aponeurosis in Dupuytren's contracture.

immature cross-links decrease with healing, whereas in DD, the levels remain high even after 10 to 15 years, indicating a failure of the lesion to mature due to the constant remodeling. The high metabolic rate of collagen was confirmed by high PICP levels for synthesis, and high concentrations of the metalloproteinases and cathepsins for degradation, in the long-term (15–20 years) DD tissues.

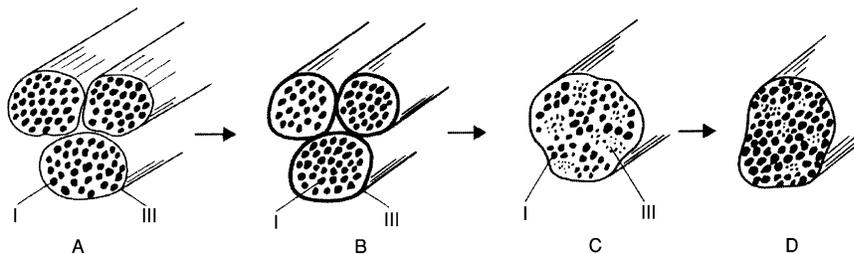
Normally, the aponeurosis possesses a fascicular structure, with the type I fibers being surrounded by a thin sheath of type III fibers (Fig. 2A). In DD, immunohistology reveals a thickening of the type III fibrous sheath, and the nodules have a disintegrating fascicular lining and an overall increase in type III fiber distribution (Fig. 2B). In the cords, the type I and III fibers are tightly packed, and there is complete

loss of the fascicular sheath (Fig. 3D). The disease is clearly evident in the apparently unaffected aponeurosis, consistent with the observation that DD often recurs within the same aponeurosis following surgery, presumably due to failure of a complete removal of affected tissue. We have suggested that the disease can be initiated and/or propagated by myofibroblast cells migrating along the fascicular sheaths. The initial stimulation for fibrosis remains unknown but appears to be retained in the tissue, as evidenced by continued collagen turnover. A candidate for triggering this response is the superoxide free radical. TGF- $\beta$  is also thought to play a role in DD due to its reported actions on collagen metabolism and its influence on fibroblast proliferation and myofibroblast differentiation.

The mechanism of palmar contraction remains elusive, but the myofibroblast is critical in this process. Modified fibroblasts certainly have the ability to contract fibrous gels and do so through architectural changes in intracellular actin filaments. These filaments attach to the cell membrane through integrins, and fibronectin in turn is thought to act as a link between the collagen fiber and the integrin molecule. Contraction is largely restricted to the highly cellular nodular regions, whereas collagen accumulation is greatest in the resulting cords.

5-Fluoro-uracil has been shown to suppress myofibroblast differentiation, and the use of  $\gamma$ -interferon was shown to inhibit the transcription of collagen mRNA. Both of these might be useful in blocking the excess deposition of collagen in DD.

DD is often linked to Peyronnie's disease. In Peyronnie's disease, the fibrosis results in curvature of the penis. Unlike DD, the fibrotic plaque of the tunica albuginea tends to harden and even mineralize, suggesting plaque maturation rather than a continuing turnover as in DD. The nature of the collagen in



**Figure 3** Schematic representation of the development of Dupuytren's disease. (A) Normal aponeurosis: fascicular structure with type I fibers sheathed in type III collagen. (B) Apparently uninvolved: thickening of the fascicular sheath. (C) Nodules: loss of fascicular structure and formation of new fine type III collagen fibers. (D) Bands or cords: tightly packed fibrils of types I and III collagen but no fascicular structure. Reproduced from Bailey (1994), with permission.

Peyronnie's disease has not been determined, but confirmation of its maturity would indicate a therapeutic approach different from that applied for DD.

### Keloids and Hypertrophic Scars

Keloids and hypertrophic scars involve the proliferation of collagen in the dermis, producing a dense collagenous network. Type I collagen mRNA levels have been reported to correlate with increased collagen production, and there is the usual increase in TGF- $\beta$ . Both of these conditions are similar to DD in that the turnover of collagen is maintained at a high level and does not return to the basal level of the normal dermis, as is the case in healing cutaneous wounds following the initial fibrosis. The stimulus is clearly present throughout these scars, but its nature is unknown.

### Scleroderma

Altered collagen metabolism resulting in the hardening of the skin and internal organs gives rise to a group of conditions generally referred to as scleroderma. Patients generally present with the disease when the processes of altered collagen metabolism are well established. Consequently, most studies have concentrated on established lesions, and the initiating factors in scleroderma remain unknown. The earliest lesions appear to be vascular and inflammatory, leading to heightened collagen synthesis in the affected areas akin to wound healing. Fibroblasts from the deep dermis display raised collagen synthesis, so the accumulation of matrix collagen is attributed to the synthetic rate rather than an increase in cell proliferation. Scleroderma fibroblasts appear to have lost their ability to regulate collagen metabolism, which would otherwise be exquisitely controlled during remodeling. Unlike DD, which is highly localized, scleroderma is progressive and spreads across the dermis and into other tissues, for example, extensive fibrosis of the heart, lung, and kidney, all of which can lead to death.

There are temporal changes in the expression of type III collagen in scleroderma. The initial stages of the disease are characterized by raised levels of this collagen type, but as the condition advances, there is a reduction in type III collagen to levels comparable to those of normal skin. This initial increase in type III collagen is also observed in the kidney, heart, and lung. Of additional significance in the case of the kidney is the increased expression of type IV collagen

in the glomerular basement membrane and capillaries, resulting in membrane duplication. In contrast to DD, which maintains a high metabolic rate, scleroderma lesions mature in a manner to that of wounds, that is, a reduction in type III collagen, increased concentrations of mature collagen cross-links, and an overall decline in turnover. Collectively, these metabolic changes result in a thickened hardened skin.

The capillaries are a potential initiating site, and attempts to improve the microcirculation through the application of vasodilators and the use of anti-inflammatory drugs have been considered. The use of agents that regulate collagen synthesis is also being investigated.

## REDUCED OVERALL METABOLISM

### Diabetes Mellitus

The metabolism of collagen is reduced in diabetes mellitus, allowing the collagen to mature and making it even less susceptible to turnover. It has been suggested that this represents "accelerated aging" of the tissues. This suggestion is supported by the high levels of glycation cross-links formed such as pentosidine and imidazolone cross-links described previously. These cross-links render the tissues stiffer, reducing their optimal function and metabolism.

Animal models that have been made diabetic (plasma glucose concentration of 300–400 mg/dl) through the application of streptozotocin or alloxan, or the diabetic BB rat, exhibit decreased bone formation. Reduced bone matrix formation has been reported for endocortical, periosteal, and trabecular–endosteal surfaces. The actual resorption of bone collagen is also reduced in these animal models. Taken together, the overall turnover of bone collagen is at a standstill in situations of severe diabetes.

Physical measurements of bone toughness support the presence of a more brittle bone tissue in diabetic rat models. This change in matrix quality may be attributed to the hypermineralization of the collagen and/or collagen glycation.

## POTENTIAL THERAPEUTIC STRATEGIES

The major questions that are being addressed are as follows. First, what are the stimuli responsible for the initiation of excess collagen synthesis or excess collagen catabolism? Second, what are the factors controlling the subsequent imbalance in metabolism? Little progress has been made on the initiating factors,

so studies have concentrated on the mechanisms of regulation.

The current fundamental research emphasis is to control the expression of the collagen genes. It is generally believed that the expression of collagen genes is coordinated under normal physiological conditions. A loss of this coordination could result in fibrosis or resorption; therefore, attempts have been made to identify these controlling factors. The control may involve several *cis* elements (specific DNA sequences in the promoter region or within the 5' end of the region required for efficient transcription) and the *trans*-acting factors (cell specific). Cells studied *in vitro* demonstrated that synthesis could be inhibited by nucleotides and by antisense genes, but the results are highly variable and delivery would be difficult. The absolute quantitation of collagen production appears to be modulated by growth factors, with these changes being mediated at the transcriptional level. Therefore, attention has focused on the many cytokines and growth factors that have been localized and identified at sites exhibiting heightened collagen synthesis.

Control at the translational level has been under study for several decades. The many specific enzymic stages in the synthesis of collagen could theoretically be inhibited to control overproduction. Prolyl 4-hydroxylase inhibition would prevent the formation of a stable triple helix, C- and N-protease inhibition would prevent fiber formation, and lysyl oxidase inhibition would result in the formation of readily

degradable non-cross-linked fibers. The inhibition of the isomers of prolyl 4-hydroxylase and its cofactors has been investigated extensively and currently appears to be the most likely candidate for the development of an antifibrotic drug.

### See Also the Following Articles

Collagen Metabolism • Graves' Disease, Hyperthyroidism in • Hypercortisolism and Cushing's Syndrome • Hyperparathyroidism, Primary • Paget's Disease of Bone

### Further Reading

- Albright, F., and Reifenstein, E. C. (1948). "The Parathyroid Glands and Metabolic Bone Disease: Selected Studies." Williams & Wilkins, Baltimore, MD.
- Bailey, A. J. (1994). Collagen changes in Dupuytren's disease. *In* "Dupuytren's Disease" (A. S. Berger, A. Delbrook, P. Brenner, and R. Hinzmann, eds.), pp. 37–48. Springer-Verlag, Berlin.
- Bailey, A. J., and Black, C. M. (1988). The role of connective tissue in the pathogenesis of scleroderma. *In* "Systemic Sclerosis (Scleroderma)" (M. I. V. Jayson and C. M. Black, eds.), pp. 75–105. Wiley, Chichester, UK.
- Henderson, J. E., and Goltzman, D. (2000). "An Osteoporosis Primer." Cambridge University Press, Cambridge, UK.
- Lijnen, P., and Petrov, V. (2000). Induction of cardiac fibrosis by aldosterone. *J. Mol. Cell. Cardiol.* **32**, 865–879.
- Lukert, B. P., and Cream, B. E. (1996). Clinical and basic aspects of glucocorticoid action in bone. *In* "Principles of Bone Biology" (J. P. Bilezikian, L. G. Raisz, and G. A. Rodan, eds.), pp. 533–548. Academic Press, San Diego.
- Mansell, J. P., and Bailey, A. J. (1998). Abnormal cancellous bone collagen in osteoarthritis. *J. Clin. Invest.* **101**, 1596–1603.



# Congenital Adrenal Hyperplasia, Prenatal Diagnosis and Therapy

Maguelone G. Forest  
Hôpital Debrousse, Lyon, France

## Glossary

**allele** Any of two or more alternative forms of a gene occupying the same chromosomal locus.

**glucocorticoid** Any of a group of adrenocortical steroid hormones whose metabolic effects include stimulation of gluconeogenesis, increased catabolism of proteins, and mobilization of free fatty acids; also a potent inhibitor of the inflammatory response (allergic response).

**mineralocorticoid** A hormone secreted by the adrenal cortex, such as aldosterone, that maintains normal blood volume, promotes sodium and water retention, and increases urinary excretion of potassium and hydrogen ions.

**mutation** The process by which a gene or some other DNA sequence undergoes a change in structure.

The biosynthesis of cortisol, a hormone necessary for survival, occurs in the adrenal glands under the stimulus of adrenocorticotrophic hormone (ACTH). The biosynthesis of all adrenal steroids is regulated by a negative feedback, but cortisol is the only adrenal steroid to exert a significant feedback control on ACTH secretion. Thus, when cortisol secretion is insufficient, the feedback loop opens and ACTH rises. A defect in any of the five enzymes necessary for the biosynthesis of cortisol from cholesterol results in congenital adrenal hyperplasia (CAH). These disorders are so named because the adrenal glands are hyperplastic at birth due to unrestrained ACTH stimulation already in fetal life.

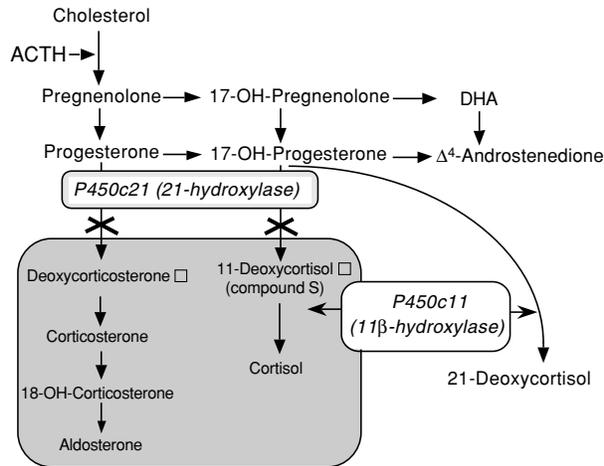
## INTRODUCTION

Steroid 21-hydroxylase (21-OH) deficiency (Fig. 1) is by far the most frequent cause (>90%) of congenital adrenal hyperplasia (CAH) and is the only cause of CAH discussed in this article. The disorder has two

major consequences: a state of cortisol deficiency and a hyperproduction of adrenal androgens due to adrenocorticotrophic hormone (ACTH) overstimulation. Steroid 21-OH deficiency has a wide spectrum of clinical variants that are not different diseases but represent points of a spectrum of disease severity. Clinical forms of 21-OH deficiency are divided into classic (severe, excongenital) and nonclassic (NC) (less severe, either symptomatic or asymptomatic, termed “cryptic”), which do not present any virilization at birth (Fig. 2).

Steroid 21-OH deficiency is one of the most common autosomal recessive genetic disorders. The classic forms occur in approximately 1 in 14,500 live newborns, whereas the NC forms occur in approximately 1% of the general population. In the classic forms, whether salt-wasting (SW) or simple virilizing (SV) type, the excessive amount of  $\Delta^4$ -androstenedione produced by the fetal adrenals is transformed into testosterone in peripheral tissues, including the genital tubercle of the growing embryo, and is responsible for the virilization of affected female fetuses, whereas the affected males have normal genitalia. However, virilization of the females does not correlate with salt loss and ranges from clitoromegaly to a fully masculinized urethra (classified in five Prader stages). The diagnosis is easy by measuring the plasma level of 17-hydroxyprogesterone (17-OHP) in antenatal and perinatal/postnatal samples. The NC presentations are less severe and are clinically expressed later during life by nonspecific hyperandrogenic symptoms, and their diagnosis requires an ACTH test.

Sexual ambiguity is a major complication of the disease, bearing the risk of sex misassignment, requiring difficult reconstructive surgery, and eventually leading to impaired long-term quality of life. This is why prenatal therapy has been proposed for preventing the *in utero* virilization of CAH females. The concept was that of David and Forest, who realized

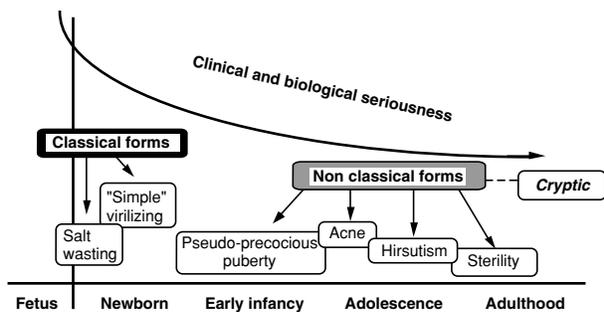


**Figure 1** Steroid 21-OH deficiency. Steroids above the enzymatic block are produced in high amounts, whereas those below the block are abnormally low (in the dotted box). The pathway of 21-deoxycortisol is illustrated.

their first prenatal treatment of CAH females in 1979 (published in 1984). Several pediatric centers have followed their protocols, which have been used throughout the world progressively. In any cases, the responsible mutations on the gene CYP21 are now well known and could be identified easily by molecular biology techniques. By combining hormonal and molecular tests, it is now possible to predict the clinical form of the disease in the context of a prenatal diagnosis, which can lead to a prenatal treatment.

**GENETIC BACKGROUND**

Steroid 21-OH deficiency is a monogenic autosomal recessive condition manifested as heterogeneous phenotypes, all caused by mutations in the CYP21 gene on chromosome 6p21.3, which lies in the class III



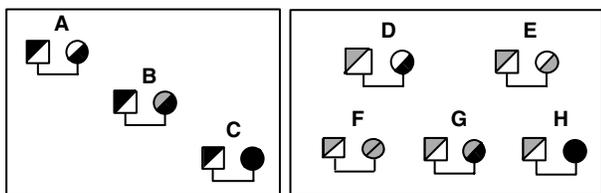
**Figure 2** Clinical forms of 21-OH deficiency. Severity and onset of symptoms are classified into two main groups: classic and non-classic.

human leukocyte antigen (HLA) region, a very peculiar region of the genome containing several duplicated genes. Indeed, there are two 21-OH genes 30 kb apart: a pseudogene 21-OHA (now called CYP21P) and the active gene 21-OHB (now called CYP21). The genes encoding the fourth component of the complement (C4A and C4B) and the 21-OH genes form a tandem of duplicated genes that are sometimes repeated more than twice.

The wide range of CAH phenotypes is associated with multiple mutations known to affect 21-OH enzyme activity. As of 2003, 56 different CYP21 mutations have been reported (and nearly 98 new ones were described in a recent thesis); these are mostly point mutations, but small deletions or insertions have been described too, as have complete gene deletions. A total of 15 mutations, constituting 90 to 95% of alleles, are derived from intergenic recombination of DNA sequences between the CYP21 gene and the highly homologous CYP21P pseudogene, whereas the remaining ones are spontaneous mutations. A reliable and accurate detection of CYP21 mutations is important not only for clinical diagnosis but also for carrier detection because there is no significant difference in 17-OHP basal levels between normal and heterozygous individuals. Several strategies based on polymerase chain reaction (PCR)-driven amplification with allele-specific oligonucleotides to the CYP21 gene have been developed. It has been demonstrated that one reaction for PCR amplification of the CYP21 gene and the chimeric CYP21P/CYP21 gene using mixed primers in combination with nested PCR and single-strand conformation polymorphism is considered highly efficient and accurate for molecular diagnosis of CAH due to 21-OH deficiency.

Gross abnormalities (e.g., 30-kb or larger gene deletions, gene conversions) are found in some patients. These rearrangements (gene conversion or deletion) producing new alleles with no enzyme activity (null alleles) account for 20 to 25 of classic 21-OH alleles, whereas small deletions and point mutations make up the rest, even if some rearrangement is associated (e.g., deletion of C4A, deletion of the pseudogene with or without a deletion of C4B). Most of them (9–12 of the 15 more frequent mutations) do normally occur in the pseudogene. The others are very rare mutations and/or are found only in certain ethnicities.

The important finding was the association between given mutations and the clinical forms of the disease. Severe mutation produces null alleles, whereas alleles carrying mild mutation have some 21-OH activity (20 to 50%).



**Figure 3** Pedigrees found in CAH families. Severe mutations are illustrated in black, whereas mild mutations are depicted by hatched areas. The classic forms of the disease are due to the combination of severe/severe mutations, whereas the combination of mild/severe mutations is associated with nonclassic forms. If a partner is known to have a severe mutation, a couple is at risk for having a child with a classic form of CAH if the other partner has at least one severe mutation, whether heterozygote (A), CAH affected with a nonclassic form of the disease, (B), or CAH affected with a classic form (C). In contrast, when one partner is known to be heterozygote for a mild mutation, there is no risk for the couple having a child affected with a classic CAH regardless of the genotype of the other partner, whether heterozygote for severe (D) or mild mutation (E), CAH affected with a nonclassic form (F, G), or even CAH affected with a classic form (H).

The clinical form of the disease is dictated by the least severe mutation carried on one allele. Patients with two alleles carrying a severe mutation (null allele), either identical (homozygous) or different (heterozygous), do have a classic SW form of the disease, whereas patients with a severe mutation on one allele and a mild mutation on the other (compound heterozygote) have an NC CAH (Fig. 3). The mutation A (or C) to G near the end of intron 2, causing premature splicing of the intron and a shift in the translational reading frame, is the single most frequent mutation in classic forms. There is a remarkable association between the mutation V281L, either in the homozygous or heterozygous condition, with NC forms. Also, the mutation I172N (either homozygous or compound heterozygous with a null allele) is associated with the SV form. Although this form is not so frequent, it is rather typical (sexual ambiguity in affected girls or early pseudo-precocious puberty in affected boys but no clinical salt loss). Concordance between genotype and phenotype is sufficiently robust to be of significant value in the diagnosis and management of the disorder. Knowledge of the genotype is essential in planning a strategy for prenatal diagnosis and treatment.

## GENETIC COUNSELING

When considering prenatal diagnosis and/or treatment, the first step is to offer genetic counseling. The role of genetic counseling is to give the parents

all available information about the disease, the clinical forms and their consequences, and the possibility of prenatal diagnosis and prenatal treatment.

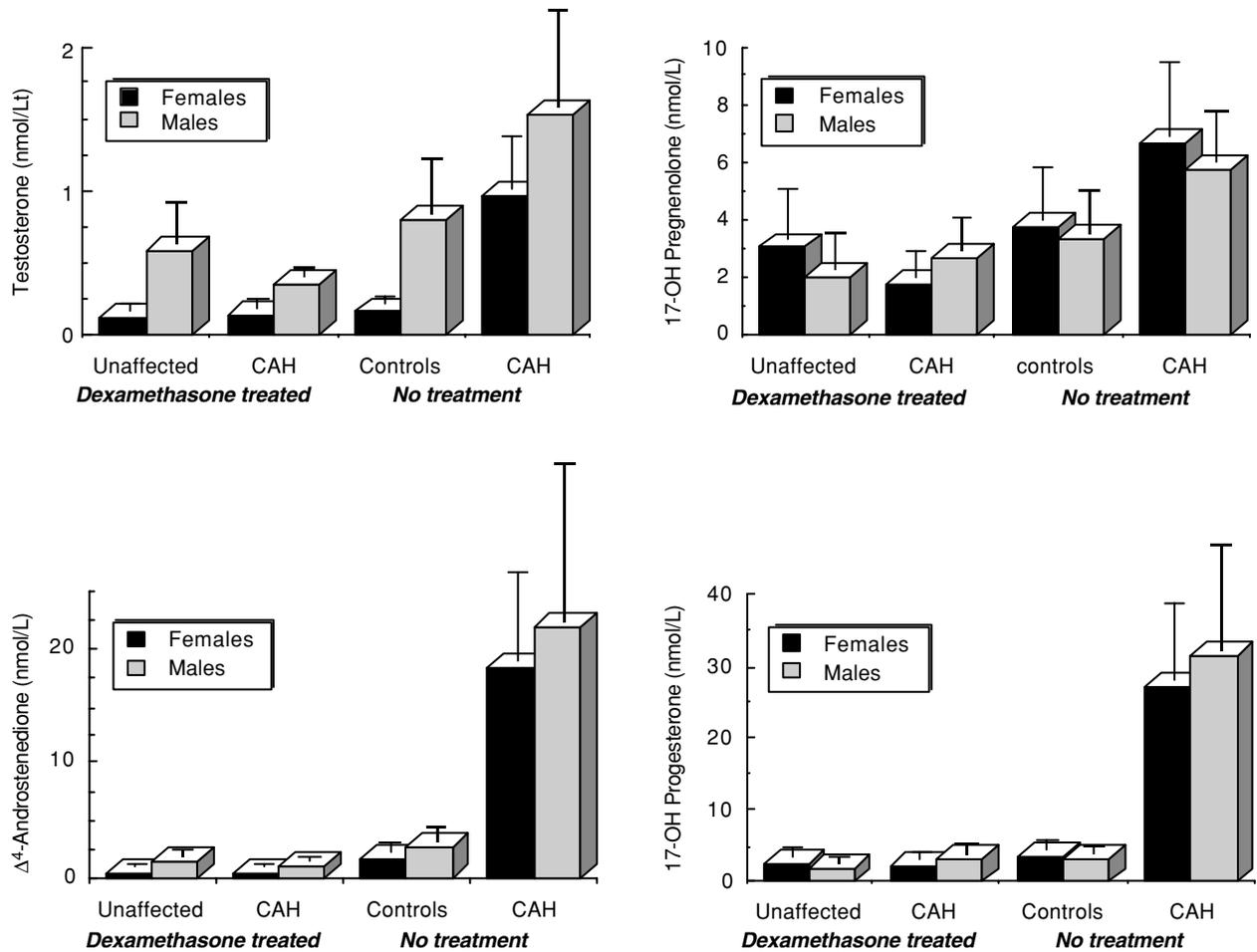
In the past, the genetics of the disease was thought to be “simple” autosomal recessive, with each parent carrying one affected and one normal allele (heterozygote). Then it became difficult to understand a misdiagnosis when prenatal diagnosis did rely on HLA typing because pedigrees were at times more complex than one thought and the parents were not always simple heterozygotes. As illustrated in Fig. 3, when individuals are compound heterozygotes (combination of mild and severe mutations), they can transmit a severe form of the disease even though they are affected with an NC or a cryptic form of the disease (couple G in Fig. 3). In contrast, one CAH-affected parent with a classic form might not transmit the severe form of the disease if the other parent carries one or two mild mutations (couple H in Fig. 3). Thus, prediction of the clinical form of the disease in a couple at risk is possible nowadays provided that the complete genotype of the future parents is available.

## PRENATAL DIAGNOSIS

In the past, prenatal diagnosis has been first possible by measuring 17-OHP and other steroid hormones (e.g.,  $\Delta^4$ -androstenedione, testosterone, 21-deoxycortisol) in the amniotic fluid (AF). The rise in 17-OHP and  $\Delta^4$ -androstenedione in CAH fetuses is clear-cut (Fig. 4). The procedure is simple and reliable and does not require the study of the index case, but it permits picking up only the classic forms, not the NC forms. One could argue that the aim of a prenatal diagnosis is to identify the classic forms of the disease, not the NC forms. However, prenatal diagnosis based on steroid analysis is still valuable when there is no other alternative provided that reliable techniques (including chromatographic purification) are available.

When HLA was found to be in closed linkage with CAH, prenatal diagnosis was made by HLA typing (associated or not with the measurement of 17-OHP). The procedure required the study of a previously affected sibling (index case). Unfortunately, the method resulted in many wrong diagnoses due to recombination, haplotype sharing, or complex pedigrees.

Nowadays, prenatal diagnosis is performed during the first trimester and relies almost exclusively on DNA analysis of mutant 21-OH alleles on tissue obtained by chorionic villus sampling (CVS) at 10 to 11 weeks of amenorrhea (WA) providing a complete



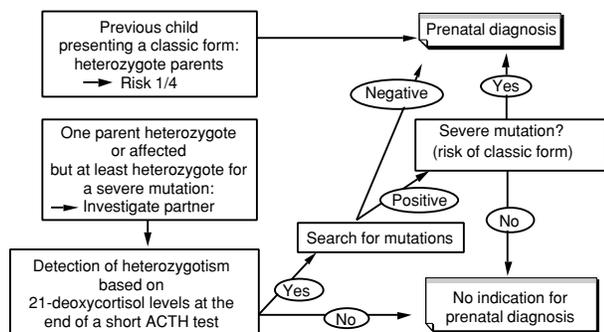
**Figure 4** Amniotic fluid (AF) steroid levels. Diagnosis of CAH is made on the high levels of 17-OHP,  $\Delta^4$ -androstenedione in both sexes, and high levels of testosterone in female fetuses. In treated pregnancies, AF steroid levels are normalized in the case of CAH-affected fetuses of both sexes.

genotype of the parents. In addition, prenatal diagnosis includes fetal sexing, which can be performed either by karyotyping on fetal cells obtained by CVS or by searching the sex-determining region of the Y chromosome (SRY) on the same cells.

In the past, prenatal diagnosis of CAH was made exclusively in couples having an affected child, but many new situations are emerging in couples where one partner is a relative of CAH patients and also when the index case is one of the parents. Therefore, if the genotype of the CAH-affected partner (presenting with either a classic form or an NC form) is known, the genetic counseling has to determine the risk for the couple to have a fetus affected with classic CAH. The prediction is based on the investigation of the other partner, in whom it is important to detect a possible heterozygotism. In our experience, this can be done in two steps. The first step involves hormonal

studies, that is, measuring 21-deoxycortisol during an ACTH test. Steroid 21-deoxycortisol is a bypass product of the hydroxylation of 17-OHP by the enzyme 11 $\beta$ -hydroxylase, which is normal (Fig. 1). This compound, normally produced in minute amounts, increases significantly during the ACTH stimulation test (which uses supraphysiological doses). In heterozygote individuals, the increase is significantly greater than that in homozygous normal individuals. In our experience of testing more than 500 people, the accuracy of the prediction of heterozygotism is 90 to 95%, and more than half of them have a normal 17-OHP response. There were fewer than 5% false-negatives, including *de novo* mutations in one parent. The algorithm is outlined in Fig. 5.

Things are not always as simple as the theory. There may be two or more mutations on the same allele (double or triple heterozygote), and when one



**Figure 5** Flow chart of prenatal diagnosis in a couple where one partner is known to have a severe mutation in the CYP21 gene on at least one allele.

mild mutation is found, one has to make sure that the same allele does not carry another mutation. Indeed, two or more mild mutations on the same chromosome result in a null allele. This implies that both parents should be genotyped to identify the genotype correctly. Performing clear-cut informative molecular studies of this disease might require hard work (e.g., Southern blot, sequencing to identify rare mutations). In some cases (< 5%), no mutation is found. Thus, when the search for mutations is negative, a prenatal diagnosis is still performed because one does not want to risk making a wrong prediction.

## PRENATAL TREATMENT

This treatment has been proposed for preventing the *in utero* virilization of CAH females. It does not cure the disease itself. After birth, there is still a need for glucocorticoid (and mineralocorticoid) replacement therapy. The rationale of the treatment involves providing the fetus with sufficient glucocorticoid levels to suppress the excessive ACTH stimulation, being mindful that the mechanism that makes the baby become virilized after birth is the same.

Protocols for prenatal treatment used by all groups is based on the work of David and Forest. The latter chose dexamethasone because it has a long half-life, crosses the placenta efficiently because it is not corticosteroid binding globulin (CBG) bound, and is not metabolized by the placental 11 $\beta$ -hydroxysteroid dehydrogenase (11 $\beta$ -HSD). The optimal dose was established by Forest and colleagues, who compared the results of various doses (fixed and then related to maternal size). The dose of 20  $\mu$ g/kg of maternal weight at the start of treatment was found to be efficacious to

bring down AF levels of 17-OHP back to normal (Fig. 4). We advise using a dose of no more than 20  $\mu$ g/kg/day, given in three divided doses to ensure a constant level in the fetus. In the French multicentric study, dexamethasone acetate has been used in all but two cases (associated with an unaffected fetus), whereas dexamethasone is more generally used elsewhere.

The second most important parameter is to start treatment early, that is, before the eighth week of pregnancy, when the fetal anlage is sensitive to the virilizing action of testosterone. If the fetus is found to be a male by karyotype or is found to be an unaffected female by DNA analysis, treatment is discontinued. Otherwise, dexamethasone treatment is continued to term.

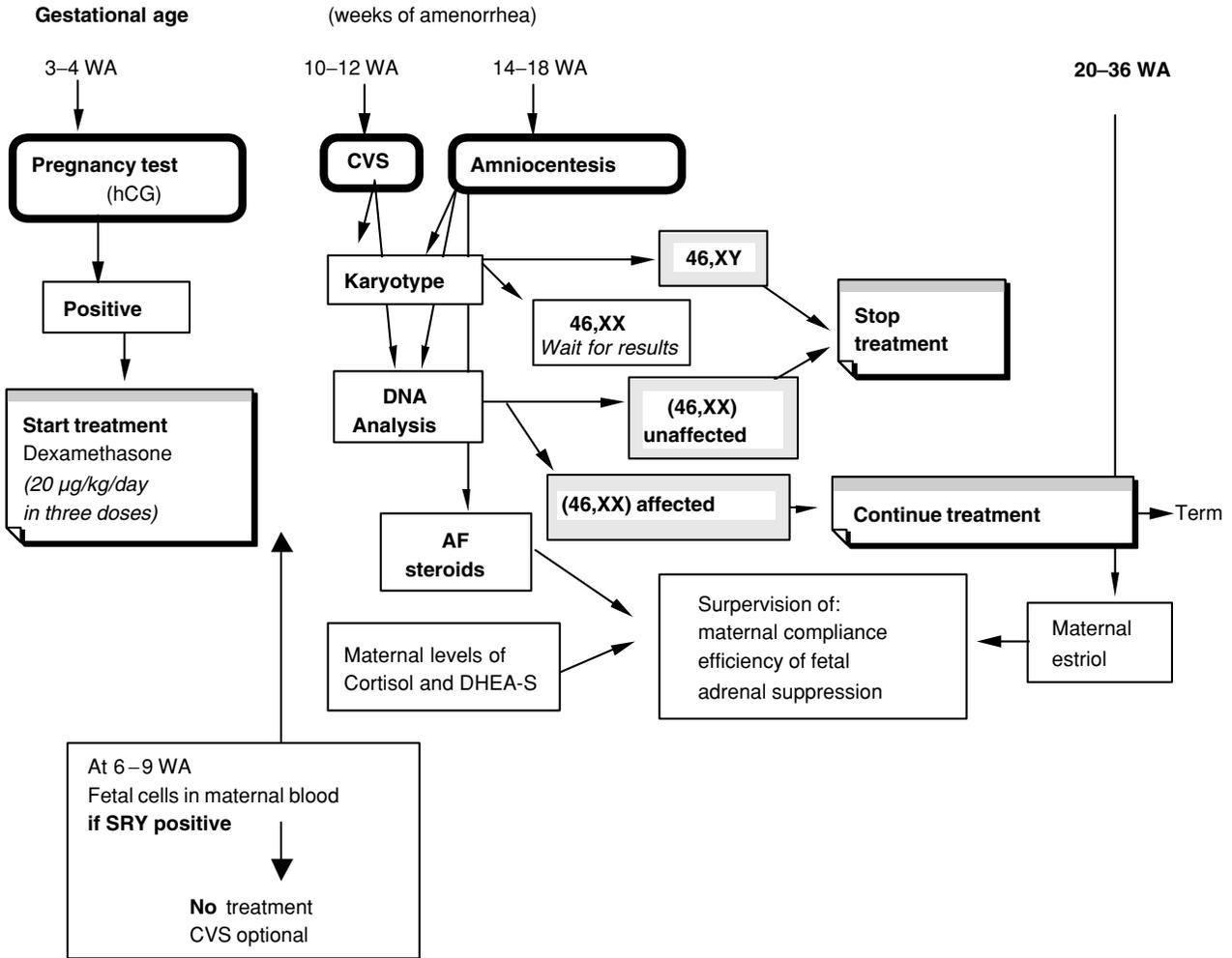
The third important part of our protocol is not to stop treatment abruptly but rather progressively by tapered doses (half the dose every 5–6 days).

Finally, our protocol includes the treated mothers a low-salt diet and asking them to pay careful attention to their weight gain.

The current protocol combines prenatal treatment with an early prenatal diagnosis and stopping maternal treatment as soon as possible if it is unnecessary. Our algorithm (Fig. 6) is very close to that proposed by New and colleagues. Treatment is started when pregnancy is diagnosed, and prenatal diagnosis is usually performed during the first trimester. However, if CVS is not available or is refused by the parents, amniocentesis can still be performed and be helpful.

It is important to monitor the compliance of the mother and the efficacy of treatment. Measuring maternal levels of cortisol and dehydroepiandrosterone sulfate (DHEA-S), which should be suppressed, can follow this. Fetal adrenal suppression is also attested by a normalization of 17-OHP levels in the AF of pregnancy associated with CAH fetuses (Fig. 4). Another way in which to look at fetal adrenal secretions is to measure maternal estriol levels, but because this metabolite rises significantly only during the second half of pregnancy, it is useful only when treatment is continued to term in case of an affected female fetus. At that time, suppressed maternal estriol levels are the best indicator of both fetal suppression and maternal compliance (Fig. 7).

The personnel involved with prenatal treatment include an obstetrician to treat the mother taking dexamethasone. Throughout the pregnancy, the mother should be monitored for weight gain, blood glucose, blood pressure, and any other complications. Treatment is discontinued if toxemia occurs. A genetic counselor with expertise in CAH must be available for the family prior to initiating prenatal diagnosis



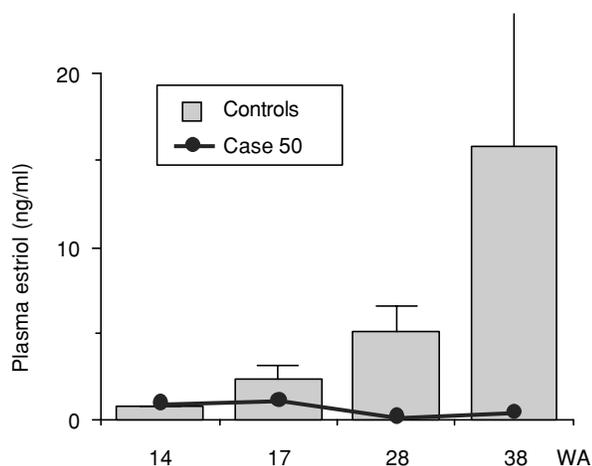
**Figure 6** Flow chart of prenatal treatment in a couple at risk for CAH due to 21-OH deficiency. Note the hope of nontreating male CAH-affected fetuses by early fetal sexing on fetal cells taken in maternal blood.

and treatment. A pediatric endocrinologist who will treat the newborn if he or she is affected should be consulted at the initiation of prenatal treatment, and a psychoendocrinologist should be integral to the follow-up of an affected child.

After the birth of a full-term or partially treated newborn, the birthweight, length, head circumference, and any pertinent physical findings (including degree of genital ambiguity and Prader score) should be recorded. At day 3 of life, measurements of serum 17-OHP levels, electrolytes, and plasma renin activity should be taken in affected infants. A blood sample for a repeat DNA analysis also is suggested. In addition, the protocol for prenatal therapy should include the requirement of developmental and psychological follow-up for affected and unaffected individuals who were prenatally treated.

### Indications of Prenatal Treatment

This treatment is indicated only in mothers at risk for having a fetus with a classic form of the disease (not with an NC form) and who are willing to continue pregnancy regardless of whether the fetus is affected. Prediction of the clinical form of the disease is now possible because, as mentioned previously, the phenotype is dictated by the least severe mutation carried by one allele. A fetus is at risk for classic CAH if a severe gene lesion (deletion, macroconversion, or point mutation that abolishes 21-OH activity) is present on at least one allele of each parent. This is the case for a couple of heterozygotes carrying a severe mutation on both the maternal and paternal sides (couple A in Fig. 3) or if one partner is heterozygote for a severe mutation and the other partner is CAH affected with



**Figure 7** Example of prenatal treatment efficacy (and/or maternal compliance) by measuring maternal estriol (here unconjugated plasma levels) during the second semester of pregnancy.

an NC/compound heterozygote (couple B in Fig. 3). When the mother (or the father) is affected with a classic form of CAH, the couple is at risk for having a child with the same disease only if the partner is heterozygote for a severe mutation (couple C in Fig. 3) but never if the partner has a mild mutation (couple H in Fig. 3).

When the mother is CAH affected, the treatment protocol is a somewhat different situation in that she undergoes a substitutive treatment involving glucocorticoids (hydrocortisone or dexamethasone) and mineralocorticoids most of the time. It is important to determine the risk for the fetus in the couple. If the couple is found to be at risk for the classic form, maternal treatment is adapted or changed for the current protocol of prenatal treatment (total daily dose of 20  $\mu\text{g}/\text{kg}$  body weight of dexamethasone divided into two or three doses) with no change in mineralocorticoids. However, if the couple is at risk only for an NC form, maternal treatment does not need to be changed.

### Informed Consent

Informed consent must be obtained from mothers participating in prenatal treatment with dexamethasone. Parents are asked to give informed consent for both treatment and protocols. Yet is their consent really informed? Informed consent, as it relates to the ethical principle of respect for autonomy, presents nurses' perceptions of informed consent, and there are variations in national rules about prenatal diagnosis and/or treatment. For example, in the United States,

the function of the institutional review board (IRB) for CAH is vital in research centers that conduct prenatal treatment for 21-OH deficiency. In France, informed consent should also by law include signed documents by centers that are empowered to conduct both genetic counseling and prenatal diagnosis as well as prenatal treatment.

However, it seems reasonable to say that the basis of informed consent includes the following:

- Parents must be appropriately and adequately informed in lay language.
- The scope and detail of the information supplied should be based on a reasonable parent's need to know rather than on the actions of a reasonable doctor.
- The doctor must take care to ensure that the information is understood by the parents and must use a proper method (explaining the "medical word" used by the "doctors") to present information.
- The benefit/risk ratio should be clearly explained in the only situation where beneficence is appropriate (CAH-affected female), as should the lack of benefit to males and unaffected females.
- The risk of complications with CVS and amniocentesis (0.6–0.8%) should also be included, although most institutions will have separate consent forms for those procedures.

A request by an individual to withdraw from the study must be honored, and participants should come from diverse socioeconomic and ethnic backgrounds. However, if prenatal treatment is started and the mother decides to interrupt pregnancy, this can be viewed as a failure of prenatal counseling.

The parents also have rights to confidentiality, and after they are informed, they have the right to freely exercise their choice and to opt not to have prenatal treatment and to have only prenatal diagnosis.

### HAVE PRENATAL DIAGNOSIS AND TREATMENT ACCOMPLISHED THEIR GOAL?

It is possible to answer this question after many years of experience in the field. Prenatal diagnosis is accurate, and prenatal treatment is effective. Approximately 70% of prenatally treated females are born with normal or only slightly virilized genitalia. The degree of success is variable according to the series (60–85%), but it is linked to maternal compliance and the rationale in the protocol used (e.g., divided daily doses). Treatment failures (i.e., affected females requiring genital reconstruction) have been attributed to late

onset of treatment, cessation of therapy during mid-gestation, noncompliance, and/or suboptimal dosing, whereas others had no ready explanation.

No teratogenic effect has been reported. The infants in large series are born at term with normal length, weight, and head circumference. However, there has been controversy concerning the efficiency of treatment, the estimation of fetal adrenal suppression, and the safety of treatment for the mother and child.

### Therapeutic Risks to the Mother

The incidence of maternal complications has varied among investigators. Overall, it is about 10%. In evaluating such data, it is important to correlate the types of adverse side effects observed and the duration of therapy. Both American and European investigators have found a higher incidence of side effects in women treated from the first through third trimesters. In the larger American study, no statistical difference between treated and nontreated women was noted except for weight gain, edema, and striae, which were greater in the treated groups and most often resolved with discontinuation of treatment. In the Scandinavian study, a significant increase in weight gain was observed during early pregnancy when treatment was initiated, but this initial rapid weight gain declined during late pregnancy or when treatment was terminated. Similar findings were observed in the largest European study except in three women, as in a small three-case American study in which severe side effects were reported: Cushingoid features, excessive weight gain, severe striae, hypertension, and hyperglycemia. Most of these women were overweight, prediabetic, and at risk for glucocorticoid exposure. In fact, when considering some of the side effects that have been related to glucocorticoid treatment, one might remember that during pregnancy there is a hypercortisolism state. Many pregnant women with no glucocorticoid treatment do have side effects that include marked weight gain, Cushing's signs, striae, edema, pre-eclampsia, and fatigue. In any event, weight, blood pressure, and glucose tolerance should be monitored closely in all treated women to term. These side effects notwithstanding, many parents of affected girls still opt for prenatal medical treatment due to the severe psychological impact of ambiguous genitalia on the children and families, and in the European study several mothers opted for two or three successive treated pregnancies.

Maternal urinary estriol measurements have also been suggested as a guide to adjusting maternal treatment (Fig. 7). A gradual decrease in the dose of

dexamethasone later during gestation might decrease the incidence of maternal side effects, but there are only little data suggesting the possible efficacy of such a regimen.

In the American study, fetal wastage was not statistically different between dexamethasone-treated and untreated patients. In the large European study, the incidence of miscarriages according to age was less than in the general population and was less than in the same mothers who had miscarriages previously during untreated pregnancies. In addition, there were no maternal side effects at delivery or during the postpartum period. All babies were full term.

These findings reinforce our position as to the contraindication to prenatal treatment in mothers at risk for diabetes or having diabetes, hypertension, obesity, insulin resistance, or cardiovascular disease. Most of the time, these are not good conditions for having babies because all of these situations are aggravated by glucocorticoids.

### Therapeutic Risks to the Fetus

In the European study, there were no side effects in most of the fetuses. Intrauterine growth retardation (IUGR) and unexplained fetal death (several months after stopping treatment) have been observed in less than 2% of treated pregnancies, a figure similar to that in the general population. IUGR was seen in only two of the children, and three neonates experienced failure to thrive, although they did catch up in growth after a while. Observed congenital malformations, such as cardiac septal hypertrophy, hydrometrocolpos, and hydrocephalus, were seen in 3 instances among 600 to 800 treated pregnancies; it is difficult to ascertain whether these were due to the treatment.

### Therapeutic Risks to the Neonates

In all studies, the dexamethasone-treated fetuses demonstrated normal pre- and postnatal growth compared with matched controls. However, in the Swedish study, several adverse events, such as failure to thrive and delayed psychomotor development, were reported among the treated infants. The three neonates who experienced failure to thrive did catch up in growth after a while.

On the whole, the short-term outcome appears to be rather successful. However, the doses used are thought to be supraphysiological given that dexamethasone is not metabolized rapidly by the placenta, as is cortisol into cortisone (11 $\beta$ -HSD). Also, the mechanism of

dexamethasone action in the fetus is unclear, and no long-term follow-up studies have been done.

## OTHER EFFECTS OF PRENATAL TREATMENT

We believe that prenatal treatment has other beneficial effects as well. In CAH-affected females, prenatal treatment not only prevents sexual ambiguity but also may prevent the consequences of the *in utero* exposure to androgens that might be deleterious for brain programming and induce polycystic ovary syndrome (PCOS). In both sexes, prenatal treatment prevents adrenal hypertrophy. Thus, the postnatal treatment with hydrocortisone is easier and more rapid to adjust at smaller doses.

Is prenatal treatment susceptible to long-term effects in unaffected fetuses? To be effective, the treatment would need to be started by week 6 of gestation, but the genetic diagnosis cannot be made until week 11 or 12. If the mother has had a previous CAH child, only one of four pregnancies will be affected and only female fetuses would benefit from treatment; thus, seven of eight fetuses will be treated needlessly.

In view of these and other concerns, the prenatal treatment of CAH remains an experimental therapy; hence, it must be done only with fully informed consent in controlled prospective trials approved by human experimentation committees at centers that see enough of these patients to collect meaningful data.

From reports of animal studies in the literature, risks for growth retardation, brain abnormalities, and/or hypertension during adulthood are questionable. The question of whether there are long-term side effects in unaffected human infants who were submitted to shorter prenatal treatment nowadays is more difficult to answer. A preliminary study did not show any side effects, but again, it is very difficult to conduct such a study because a control group is not easily available. At this time, we cannot answer such questions because we have to follow these patients or these individuals for at least 20 years. This is why it is believed that a registry should be open for those treatments. Others emphatically disagree with a recent suggestion that prenatal treatment is so experimental as to require approval by institutional review boards.

## NEW HOPES IN EARLY FETAL SEXING

It has been known for at least a decade that fetal cells are found in maternal blood. Fetal sex prediction can be achieved using PCR targeted at the SRY gene by

analyzing cell-free fetal DNA in maternal serum. Unfortunately, the results reported show a lack of sensitivity, especially during the first trimester of pregnancy. A new, highly sensitive, real-time PCR was developed to detect an SRY gene sequence in maternal serum during the first trimester (6–11 weeks pregnancy). No false-negative results were observed. Furthermore, no false-positive results occurred, although 27 women who carried a female fetus during their current pregnancies had at least one previous male-bearing pregnancy. Thus, analysis of cell-free fetal DNA in maternal plasma for fetal sex determination would abolish the need for corticosteroid administration and CVS in women with a male fetus at risk for 21-OH deficiency. Then, only three of eight fetuses at risk would undergo unnecessary treatment.

## LONG-TERM SIDE EFFECTS OF CAH FEMALES AT RISK

In Lyon, France, we have started a small long-term study with regular follow-up and found normal fundi, normal head circumference (as a testimony of brain development), normal growth, and normal school performance. A control group is very difficult to establish because CAH females are treated with glucocorticoids after birth anyway.

## CONCLUSION

Prenatal treatment is efficient in preventing sexual ambiguity in the female CAH-affected individuals. All cases treated should be followed closely and reported. It is of utmost importance that prenatal treatments be in the hands of very specialized teams of clinicians, biologists, and geneticists.

The main benefit of prenatal therapy is to ameliorate potential genital ambiguity in affected females, and it does so successfully. The risks of unnecessarily treating unaffected pregnancies, which seem small at this time, might not be fully elucidated for many years. Prenatal treatment must be done under careful, centralized, and (ideally) long-term medical supervision.

A long-term European study (PREDEX), piloted by the Swedish group, has been started, and we will have some answers in a couple of decades or so.

## See Also the Following Articles

Adrenarche, Premature • Beckwith-Wiedemann Syndrome (BWS) • Congenital Adrenal Hypoplasia Syndromes • Congenital Lipoid Adrenal Hyperplasia • Gender Assignment

and Psychosocial Management • Growth and Glucocorticoids • 11 $\beta$ -Hydroxylase Deficiency • 21-Hydroxylase Deficiency, Genetics of • 3 $\beta$ -Hydroxysteroid Dehydrogenase Deficiency • Newborn Ambiguous Genitalia Management

### Further Reading

- David, M., and Forest, M. G. (1984). Prenatal treatment of congenital adrenal hyperplasia resulting from 21-hydroxylase deficiency. *J. Pediatr.* **105**, 799–803.
- Forest, M. G., Bétuel, H., and David, M. (1989). Prenatal treatment in congenital adrenal hyperplasia due to 21-hydroxylase deficiency: Update 88 of the French Multicentric Study. *Endocrine Res.* **15**, 277–301.
- Forest, M. G., and Dörr, H. G. (2003). Prenatal therapy in congenital adrenal hyperplasia due to 21-hydroxylase deficiency: Retrospective follow-up study of 253 treated pregnancies in 215 families. *The Endocrinologist* **13**, 252–259.
- Forest, M. G., Morel, Y., and David, M. (1998). Prenatal treatment of congenital adrenal hyperplasia. *Trends Endocrinol. Metab.* **9**, 284–289.
- Guidollet-Tardy, V. (2002). Le déficit en 21-hydroxylase: Mise au point de l'exploration moléculaire de 1538 patients et 188 hétérozygotes—Corrélations génotype-phénotype. Ph.D. thesis, University of Lyon I.
- Kay, H. H., Bird, I. M., Coe, C. L., and Dudley, D. J. (2000). Antenatal steroid treatment and adverse fetal effects: What is the evidence? *J. Soc. Gynecol. Invest.* **7**, 269–278.
- Lajic, S., Wedell, A., Bui, T. H., Ritzen, E. M., and Holst, M. (1998). Long-term somatic follow-up of prenatally treated children with congenital adrenal hyperplasia. *J. Clin. Endocrinol. Metab.* **83**, 3872–3880.
- New, M. I., Carlson, A., Obeid, J., Marshall, I., Cabrera, M. S., Goseco, A., Lin-Su, K., Putnam, A. S., Wei, J. Q., and Wilson, R. C. (2001). Prenatal diagnosis for congenital adrenal hyperplasia in 532 pregnancies. *J. Clin. Endocrinol. Metab.* **86**, 5661–5667.
- Trautman, P. D., Meyer-Bahlburg, H. F., Postelnek, J., and New, M. I. (1996). Mother's reactions to prenatal diagnostic procedures and dexamethasone treatment of congenital adrenal hyperplasia. *J. Psychosom. Obstetr. Gynaecol.* **17**, 175–181.
- White, P. C., and Speiser, P. W. (2000). Congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Endocrine Rev.* **21**, 245–291.



# Congenital Adrenal Hypoplasia Syndromes

Constantine Tsigos

*Hellenic National Diabetes Center, Athens, Greece*

## Glossary

**hormone insensitivity** Inability of the body to respond to the physiological action of a hormone.

**hypothalamic–pituitary–adrenal axis** The hormonal axis that regulates the production of adrenal steroid hormones, including glucocorticoids, androgens, and mineralocorticoids.

**hypoplasia** Atrophy; lack of normal development of an organ or a tissue.

**orphan receptor** A receptor for which no endogenous ligand has been identified as yet.

The adrenal cortex secretes steroid hormones, including glucocorticoids, mineralocorticoids, and androgens. The glucocorticoid cortisol is secreted by the cells of the intermediate zona fasciculata and its secretion is tightly regulated in a cascade that is called the hypothalamic–pituitary–adrenal axis. Glucocorticoids are crucial for the maintenance of metabolic, cardiovascular, and immune homeostasis. The mineralocorticoid aldosterone is produced by the outer adrenal zona fasciculata. This steroid helps to maintain water and electrolyte homeostasis and its secretion is under the principal control of the renin–angiotensin axis and is only weakly influenced by adrenocorticotrophic hormone (ACTH). Adrenal androgens with 19 carbon atoms, such as dehydroepiandrosterone, dehydroepiandrosterone sulfate, and androstenedione, are secreted by the inner zona reticularis and are under the control of ACTH.

## ADRENOCORTICOTROPIN INSENSITIVITY SYNDROMES

Familial isolated glucocorticoid deficiency is a form of potentially lethal hereditary unresponsiveness to ACTH that manifests as primary adrenal insufficiency, usually without mineralocorticoid deficiency. Affected children commonly present within the first 2 years of life with hyperpigmentation, recurrent hypoglycemia, chronic asthenia, and failure to thrive.

Typically, they have deficient production of cortisol and adrenal androgens, in the presence of markedly elevated ACTH levels, whereas renin and aldosterone levels are usually normal and responsive to activation of the renin–angiotensin axis. In some cases, the isolated glucocorticoid deficiency is accompanied by alacrima (lack of tears) and achalasia of the esophagus, a triad called the triple A syndrome.

## Isolated Glucocorticoid Deficiency

Familial isolated glucocorticoid deficiency is a rare autosomal-recessive disorder that manifests as primary adrenal insufficiency, usually without mineralocorticoid deficiency. From the first description of the syndrome in 1959, it became apparent that affected individuals suffer from a form of hereditary unresponsiveness to ACTH. Indeed, affected children commonly present within the first 2 to 3 years of life with hyperpigmentation, recurrent hypoglycemia that can lead to convulsions or coma, chronic asthenia, and failure to thrive or even death, if it remains undiagnosed. Biochemically, they have deficient production of cortisol and adrenal androgens, in the presence of markedly elevated ACTH levels. Circulating cortisol levels, which are undetectable in the majority of cases, do not increase with short or prolonged stimulation with pharmacological doses of ACTH. Aldosterone and adrenal androgen responses to ACTH are also lost in these patients, suggesting that the underlying unresponsiveness to ACTH is generalized. Renin and aldosterone levels are usually normal, however, and respond appropriately to activation of the renin–angiotensin axis by orthostasis, salt restriction, or furosemide-induced diuresis. Treatment of hereditary isolated glucocorticoid deficiency consists of glucocorticoid replacement therapy. Affected subjects achieve normal growth and development with steroid replacement and live an otherwise normal life.

Histological postmortem studies of the adrenal glands in affected patients have revealed that the

ACTH-dependent zonae fasciculata and reticularis are extremely atrophic, reduced to a narrow band of fibrous tissue. In contrast, the angiotensin II-dependent zona glomerulosa is relatively well preserved, suggesting that the defect is limited to the ACTH-dependent zonae of the adrenal cortex. The etiological involvement of the ACTH receptor gene in familial glucocorticoid deficiency has been proposed since the first description of the syndrome in the late 1950s, but it was only after the cloning of the ACTH receptor gene in 1992 that this hypothesis could be tested. The ACTH receptor [melanocortin receptor 2 (MC2)] consists of 297 amino acids and is encoded by an intronless gene mapped to the distal end of chromosome 18. It belongs to the distinct melanocortin subfamily of the G protein-coupled receptors that couple to G<sub>s</sub> and adenyl cyclase to generate cyclic AMP as a second messenger. More than 15 point mutations and frameshift mutations have been reported as homozygote or compound heterozygote mutations in different pedigrees with isolated glucocorticoid deficiency. These mutations are scattered throughout the ACTH receptor molecule and affect all aspects of receptor function.

Mutations within the coding region of the ACTH receptor gene have not been found in all clinically defined cases of hereditary isolated glucocorticoid deficiency. Furthermore, using pairs of polymorphic dinucleotide repeats that are localized in the same region of chromosome 18, to which the human ACTH receptor gene has been mapped, no apparent linkage was demonstrated between the disease and the ACTH receptor gene in the majority of these families. These findings suggest that the etiology of isolated glucocorticoid deficiency might be heterogeneous and that a gene(s) other than the ACTH receptor gene might produce the same phenotype. Reverse genetics and linkage analysis might provide the means to localize the putative gene(s).

### Triple A Syndrome

A subset of the patients with hereditary unresponsiveness to ACTH, in addition to hypocortisolism, develop alacrima (lack of tears) and achalasia of the esophagus (leading to difficulty in swallowing). This constellation of symptoms is referred to as triple A syndrome, first described in 1978. Low tear production, usually present from early infancy on, may be confirmed by the Shirmer test, whereas esophageal dysmotility can be demonstrated by barium swallow and/or endoscopic examination. In some cases, the

diagnosis of achalasia may precede the diagnosis of cortisol deficiency. Occasional patients may also develop a variable degree of mineralocorticoid deficiency. It has become apparent that progressive and variable neurologic impairment that involves both central and peripheral neurons is also frequently associated with the triple A syndrome. Neurological defects may include autonomic and peripheral neuropathy, ataxia, and mental retardation and may thus result in a severely disabling disease.

It had been originally proposed that the ACTH receptor may also be defective in the triple A syndrome, but no mutations were found in the entire coding region of this gene in several families with the triple A syndrome. Homozygote or compound heterozygote mutations were found in the AAAS gene on 12q13 in families with the triple A syndrome. AAAS codes for the WD repeat-containing protein ALADIN (alacrima–achalasia–adrenal insufficiency–neurologic disorder). It is of note that no AAAS mutations have been found in testing of several families with isolated glucocorticoid deficiency with a normal ACTH receptor gene, suggesting that isolated glucocorticoid deficiency without MC2 receptor gene mutations is not a *forme fruste* of the triple A syndrome. The function of the AAAS gene must be clarified before further information on the effect of the individual genetic defects can be acquired. ALADIN is postulated to be involved either in cytoplasmic trafficking, like other proteins with WD repeats, or in peroxisomal activities.

### X-LINKED CONGENITAL ADRENAL HYPOPLASIA

Congenital X-linked adrenal hypoplasia is a rare disorder that is characterized by primary adrenal insufficiency and hypogonadotropic hypogonadism. The gene responsible for this disorder, DAX-1 (for dosage-sensitive sex reversal critical region at the adrenal hypoplasia locus on the X chromosome), was localized on the short arm of the X chromosome (Xp21). DAX-1 gene encodes a 470-amino-acid member of the nuclear receptor superfamily with an unknown ligand (orphan receptor). This orphan nuclear hormone receptor has a novel DNA-binding domain that has a unique structure that does not resemble the classic zinc-finger DNA-binding domain of these receptors. The DAX-1 gene is expressed not only in the adrenal gland but also in most of the reproductive tissues, such as the hypothalamus, the pituitary gonadotropic cells, and the gonads, and it plays an

important role in the development and functions of both the adrenal glands and the gonads.

More than 60 different mutations in DAX-1 have been reported in patients with X-linked congenital adrenal hypoplasia. Missense mutations in the DAX-1 gene have been identified in the C-terminal ligand-binding domain, whereas frameshift or nonsense mutations have been described in the N-terminal domain, the majority being frameshift or nonsense mutations resulting in premature truncation of the DAX-1 protein.

Patients with DAX-1 mutations develop primary, complete adrenal insufficiency, presenting as salt-wasting and hypoglycemic convulsions in infancy or childhood, as a result of a failure of formation of the adrenal cortex. Plasma cortisol and aldosterone are at a low level and do not respond to exogenous ACTH administration. Because of early diagnosis and treatment of adrenal failure, an increasing number of patients with X-linked congenital adrenal hypoplasia survive. In these patients, hypogonadotropic hypogonadism arises later in life and is recognized as a universal feature of this syndrome. It presents as pubertal delay, possibly due to defective production of hypothalamic gonadotropin-releasing hormone (GnRH) and impaired responsiveness of pituitary gonadotropes to GnRH. Indeed, pulsatile GnRH has proved ineffective at inducing puberty in these patients. Nevertheless, spontaneous onset of puberty, but with incomplete pubertal development, has been reported. More than 10% of patients have bilaterally undescended testes at birth.

Administration of human chorionic gonadotropin (HCG) stimulates testosterone concentrations into the normal range in most patients. However, exogenous gonadotropins are ineffective at inducing spermatogenesis, probably reflecting a direct effect of DAX-1 mutations on Sertoli cell function. Interestingly, an I496S missense mutation in DAX-1 was found in a man who presented with incomplete hypogonadotropic hypogonadism in his twenties and mild adrenal failure. Functional studies showed that this mutation caused a partial loss of DAX-1 function, suggesting that partial loss-of-function mutations in DAX-1 can present with hypogonadotropic hypogonadism and covert adrenal failure in adulthood.

A remarkable discrepancy indeed exists in genotype-phenotype relationships. Thus, the same mutation resulted in two brothers with the complete syndrome, an unaffected grandfather, and a homozygote aunt with only hypogonadotropic hypogonadism and normal adrenal function. It is of note that female heterozygotes for DAX-1 mutations may present with delayed puberty but normal fertility.

## PRIMARY ADRENAL INSUFFICIENCY CAUSED BY STEROIDOGENIC FACTOR-1 MUTATIONS

Steroidogenic factor-1 (SF-1) is an orphan nuclear receptor, formally isolated as a pivotal factor in the tissue-specific expression of cytochrome P450 steroid hydroxylases, which are essential for the synthesis of steroid hormones. SF-1 is expressed in the adrenal glands and gonads and is essential for the development and function of these organs and for normal sexual differentiation. In the adrenal glands, SF-1 not only regulates the levels of cytochrome P450 steroid hydroxylases, but also modulates the expression of  $3\beta$ -hydroxysteroid dehydrogenase, steroidogenic acute regulatory protein (StAR), and the ACTH receptor. SF-1 is also expressed in the pituitary gland and hypothalamus, contributing to the differentiation of pituitary primordial cells into gonadotropes.

Two cases of SF-1 gene mutations have been described. The first mutation was reported in a 46,XY phenotypic female, who presented with primary adrenal failure during the first weeks of life with low circulating cortisol and aldosterone and high ACTH. Although the karyotype of the patient was 46,XY, she retained normal Müllerian structures and streak-like gonads, containing poorly differentiated seminiferous tubules and connective tissue. The patient had a heterozygote two-codon replacement in exon 3 of one of the SF-1 genes, causing substitution of glycine at amino acid 35 by glutamate in the DNA-binding domain of the protein, abolishing its DNA-binding activity. Pituitary gonadotropins responded to GnRH but testosterone did not respond to exogenous HCG administration, suggesting defective gonadal function. After introduction of estrogen and progesterone, the uterus grew and regular menstruation occurred. Based on this case, it seems that SF-1 is essential for sex determination, steroidogenesis, and reproduction. The second heterozygote SF-1 mutation was described in a 27-month-old 46,XX female with normal external genitalia and adrenal failure. Her pubertal development remains to be observed in future years. However, impaired ovarian development and/or steroidogenesis would be expected for this patient.

### See Also the Following Articles

ACTH (Adrenocorticotrophic Hormone) • Adrenal Insufficiency • Adrenarche, Premature • Congenital Adrenal Hyperplasia, Prenatal Diagnosis and Therapy • Congenital Lipoid Adrenal Hyperplasia • Constitutional Delay of

## Growth and Puberty (CDGP) • Gonadotropin-Releasing Hormone Deficiency, Congenital Isolated

## Further Reading

- Aschermann, J. C., Ito, M., Hindmarsch, P. C., and Jameson, J. L. (1999). A mutation in the gene encoding steroidogenic factor-1 causes XY sex-reversal and adrenal failure in humans. *Nat. Genet.* **22**, 125–126.
- Aschermann, J. C., Meeks, J. J., and Jameson, J. L. (2001). Phenotypic spectrum of mutations in DAX-1 and SF-1. *Mol. Cell. Endocrinol.* **185**, 17–25.
- Clark, A. J., and Weber, A. (1998). Adrenocorticotropic insensitivity syndromes. *Endocr. Rev.* **19**, 828–843.
- Pelet-Tullio, A., Salomon, R., Hadj-Rabia, S., *et al.* (2000). Mutant WD-repeat protein in triple A syndrome. *Nat. Genet.* **26**, 332–335.
- Reutens, A. T., Achermann, J. C., Ito, M., *et al.* (1999). Clinical and functional effects of mutations in the DAX-1 gene in patients with adrenal hypoplasia congenita. *J. Clin. Endocrinol. Metab.* **84**, 504–511.
- Sandrini, F., Farmakidis, C., Kirschner, L. S., *et al.* (2001). Spectrum of mutations of the AAAS gene in Allgrove syndrome. Lack of mutations in six kindreds with isolated resistance to corticotropin. *J. Clin. Endocrinol. Metab.* **86**, 5433–5437.
- Tsigos, C. (1999). Isolated glucocorticoid deficiency and ACTH receptor mutations. *Arch. Med. Res.* **30**, 475–480.
- Tsigos, C., and Chrousos, G. P. (1999). ACTH resistance syndromes. In “G Proteins, Receptors and Disease” (A. M. Spiegel, ed.), pp. 261–269. Humana Press, Totowa, NJ.
- Tsigos, C., and Chrousos, G. P. (2002). Hypothalamic–pituitary–adrenal axis, neuroendocrine factors and stress. *J. Psychosom. Res.* **53**, 865–871.



# Congenital Lipoid Adrenal Hyperplasia

Delphine Mallet and Yves Morel  
Hôpital Debrousse, Lyon, France

## Glossary

**congenital lipoid adrenal hyperplasia** One of the rarest forms of congenital adrenal hyperplasia (CAH). Adrenal hyperplasia is due to adrenocorticotropin hypersecretion, which is common to all defects of cortisol biosynthesis. The term lipoid refers to the accumulation of lipids in adrenal cells, specific to this form of CAH.

**CYP11A** Gene encoding the cytochrome P450<sub>scc</sub>, a steroidogenic enzyme member of a large family of cytochromes.

**cytochromes** Heme-containing enzymes responsible for protein oxidation, involved in either biosynthesis or degradation. All steroidogenic enzymes belong to the cytochrome family, except for 3 $\beta$ -hydroxysteroid dehydrogenase and 17 $\beta$ -hydroxysteroid dehydrogenase.

**male pseudohermaphroditism** Defect of virilization of external genitalia in 46,XY individuals with well-differentiated testes.

**StAR** steroidogenic acute regulatory protein: protein implicated in mitochondrial cholesterol transfer, whose gene is mutated in most cases of congenital lipoid adrenal hyperplasia.

Congenital lipoid adrenal hyperplasia is a rare autosomal recessive disease caused by defective production of all adrenal and gonadal steroids, responsible for severe adrenal insufficiency and female phenotype in both sexes. Impairment of steroidogenesis is due to the inability of affected cells to convert cholesterol into pregnenolone.

## BIOLOGICAL AND CLINICAL DATA FOR CONGENITAL LIPOID ADRENAL HYPERPLASIA

Congenital lipoid adrenal hyperplasia was first described by Andrea Prader in 1955 and is characterized by severe adrenal failure and female phenotype in both sexes (Fig. 1). Production of all adrenal and

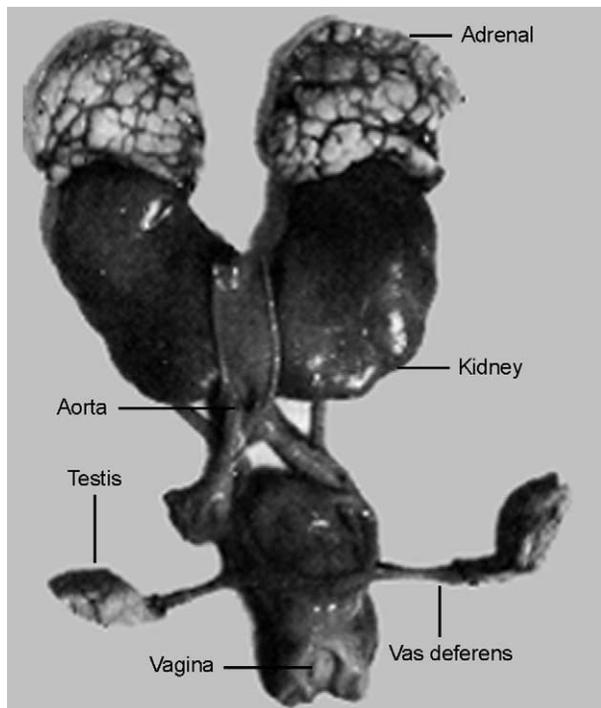
gonadal steroids is severely impaired and can not be stimulated by adrenocorticotropin (ACTH) or human choriogonadotrophic hormone. Mineralocorticoid and glucocorticoid defects are potentially lethal within the first days or weeks after birth in the absence of hormonal replacement therapy because of severe salt loss, vomiting, and dehydration. 46,XY patients present with male pseudohermaphroditism (female phenotype) due to defective testosterone production. Because testicular differentiation is normal, anti-Müllerian hormone is physiologically secreted, which explains the absence of Müllerian-derived structures, such as the uterus or fallopian tubes, in these patients. Skin hyperpigmentation at birth, which occurs in a large number of patients, is due to ACTH hypersecretion during fetal life.

Adrenal enlargement due to the accumulation of lipid droplets is not systematically present or detected by imaging studies. Some symptoms vary, such as age at apparition of first symptoms and degree of hyponatremia. Respiratory distress and hypoglycemia have also been described.

## FIRST STEP OF STEROIDOGENESIS

Steroids are all produced from a common precursor, cholesterol. The first enzymatic step of steroidogenesis produces pregnenolone by cleaving the cholesterol side chain (Fig. 2). Cytosolic cholesterol migrates from the outer to the inner mitochondrial membrane, where an enzymatic complex achieves cholesterol hydroxylation and side chain cleavage by transferring electrons and H<sup>+</sup> ions from P450<sub>scc</sub> to cholesterol.

This complex is an electron-transfer chain in which NADPH gives electrons to a flavoprotein called adrenodoxin reductase, which transfers them to an iron-sulfur protein called adrenodoxin. Finally, adrenodoxin transfers electrons to P450<sub>scc</sub> (Fig. 3). Cholesterol has different origins: One part is endogenously synthesized from acetyl-CoA, another part derives from the low-density lipoproteins, and the last



**Figure 1** The adrenals and the urogenital tract of a 46,XY individual with congenital lipoid adrenal hyperplasia. Note adrenal enlargement and circumvolutions due to lipid accumulation and the weak virilization of the internal genitalia, vas deferens, and epididymis. From Prader (1955).

part derives from lipid droplets in which cholesterol is stored as cholesterol esters.

## SEARCHING FOR CANDIDATE GENES

For a long time, it was thought that congenital lipoid adrenal hyperplasia was due to an enzymatic defect, like other forms of CAH. Cytochrome P450<sub>scc</sub> was first suspected, but no mutation of the *CYP11A* gene encoding P450<sub>scc</sub> was identified in several patients studied. This hypothesis was discarded, although it has been demonstrated that in rabbit congenital lipoid adrenal hyperplasia is due to a deletion of the *CYP11A* gene. Furthermore, it was demonstrated that progesterone production is essential for maintaining mammalian pregnancy and avoiding spontaneous abortion.

During pregnancy, progesterone can be produced either by the corpus luteum of the mother's ovary or by the placenta. In rabbit, the female's ovary assumes progesterone production throughout pregnancy, whereas in humans, it produces progesterone only during the first trimester. Therefore, placental progesterone production is necessary to maintain pregnancy in humans. Because the placenta is a fetal

tissue, a fetus homozygous for a mutation abolishing all P450<sub>scc</sub> activity would abort. In contrast, in rabbit, a female's ovary would continue to produce progesterone, avoiding abortion. Thus, it was determined that the *CYP11A* gene defect could not be responsible for congenital lipoid adrenal hyperplasia in humans, and that the candidate gene is not expressed in the placenta.

Genes encoding adrenodoxin and adrenodoxin reductase were also candidates, but no mutations were found in several patients. This result was not surprising because these proteins are implicated in other enzymatic systems. Therefore, research has been focused on proteins implicated in cholesterol transfer from the cytosol to the mitochondria.

Genes encoding sterol carrier protein-2, its precursor, a 78-amino acid peptide produced from the C-terminal end of the glucose-related protein called GRP-78, the protein presumably implicated in mitochondrial transport of cholesterol, and the benzodiazepine receptor and its presumed ligand, endozepine, became good candidate genes. Unfortunately, no mutation of these genes was detected. Finally, in 1994, Stocco and colleagues identified a protein in mouse Leydig MA-10 cells that was specifically expressed in gonads and adrenals but not in placenta. Its synthesis was rapidly induced by ACTH via cAMP, and this protein was able to increase steroidogenesis when it was transfected in MA-10 cells, accounting for the acute steroid production in response to hormonal stimulation. Therefore, this protein was named the steroidogenic acute regulatory protein (StAR). This protein seemed to be a good candidate to explain the pathogenesis of congenital lipoid adrenal hyperplasia, which has been further confirmed by the detection of mutations of StAR genes in several patients.

## STRUCTURE AND FUNCTION OF THE StAR PROTEIN

The StAR protein is 285 amino acids long. The first 26 amino acids of the N-terminal end are presumably implicated in mitochondrial signaling. StAR is expressed in adrenal and gonads but not in the placenta. The mechanism by which StAR allows cholesterol transfer is not clearly understood. Two models exist. The first model was deduced from the study of StAR deletion mutants. StAR lacking the 62 N-terminal amino acids has been shown to be active, whereas it could not enter the mitochondria due to the lack of the signaling sequence. On the other hand, mutants deleted in the C-terminal end lose activity. Recently, a protein

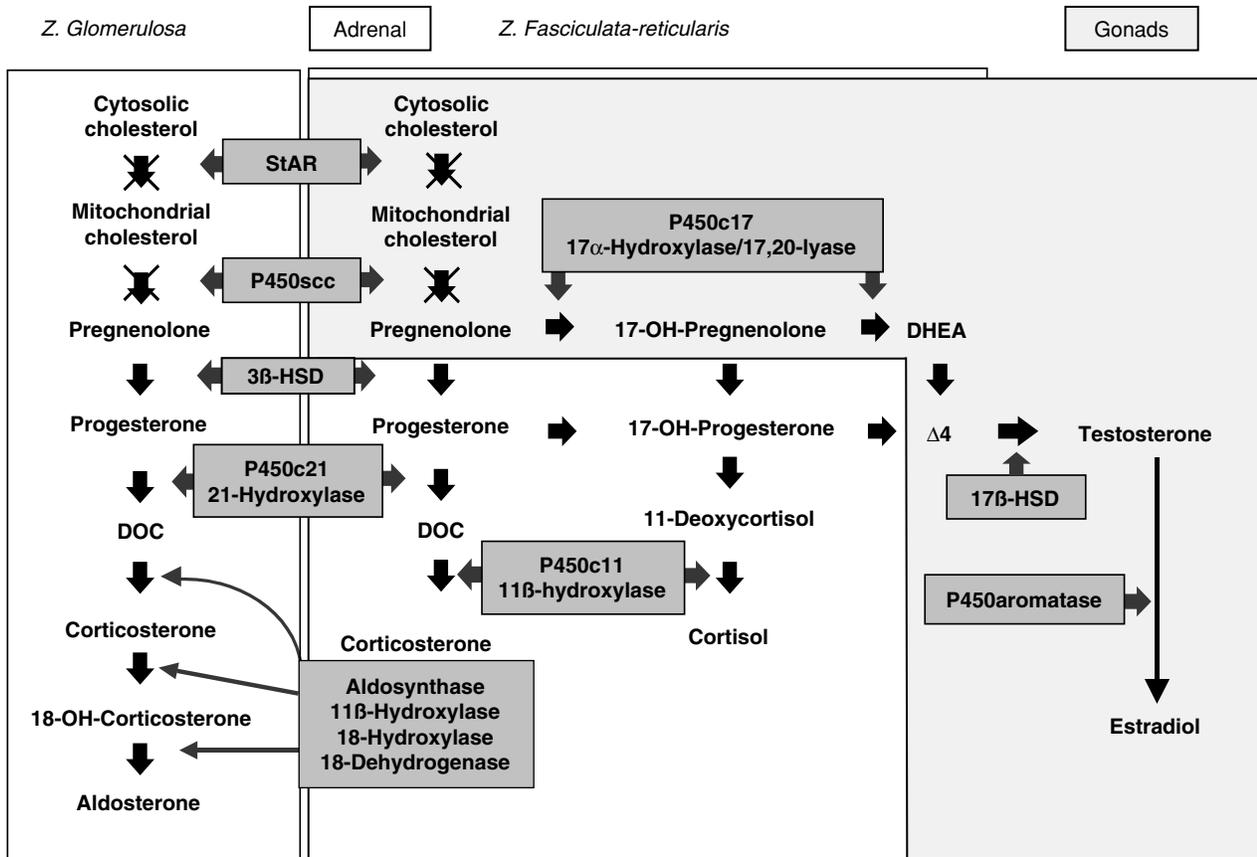


Figure 2 Adrenal and gonadal steroidogenesis. The adrenal contains two zones—the zona glomerulosa, which produces mineralocorticoids, and the zona fasciculata-reticularis, which produces glucocorticoids. Gonads make only sex steroids, using preferentially the  $\Delta 5$  pathway.

called MLN64, with an amino acid sequence similar to that of the C-terminal end of StAR, has been identified in placenta. This protein is able to enhance steroidogenesis *in vitro*. Therefore, MLN64 seems to play the StAR role in the placenta, according to the hypothesis that the defective protein in congenital lipoid adrenal hyperplasia is not expressed in this tissue. Similarities between StAR and MLN64 C-terminal ends support the importance of this region. Sequences homologous to the StAR C-terminal region were determined to be a new domain called START (StAR-related lipid transfer domains) because they are able to bind lipids. Study of the tertiary structure of this domain has shown that it forms a hydrophobic tunnel, to which a molecule of cholesterol can bind. Therefore, the first model proposes that cholesterol enters the mitochondria through this hydrophobic tunnel.

The second model was deduced from observations that StAR could form a structure called molten globule at pH 3.5–4. Molten globules consist of proteins that lose some of their tertiary structure but retain their

secondary structure. This model hypothesizes that a StAR conformational change induces a modification of the outer mitochondrial membrane structure, allowing cholesterol transfer.

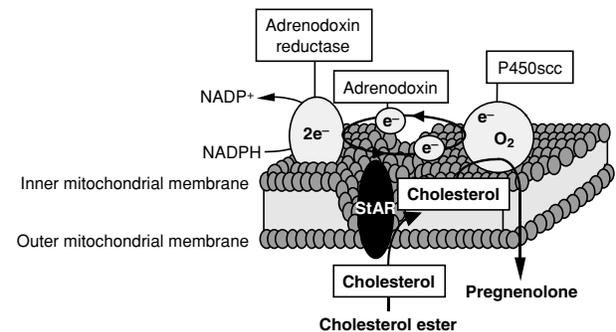


Figure 3 Schematic representation of the first step of steroidogenesis. StAR transfers cholesterol from the outer to the inner mitochondrial membranes and to the side chain cleavage complex—adrenodoxin reductase, adrenodoxin, and cytochrome P450scc. The cholesterol side chain is then cleaved to form pregnenolone.

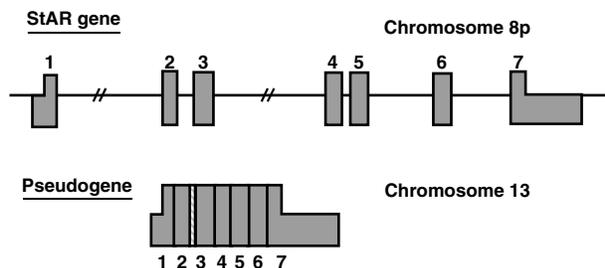


Figure 4 StAR gene and pseudogene.

Some issues remain unresolved. For example, only one molecule of cholesterol can pass through the hydrophobic tunnel at a time, which seems inconsistent with the acute steroidogenic response to stimulation. Concerning the molten globule hypothesis, it is not clear how the pH around StAR becomes acidic. Nevertheless, these two models are at the present time the most relevant.

## CONGENITAL LIPOID ADRENAL HYPERPLASIA AND THE StAR GENE

Congenital lipoid adrenal hyperplasia has an autosomal recessive mode of inheritance and is essentially due to StAR gene mutations. The gene encoding StAR is located on chromosome 8p, 8 kb long, and composed of seven exons. Amplification of the StAR gene requires the use of primers located in introns due to the existence of a pseudogene on chromosome 13 that does not have introns and contains a large number of rearrangements (Fig. 4). Approximately 30 StAR gene mutations have been described in 57 patients (Table I). Two mutations are more frequent in specific ethnic groups: the *Q258X* mutation in the Japanese population and the *R182L* mutation in the Palestinian population. By studying the clinical data of the 57 patients described as having StAR gene mutations, it appears that there are variations in age at onset of first symptoms. Adrenal insufficiency usually occurs in the first few days after birth, but some patients underwent adrenal insufficiency later and survived without substitutive hormonal treatment for several weeks or months. Differences also exist in gonadal lesions between sexes. Testicular lesions appear early in the development of the fetal testis, as demonstrated by the existence at birth of female phenotype in 46,XY patients.

In contrast, ovarian damage seems to be less severe until puberty since spontaneous breast development at puberty has been observed in some affected girls. A model describing the consequences of StAR gene mutations has been proposed to explain these differences:

the two-hit model. The absence of cholesterol transfer from the outer to the inner mitochondrial membrane results in no response of StAR-dependent steroidogenesis to stimulation, causing the first hit. However, a system exists that allows production of small amounts of steroids by a StAR-independent mechanism that is not clearly understood. Second, accumulation of unmetabolized cholesterol becomes toxic for cellular metabolism, leading to cell death. Because testes are highly stimulated early in fetal life to induce testosterone biosynthesis, engorgement of testicular cells by unmetabolized cholesterol occurs rapidly and leads to cell destruction, which explains the absence of virilization in 46,XY fetus. In contrast, the ovary is not really stimulated before puberty. During the short period preceding engorgement, affected cells are able to synthesize small amounts of estrogens, allowing some degree of feminization at puberty.

Also interesting with regard to congenital lipoid adrenal hyperplasia is the discrepancy between the weak virilization of the vas deferens and epididymis and the absence of virilization of external genitalia described in some 46,XY patients (Fig. 1) and in StAR knockout mice. This suggests that the testicular StAR-independent synthesis of steroids should be sufficient to weakly virilize and develop the nearest internal genitalia by a mechanism of local diffusion. External genitalia, which are more distant from the testes, do not receive enough androgens to be virilized. This dissociated action of androgens has also been described in 17 $\beta$ -hydroxysteroid dehydrogenase deficiency, but in this case the testosterone derives from the local conversion of precursors achieved by a peripheral isoenzyme.

Although StAR gene mutations have been identified in most patients with congenital lipoid adrenal hyperplasia, 1 patient has been reported without a StAR gene mutation and 2 patients have been reported to be heterozygous. We were unable to identify any StAR gene mutations in 5 of 13 Caucasian patients studied. Research on new candidate genes had been unfruitful until recent reports of *CYP11A* gene mutations in 2 unrelated patients.

## CONGENITAL LIPOID ADRENAL HYPERPLASIA AND THE CYP11A GENE

In 2001, the role of the *CYP11A* gene was revisited after the identification of *CYP11A* gene mutations in two patients. One patient is heterozygous for a *de novo* 6-base pair insertion. *In vitro* study showed that mutant protein is totally inactive, suggesting that

Table 1 Clinical and Biological Data from Patients with Congenital Lipoid Adrenal Hyperplasia Carrying StAR Gene Mutations<sup>a</sup>

Patient No.	Ethnic origin <sup>b</sup>	Age at onset of first symptoms	Age at beginning of treatment	Adrenal	External genitalia	Glucose (nM)	Na (nM)	K (nM)	Respiratory problems	StAR gene mutations
<b>46,XY patients</b>										
1	Japanese	Birth			Female	0.77	139 mEq/liter	4.6 mEq/liter		<i>A218V/R217T</i>
2	Vietnam	8 weeks	10 weeks	Enlarged	Female	5.11	118	8.9		<i>IVS4-11T&gt;A</i> homozygous
3	Japanese	19 days	19 days		Female					<i>Q258X/Q258X</i>
4	Japanese	1 month	2 months		Female	5.11	137	5.8		<i>Q258X/247insG</i>
5	Japanese	4 days	3 months		Female		130	6.7		<i>E169K/Q258X</i>
6	Palestinian	1 day	1 day		Female	3.11	124	7.8	Yes	<i>E169G/E169G</i>
7	Palestinian <sup>†</sup>	5 days	5 days		Female		133	6.6		<i>R182L/R182L</i>
8	Palestinian	7 days	12 days		Female	0.72	120	8		<i>R182L/R182L</i>
9	Palestinian	10 days	22 days		Female	3.9	123	8.1	Yes	<i>650delC, 593delTT/R182L</i>
10	Palestinian	2 weeks	3 weeks		Female	2	126	6.8	Yes	<i>593delTT/R182L</i>
11	Mexican	8 days	6 weeks		Female	0.6	126	5.6		<i>ΔR272/ΔR272</i>
12	British		6 months		Female	4.9	118	5.9		<i>insDNA/ND</i>
13	Canadian	2 months	4 months		Female	3.5	123	6.9		<i>A218V/L275P</i>
14	Danish	3 weeks	2 months		Female		107	7.3		No mutation
15	Egyptian	1 week	16 days		Female		124	8.9	Yes	<i>548insTT/548insTT</i>
16	Korean <sup>‡</sup>	10 days			Female	normal	113	9		<i>Q258X/Q258X</i>
17	Korean <sup>‡</sup>									<i>Q258X/Q258X</i>
18	Japanese	4 weeks	4 weeks		Female		128	6.9		<i>Q258X/Q258X</i>
19	American	1 day	9 days		Female	0.72	128 <sup>d</sup>	7.1	Yes	<i>R193X/R193X</i>
20	Japanese		31 days							<i>Q258X/189delG</i>
21	Japanese		14 days							<i>246insG/246insG</i>
22	Japanese		30 days							<i>246insG/A218V</i>
23	Japanese <sup>o</sup>		9 days							<i>Q258X/A218V</i>
24	Japanese		10 months		Clitoro-megaly					<i>Q258X/M225T</i>
25	Japanese		2 days							<i>Q258X/Q212X</i>
26	Japanese		14 days							<i>Q258X/564del13bp</i>
27	Japanese		13 days							<i>Q258X/A218V</i>
28	Japanese		14 days							<i>Q258X/Q258X</i>
29	Japanese		13 days							<i>Q258X/246insG</i>
30	Serbian	2-3 weeks		Normal	Female		131	9.6		<i>W250X/W250X</i>
31	Japanese				Female					<i>R182L/R182L</i>
32	Japanese				Female					<i>Q258X/163insA</i>
33	Japanese				Female					<i>Q258X/643insT</i>
34	Guatemala	2 months		Normal?	Partial posterior fusion	Hypo				<i>677insA/677insA</i>
35	Libyan	Birth			Female					<i>IVS1+2T&gt;G</i> homozygous
36	Katar	3 weeks			Female		122	6.1		<i>R182H/R182H</i>

(continues)

10005 Table 1 (continued)

Patient No.	Ethnic origin <sup>b</sup>	Age at onset of first symptoms	Age at beginning of treatment	Adrenal	Puberty	Glucose (nM)	Na (nM)	K (nM)	Respiratory problems	StAR gene mutations
<b>46,XX patients</b>										
37	Palestinian <sup>¶</sup>	1 day	5 months			Low	117 <sup>c</sup>	7.1	Yes	R182L/R182L
38	Greek	10 days	35 days				132	7.2		947insA/947insA
39	Korean <sup>‡</sup>	4 weeks								Q258X/Q258X
40	Japanese <sup>*</sup>		40 days		15 years					Q258X/838delA
41	Japanese <sup>*</sup>		8 days		12.4 years					Q258X/838delA
42	Japanese <sup>°</sup>		35 days							Q258X/A218V
43	Japanese <sup>†</sup>		30 days							Q258X/Q258X
44	Japanese <sup>†</sup>		7 days		12 years					Q258X/Q258X
45	Japanese		14 days		21 years					Q258X/Q258X
46	Japanese		60 days		20 years					Q258X/Q258X
47	Japanese		11 days							246insG/ND
48	Japanese		14 days		20 years					Q258X/Q258X
49	Korean		1 year?				136 mEq/ liter	4.5 mEq/ liter		IVS2+2insT homozygous
50	Indian <sup>#</sup>	4 days		Enlarged		Hypo				
51	Indian <sup>#</sup>	4 months	4 months		11.5 years	Low	109 mEq/ liter	5.9 mEq/ liter		261delT/261delT
52	Japanese	2 months	2 months		12–15 years		131 mEq/ liter	6.9 mEq/ liter		Q258X/Q258X
53	Japanese <sup>§</sup>		4 weeks		12 years					Q258X/838delA
54	Japanese <sup>§</sup>	8 days								Q258X/838delA
55	Japanese									Q258X/189delG
56	Israel	2 weeks								R193X/R193X
57	Japanese				12 years	Hypo				Q258X/Q258X
58	Katar	Birth								IVS1+2T>G homozygous

ND, no StAR gene mutation detected.

<sup>a</sup>Missing were not provided in original publications.

<sup>¶</sup>, <sup>‡</sup>, <sup>°</sup>, <sup>\*</sup>, <sup>#</sup>, <sup>§</sup>Patients with the same symbol are related.

<sup>c</sup>Na and K values are from Bose *et al.* (1996). Confirmation of these low Na values at day 1 was not provided in the original report.

this patient suffers from congenital lipoid adrenal hyperplasia because of *CYP11A* haploinsufficiency. The other patient is compound heterozygous for a missense mutation and a *de novo* splicing mutation. The mutant protein bearing the missense mutation retains some enzymatic activity, whereas the splicing mutation abolishes all activity. Thus, in these two cases, a residual activity during fetal life produced enough progesterone to avoid spontaneous abortion. We have also identified *CYP11A* gene mutations in three of five patients affected by congenital lipoid adrenal hyperplasia without StAR gene mutations. Studies of other patients with *CYP11A* gene mutations would help to determine the relationship between genotype and phenotype. Particularly, is the

detection of a heterozygous *CYP11A* gene mutation in a patient enough to explain this disease?

## CONCLUSION

The pathogenesis of congenital lipoid adrenal hyperplasia is more complex than that of other steroidogenic enzyme defects. This disorder is an enzymatic disease but also a storage disease, which explains the various phenotypic aspects. The majority of mutations do not affect the *CYP11A* gene encoding P450scc responsible for the enzymatic step, but they do affect the StAR protein, which is involved in cholesterol transport. The study of StAR in the pathogenesis of congenital lipoid adrenal hyperplasia has resulted in a better understanding

of some fundamental aspects of steroidogenesis: its role in transport of cholesterol, the existence of two mechanisms of steroids biosynthesis (one independent of StAR and the other StAR dependent) and the relationship between these two mechanisms, and the phenotype of affected patients. Recently, *CYP11A* gene mutations have been described in two patients with congenital lipoid adrenal hyperplasia, indicating that this disease may be due to either a StAR or a P450scc defect.

The predominance of StAR gene mutations versus *CYP11A* gene mutations may be explained by the fact that complete abolition of StAR activity is compatible with pregnancy, whereas complete P450scc inactivation is lethal for the fetus.

### See Also the Following Articles

Congenital Adrenal Hyperplasia, Prenatal Diagnosis and Therapy • Congenital Adrenal Hypoplasia Syndromes • Prader-Willi Syndrome

### Further Reading

- Bose, H. S., *et al.* (1996). The pathophysiology and genetics of congenital lipoid adrenal hyperplasia. *International Congenital Lipoid Adrenal Hyperplasia Consortium. N. Engl. J. Med.* **335**, 1870–1878.
- Clark, B. J., *et al.* (1994). The purification, cloning, and expression of a novel luteinizing hormone-induced mitochondrial protein in MA-10 mouse Leydig tumor cells. Characterization of the steroidogenic acute regulatory protein (StAR). *J. Biol. Chem.* **269**, 28314–28322.
- Katsumata, N., *et al.* (2002). Compound heterozygous mutations in the cholesterol side chain cleavage enzyme gene (*CYP11A*) cause congenital adrenal insufficiency in humans. *J. Clin. Endocrinol. Metab.* **87**, 3808–3813.
- Miller, W. L. (1997). Congenital lipoid adrenal hyperplasia: The human gene knockout for the steroidogenic acute regulatory protein. *J. Mol. Endocrinol.* **19**, 227–240.
- Prader, A. G. (1955). Das syndrom des pseudohermaphroditismus masculinus bei kongenitaler nebennierendehyperplasie ohne androgenuberproduktion (adrenaler pseudohermaphroditismus masculinus). *Helvetica Paediatr. Acta* **10**, 397–412.
- Stocco, D. M. (2001). Tracking the role of a star in the sky of the new millennium. *Mol. Endocrinol.* **15**, 1245–1254.



# Conn's Syndrome

Franco Mantero and Oliver Bedendo  
University of Padua, Padua, Italy

## Glossary

**aldosterone-producing adenoma** A benign adrenal tumor that produces aldosterone in excess; it is usually small (0.5–2.0 cm) and solitary.

**hypertension** This condition, when due to an excess of aldosterone, is characterized by water retention and an expansion in extracellular volume, resulting in low or suppressed plasma renin activity values and high absolute or relative plasma aldosterone levels.

**hypokalemia** The excessive preservation of sodium at the expense of potassium loss, as occurs in hyperaldosteronism; however, only < 40% of patients affected by primary aldosteronism are hypokalemic.

**primary aldosteronism** The most common cause of mineralocorticoid-based hypertension. It is due to an excess of aldosterone, usually caused by an aldosterone-producing adenoma (Conn's syndrome) or bilateral adrenal hyperplasia (idiopathic hyperaldosteronism).

First described in 1955 by Jerome Conn, Conn's syndrome is a clinical syndrome characterized by hypertension and hypokalemia, with low or suppressed plasma renin activity and increased aldosterone excretion, due to a circulating excess of aldosterone, autonomously produced by a benign adrenal tumor.

## EPIDEMIOLOGY

Idiopathic hyperaldosteronism due to bilateral adrenal hyperplasia and Conn's syndrome are the most common forms of primary aldosteronism. Conn's syndrome itself accounts for more than 50% of cases of primary aldosteronism.

Primary aldosteronism is the most common cause of mineralocorticoid-based hypertension; initial studies reported prevalence rates of less than 1% among hypertensive patients, but these were largely based on detection of unexplained hypokalemia. Since measurement of the plasma aldosterone concentration

to plasma renin activity ratio was introduced as a screening test and became widely used, higher prevalence rates of 2–12% have been reported, with normokalemic patients representing up to 60% of cases.

## PATHOPHYSIOLOGY

Elevated plasma aldosterone levels promote excessive preservation of sodium at the expense of hydrogen and potassium loss in the distal nephron. Sodium retention promotes water retention and an expansion in extracellular volume, resulting in hypertension and suppression of plasma renin.

In patients with an aldosterone-producing adenoma, hypokalemia can be absent or can be present in different grades, sometimes with metabolic alkalosis, mostly depending on the degree and duration of aldosterone excess.

## CAUSES

Conn's syndrome usually derives from a small (0.5–2.0 cm), solitary, aldosterone-producing adrenal adenoma. However, more rarely, the adenoma can be multiple or bilateral.

The etiology of Conn's syndrome is still unknown, even if there are rare cases with a genetic basis, as a component of multiple endocrine neoplasia type 1 or manifested in a glucocorticoid- or nonglucocorticoid-remediable form of familial hyperaldosteronism (familial hyperaldosteronism types I and II).

Typically, the tumor is yellow in color due to its high cholesterol content. Based on morphological and biochemical patterns, aldosterone-producing adenomas can be divided into two groups, as follows:

### Angiotensin II-Unresponsive Aldosterone-Producing Adenomas

Angiotensin II-unresponsive aldosterone-producing adenomas are predominantly composed of fasciculated-like cells. Aldosterone production is increased

when adrenocorticotropic hormone is administered, but does not respond to low-dose angiotensin II infusion or to a change from a supine to an upright position (postural test). The so-called “hybrid” steroids, 18-oxocortisol and 18-hydroxycortisol, are produced in excess. These tumors are more common among females.

### Angiotensin II-Responsive Aldosterone-Producing Adenomas

Angiotensin II-responsive aldosterone-producing adenomas are composed predominantly of glomerulosa-like cells. Aldosterone secretion is increased when low doses of angiotensin II are infused or after a change from a supine to an upright position (postural test). Expression of renin mRNA is typically increased in these adenomas.

### CLINICAL FEATURES

Clinical features (and laboratory findings) are usually more severe in Conn's syndrome than in other forms of primary aldosteronism. Moreover, patients are usually younger, with age at presentation typically being less than 50 years.

#### Hypertension

Similar to other forms of primary aldosteronism, hypertension is often resistant to antihypertensive drugs; however, patients with an aldosterone-secreting adenoma may have more severe hypertension.

#### Hypokalemia

In Conn's syndrome, hypokalemia may be more profound, with values of less than 3 mEq/liter, compared to all forms of primary aldosteronism.

If hypokalemia is severe, symptoms or signs related to metabolic alkalosis, such as tetany or positivity of Chvostek or Trousseau sign, symptoms or signs related to myopathy, such as weakness, or symptoms or signs related to nephrogenic diabetes insipidus, such as polyuria and nocturia, can be present.

The presence of an aldosterone-producing adenoma is possible even in normokalemic patients.

### INVESTIGATIONS

In patients affected by Conn's syndrome, plasma and urinary aldosterone levels are usually higher than in other forms of primary aldosteronism.

Diagnosis is made based on clinical suspicions (hypokalemia, refractory hypertension, young age) in a hypertensive patient, a screening test to reveal low or suppressed plasma renin activity and inappropriately high plasma aldosterone levels, a test to confirm autonomous aldosterone secretion with nonsuppressible plasma aldosterone levels after salt or water loading, and adrenal imaging to reveal the presence of the tumor.

Adrenal vein sampling is indicated to assess lateralized aldosterone excess from solitary adenoma. There remains a debate as to whether adrenal vein sampling should be performed in all cases of primary aldosteronism or only when all previous diagnostic studies are inconclusive.

### TREATMENT

Unilateral adrenalectomy is the treatment of choice for patients with an aldosterone-secreting adenoma, resulting in normalization of the serum potassium concentration in 100% of cases, a decrease in blood pressure values in the majority of cases, and a return to normal blood pressure values in 30–35% of cases, after successful excision. Laparoscopic adrenalectomy is as effective as open adrenalectomy, but with lower intra- and postoperative risks.

If surgical therapy is declined or if the patient is not an appropriate candidate for surgery, long-term medical therapy with mineralocorticoid receptor antagonists or epithelial sodium-channel blockers can be successfully undertaken.

### See Also the Following Articles

Conn's Syndrome, Diagnosis of • Hypertension, Endocrine • Hypertension, Overview • Hypertension, Renin and • Primary Aldosteronism (PAL) • Tissue Renin-Angiotensin-Aldosterone System • Toxic Adenoma

### Further Reading

- Conn, J. W. (1955). Primary aldosteronism, a new clinical entity. *J. Lab. Clin. Med.* **45**, 3–17.
- Fardella, C. E., et al. (2000). Primary hyperaldosteronism in essential hypertensives: Prevalence, biochemical profile, and molecular biology. *J. Clin. Endocrinol. Metab.* **85**, 1863–1867.
- Gordon, R. D., et al. (1995). Primary aldosteronism—Some genetic, morphological, and biochemical aspects of subtypes. *Steroids* **60**, 35–41.
- Stowasser, M. (2000). Primary aldosteronism: Revival of a syndrome. *J. Hypertension* **18**, 363–366.



# Conn's Syndrome, Diagnosis of

Franco Mantero and Oliver Bedendo  
*University of Padua, Padua, Italy*

## Glossary

**adrenal computed tomography** Noninvasive radiological investigation providing morphological data on the adrenals. A high-resolution scan with contiguous <5 mm cuts is usually sensitive enough to detect Conn's adenoma.

**adrenal vein sampling** Invasive radiological investigation consisting of cannulation of the adrenal veins and subsequent sampling with measurement of plasma cortisol and plasma aldosterone concentrations. It is considered the gold standard test to assess, or to exclude as a diagnosis, a lateralized excess of aldosterone from the adrenals.

**fludrocortisone (9 $\alpha$ -fluorohydrocortisone) suppression test** This investigation assesses whether suppression of plasma aldosterone by mineralocorticoids and sodium loading is present ("essential" hypertension) or absent (primary aldosteronism). It is considered the gold standard test to confirm primary aldosteronism when the plasma aldosterone concentration to plasma renin activity ratio is increased.

**plasma aldosterone concentration to plasma renin activity ratio** Screening test to select (when the ratio is increased) hypertensive patients that should undergo further investigations for primary aldosteronism. Plasma aldosterone concentration and plasma renin activity are measured when the patient is in an upright position.

**D**iagnosis of Conn's syndrome includes a screening test and a confirmatory test to assess inappropriately high and non-suppressible plasma aldosterone levels, radiological imaging to localize the adrenal tumor, and sometimes adrenal vein sampling to assess lateralization of aldosterone excess.

## SCREENING

All hypertensive patients with refractory hypertension or hypokalemia or who are younger than 50 years of age should be screened for Conn's syndrome.

Measurements of serum potassium or plasma or urinary aldosterone concentration can be useful, but they are not sensitive enough to confirm the diagnosis, as patients affected by primary aldosteronism may sometimes exhibit levels that are in the normal range. However, when an aldosterone-producing adenoma is present, plasma aldosterone and urinary aldosterone are usually higher than in other forms of primary aldosteronism, with values greater than 25 ng/dl and 30  $\mu$ g/24 h, respectively. The isolated measurement of plasma renin activity is also inadequate for screening, as it can show low values even in "essential" hypertensives or in patients with chronic renal failure.

A widely accepted, valid screening test is the measurement of the plasma aldosterone concentration to plasma renin activity ratio taken when the patient is in an upright position; the ratio is typically increased in primary aldosteronism.

Due to their influence on the renin-angiotensin system, antihypertensive medications should be discontinued for 2 weeks before the test is taken and anti-aldosterone drugs should be discontinued for at least 6 weeks prior to testing; in hypokalemic patients, potassium stores must be replenished for 2 weeks prior to the test. However, when complete cessation of medications may not be safe, the ratio seems to be a valid screening test even in patients still taking antihypertensive drugs, except for anti-aldosterone antagonists and beta-blockers.

Although the plasma aldosterone concentration to plasma renin activity ratio has a high level of sensitivity, specificity is poor, as renin levels are sometimes low simply due to ethnicity, age, or kidney disease. A suppression dynamic test to confirm an inappropriate and autonomous aldosterone secretion is thus needed.

## CONFIRMATION

A confirmatory test is usually performed after a positive screening test. The fludrocortisone suppression test, saline infusion test, and oral salt-loading test are

widely used to confirm inappropriately high plasma aldosterone levels, which are typically not suppressed by volume or salt loading in Conn's syndrome.

The fludrocortisone suppression test, possibly associated with oral salt loading, has been proposed as the gold standard, as it has the highest level of sensitivity of the available tests, but the patient may require hospitalization when the test is performed.

## LOCALIZATION

Radiological investigation is required to detect the adenoma. High-resolution adrenal computed tomography (CT) scanning with contiguous <5 mm cuts seems to be more sensitive than adrenal magnetic resonance imaging (MRI).

Adrenal scintigraphy with iodo-cholesterol and under dexamethasone suppression provides both morphological information and functional information, but sensitivity is poor for tumors smaller than 1.5 cm.

Despite the high degree of sensitivity of CT and MRI, a very small (less than 5 mm) adenoma could escape radiological detection. If the studies are inconclusive, with dubious adrenal imaging features, but there is a high suspicion of adenoma, patients should be considered for adrenal vein sampling.

## LATERALIZATION

Despite technical difficulties, such as cannulation of the right adrenal vein, and risks, such as adrenal vein hemorrhage or thrombosis, adrenal vein sampling is the only reliable test to assess lateralization of aldosterone excess.

Cortisol from each adrenal is measured to confirm the correct location of the cannula and the aldosterone to cortisol ratio of the two adrenal veins is compared to assess lateralization. If lateralization is present, contralateral adrenal aldosterone secretion is typically inhibited.

Due to the high prevalence of adrenal incidentalomas, it is still a matter of debate as to whether sampling should be performed in all cases of confirmed primary aldosteronism.

## DIFFERENTIAL DIAGNOSIS

Conn's syndrome should always be differentiated from other forms of primary aldosteronism, as this has important therapeutic implications.

## Conn's Syndrome versus Idiopathic Hyperaldosteronism

Since angiotensin II-responsive aldosterone-producing adenomas have been recognized, a postural test cannot be used to distinguish between adenoma and bilateral adrenal hyperplasia. Angiotensin II-unresponsive aldosterone-producing adenomas do not respond to a change from a supine to an upright position in terms of aldosterone production, whereas both bilateral adrenal hyperplasia and angiotensin II-responsive adenomas respond (resulting in an increase in plasma aldosterone of greater than 30%)

Negative imaging and lack of lateralization of adrenal vein sampling indicate a diagnosis of idiopathic hyperaldosteronism.

## Conn's Syndrome versus Unilateral Adrenal Hyperplasia

Unilateral adrenal hyperplasia, a rare case of primary aldosteronism with a lateralized aldosterone excess, shares the adrenal vein sampling result and therapeutic approach with Conn's syndrome. Radiological imaging can help in the differential diagnosis.

## See Also the Following Articles

Conn's Syndrome • Hypertension, Endocrine • Hypertension, Overview • Hypertension, Renin and • Primary Aldosteronism (PAL) • Tissue Renin-Angiotensin-Aldosterone System • Toxic Adenoma

## Further Reading

- Doppman, J. L., *et al.* (1996). Hyperaldosteronism: Sampling the adrenal veins. *Radiology* **198**, 309–312.
- Gordon, R. D., *et al.* (2001). Primary aldosteronism: Are we diagnosing and operating too few patients? *World J. Surg.* **25**, 941–947.
- Magill, S. B., *et al.* (2001). Comparison of adrenal vein sampling and computed tomography in the differentiation of primary aldosteronism. *J. Clin. Endocrinol. Metab.* **86**, 1066–1071.
- Rossi, G. P., *et al.* (2001). Identification of the etiology of primary aldosteronism with adrenal vein sampling in patients with equivocal computed tomography and magnetic resonance findings: Results in 104 consecutive cases. *J. Clin. Endocrinol. Metab.* **86**, 1083–1090.
- Young, W. F., Jr., *et al.* (1996). Primary aldosteronism: Adrenal venous sampling. *Surgery* **120**, 913–920.
- Young, W. F., Jr. (2003). Minireview: Primary aldosteronism—Changing concepts in diagnosis and treatment. *Endocrinology* **144**, 2208–2213.



# Constitutional Delay of Growth and Puberty (CDGP)

Joost Rotteveel and Henriette A. Delemarre-Van de Waal  
Vrije University Medical Center, Amsterdam, The Netherlands

## Glossary

**central intrinsic restraint** Inhibition of gonadotropin-releasing hormone generator by structures above the hypothalamic level.

**follicle-stimulating hormone (FSH)** A gonadotropin produced by the pituitary that stimulates the testicular Sertoli cells to produce semen in men.

**gonadotropin-releasing hormone (GnRH)** A hypothalamic hormone that stimulates the pituitary to produce the gonadotropins LH and FSH.

**luteinizing hormone (LH)** A gonadotropin produced by the pituitary that stimulates the testicular Leydig cells to produce testosterone in men.

**puberty** GnRH generator activity leading to the development of sexual characteristics.

**target height range** A zone of  $\pm 1.3$  standard deviation score (SDS) from target height in which a person is expected to grow; an estimation of the genetic potential for growth.

Children with constitutionally delayed puberty and maturation present with height standard deviation scores below the target height range, reduced height velocity, and delayed bone age in combination with late onset of puberty. The first manifestation of sexual maturation occurs after 13 years for girls and 14 years for boys.

## CHARACTERISTICS OF PUBERTY

Puberty is a maturational process of the hypothalamus–pituitary–gonadal axis resulting in the pubertal growth spurt and development of sex characteristics. At birth, gonadotropin levels are low. This is followed by a transient increase in luteinizing hormone (LH) and follicle-stimulating hormone (FSH) levels during the first months—the postnatal peak. During childhood, the gonadotropin-releasing hormone (GnRH) generator is suppressed by a central intrinsic restraint

system until puberty begins with pulsatile GnRH and subsequent pulsatile gonadotropin secretion. Pituitary gonadotropins stimulate gonadal growth and activity. Subsequently, the gonads start to produce sex steroids, which are responsible for the pubertal growth spurt and development of sex characteristics. By the end of this process, reproductive function is attained and major psychological changes have occurred.

Prior to the 1990s, there was a trend toward earlier onset of puberty and increased final height. A halt in this trend has been described in Western countries. However, an unconfirmed trend in children in the United States toward earlier onset has been described.

At the population level, there is a large variation in the timing of onset of puberty most commonly due to differences in the maturational program of GnRH secretion. The onset of puberty takes place at the age of 11.5 years (range, 9–14 years) in boys and 10.7 years (range, 8–13 years) in girls. Onset of puberty after these ranges is considered abnormal. The duration of puberty varies considerably. By definition, an equal number of males and females have delayed puberty. However, boys are overrepresented as patients with delayed puberty who seek the help of a clinician. Male puberty occurs later than female puberty. Furthermore, boys with delayed puberty are smaller than their peers due to their prepubertal growth dip. Boys are often extremely distressed about their short stature and absence of puberty.

## PUBERTAL DELAY

Delayed puberty is a failure to manifest the initial characteristics of sexual maturation by an age that is more than two standard deviations above the mean for the population. In girls, this is breast development after 13 years of age or menarche after 16 years of age. In boys, this is testicular volume  $> 4$  ml after 14 years of age. Pubic hair development, dependent on the secretion of adrenal and testicular androgens, is often delayed in the constitutional delay of growth and puberty (CDGP) syndrome. However, staging of

pubic hair provides no information about the onset of central puberty: The timing of gonadarche and adrenarche is only loosely related.

## CAUSES OF PUBERTAL DELAY

### Delayed Puberty without Underlying Abnormality: CDGP

The vast majority of children presenting with delayed puberty have no underlying pathology, and CDGP is the most frequent diagnosis. The fact that other diagnoses have to be excluded does not mean that it is a vague diagnosis. In fact, growth, history, and physical examination reveal a characteristic pattern supporting the diagnosis.

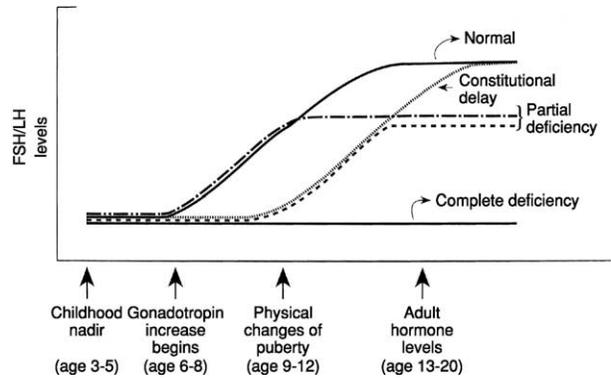
Often, one or both parents had delayed puberty. Growth charts from patients show a low height velocity from childhood that is often lower than the target height range. Boys often present to the clinician when height velocity declines even more during their prepubertal growth dip. Bone age is often delayed but within the target height range.

### Hypogonadotropic Hypogonadism

Ruling out hypogonadotropic hypogonadism is the main diagnostic challenge in patients with pubertal delay. Hypogonadotropic hypogonadism is characterized by low levels of gonadotropins. Impaired secretion of GnRH can be functional, such as in CDGP, hypothyroidism, hyperprolactinemia, chronic illness, anorexia nervosa, or malnutrition. In other conditions, GnRH is truly defective. This may be congenital, as in septo-optic dysplasia and Kallmann's syndrome alone or in combination with congenital adrenal hypoplasia (DAX gene mutation), or it may be caused by local factors, such as tumors at the hypothalamus–pituitary level, especially craniopharyngioma. Pituitary defects can be congenital, such as in panhypopituitarism or partial gonadotropin deficiency, caused by local factors such as tumors, or occur after radiation. In partial gonadotropin deficiency, there is a wide degree of hormone deficiency: Pubertal onset may be timely but the progression of sexual maturation incomplete (Fig. 1).

### Hypergonadotropic Hypogonadism

If basal concentrations of LH and FSH are raised, the diagnosis is primary gonadal failure. Radiotherapy and chemotherapy can result in gonadal failure.



**Figure 1** Patterns of gonadotropin production. In CDGP, normal levels of gonadotropins are reached at a later stage. Partial gonadotropin deficiency may result in late onset or incomplete progression of the expected gonadotropin increments. Therefore, follow-up includes repeated measurements of gonadotropin levels. From Kulin (1996). *J. Clin. Endocrinol. Metab.* **81**, 3460–3464. © The Endocrine Society.

Gonadal insufficiency in girls may be caused by chromosomal abnormalities, such as in Turner's syndrome or variants of gonadal dysgenesis, or by autoimmune ovarian failure. Gonadal failure in boys can be caused by chromosomal abnormalities (Klinefelter's syndrome), anorchia, undescended testes, infections, trauma, or steroid enzymatic deficiencies.

## DIAGNOSIS

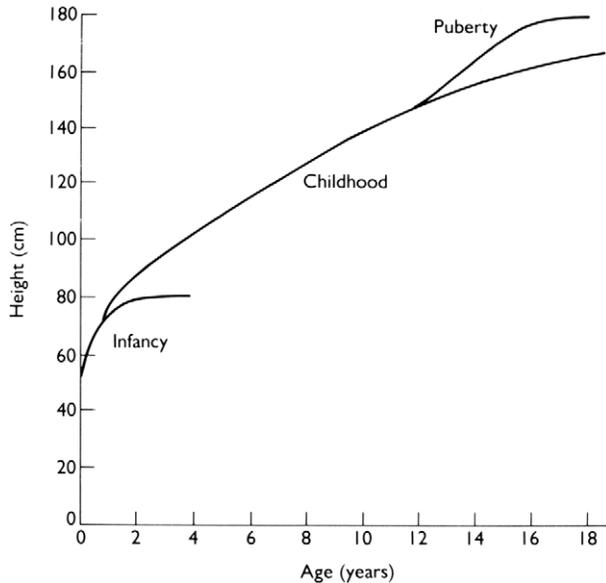
The main diagnostic concern is not only whether puberty has begun or will begin but also whether puberty will progress. Sequential follow-up over years may be required for a correct diagnosis.

## History

Patients with constitutional delay of puberty typically present with delayed linear growth and sexual development. A characteristic growth pattern may be recognized in all three phases of human growth: infancy, childhood, and puberty (Fig. 2).

Height and weight are usually normal at birth. During infancy, children with CDGP often show a decrease in height standard deviation score (SDS), which may result in growth below their target height range. Height SDS will remain low during childhood. The pattern of childhood will continue until puberty starts.

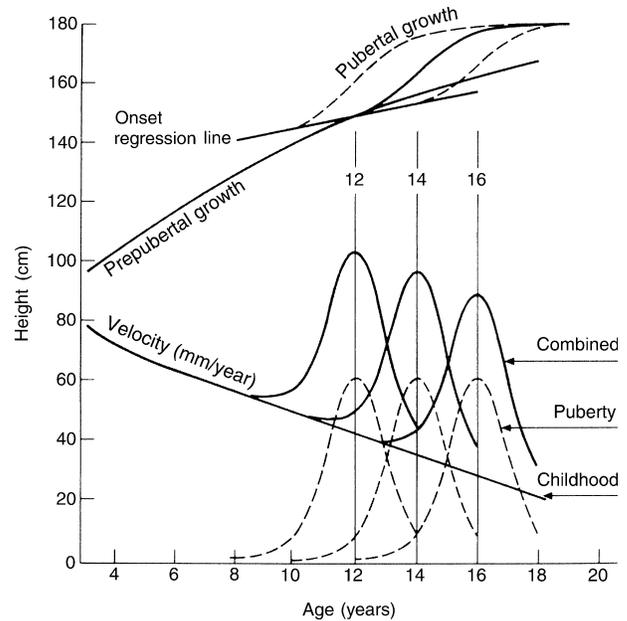
Height velocity will not increase again unless puberty progresses (Fig. 3). Growth during puberty is determined by both sex steroids and growth hormone.



**Figure 2** The infancy-childhood-puberty (ICP) model of growth for boys comprises three components. The first 2 or 3 years during infancy are characterized by rapid growth. Thereafter, during the growth hormone-dependent childhood component, growth velocity decreases to a prepubertal rate of 4 or 5 cm/year. Finally, growth velocity during puberty increases again due to growth hormone augmentation by sex hormones. Data are the mean values for age. From Karlberg, J., *et al.* (1987). Analyses of linear growth using a mathematical method. *Acta Paediatr. Scand.* 76(Suppl.), 478-488.

Therefore, delay in the production of sex steroids results in declining height velocity, and patients often present with short stature at an age when puberty can be expected. This is called the prepubertal growth dip (Fig. 4). The magnitude of the growth spurt is determined by the timing of puberty. In CDGP, the growth spurt is decreased (Fig. 5). Final heights may be comparable with those of early maturers due to the longer period of prepubertal growth, and final height is usually within the target height range. However, some researchers have reported that a proportion of these boys may not achieve full final height appropriate to their parental percentiles, probably because of detrimental effects of late puberty on spinal growth.

The main focus of the history should be the pattern of growth. If possible, old growth charts should be obtained. History should also focus on the start and progression of pubertal development. In 75% of cases, family history reveals a pattern of slow maturation in parents or siblings. Often, parents were short statured as children, and after a late onset of puberty attained normal final height. Women should be asked about menarche and men about first spontaneous ejaculation or the age of first shaving. Some parents have been treated for late puberty. The absence of a



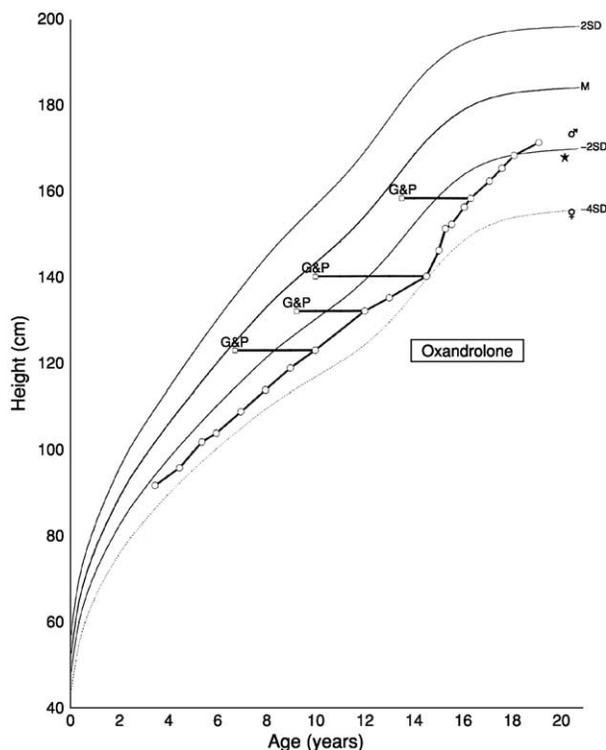
**Figure 3** Velocity chart from the ICP model showing the contribution of the pubertal component. From Karlberg, J., *et al.* (1987). Longitudinal analysis of infantile growth in an urban area of Lahore, Pakistan. In "Modelling of Human Growth" (J. Karlberg, ed.). University of Goteborg, Goteborg, Sweden.

positive family history makes it important for other diagnostic possibilities to be excluded.

One should specifically ask about nutritional habits. Malnutrition, especially anorexia nervosa, is an important cause of maturational delay in girls. Exercise intensity is important in athletes and ballet dancers. Regarding prior medical illness, one should ask about congenital defects. Signs of diseases such as hypothyroidism, growth hormone deficiency, inflammatory bowel disease, Noonan syndrome, and Turner syndrome should be sought for actively.

### Physical Examination

Important in the physical examination is the measurement of height. Short stature, well below target height range, is a common finding. If possible, height velocity should be calculated. Typically, sitting height is low in comparison to leg length due to the inhibitory effect of delayed puberty on spinal growth. This may result in a disproportionate adult stature. Disproportion due to low sitting height is a very reliable sign. Its absence points to an alternate diagnosis. Weight for height should be recorded to evaluate nutritional status. One should search for signs of congenital conditions (midline defects) and chronic illnesses.

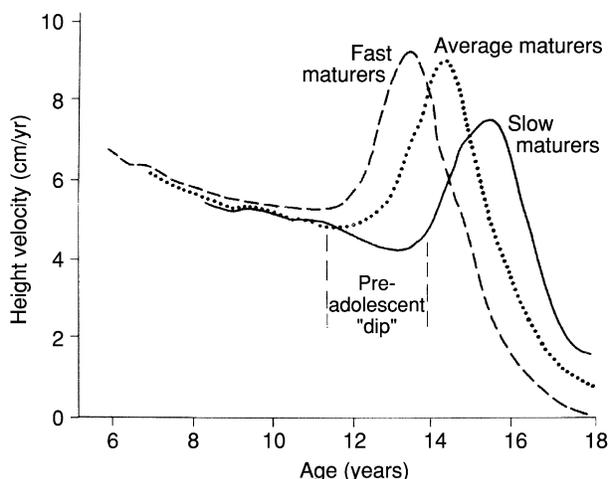


**Figure 4** Prepubertal loss of height velocity (prepubertal dip) in a boy with CDGP. He was successfully treated with oxandrolone, a weak nonaromatizable androgen.

Sexual development should be recorded as Tanner stage. Stage B2 in girls is breast budding. In boys, stage G2 or a testicular volume of more than 4 ml indicates onset of puberty. Asymmetrical testes are found in tumors that may accompany intersex disorders presenting at puberty.

### Bone Age

An X-ray of the left hand and wrist should be obtained at the first visit. During follow-up, this can be repeated. Bone age will usually be delayed more than 2 years. The predicted final height should fall within the parental target height range. Bone age at presentation is often 12 or 13 years. Without the influence of sex steroids, bone age is unlikely to progress beyond the age at presentation. Bone age may be delayed from childhood on. In contrast, bone age in hypogonadotropic hypogonadism can progress beyond 13 years. Therefore, a bone age of more than 13 years indicates the latter diagnosis in a boy with an absence of pubertal development.



**Figure 5** Timing of the pubertal growth spurt for early and late male maturers. Note the relatively slow height velocity (<5 cm per year) for the slow maturer from 11 to 14 years of age. This is the period of most concern for the slow maturer, during which the physical discrepancy from his peers becomes more dramatic. Note also that the peak height velocity of the slow maturer is similar to those of the fast and the average maturers. Final height attainment was achieved by the fast maturers by 18 years, whereas the slow maturers were still growing at this age. Thus, although the slow maturer is smaller as a boy, and especially as a young teenager, the adult stature is usually normal. Adapted from Hammar, S. L., and Eddy, J. A. K. (1968). "Nursing Care of the Adolescent." Heinemann Medical, London.

### Laboratory Findings

Random measurement of gonadotropins and estrogens should be done initially. As in pubertal failure, gonadotropin levels are low in CGDP. Follow-up will show an increase in gonadotropin levels in CDGP but not in hypogonadotropic hypogonadism. Therefore, it is not possible to differentiate CDGP from hypogonadotropic hypogonadism during puberty. Even in adulthood, a major diagnostic difficulty is partial gonadotropin deficiency.

No single test can reliably discriminate CGDP from hypogonadotropic hypogonadism. Several investigators have tried to develop a test that can differentiate between CDGP and gonadotropin deficiency. Although in general GnRH agonists cause higher levels of gonadotropins in patients with CDGP than in patients with hypogonadotropic hypogonadism, children with hypogonadotropic hypogonadism may respond and children with CDGP may not. Therefore, there may be no advantage in performing a GnRH test. In a sleep test, increased LH levels can be measured for CDGP but not for hypogonadotropic hypogonadism. However, no single test seems to be

completely reliable. Therefore, clinical symptoms are far more important. Establishing the diagnosis depends mainly on the growth curve, testicular volume, and bone age

Hyperprolactinemia should be excluded since this condition may induce late onset of puberty or a halt in pubertal progress. Prolactin levels can be measured randomly. Hypothyroidism needs to be excluded by measuring thyrotropin and thyroxine levels.

### Additional Tests

In the case of hypergonadotrophic hypogonadism, a karyotype should be performed to rule out Klinefelter's syndrome in boys and Turner's syndrome in girls.

The spontaneous growth hormone secretion of insulin-like growth factor-1 (IGF-1) and IGFBP-3 is diminished in many boys with CDGP. Also, during growth hormone stimulation tests, stimulated levels of growth hormone may be in the deficient range. However, a diagnosis of growth hormone deficiency cannot be established on these grounds. Boys older than age 10 should be primed with testosterone 5 days before the growth hormone stimulation test; otherwise, growth hormone deficiency will be falsely diagnosed. Furthermore, there is no need to investigate the growth hormone axis when the diagnosis of CDGP is considered. When puberty begins or testosterone therapy is initiated, growth hormone secretion normalizes completely.

### TREATMENT

The treatment of CDGP should be aimed at the different problems that these boys encounter. Reassurance that puberty will eventually occur is usually not enough. Boys with pubertal delay are often extremely distressed and feel uncomfortable with their peers. School performance and psychosocial well-being may be seriously affected. Depression, oppositional behavior, low self-esteem, reduced peer contact, aggression toward peers, and general social immaturity can all be presenting signs. Other problems to consider are the deleterious effects of excessively delayed puberty on spinal growth, leading to a disproportionate stature in adulthood. Furthermore, a delay in onset of puberty may lead to a delay in peak bone mineral density and herald an increased risk of osteoporosis in later life. This has not been confirmed, and normal volumetric bone mineral density and bone turnover in young men with histories of constitutional delay of puberty have been reported.

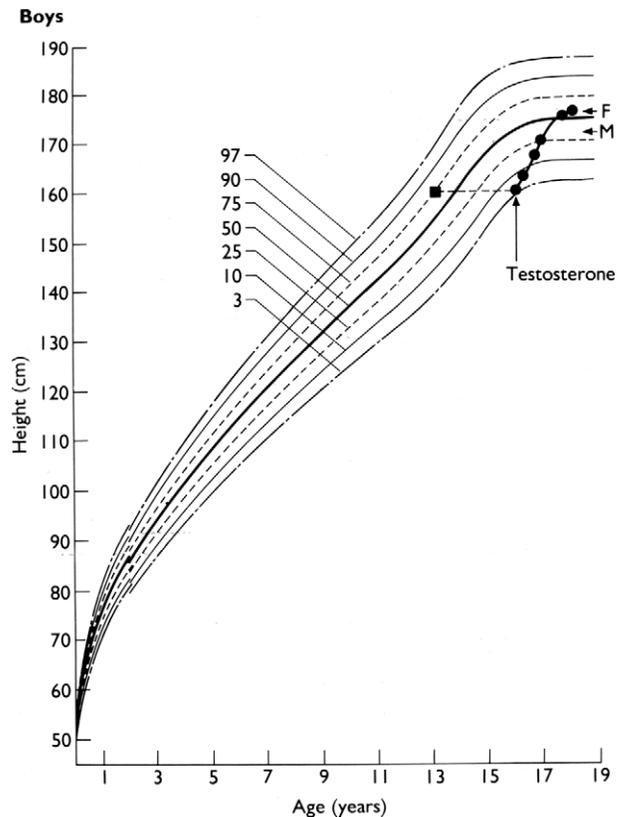
### Testosterone Treatment

#### Short-Term Low-Dose Testosterone

Treatment with sex steroids is logical for CDGP. Treatment at a normal pubertal age ( $\leq 13$  years) provides a better chance of preventing segmental body disproportion, and short-term testosterone therapy in boys may increase bone mineral density. Self-confidence may return even after a short period of low-dose testosterone therapy.

Short-term low-dose therapy in most boys accelerates growth (Fig. 6). Final height is not diminished since bone maturation does not advance with this treatment regimen. However, final height is not improved either.

The optimum time to begin testosterone therapy is probably the normal pubertal age, but bone age maturation may be equally important. Testosterone therapy in boys with a bone age  $< 10$  years should be avoided. Dosage is 25–100 mg/month intramuscularly



**Figure 6** Induction of puberty with testosterone. This boy presented with no signs of puberty at 16 years and was treated with testosterone. Virilization and a normal pubertal growth spurt were induced. Testicular volume did not increase to  $> 6$  ml, and testosterone levels decreased when treatment was stopped. The diagnosis was hypogonadotrophic hypogonadism. From Brook, C. D. G. (1998). "Clinical Paediatric Endocrinology." Blackwell, Oxford, UK.

for 3 months. Testosterone is usually injected at intervals of 2–4 weeks. Monthly intramuscular treatment provides 2-week periods of action. In between, testicular growth stimulated by LH and FSH can be assessed, indicating when testosterone therapy can be discontinued. Although this treatment regimen has been proven to be safe, the progression of bone age during testosterone therapy should be monitored. A second short-term low-dose course of testosterone may be given if puberty does not proceed as expected after the first course. However, the diagnosis should be reconsidered if puberty does not progress at all.

#### **Longer Term Courses of Testosterone**

Longer term courses or higher dosages of testosterone have been associated with major advances in bone age, resulting in final heights below the genetic potential. However, the optimum dose and duration of a testosterone course that provides a good balance between virilization, growth, and bone age resulting in a normal final height are not known. Until course duration and dosages can be compared, we advise short-term low-dose courses only.

#### **Oxandrolone**

The synthetic androgen oxandrolone is not aromatizable. Theoretically, this offers the possibility to exert an androgen-mediated, growth-promoting effect without increasing bone age, thereby improving final height. In clinical practice, however, little or no difference is seen between oxandrolone and short-term low-dose testosterone in terms of growth stimulation and bone age. Both treatments result in little or no increase in final height. However, oxandrolone may have a positive effect on the prepubertal growth dip (Fig. 4). If the main goal is to achieve height gain during this phase, then oxandrolone may actually improve final height. If virilization is the main goal, then testosterone should be used.

The best time to start oxandrolone therapy is when height velocity starts to decline (i.e., at the time of the prepubertal growth dip). Oxandrolone can be given once daily in a dosage of 1.25–2.5 mg. Testicular volume should be followed until the volume is 10–12 ml. The endogenous production of testosterone is capable of inducing the pubertal growth spurt. Testosterone levels should be checked before discontinuing oxandrolone to prevent loss of height velocity due to relatively low levels of testosterone. If levels are low, oxandrolone therapy should be continued until

testosterone levels reach the adult reference range. The progression of bone age should be followed during therapy on a 6-month basis.

#### **Aromatase Inhibitors**

Estrogen causes epiphyseal closure in both sexes. In males, estrogen is derived from testosterone via the aromatase system. When testosterone is combined with an aromatase inhibitor, the former causes virilization and a growth spurt, whereas the latter delays epiphyseal closure. A significantly higher predicted adult height has been reported with the use of aromatase inhibitors. However, clinical trials are still under way, only small groups of patients have been studied, and actual final heights have not been reported.

#### **Growth Hormone Therapy**

There is no indication for the use of growth hormone in boys with delayed puberty. No studies have found a positive effect on growth. Short-term use can cause a transient acceleration in growth velocity, but long-term use has not been found to improve growth or adult height.

#### **FOLLOW-UP**

During therapy, the progress of puberty should be followed carefully by repeated measurement of testicular volume. An increase in gonadotropin levels provides additional information. Especially when treating with androgens, X rays of the left hand and wrist for bone age estimations should be obtained regularly.

Before discharging the patient, the clinician must be sure that the endpoints of puberty have been reached. For boys, the endpoint is an adult level of testosterone, and for girls it is the onset of a normal menstrual cycle. Adult-sized testes is not a sufficient endpoint since the bulk of gonadal constituents are of germinal cell origins.

#### **See Also the Following Articles**

Congenital Adrenal Hypoplasia Syndromes • Delayed Puberty and Hypogonadism, Female • Delayed Puberty and Hypogonadism, Male • Delayed Puberty, Male • Eating Disorders and the Reproductive Axis • Gonadotropin-Releasing Hormone Deficiency, Congenital

Isolated • Gonadotropin-Releasing Hormone (GnRH) Actions • Kallmann's Syndrome and Idiopathic Hypogonadotropic Hypogonadism • Noonan Syndrome • Puberty: Physical Activity and Growth • Turner Syndrome

### Further Reading

- Arslanian, S. (1997). Testosterone treatment in adolescents with delayed puberty: Changes in body composition, protein, fat, and glucose metabolism. *J. Clin. Endocrinol. Metab.* **82**(10), 3213–3220.
- Bertelloni, S. (1998). Normal volumetric bone mineral density and bone turnover in young men with histories of constitutional delay of puberty. *J. Clin. Endocrinol. Metab.* **83**(12), 4280–4283.
- De Luca, F. (2001). Management of puberty in constitutional delay of growth and puberty. *J. Pediatr. Endocrinol. Metab.* **14**, 953–957.
- Kulin, H. E. (1996). Extensive personal experience: Delayed puberty. *J. Clin. Endocrinol. Metab.* **81**(10), 3460–3464.
- Palmert, M. R. (2001). Variation in the timing of puberty: Clinical spectrum and genetic investigation. *J. Clin. Endocrinol. Metab.* **86**(6), 2364–2368.
- Schroor, E. (1995). The effect of prolonged administration of an anabolic steroid (oxandrolone) on growth in boys with constitutionally delayed growth and puberty. *Eur. J. Pediatr.* **154**(12), 953–957.
- Terasawa, E. (2001). Neurobiological mechanisms of the onset of puberty in primates. *Endocr. Rev.* **22**(1), 111–151.
- Wickman, S. (2001). A specific aromatase inhibitor and potential increase in adult height in boys with delayed puberty: A randomised controlled trial. *Lancet* **357**, 1743–1748.



# Contraception, Male

C. Hay and F. C. W. Wu

University of Manchester, Manchester, United Kingdom

## Glossary

**azoospermia** The absence of sperm in the ejaculate.

**gonadotropins** The pituitary hormones luteinizing hormone and follicle-stimulating hormone.

**oligozoospermia** The presence of <3 million sperm/ml in the ejaculate.

**severe oligozoospermia** The presence of  $\leq 1$  million sperm/ml in the ejaculate.

Development of a male hormonal contraceptive would ideally provide a safe, effective, and alternative birth control method for couples. One-third of couples worldwide choose a male contraceptive method; only two effective options are available. Vasectomy is very effective but essentially irreversible. Condoms as long-term contraception have a relatively high failure rate. In addition to the impact on the contraceptive choice for individual couples, at a global level there is a need for population control. The world population in 1999 reached 6 billion people for the first time, representing an increase in population of 4.4 billion since 1900. If the current birth rate is maintained, the world population will approach 19 billion by the year 2100. Such a population size would be unsustainable and result in immense poverty, overcrowding, resource depletion, and increased pollution.

## MECHANISM OF ACTION

The aim of a male hormonal contraceptive is to suppress spermatogenesis and therefore prevent pregnancies. Spermatogenesis is dependent on the pituitary gonadotropins both directly and via intratesticular testosterone (T) production (see Fig. 1). The intratesticular T concentration in the seminiferous tubules is critical to the maintenance of spermatogenesis. The drive for T production is provided by luteinizing hormone action on Leydig cells in the testis. Gonadotropin secretion requires pulsatile gonadotropin-releasing hormone (GnRH) secretion from the hypothalamus; therefore, either GnRH deficiency or

nonpulsatile stimulation inhibits the pituitary-gonadal axis. Inhibitors of the hypothalamic-pituitary-gonadal axis are shown in Fig. 1. With the resultant inhibition of intratesticular testosterone production, systemic levels will also fall. Symptoms and signs of hypogonadism would therefore result if systemic testosterone were not administered in any treatment regimen. Successful treatments used for suppression of spermatogenesis have included androgen-only regimens and androgen combinations with other agents, i.e., progestogens.

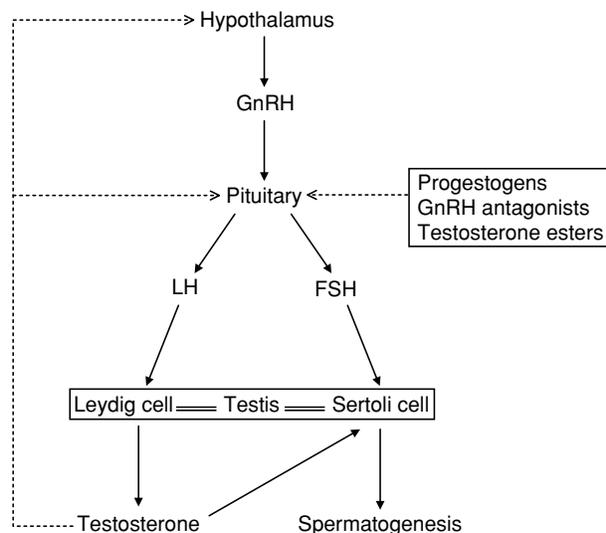
## CONTRACEPTIVE REGIMENS

### Androgen Only

Testosterone is rapidly degraded by first-pass metabolism in the liver following oral or systemic administration. Chemical modification of androgen preparations is therefore necessary to reduce hepatic breakdown and prolong the duration of therapeutic action. The oral preparations that are available have a relatively short half-life and frequent dosage is required. Prolongation of the half-life is achieved by the formulation of T esters in lipid vehicles that are administered by intramuscular (im) injection.

### Testosterone Enanthate

Studies using im testosterone enanthate (TE) have confirmed that supraphysiological doses of testosterone induced azoospermia in 40–70% of Caucasian males. Efficacy studies subsequently demonstrated a relationship between the extent of spermatogenesis suppression (sperm concentration) and contraceptive failure (pregnancy) rate. The pregnancy rate for azoospermia was only 0.8 per 100 person years (comparable to female injectables) and 8.1 per 100 person years for oligozoospermia. The combined failure rate was 1.4 pregnancies per 100 person years, which is comparable to the female oral contraceptive pill and is better than the typical first-year failure rate of condoms (12%). These studies demonstrated that



**Figure 1** The hypothalamic–pituitary–gonadal axis. Solid arrows show the pathway of spermatogenesis promotion and the dashed arrows show inhibitory influences. Endogenous testosterone acts via negative feedback pathways to inhibit both GnRH release from the hypothalamus and gonadotropin release from the pituitary. Exogenous sex steroids and GnRH analogues reversibly manipulate this physiological response. LH, luteinizing hormone; FSH, follicle-stimulating hormone.

hormonal suppression of spermatogenesis can be an efficacious and reversible contraceptive for men. They also demonstrate that targets of suppression to ensure effective contraceptive protection are azoospermia and severe oligozoospermia (<1 million sperm/ml).

This approach had two drawbacks: (1) unsatisfactory pharmacokinetics of TE resulted in widely fluctuating levels of testosterone and frequent weekly injections and (2) high doses (causing supraphysiological T levels) were required to induce and maintain adequate suppression of spermatogenesis. Side effects encountered in these studies were androgen-related effects including acne, weight gain, behavioral effects, lowered high-density lipoprotein cholesterol (HDL-C), and increased hematopoiesis. For these reasons, longer acting testosterone esters and alternative strategies were sought.

### Testosterone Undecanoate

Testosterone undecanoate (TU) is an unsaturated, aliphatic, fatty acid ester of T that is partially absorbed from the gut lymphatics following oral administration. The only available oral TU (Restandol) has low bioavailability and requires twice or thrice daily dosing.

The formulation of TU in tea seed oil (in China) and castor oil (in Germany) for intramuscular use has yielded longer acting testosterone depots with more

favorable pharmacokinetics. TU alone (tea seed oil) was found to induce azoospermia in 96% (23/24) of Asian men (500 or 1000 mg in tea seed oil every 4 weeks) and in 57% (8/14) of Caucasian men (1000 mg in castor oil every 6 weeks) studied. Phase III contraceptive efficacy trials using TU alone are in progress in China. A combination of TU with a progestogen has been studied in Caucasian subjects (see [Table I](#)).

### Alternative Androgen Preparations

#### Testosterone Implants

Testosterone implants are small cylindrical pellets that are inserted into the subcutaneous tissue of the lower abdominal wall. Insertion requires a minor surgical procedure under local anesthesia. In Australia, a dose of 1200 mg (six pellets) has been shown to cause spermatogenic suppression equal to that for TE alone with similar or fewer metabolic side effects.

#### Testosterone Patches

Testosterone patches provide noninvasive, transdermal, self-administered delivery of testosterone when applied daily to the body or scrotal skin. The main limitations include a high incidence (40–60% of subjects) of skin irritation related to the alcoholic excipients employed, the frequent shaving of the scrotum, and poor adhesiveness, particularly in hot weather. Two studies have used a body patch in combination with a progestogen (oral levonorgestrel and desogestrel).

#### 19-Nortestosterone

19-Nortestosterone (19-NT) and its derivative 7 $\alpha$ -methyl-19-nortestosterone (MENT) have both been studied in healthy men, with encouraging results. 19-NT alone was studied in Indonesian men and rates of sperm suppression were equal to those for TE. Subdermal MENT implants have also been shown to suppress to azoospermia in 70% of cases.

#### Testosterone Buciclate

Testosterone buciclate is a long-acting ester. Due to problems with the formulation, few data are available.

#### Testosterone Microspheres

Testosterone microspheres are biodegradable polyacrylate–glycolide spheres containing testosterone and they exhibit first-order absorption kinetics. They have been studied in hypogonadal men via both intramuscular and subcutaneous administration. The results are encouraging and application in a hormonal contraceptive is feasible in the future.

Table I Progesterone/Androgen Combinations for Male Hormonal Contraception

Reference	Duration of Rx	Progesterone	Androgen	Azoospermia	Severe oligozoospermia
Handelsman <i>et al.</i> (1996)	1 year	DMPA 300 mg	T implants 800 mg	9/10 (90%)	10/10 (100%) <sup>d</sup>
Meriggiola <i>et al.</i> (1996)	16 weeks	CPA 50 or 100 mg	TE 100 mg	5/5 (100%)	5/5 (100%)
Meriggiola <i>et al.</i> (1997)	16 weeks	CPA 12.5 mg	TU 80 mg bid	1/8 (13%)	3/8 (38%)
Meriggiola <i>et al.</i> (1998)	16 weeks	CPA 12.5 mg	TE 100 mg	3/5 (60%)	
		CPA 25 mg	TE 100 mg	5/5 (100%) <sup>b</sup>	5/5 (100%)
Bebb <i>et al.</i> (1996)	24 weeks	LNG 500 µg	TE 100 mg	12/18 (67%)	14/18 (78%)
Anawalt <i>et al.</i> (1999)	24 weeks	LNG 250 µg	TE 100 mg	14/18 (78%)	16/18 (89%)
		LNG 125 µg	TE 100 mg	11/18 (61%)	16/18 (89%)
Buchter <i>et al.</i> (1999)	24 weeks	LNG 250–500 µg <sup>c</sup>	T patch 5 mg	2/11 (18%)	5/11 (45%) <sup>d</sup>
Kamischke <i>et al.</i> (2000)	24 weeks	LNG 250 µg	TU 1000 mg	7/14 (50%)	13/14 (92%)
Wu <i>et al.</i> (1999)	24 weeks	DSG 300 µg	TE 100 mg	6/8 (75%)	6/8 (75%)
		DSG 150 µg	TE 100 mg	4/7 (57%) <sup>d</sup>	7/7 (100%)
		DSG 300 µg	TE 50 mg	8/8 (100%)	8/8 (100%)
Anawalt <i>et al.</i> (2000)	24 weeks	DSG 150 µg	TE 50 mg	4/7 (57%) <sup>e</sup>	6/9 (67%)
		DSG 150 µg	TE 100 mg	7/7 (100%)	7/7 (100%)
		DSG 300 µg	TE 100 mg	7/8 (88%)	7/8 (88%)
Hair <i>et al.</i> (2001) <sup>f</sup>	24 weeks	DSG 300 µg	T patch 5 mg	4/7 (57%)	5/7 (71%)
Kinniburgh <i>et al.</i> (2002)	24 weeks	DSG 150 µg	T implants	22/31 (71%)	28/31 (90%) <sup>g</sup>
		DSG 300 µg	T implants 400 mg	28/28 (100%)	28/28 (100%)
Kamischke <i>et al.</i> (2000)	24 weeks	NETE 200 mg	TU 1000 mg	13/14 (93%)	
Anderson <i>et al.</i> (2002)	24 weeks	ENG 1 rod	T implants	9/14 (64%)	10/14 (71%)
		ENG 2 rod	T implants	9/12 (75%) <sup>b</sup>	13/14 (93%)

Note. DMPA (depot medroxyprogesterone acetate) was administered by a single im injection. T implants (testosterone implants) were administered once (800 mg dose) or at weeks 0 and 12 (400 mg). CPA (cyproterone acetate) was administered orally once daily. TE (testosterone enanthate) at 100 mg was administered by im injection once weekly. LNG (levonorgestrel) was administered orally once daily. TU (testosterone undecanoate) was administered either orally (80 mg) twice daily or im (1000 mg) every 6 weeks. T patch (testosterone patch) delivers 5 mg every 24 h. DSG (desogestrel) was administered orally once daily. NETE (norethisterone enanthate) was administered im every 6 weeks.

<sup>a</sup>Severe oligozoospermia in this report was defined as <3 million/ml.

<sup>b</sup>One subject was azoospermic at week 12 but had a count of 0.1 million/ml at week 16.

<sup>c</sup>Subjects were commenced on 250 µg daily until week 12. If they had not suppressed to azoospermia at week 12, the dose was increased to 500 µg daily.

<sup>d</sup>One subject dropped out of the trial.

<sup>e</sup>Two subjects dropped out of the trial for personal reasons.

<sup>f</sup>Only one dose group of the three in the study is included.

<sup>g</sup>Three Chinese men maintained sperm counts >3 million/ml.

<sup>h</sup>Two men withdrew at weeks 12 and 16 (counts <1 million/ml).

## GnRH Antagonists

GnRH antagonists cause suppression of gonadotropins, which is rapidly achieved within hours. Studies have shown spermatogenic suppression at least equal to that for TE. Furthermore, when the GnRH antagonist Nal-Glu was used to initiate suppression, this suppression could then be maintained with TE alone. Although these synthetic peptide compounds clearly have contraceptive potential, the disadvantages are their expense, their short half life, and the need for subcutaneous injection to avoid intestinal breakdown. Side effects encountered have included local skin reactions and pruritis at the site of injection. New long-acting depot preparations of potent GnRH

antagonists may have a place in male contraception where rapid induction of spermatogenic suppression can subsequently be maintained by testosterone alone.

## Progesterone/Androgen Combinations

Exogenous progestogens can inhibit gonadotropin secretion, reduce systemic testosterone levels, and suppress spermatogenesis in men. Combining a progestogen with testosterone exploits the synergistic actions of the two steroids that can be used at lower doses for spermatogenesis suppression. A number of different progesterone/androgen combinations have been studied (see Table I).

### **Depot Medroxyprogesterone Acetate**

Depot medroxyprogesterone acetate (DMPA) has been combined with 19-NT, TE, and T implants. Trials with 19-NT in combination with DMPA in Indonesian men resulted in an azoospermia rate of 98% compared to 96% for TE alone. In Caucasian men, the combination of DMPA with T implants achieved azoospermia rates exceeding those for TE alone. No significant weight gain or metabolic effects (e.g., on lipid parameters) resulted. The main limitation with DMPA use is the prolonged period (up to 6 months) that is often necessary for sperm recovery following cessation of treatment.

### **Cyproterone Acetate**

Cyproterone acetate (CPA) has been combined with im TE and oral TU. The anti-androgenic properties of CPA are effected by the blockade of T and DHT at the receptor level. There is also gonadotropin inhibition at the pituitary. The combination of two oral preparations in a true male pill was assessed with oral CPA and TU. Unfortunately, azoospermia occurred in less than 20% of subjects. The reason for this is most likely related to the inadequate androgen replacement provided by oral TU. CPA at a higher dose (25–100 mg) with TE did result in azoospermia in 100% of men. Furthermore, the time to azoospermia was half that of the TE-alone group (49 days versus 98 days). Though encouraging, dose-dependent decreases in hemoglobin and body weight were observed in those subjects receiving CPA. These side effects are very likely anti-androgen related. These effects may limit future application of this combination.

### **Levonorgestrel**

Levonorgestrel (LNG) has been extensively studied with a variety of androgens. Initially, a dose of 500 µg daily with TE confirmed greater sperm suppression than with TE alone (azoospermia in 67% versus 33% of subjects). Further study with lower doses of LNG (250 and 125 µg) did not compromise sperm suppression rates. Adverse effects were dose dependent and included lowered HDL-C and weight gain. A combination of TU/LNG was compared to TU alone. There was no difference in suppression to azoospermia (50%). Testosterone patches have been studied with oral LNG (250 µg daily) and with long-acting LNG implants. Relatively poor sperm suppression (severe oligozoospermia in <60% of subjects) was observed in both studies. A likely reason for the relatively disappointing results may relate to the unreliable administration or absorption of T so that

circulating levels in only the low-normal range can be achieved. This highlights the critical role of testosterone in achieving efficient spermatogenesis suppression: a sufficient but not supraphysiological amount of T conveniently delivered with kinetics that can sustain stable plasma levels is required. The transdermal preparations that are available fall short of these requirements.

### **Desogestrel and Etonogestrel**

Desogestrel (DSG) is an oral third-generation progestogen with potent progestational activity and lower androgenicity. These potentially favorable properties led to the study of DSG in combination with TE. Oral DSG in combination with various TE doses produced 100% azoospermia (Table I). These results demonstrate the synergistic action of this steroid combination. A lowering of HDL-C of 20–25% was observed in these studies. A cross-cultural study assessed DSG in combination with 400 mg testosterone implants. At 24 weeks, there was a significant difference between dose groups with 100% (300 µg group) versus 71% (150 µg group; 11/13 Caucasian men, 11/18 Chinese men) azoospermia. A significant decline in HDL-C independent of dose was observed in the Caucasian men only. DSG has been combined with a T patch but similar to the LNG/patch study, sperm suppression was inferior to that seen with other regimens.

Etonogestrel (ENG) is the active metabolite of DSG and has been formulated as a subdermal implant for female contraception (Implanon, NV Organon). One study has assessed ENG rods (68 mg ENG per rod) in combination with T implants. Azoospermia was achieved in 9/14 men in both groups. However, suppression was inconsistent in the lower dose group. Further study of this promising progestogen implant is planned for the future.

### **Norethisterone Enanthate**

Norethisterone enanthate (NETE) is a depot progestogen administered by intramuscular injection converting to the metabolically active norethisterone. NETE and TU have been combined in a 6-week regimen. The results have been very encouraging (see Table I), with azoospermia occurring in 93% of men. Adverse effects of the TU/NETE combination have included moderate increases in hemoglobin (within the normal range) and decreases in HDL-C.

## **ETHNIC DIFFERENCES**

There are clear differences in the rates of azoospermia achieved in Asian populations compared to Caucasian

populations. In studies using androgens alone (TE and TU), azoospermia was achieved in 90 and 96% of Asian men versus 74.3 and 57% of European or American men. The possible explanations for these differences include genetic and environmental influences that are not yet fully understood. Intraethnic variation is also seen within Caucasian populations. Suppression to azoospermia occurred in 40–70% of men in the World Health Organization trials using TE alone. No correlations were found to explain this. A possible explanation is intraethnic heterogeneity in the biological response to androgens.

## SUMMARY AND CONCLUSIONS

There is evidence to support couple dissatisfaction with available contraceptive methods. A cross-cultural survey found that women felt the contraceptive burden too frequently falls to them and they welcomed the idea of a male contraceptive pill. An important aspect of any male hormonal method is that from the start of treatment there is a time lag until sperm suppression is sufficient for contraceptive efficacy. This lag time occurs because sperm production takes 2–3 months and therefore following gonadotropin suppression a similar period of time is necessary to reach severe oligozoospermia. However, a similar lag time also applies to vasectomy (irreversible) and a lag time of 1 month applies to the female oral contraceptive pill. An important point is that hormonal methods for either the male or female should be suggested as sole contraception only when the user is in a stable relationship. In this situation, the choice of contraceptive is often planned and a lag time is unlikely to deter users.

Public-funded research has reached a stage where realization of male hormonal contraception is within reach. The beginnings of participation by the pharmaceutical industry are encouraging and a necessary step leading to product development. A progestogen/androgen combination is most likely to be the first product to be marketed. Further development of long-acting depot formulations or oral and nonsteroidal compounds will provide a variety of contraceptive formulations that will allow men to have a widened choice and encourage acceptability and usage.

## See Also the Following Articles

Fertility in Men with Spermatogenesis Abnormalities • Gonadotropin-Releasing Hormone (GnRH) Actions • Sexual Function and Androgens • Spermatogenesis, Endocrine Control of

## Further Reading

- Anderson, R. A., Kinniburgh, D., and Baird, D. T. (2002). Suppression of spermatogenesis by etonogestrel implants with depot testosterone: Potential for long-acting male contraception. *J. Clin. Endocrinol. Metab.* **87**, 3640–3649.
- Bebb, R. A., Anawalt, B. D., Christensen, R. B., Paulsen, C. A., Bremner, W. J., and Matsumoto, A. M. (1996). Combined administration of levonorgestrel and testosterone induces more rapid and effective suppression of spermatogenesis than testosterone alone: A promising male contraceptive approach. *J. Clin. Endocrinol. Metab.* **81**, 757–762.
- Bradley, D., Anawalt, M. D., Herbst, K. L., Matsumoto, A. M., Mulders, T. M. T., Coelingh-Bennink, H. J. T., and Bremner, W. J. (2000). Desogestrel plus testosterone effectively suppresses spermatogenesis but also causes modest weight gain and high density lipoprotein suppression. *Fertil. Steril.* **74**, 707–714.
- Buchter, D., von Eckardstein, S., von Eckardstein, A., Kamischke, A., Simoni, M., Behre, H. M., and Nieschlag, E. (1999). Clinical trial of transdermal testosterone and oral levonorgestrel for male contraception. *J. Clin. Endocrinol. Metab.* **84**, 1244–1249.
- Hair, W. M., Kitteridge, K., O'Connor, D. B., and Wu, F. C. W. (2001). A novel male contraceptive pill-patch combination: Oral desogestrel and transdermal testosterone in the suppression of spermatogenesis in normal men. *J. Clin. Endocrinol. Metab.* **86**, 5201–5209.
- Handelsman, D. J., Conway, A. J., Howe, C. J., and Mackey, M.-A. (1996). Establishing the minimum effective dose and additive effects of depot progestin in suppression of human spermatogenesis by a testosterone depot. *J. Clin. Endocrinol. Metab.* **81**, 4113–4121.
- Kamischke, A., Heuermann, T., Kruger, K., von Eckardstein, S., Schellschmidt, I., Rubig, A., and Nieschlag, E. (2002). An effective hormonal male contraceptive using testosterone undecanoate with oral or injectable norethisterone preparations. *J. Clin. Endocrinol. Metab.* **87**, 530–539.
- Kinniburgh, D., Zhu, H., Cheng, L., Kicman, A. T., Baird, D. T., and Anderson, R. A. (2002). Oral desogestrel with testosterone pellets induces consistent suppression of spermatogenesis to azoospermia in both Caucasian and Chinese men. *Hum. Reprod.* **17**, 1490–1501.
- Merrigiola, C. M., Bremner, W. J., Paulsen, C. A., Valdiserri, A., Incorvaia, L., Motta, R., Pavani, A., Capelli, M., and Flamigni, C. (1996). A combined regimen of cyproterone acetate and testosterone enanthate as a potentially highly effective male contraceptive. *J. Clin. Endocrinol. Metab.* **81**, 3018–3023.
- World Health Organization Task Force on Methods for the Regulation of Male Fertility. (1996). Contraceptive efficacy of testosterone-induced azoospermia and oligozoospermia in normal men. *Fertil. Steril.* **65**, 821–829.
- World Health Organization Task Force on Methods for the Regulation of Male Fertility. (1990). Contraceptive efficacy of testosterone-induced azoospermia in normal men. *Lancet* **336**, 995–999.
- Wu, F. C. W., Balasubramanian, R., Mulders, T. M. T., and Coelingh-Bennink, H. J. T. (1999). Oral progestogen combined with testosterone as a potential male contraceptive: Additive effects between desogestrel and testosterone enanthate in suppression of spermatogenesis, pituitary–testicular axis, and lipid metabolism. *J. Clin. Endocrinol. Metab.* **84**, 112–122.
- Zhang, G.-Y., Gu, Y.-Q., Wang, X.-H., Cui, Y.-G., and Bremner, W. J. (1999). A clinical trial of injectable testosterone undecanoate as a potential male contraceptive in normal Chinese men. *J. Clin. Endocrinol. Metab.* **84**, 3642–3647.



# Corticotropin-Releasing Hormone (CRH) and Inflammation

George Mastorakos and Ioannis Ilias

University of Athens Medical School and Aretaieion Hospital, Athens, Greece

## Glossary

**antalarmin** A pyrrolopyrimidine nonpeptide compound {*N*-butyl-*N*-ethyl[2,5,6-trimethyl-7-(2,4,6-trimethylphenyl)pyrrolo[2,3-*d*]pyrimidin-4-yl]amine(1)}, acting as a selective corticotropin-releasing hormone receptor type 1 antagonist.

**corticotropin-releasing hormone (CRH)** A 41-amino-acid peptide, secreted principally by the parvocellular neurons of the paraventricular hypothalamic nuclei. CRH is the major physiological corticotropin secretagogue.

**immune/inflammatory (I/I) response** Septic or nonseptic injurious agents lead to the immune system's reaction by means of the I/I response. The result of this response is the destruction or walling-off of the attacking microorganism or the destroyed tissue. This response is microscopically characterized at the local level by dilation and increased permeability of microvessels, both leading to increased blood flow, exudation of plasma, and migration and accumulation of leukocytes.

The immune/inflammatory (I/I) response is influenced by the brain through regulation of peripheral nervous system functions and endocrine responses. Cells of the immune system have receptors for a number of hormones, neuropeptides, and neurotransmitters. Therefore, the responses of these cells may be modulated by changes in neuroendocrine and/or autonomic activity.

## INTRODUCTION

The hypothalamic–pituitary–adrenal (HPA) axis is activated during stress and represents one of the major pathways through which the brain regulates the I/I response. Conversely, products of the I/I response influence brain function. In addition, immune cells produce a number of hormones and neuropeptides, such as corticotropin-releasing hormone (CRH) and

$\alpha$ -melanocyte-stimulating hormone, which probably act locally as autacoids during both the early and late stages of the I/I process. This locally produced CRH, hereafter called “peripheral CRH,” has been also implicated in other physiological functions, such as reproduction.

## CENTRAL CRH AND THE HPA AXIS—PERIPHERAL CRH

Sauvagine, urotensin I, urocortin, and CRH belong to a family of peptides with similar activities. In humans, the CRH and urocortin genes are located on chromosomes 8 and 2, respectively. Initially, CRH is synthesized as a larger precursor molecule (spanning 191 amino acids in humans) from which it is cleaved at flanking basic amino acid pairs. CRH is secreted—along with other adrenocorticotrophic hormone (ACTH) secretagogues, such as arginine vasopressin (AVP), cholecystokinin, met-enkephalin, and dynorphin—into the hypophyseal portal blood via projecting axons to the median eminence. The plasma half-life of CRH in humans is 4 min. The secretion of CRH is regulated by inputs from higher centers integrating the effects of the circadian pacemaker, stress, and glucocorticoid negative feedback (acting at pituitary, hypothalamic, and higher levels, such as the hippocampus). ACTH released by CRH leads to the secretion of cortisol and other adrenal steroids.

Most of the plasma CRH is apparently of nonhypothalamic origin since hypothalamic CRH is rapidly enzymatically decomposed at the pituitary level. Although the contribution of hypothalamic CRH to the total plasma CRH is small, in certain circumstances, such as insulin-induced hypoglycemia, hypothalamic CRH release leads to increments in plasma CRH concentration.

Peripheral CRH has been found in the adrenal medulla, the testes, the ovaries, the gastrointestinal tract, the pancreas, the myometrium, the endometrium, and the placenta as well as diverse inflammatory

sites. The epithelial cells of human and rodent endometrium produce CRH throughout the menstrual cycle, whereas the stroma needs to undergo decidualization in order to produce CRH. The placenta secretes CRH, leading to an increased maternal plasma CRH concentration during the third trimester of pregnancy. Only in humans, plasma CRH is bound to a high-affinity binding protein, with a concomitant reduction in its bioactivity.

Two types of G protein-coupled CRH receptors, types 1 and 2 (CRH-R1 and CRH-R2, respectively), with seven transmembrane domains each, have been described. Their genes are located on chromosome 17q12. Alternatively spliced isoforms  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$  have been identified for CRH-R1, whereas isoforms  $\alpha$ ,  $\beta$ , and  $\gamma$ , plus a stomach variant, have been identified for CRH-R2. CRH-R1 is most abundant in neocortical, cerebellar, and sensory relay structures, whereas CRH-R2 is predominantly localized in specific subcortical areas and peripheral tissues. CRH-R1 $\alpha$  is present on both epithelial and stromal cells of the human endometrium. CRH binds to type CRH-R1 of the anterior pituitary corticotrophs, resulting in adenylyl cyclase activation and increased intracellular cyclic AMP (cAMP) concentration, cAMP-dependent protein kinase A activation, increased influx of extracellular calcium via L-type calcium channels, and the production of lipoxygenase metabolites of arachidonic acid. The net result is the secretion of ACTH and other proopiomelanocortin (POMC)-derived peptides within a few seconds, whereas increased POMC gene transcription and POMC biosynthesis ensue. The number of CRH-R in corticotroph cells may modulate the ACTH response.

## THE I/I RESPONSE

The cellular components of the I/I response consist of circulating nonlymphoid leukocytes and lymphocytes and local immune accessory cells. Nonlymphoid leukocytes include the monocytes/macrophages, neutrophils, basophils, and eosinophils. Local immune accessory cells include the endothelial cells, tissue fibroblasts, resident macrophages, and macrophage-related cells, such as liver Kupffer cells, type A synovial-lining cells, and central nervous system (CNS) glial cells, as well as the basophil-related mast cells. Many substances secreted locally in the inflammatory area by the above-mentioned cells act as autocrine or paracrine regulators and/or mediators of the inflammatory response, as well as endocrine messengers between the inflammatory process and other systems

such as the CNS, HPA axis, and peripheral nervous system. These substances include the vasoactive amines, histamine and serotonin, the kallikrein/kinin system, the Hageman factor and other clotting factors, the fibrinolytic system, several components of the complement system, and eosinophil and platelet activators. They also include cytokines [such as tumor necrosis factor  $\alpha$  (TNF $\alpha$ ), TNF $\beta$ , interferon- $\alpha$ , interferon- $\beta$ , and interferon- $\gamma$ ], interleukin-1 (IL-1) through IL-14 (as well as their binding proteins and natural antagonists), many lipid and glucolipid products of triglyceride metabolism (such as several active products of arachidonic acid, including the endoperoxides, the thromboxanes, prostacyclin, and leukotrienes), and platelet-activating factor. In addition, active oxygen radicals, including nitric oxide and lysosomal constituents such as neutral proteases, participate in the inflammatory response.

## CRH AND THE I/I RESPONSE

Hypothalamic (central) CRH has been considered to act indirectly in an anti-inflammatory fashion, since the end product of the HPA axis' stimulation is cortisol, which has thoroughly studied anti-inflammatory actions. Moreover, it is known that among the pro-inflammatory cytokines, IL-6 is a potent stimulator of the human HPA axis, and a secretagogue of magnocellular AVP. The administration of IL-6 subcutaneously provokes considerable elevations in plasma ACTH, cortisol, and AVP. It has been shown that IL-6, in patients with head trauma (an aseptic inflammatory state) and the syndrome of inappropriate secretion of antidiuretic hormone, is quantitatively correlated with AVP. The latter is also known to be a potent activator of the HPA axis.

Since CRH is the main stimulator of POMC production from the pituitary, and certain POMC products (like ACTH and  $\beta$ -endorphin) can directly affect the I/I response, it has been hypothesized that CRH itself might be directly involved with the I/I response. The putative role of CRH in the I/I response was further suggested by the presence of CRH-specific binding sites in human lymphocytes secreting POMC-derived peptides (ACTH and  $\beta$ -endorphin). By employing the rat air-pouch model of acute aseptic chemical inflammation, immunoreactive CRH (IrCRH) was localized in the inflammatory tissue by immunohistochemistry. IrCRH was also found in the cytoplasm of immune accessory cells such as macrophages, endothelial cells surrounding vessels, and tissue fibroblasts. Immunoneutralization

studies *in vivo* with a highly specific anti-CRH polyclonal antiserum resulted in a significant suppression of the inflammatory response, suggesting that peripheral CRH promotes inflammation. The decrease in inflammation by anti-CRH antiserum was similar to that caused by immunoneutralization of TNF $\alpha$  (a well-known mediator of the I/I response that was used as a positive control). The effects of the combined administration of anti-CRH and anti-TNF $\alpha$  were not additive, indicating that the two antisera might interfere with a common pathway of the inflammatory response. Furthermore, locally produced somatostatin mediates the glucocorticoid anti-inflammatory effects at the inflammatory sites, whereas CRH levels at inflammatory sites are lowered in the presence of somatostatin analogues.

The mobility of this "immune" peripheral CRH is similar, as assessed by high-performance liquid chromatography, to that of r/hCRH 1–41 (the form produced by the rat and human hypothalamus as well as by the human placenta). The presence of CRH at peripheral inflammatory sites has also been demonstrated in other animal models of both acute and chronic inflammation. IrCRH was present in inflammatory cells of rat joint tissue with streptococcal cell wall- and adjuvant-induced arthritis. CRH mRNA was present in the inflamed synovia from arthritic rat joints that expressed specific CRH-binding sites. Furthermore, IrCRH was seen in the synovial-lining cell layers and blood vessels from the joints of patients with rheumatoid arthritis (RA) and osteoarthritis, whereas high levels of CRH immunoreactivity were found in the synovial fluids of patients with RA. IrCRH was also found in immune accessory cells from uveitic retinas and corpora vitrea from Lewis rats with experimentally induced autoimmune uveitis. Additionally, the local presence of CRH appears to be of pivotal importance in the process of experimental autoimmune uveoretinitis in rodents, since retinas from immunized B10.A mice treated with anti-CRH antibody showed significantly lower apoptosis and Fas and Fas ligand (FasL) expression than placebo-treated animals. Thus, CRH at inflammatory sites seems to be involved in the activation of the Fas/FasL system.

Whereas central CRH participates in the systemic endocrine inhibition of the I/I reaction, peripheral immune CRH may participate in autocrine/paracrine stimulation of inflammation. The mechanisms of the peripheral CRH-mediated component of the I/I response are still unclear, although these may be mediated by local POMC gene products with known pro-inflammatory activity and/or by inflammatory cytokines. Another possible source of peripheral

immune CRH and other neuropeptides thought to be important in inflammation (such as substance P and SMS) other than the accessory immune cells is the primary afferent (sensory) nerves as well as the sympathetic postganglionic neurons. CRH and substance P are depleted in the rat spinal cord and dorsal root ganglia in response to capsaicin, which is toxic to the sensory afferent fibers. Also, IrCRH is present in the interomedialateral sympathetic column as well as the ganglia of the sympathetic chain and, therefore, could contribute to the inflammatory process through the sympathetic postganglionic fibers. Thus, CRH is involved in opposing, and site-specific, pro- and anti-inflammatory actions.

### CRH IN THE I/I PHENOMENA OF THE FEMALE REPRODUCTIVE SYSTEM

Physiological phenomena taking place in the female reproductive tract (such as decidualization and luteolysis) bear characteristics of an aseptic inflammation. In addition, pregnancy has been explored from the immunological point of view since it can be considered a semiallograft situation.

CRH and CRH-R have been localized in the ovary in rat and in the ovary, endometrium, myometrium, and placenta in human. Further tentative actions of CRH on the female reproductive system have been suggested. CRH was shown to have an inhibitory effect on ovarian steroidogenesis, mediated through CRH and interleukin-1 receptors, and possibly linked to the processes of follicular atresia and luteolysis. Locally produced embryonic and endometrial CRH plays a role in both the aseptic inflammatory process of implantation and the anti-rejection process that protects the fetus from the maternal immune system. Early in pregnancy, the implantation sites in rat endometrium contain 3.5-fold higher concentrations of CRH than the interimplantation regions. It has been suggested that CRH of fetal and maternal origin regulates FasL production, thus affecting the invasion process through a local autocrine/paracrine regulatory loop of cytotrophoblast cells and regulating their own apoptosis. CRH decreases FasL expression in embryonic trophoblast and maternal decidua and promotes apoptosis of activated T lymphocytes. Abnormalities of maternal immune tolerance to the fetal semiallograft, and of fetal tolerance to the maternal semiallograft, have been implicated in pathological conditions of pregnancy, such as recurrent early miscarriage, preeclampsia, and eclampsia. These conditions are characterized by inflammation in the

fetal–maternal interface and/or systemic manifestations. More specifically, it has been proposed that inadequate, CRH-mediated, self-induction of FasL in extravillous trophoblasts might be involved in the pathophysiology of infertility and recurrent fetal resorption or miscarriage.

## PERSPECTIVES

Pyrolopyrimidine compounds have been developed as CRH receptor antagonists. Antalarmin has been used in investigations of the physiologic central and peripheral roles of CRH in the I/I response and reproductive function. The binding kinetics of antalarmin were determined in competitive displacement binding experiments with homogenates prepared from tissues differentially expressing CRH-R subtypes. In experiments in rats, the *in vivo* administration of antalarmin significantly antagonized both central and peripheral actions of CRH. In these experiments, the prior administration of antalarmin or neutralizing anti-CRH antibody blocked pituitary CRH receptors and the exogenous or endogenous CRH-induced ACTH release. Confirming the peripheral pro-inflammatory actions of CRH, antalarmin also suppressed the subcutaneous inflammation induced by carrageenan. Leukocyte concentrations of subcutaneous exudate were reduced by antalarmin compared to vehicle controls. This effect was dose-dependent and was comparable to that of CRH antibody, where suppression of the leukocyte concentration was observed.

In some experiments, antalarmin was administered in rats from day 1 of pregnancy and for the following 10 days. A 50% reduction was observed in the number of implantation sites in animals that received the lower dose and a 70% reduction in those that received the higher dose of antalarmin. Additionally, female rats treated with antalarmin showed diminished endometrial FasL expression. Hence, implantation in rats was prevented by antalarmin, which blocked CRH-R1. Antalarmin reduced the inflammatory-like reaction of the endometrium to the invading blastocyst. Consequently, antalarmin and analogous compounds might represent a new class of nonsteroidal inhibitors of pregnancy at its very early stages.

In sheep, hypothalamic CRH stimulates the fetal production of ACTH, which in turn leads to a surge of fetal cortisol secretion that precipitates parturition. A 10-day intravenous infusion of antalarmin in sheep fetuses significantly prolonged gestation. Thus, CRH receptor antagonism in the fetus can also delay parturition.

The therapeutic potential of pyrrolopyrimidine compounds in some forms of inflammation directly mediated by CRH-R1 is evident and hopefully will enhance the understanding of the multitude of roles that CRH plays in I/I reactions. Although the systemic toxicity of this class of compounds has not yet been fully determined, preliminary studies in rats and nonhuman primates have indicated that they are relatively safe.

## See Also the Following Articles

ACTH (Adrenocorticotrophic Hormone) • Adrenal Suppression • Chronic Fatigue Syndrome and Fibromyalgia • Corticotropin-Releasing Hormone, Family of • Corticotropin-Releasing Hormone, Placenta • Cytokine Actions, Cellular Mechanism of • Glucocorticoids and Immunity • G Protein-Coupled Receptors • Immune System, Hormonal Effects on • Interleukin-6

## Further Reading

- Chrousos, G. P. (1995). The hypothalamic–pituitary–adrenal axis and the immune/inflammatory reaction. *N. Engl. J. Med.* **332**, 1351–1362.
- Crofford, L. J., Sano, H., Karalis, K., *et al.* (1993). Corticotropin-releasing hormone in synovial fluids and tissues of patients with rheumatoid arthritis and osteoarthritis. *J. Immunol.* **151**, 1–10.
- Ghizzoni, L., Mastorakos, G., Vottero, A., *et al.* (1997). Corticotropin-releasing hormone (CRH) inhibits steroid biosynthesis by cultured human granulosa–lutein cells in a CRH and interleukin-1 receptor-mediated fashion. *J. Clin. Endocr. Metab.* **138**, 4806–4811.
- Gionis, D., Ilias, I., Moustaki, M., *et al.* (2003). Hypothalamic–pituitary–adrenal axis and interleukin-6 activity in children with head trauma and syndrome of inappropriate secretion of anti-diuretic hormone. *J. Pediatr. Endocr. Metab.* **16**, 49–54.
- Karalis, K., Sano, H., Redwine, J., Listwak, S., Wilder, R. L., and Chrousos, G. P. (1991). Autocrine or paracrine inflammatory actions of corticotropin-releasing hormone *in vivo*. *Science* **254**, 421–423.
- Makriganakis, A., Zoumakis, E., Kalantaridou, S., *et al.* (2001). Corticotropin-releasing hormone promotes blastocyst implantation and early maternal tolerance. *Nat. Immunol.* **2**, 1018–1024.
- Mastorakos, G., Chrousos, G. P., and Weber, J. S. (1993). Recombinant interleukin-6 activates the hypothalamic–pituitary–adrenal axis in humans. *J. Clin. Endocrinol. Metab.* **77**, 1690–1694.
- Mastorakos, G., Scopa, C. D., Kao, L. C., *et al.* (1996). Presence of immunoreactive corticotropin-releasing hormone in human endometrium. *J. Clin. Endocrinol. Metab.* **81**, 1046–1050.
- Mastorakos, G., Silver, P. B., Bouzas, E. A., Caspi, R. R., Chan, C. C., and Chrousos, G. P. (1992). Immunoreactive corticotropin releasing hormone in experimental uveitis. *Invest. Ophthalmol. Vis. Sci.* **33**, 933.
- Mastorakos, G., Weber, J. S., Magiakou, M. A., Gunn, H., and Chrousos, G. P. (1994). Hypothalamic–pituitary–adrenal axis activation and stimulation of systemic vasopressin secretion by

- recombinant interleukin-6 in humans: Potential implications for the syndrome of inappropriate vasopressin secretion. *J. Clin. Endocrinol. Metab.* **79**, 934–939.
- Vale, W., Rivier, C., Brown, M. R., *et al.* (1983). Chemical and biological characterization of corticotropin releasing factor. *Rec. Prog. Horm. Res.* **39**, 245–270.
- Vamvakopoulos, N. C., and Sioutopoulou, T. O. (1994). Human corticotropin-releasing hormone receptor gene (CRHR) is located on the long arm of chromosome 17 (17q12–qter). *Chromosome Res.* **2**, 471–473.
- Webster, E. L., Lewis, D. B., Torpy, D. J., Zachman, E. K., Rice, K. C., and Chrousos, G. P. (1996). *In vivo* characterization of antalarmin, a nonpeptide corticotropin-releasing hormone (CRH) receptor antagonist: Suppression of pituitary ACTH release and peripheral inflammation. *Endocrinology* **137**, 5747–5750.



# Corticotropin-Releasing Hormone, Family of

Robert J. Denver

University of Michigan, Ann Arbor, Michigan, United States

## Glossary

**corticotropin-releasing hormone (CRH)-binding protein** A secreted protein that binds CRH-like peptides with affinity that is equal to or higher than that for the CRH receptors.

**G protein-linked receptor** A polypeptide containing seven-transmembrane spanning domains. The ligand-binding domain is found on the extracellular side of the plasma membrane. These receptors influence the activity of intracellular second messenger systems by interacting with G proteins in the plasma membrane.

**hypophysiotropin** A hormone produced in the hypothalamus and transported to the anterior pituitary that modifies pituitary hormone secretion.

**neurosecretion** Release of neurohormones from modified nerve terminals directly into a vascular supply. Organs capable of neurosecretion are referred to as neurohemal organs.

**prohormone** Precursor polypeptide that is processed by proteolytic cleavage to biologically active peptides or hormones.

**stress** Physical or emotional strain or tension capable of disrupting homeostasis. Immediate endocrine responses to stress are characterized by increased plasma concentrations of glucocorticoids. Behavioral responses to stress include increased arousal, escape behavior, anxiety-like behavior, and decreased sexual behavior.

Corticotropin-releasing hormone [CRH; also referred to as corticotropin-releasing factor (CRF)] is a member of a family of related peptides in vertebrates that includes the fish urotensins-I, frog sauvagine, and the urocortin/stresscopin peptides. CRF was first named for its stimulatory effect on corticotropin [also known as adrenocorticotrophic hormone (ACTH)] secretion from the anterior pituitary gland. ACTH is the primary pituitary regulator of adrenal glucocorticoid biosynthesis and secretion. Members of the CRH family of peptides play central roles in the regulation of neuroendocrine, autonomic, and behavioral responses to physical and emotional stress.

## THE CRH FAMILY OF PEPTIDES

The chemical nature of the corticotropin-releasing factor activity present in the vertebrate hypothalamus remained elusive until 1981, when Wylie Vale and colleagues isolated a 41-amino acid (aa) peptide from porcine hypothalami that was capable of stimulating pituitary adrenocorticotrophic hormone (ACTH) secretion. Concurrently, Karl Lederis and colleagues isolated a 41-aa peptide with sequence similarity to corticotropin-releasing hormone (CRH) from the caudal neurosecretory organ (the urophysis) of the teleost fish, the common sucker (*Catostomus commersoni*), that they named urotensin-I (UI) based on its tissue of origin (urophysis) and its vasopressive activity. Since the early 1980s, several other CRH-like peptides have been isolated and some of their diverse physiological roles have been elucidated.

The vertebrate CRH family of peptides is composed of at least two distinct paralogous lineages, designated the CRH lineage and the UI/urocortin (UCN) lineage. CRH molecules, which are 41 aa in length, have been isolated from the brains of fishes, an amphibian, and mammals. Additional CRH-like peptides represented by the paralogous UI/UCN lineage were first identified in fishes; these peptides share approximately 50% sequence similarity with the CRHs. UCN, a 40-aa peptide, was isolated from the brains of rodents and human and shown to have 53–63% sequence similarity with fish UIs. Subsequently, two other UCN-like peptides that are 38 aa in length were predicted based on cDNA sequences isolated in mammals. These UCN-like sequences isolated from human were given the names stresscopin (SCP) and stresscopin-related peptide (SRP). Similar sequences were simultaneously isolated from mouse and designated urocortin-II (UCN-II) and urocortin-III (UCN-III). The human SCP and mouse UCN-III are homologous, as are the human SRP and the mouse UCN-II. The UCN-II and UCN-III peptides and similar sequences from pufferfishes (*Fugu rubripes* and *Tetraodon nigroviridis*) may

represent a separate, closely related lineage of the CRH family.

A 40-aa peptide that appears to be related to the UI/UCN lineage is sauvagine (SV), which was isolated from the skin of the frog *Phyllomedusa sauvagei*. Because this peptide has only been isolated from one frog species, and other UI/UCN-like peptides have not been identified in amphibians, the precise phylogenetic relationship of SV to the UI/UCN lineage is uncertain.

All CRH-like peptides are initially synthesized as a larger prohormone that is cleaved by prohormone convertases at dibasic residues to form the mature peptide; mature CRH-like peptides range from 38 to 41 aa residues. The cryptic peptide (N-terminal region of the prohormone) is cosecreted with CRH but has no known biological function. It is hypothesized that the cryptic peptide of the prohormone plays a role in intracellular protein folding and perhaps targeting to the secretory pathway.

## MECHANISMS OF ACTION OF CRH-LIKE PEPTIDES: CRH RECEPTORS AND BINDING PROTEIN

The actions of CRH are mediated by specific receptors expressed in target cells and localized to the plasma membrane. In mammals, a number of distinct CRH receptor isoforms deriving from two paralogous genes and one secreted binding protein (CRH-BP) have been identified. The receptors are seven-transmembrane domain G-linked proteins that transduce extracellular signals by stimulating intracellular cAMP production. The receptors exhibit differential rank order affinities for CRH peptides and tissue-specific patterns of expression. The first CRH receptor isolated in mammals, designated CRHR<sub>1</sub>, is a 415-aa protein and was originally thought to not discriminate among the CRH or UI/UCN lineages (i.e., it possessed high affinity for all). However, findings show that CRHR<sub>1</sub> exhibits very low affinity for SRP/UCN-II and no measurable affinity for SCP/UCN-III (these peptides are selectively bound by CRHR<sub>2</sub>). CRHR<sub>1</sub> is expressed in the anterior and intermediate lobes of the pituitary and in numerous sites throughout the brain.

A second CRH receptor subtype, CRHR<sub>2</sub>, shares 70% sequence similarity with CRHR<sub>1</sub>. The CRHR<sub>2</sub> has two splicing variants in rodents and three splicing variants in humans (humans: CRHR<sub>2α</sub>, 411 aa;

CRHR<sub>2β</sub>, 431 aa; CRHR<sub>2γ</sub>, 397 aa). The CRHR<sub>2</sub> proteins possess significantly higher affinity for UI/UCN-like peptides than for CRHs. Significantly, SRP/UCN-II and SCP/UCN-III are selectively bound by the CRHR<sub>2</sub> proteins. In rodents, CRHR<sub>2α</sub> is expressed in the brain within the lateral septal nuclei and regions of the hypothalamus, whereas CRHR<sub>2β</sub> is expressed in peripheral tissues (e.g., lung, skeletal muscle, ovary, cardiac myocytes, and gastrointestinal tract). In humans, all three CRHR<sub>2</sub> splicing variants are expressed in the brain, but only CRH-R<sub>2α</sub> is expressed in peripheral tissues.

Studies of transgenic mice that lack one of the two CRH receptor genes support the hypothesis that the molecules subserve different physiological functions. Compared with wild type, CRHR<sub>1</sub> knockout (ko) mice exhibit lower plasma corticosterone concentrations, blunted ACTH responses to stress, altered adrenal morphology, and decreased anxiety-like behavior. CRHR<sub>2</sub> ko mice show altered adaptation to stress, increased anxiety-like behavior, and impaired cardiovascular function. Further advances in our understanding of CRH receptor biology have been possible due to the development of selective, nonpeptide CRHR<sub>1</sub> antagonists. For example, several pyrolopyrimidine compounds (e.g., antalarmin) have been found to selectively block CRH binding to CRHR<sub>1</sub> in a noncompetitive manner; this compound does not affect binding to CRHR<sub>2</sub> or the binding of other ligands to other G protein-linked receptors. Work on peptide-based CRHR<sub>2</sub>-specific antagonists shows that this receptor can also be specifically targeted using pharmacological approaches.

A secreted CRH-binding protein (CRH-BP) was first isolated from human blood plasma. The structure of this protein has been elucidated in several mammalian species and in the frog *Xenopus laevis*. The protein ranges from 321 to 324 aa and exhibits a high degree of sequence similarity among species. Biochemical evidence indicates the presence of a CRH-BP in the brains of representatives of each vertebrate class. It is hypothesized that the CRH-BP modulates the bioavailability of CRH-like peptides. The CRH-BPs that have been characterized bind CRH-like peptides with affinities that are equal to or higher than those for the CRH receptors. This and other findings have led to the hypothesis that the primary function of the CRH-BP is to neutralize CRH activity by binding the peptide and making it unavailable to bind to receptors. It is also possible that the CRH-BP could function in targeting the peptide for clearance. Alternatively, the CRHBP could function to maintain

high levels of CRH in the blood or within tissues, thus facilitating the action of the peptide. The CRH-BP is expressed at different sites depending on the species (i.e., brain, liver, pituitary, intestine, and placenta), but all species studied express the protein in the pituitary and brain. Placental expression of the CRH-BP in humans has been implicated in regulating CRH bioavailability during late gestation; CRH is thought to play a critical role in the timing of parturition in mammals. Finally, studies of transgenic mice that either overexpress the CRH-BP or lack the CRH-BP gene indicate an important role for this binding protein in modulating CRH bioavailability.

## EXPRESSION AND PHYSIOLOGICAL ACTIONS OF CRH-LIKE PEPTIDES

The expression of vertebrate CRH-like peptides in various brain regions and in peripheral tissues is consistent with their primary role in physiological and behavioral responses to stress. The synthesis and release of CRH increases in response to stress, and the peptide is implicated in the rapid increase in glucocorticoid biosynthesis by the adrenal glands, suppression of feeding, suppression of immune function, and enhanced locomotion, among others. The expression of CRH in neurosecretory neurons in the hypothalamus and the stimulatory action of CRH on pituitary ACTH secretion are common to all vertebrates. CRH is synthesized in hypothalamic nuclei (parvocellular neurons of the paraventricular nucleus in mammals) and is released at modified nerve terminals in the median eminence into the pituitary portal circulation. CRH is transported to the anterior pituitary gland, where it binds to its receptors and controls the secretion of pituitary hormones. CRH is also a potent releasing factor for thyroid-stimulating hormone in nonmammalian species. This thyroid stimulatory role for CRH has been implicated in the control of amphibian metamorphosis, which is a thyroid-dependent process. CRH also regulates  $\alpha$ -melanocyte-stimulating hormone secretion in lower vertebrates, thus playing a role in skin pigmentary changes associated with background adaptation.

CRH-like peptides are expressed in brain regions outside of the neurosecretory cells of the hypothalamus, including the cortex, limbic system, and brainstem nuclei that are associated with autonomic function. Although there are differences in expression patterns among species, the basic brain region-specific patterns of expression are conserved among vertebrates. Prominent expression of CRH in mammals

is observed in the cerebral cortex, tegmentum, amygdala/hippocampus (important for stress adaptation), brainstem (locus coeruleus), and spinal cord. Peripheral sites of CRH expression include the gut, spleen, thymus, skin, adrenal gland, and placenta. Expression of UCN tends to be more restricted than that of CRH, being found in the Edinger–Westphal nucleus of the tegmentum, the hypothalamus, and a small group of neurons in the telencephalon. Mouse UCN-II is expressed in stress-related cell groups in the hypothalamus and brainstem, and UCN-III is expressed in the hypothalamus and the medial amygdala.

The expression of CRH-like peptides in extrahypothalamic neural circuits has been implicated in their behavioral effects, where they are thought to function as neurotransmitters/neuromodulators, thus integrating behavioral and physiological responses to stress. In mammals, the actions of CRH peptides, in addition to their hypophysiotropic role, include control of appetite (to decrease food intake), behavioral responses to stress (arousal, escape, anxiety-like behavior, and diminished sexual behavior), enhancement of learning, alterations in cardiovascular function, and modulation of immune responses. Thus, in addition to their central role in the endocrine stress response through activation of the hypothalamic–pituitary–adrenal axis, CRH-like peptides also serve to integrate the autonomic and behavioral responses to stress via their actions in the central nervous system. CRH-like peptides are potent anxiogenic and anorectic agents, and their aberrant expression has been implicated in depressive, anxiety-related, and eating disorders.

## See Also the Following Articles

ACTH (Adrenocorticotrophic Hormone) • Corticotropin-Releasing Hormone (CRH) and Inflammation • Corticotropin-Releasing Hormone, Placenta • Stress and Endocrine Physiology

## Further Reading

- Challis, J. R. G., and Smith, S. K. (2001). Fetal endocrine signals and preterm labor. *Biol. Neonate* **79**, 163–167.
- Dautzenberg, F. M., and Hauger, R. L. (2002). The CRF peptide family and their receptors: Yet more partners discovered. *Trends Pharmacol. Sci.* **23**, 71–77.
- Denver, R. J. (1999). Evolution of the corticotropin-releasing hormone signaling system and its role in stress-induced phenotypic plasticity. *Neuropeptides Structure Function Biol. Behav.* **897**, 46–53.

- Lovejoy, D. A., and Balment, R. J. (1999). Evolution and physiology of the corticotropin-releasing factor (CRF) family of neuropeptides in vertebrates. *Gen. Comp. Endocrinol.* **115**, 1–22.
- Seasholtz, A. F., Burrows, H. L., Karolyi, I. J., and Camper, S. A. (2001). Mouse models of altered CRH-binding protein expression. *Peptides* **22**, 743–751.
- Seasholtz, A. F., Valverde, R. A., and Denver, R. J. (2002). Corticotropin-releasing hormone-binding protein (CRH-BP): From fishes to mammals. *J. Endocrinol.* **175**, 89–97.
- Speert, D. B., and Seasholtz, A. F. (2001). Corticotropin-releasing hormone receptors and ligands in stress: Who's on first? *Curr. Opin. Endocrinol. Diabetes* **8**, 161–165.



# Corticotropin-Releasing Hormone, Placenta

Katia P. Karalis and Joseph A. Majzoub  
*Children's Hospital, Boston, Massachusetts, United States*

## Glossary

**cytotrophoblast, syncytiotrophoblast** Placenta-specific cell types. The former refers to single cells that constitute the predominant trophoblast cell in the first part of pregnancy and are replaced by syncytiotrophoblast as pregnancy advances. Syncytiotrophoblast is the main secretory trophoblast cell.

**hypothalamic–pituitary–adrenal (HPA) axis** The main endocrine axis mediating the stress response. Activation of the HPA axis leads to stimulation of hypothalamic corticotropin-releasing hormone, which stimulates pituitary ACTH secretion, leading to glucocorticoid secretion from the adrenal.

Corticotropin-releasing hormone (CRH) is a 41-amino acid neuropeptide identified in the central nervous system as well as in several peripheral tissues. CRH has been found in the placentae of different primates. In humans, CRH can be measured in the maternal circulation, with very high levels as labor approaches, suggesting its usefulness as a marker of the timing of parturition. In this article, the regulation of placental CRH during human pregnancy, its maternal effects, and its putative significance for the fetus are discussed.

## INTRODUCTION

Corticotropin-releasing hormone (CRH), first identified in the hypothalamus as one of the major regulators of the hypothalamic–pituitary–adrenal (HPA) axis activation in response to stress, is also found in a variety of extrahypothalamic sites, including placenta. Among mammals, CRH is expressed only in primate placenta. However, the gorilla and chimpanzee are the only primates in which the pattern of secretion of placental CRH during pregnancy resembles that of humans. In humans, CRH is detectable in maternal circulation at the end of the first trimester of pregnancy, with levels increasing up to 100-fold during the last month of pregnancy and declining to nondetectable levels postpartum. The pattern of

secretion of placental CRH into maternal blood is pulsatile, with a constant pulse frequency throughout gestation and increasing pulse amplitude as pregnancy advances. What stimulates the expression of placental CRH in primates but not in other mammals, as well as the basis for the differential expression and secretion patterns among primates, remains unclear.

## HUMAN PLACENTAL CRH: EXPRESSION AND REGULATION

Hypothalamic and placental CRH are products of the same gene, which is well conserved among different vertebrates. CRH has been identified in cytotrophoblasts in placentae from first- and second-trimester pregnancies. During the third trimester of pregnancy, placental CRH expression is abundant in the syncytiotrophoblast cells, compatible with the high levels of circulating CRH during that gestational period. In humans, significant CRH expression has been found in other tissues of the maternal–fetal unit, including the amnion, chorion, and decidua.

A variety of factors shown to regulate hypothalamic CRH gene expression have similar effects on the placental CRH gene. Thus, interleukin-1, prostaglandins, and catecholamines stimulate, whereas sodium nitroprusside inhibits, its release. Glucocorticoid, an inhibitor of hypothalamic CRH synthesis and secretion, paradoxically stimulates placental CRH mRNA and peptide expression. Progesterone, a steroid synthesized by trophoblasts, also inhibits placental CRH gene expression. Progesterone and glucocorticoid are antagonistic steroids based on their competition for binding on the glucocorticoid receptor. The stimulatory effect of glucocorticoid on placental CRH may be explained by the reversal of the inhibitory effect of progesterone in trophoblasts, which express glucocorticoid receptor but not progesterone receptor. Furthermore, mifepristone (RU 486), another antiprogesterin, also stimulated placental CRH gene expression.

In humans, increased placental CRH expression parallels the increasing level of cortisol in late gestation, suggestive of a positive feedback relationship between placental CRH and glucocorticoid, unlike the negative feedback loop operating between the latter and hypothalamic CRH. Administration of  $\beta$ -methasone, a potent glucocorticoid analog resistant to placental inactivation, in pregnancies less than 32 weeks' gestational age, results in suppression of placental CRH secretion, whereas in more advanced pregnancies suppression is not found and stimulation may occur. The change in the response of placental CRH to glucocorticoid during gestation might be related to the relative level of progesterone associated with the gestational age, consistent with *in vitro* findings.

### PARADIGMS OF ABNORMAL PLACENTAL CRH EXPRESSION IN HUMANS

The highest expression of CRH is achieved as labor approaches, and it decreases to prepartum levels within 24 h following delivery, supporting the placental origin of circulating CRH during pregnancy. Inappropriately high CRH levels, compared to those in gestational age-matched normal pregnancies, have been described in pregnant women with pregnancy-induced hypertension or in intrauterine growth retardation. In several cases of complicated pregnancies, such as pregnancy-induced hypertension, increased glucocorticoid levels coexisted with inappropriately high for the gestational age placental CRH levels. In preterm labor, both high and normal levels of CRH have been reported; this discrepancy is most likely attributed to the wide spectrum of pathology associated with this condition.

### MATERNAL EFFECTS OF PLACENTAL CRH

CRH made in the placenta is secreted into both the fetal and the maternal circulation. Biological actions of CRH in humans may be limited by its binding to a high-affinity binding protein, CRH-binding protein (CRH-BP), an approximately 40-kDa glycoprotein expressed in the brain of all species studied. The presence of high amounts of CRH-BP in human maternal plasma up to the last weeks of gestation may serve to protect the maternal pituitary from the increased circulating CRH levels and thus to prevent development of Cushingoid features. The mild Cushingoid appearance of pregnant women might

result from the escape of CRH from binding by the CRH-BP, which could thereby stimulate maternal pituitary ACTH release and cortisol secretion and contribute to pregnancy-associated hypercortisolemia. In further support of a biological effect of placental CRH in maternal physiology is the suggestion that the postpartum decline in maternal HPA axis activity might result from the withdrawal of placental CRH and possibly account for the occurrence of postpartum mood changes.

In addition to its synthesis in placenta, CRH is expressed in the decidua and the fetal membranes. CRH has been reported to have synergistic effects with oxytocin or prostaglandin on the induction of myometrial contractility. It has also been shown to cause vasodilation of the uterine vasculature through a direct effect on smooth muscle contractility. Thus, CRH may directly affect the mechanical component of parturition, in addition to its regulatory effect on the hormonal pathways related to the initiation of labor.

### PUTATIVE SIGNIFICANCE OF PLACENTAL CRH

Unlike the placentae of nonprimate mammals such as the sheep, the human placenta does not express P450 17,20-lyase; thus, the increased fetal cortisol production in late human pregnancy is unable to inhibit the production of placental progesterone. Since the human placenta cannot directly synthesize estradiol, the fetal adrenal zone is the predominant source of its precursor, dehydroepiandrosterone (DHEA). Fetal adrenal DHEAS secretion, leading to placental estradiol production, likely plays a major role in the initiation of labor via stimulation of oxytocin, oxytocin receptor, gap junction, and prostaglandin synthesis. Placental CRH may have evolved in primates to stimulate fetal adrenal activity and possibly ACTH secretion, thereby satisfying the high demand for DHEA synthesis.

The increase in fetal cortisol secretion is necessary for the maturation of the fetal lung and other systems required for extrauterine survival. Consistent with the hypothesis that placental CRH stimulates the fetal pituitary-adrenal axis is the parallel increase in placental CRH and fetal cortisol levels in the last weeks of gestation. It is possible that placental CRH acts directly on fetal adrenals, in addition to its stimulatory effect on fetal pituitary. This hypothesis is supported by a report describing the expression of CRH-R1 mRNA in early midgestation human fetal adrenal tissue as well as *in vitro* evidence for direct stimulation of steroidogenesis by CRH.

## CONCLUSION

CRH is expressed in the primate placenta. In humans, its expression is significantly increased in the last trimester of pregnancy. Placental CRH expression is further stimulated in pregnancies associated with preterm labor, such as when complicated by intrauterine growth retardation or pregnancy-induced hypertension, or in twin pregnancies. Placental CRH may be important for both the maturation of the fetus and the completion of gestation.

## See Also the Following Articles

Corticotropin-Releasing Hormone (CRH) and Inflammation • Corticotropin-Releasing Hormone, Family of • Feto-Placental Unit • Pregnancy Endocrinology

## Further Reading

- Challis, J. R., Lye, S. J., Gibb, W., Whittle, W., Patel, F., and Alfaidy, N. (2001). Understanding preterm labor. *Ann. N. Y. Acad. Sci.* **943**, 225–234.
- Karalis, K., Goodwin, G., and Majzoub, J. A. (1996). Cortisol blockade of progesterone: A possible molecular mechanism involved in the initiation of human labor. *Nature Med.* **2**(5), 556–560.
- Karteris, E., Grammatopoulos, D. K., Randevara, H. S., and Hillhouse, E. W. (2001). The role of corticotropin-releasing hormone receptors in placenta and fetal membranes during human pregnancy. *Biol. Reprod.* **64**(4), 1019–1032.
- Magiakou, M. A., Mastorakos, G., Rabin, D., Margioris, A. N., Dubbert, B., Calogero, A. E., Tsigos, C., Munson, P. J., and Chrousos, G. P. (1996). The maternal hypothalamic–pituitary–adrenal axis in the third trimester of human pregnancy. *Clin. Endocrinol.* **44**(4), 419–428.
- Majzoub, J. A., McGregor, J. A., Lockwood, C. J., Smith, R., Taggart, M. S., and Schulkin, J. (1999). A central theory of preterm and term labor: Putative role for corticotropin-releasing hormone. *Am. J. Obstet. Gynecol.* **180**(1 Pt. 3), S232–S241.
- McLean, M., Bisits, A., Davies, J., Woods, R., Lowry, P., and Smith, R. (1995). A placental clock controlling the length of human pregnancy. *Nature Med.* **1**(5), 460–463.
- Whittle, W. L., Patel, F. A., Alfaidy, N., Holloway, A. C., Fraser, M., Gyomory, S., Lye, S. J., Gibb, W., and Challis, J. R. (2001). Glucocorticoid regulation of human and ovine parturition: The relationship between fetal hypothalamic–pituitary–adrenal axis activation and intrauterine prostaglandin production. *Biol. Reprod.* **64**, 1019–1032.



# Craniopharyngiomas

David Anthony Price

Royal Manchester Children's Hospital, Manchester, United Kingdom

## Glossary

**chondrification** The formation of cartilage or a conversion into cartilage.

**fossa** A trench or channel.

**papilloedema** Swelling of the optic disk, often due to increased intracranial pressure.

Craniopharyngiomas are intracranial tumors derived from remnants of Rathke's pouch, an embryological structure formed as a diverticulum of the stomodeum. These tumors produce clinical symptoms that are caused by increased intracranial pressure, damage to contiguous neural structures, and destruction of the pituitary gland. Although nonmalignant in nature, they are locally invasive, recurrent, and frequently a mixture of solid and cystic components. Cushing introduced the term craniopharyngioma in 1932.

## EMBRYOLOGY

In the fourth week of embryological life, a diverticulum is formed in the roof of the primitive mouth cavity and meets a downgrowth from the floor of the forebrain. The diverticulum is termed the hypophysial pouch, ectodermal in origin, and the precursor of the anterior pituitary gland. The forebrain derivative is the neurohypophysial bud, neuroectodermal in origin, and will form the posterior pituitary gland, the infundibular stem, and the median eminence.

The ectodermal diverticulum is also known as Rathke's pouch and its stalk passes between the chondrification centers of the presphenoid and basisphenoid bones. The connection between Rathke's pouch and the pharynx degenerates in the sixth week, although a remnant, the basipharyngeal canal, may be detected in the sphenoid bone at birth. During the formation of the pituitary, there is an apparent rotation resulting in other remnants of Rathke's pouch lying above the gland and pituitary fossa.

It is from these remnants that craniopharyngiomas are considered to arise.

## INCIDENCE

The incidence of new cases of craniopharyngioma is 1 to 2 per 1 million per year, and they account for 1–5% of all intracranial tumors. They occur more commonly in children and adolescents than in adults. No consistent sex bias has been reported, but some studies have shown a slight male preponderance.

## CLINICAL PRESENTATION

Symptoms and signs of increased intracranial pressure, such as headache, vomiting, and papilloedema, may be the presenting features. Visual disturbances, especially visual field defects, and features of endocrine dysfunction are common. Children are more prone to endocrine problems, for example, poor growth and delayed or arrested puberty. Symptoms of diabetes insipidus can present but may be obscured by concomitant adrenal insufficiency. There may be cognitive dysfunction and behavioral and personality changes. Owing to the subtle and insidious nature of endocrine and cognitive changes, there is often a protracted period of symptomatology that can last for years before diagnosis.

## RADIOLOGICAL FEATURES

Children are very likely to have abnormalities detected on plain skull radiographs that include enlargement, erosion, and destruction of the sella turcica and tumor calcification. Adults also show abnormalities of the sella and tumor calcification but to a lesser extent, which probably reflects the relative preponderance of adamantinomatous and papillary types of tumor. Computed tomography (CT) and magnetic resonance (MR) scans demonstrate localization and characteristics of these tumors. Most are suprasellar



**Figure 1** MR scan with gadolinium (T<sub>1</sub> weighted) of craniopharyngioma in a child. Solid components are white. The tumor is mainly above the sella and extends behind the clivus, invading the hypothalamus above and displacing the pons behind. Tumor extends into the sella.

with intrasellar elements. CT scans show any calcification and, with contrast, the solid component and capsule of the cystic component. MR scans (Fig. 1) also allow distinction between the solid and cystic components.

## HISTOLOGY

In the World Health Organization's "International Histological Classification of Tumors," craniopharyngiomas are grade I. Two histological types are described, adamantinomatous and papillary.

Adamantinomatous craniopharyngiomas have a solid component that is a loosely structured, multi-layered squamous epithelium and has some similarity with developing enamel cells or adamantoblasts. Nodular wet keratin is typical and has a tendency to calcify. The cystic element is usually multilocular but can be singular, and it is mostly lined by a stratified squamous epithelium. Cysts contain cholesterol,

and the viscous yellow or brown fluid is variously described as motor oil or engine oil.

Papillary craniopharyngiomas consist of epithelial strands forming papillae and lack calcification and the oily fluid. This histological type is rarely seen in children.

## TUMOR MANAGEMENT

Surgery and irradiation are the main forms of treatment; the role of chemotherapy is not established. Surgical interventions are dependent on localization and size of the tumor. Resection may be partial or radical, and ventricular shunting is needed if there is secondary hydrocephalus. Recurrence can occur after apparent total resection. Irradiation may be given if there is residual tumor after operation. The role of postoperative irradiation is debatable; there are strong advocates for its use to prevent or delay recurrence. Of major concern is the severe morbidity following hypothalamic damage after surgery, and many authorities prefer more limited resection of large tumors invading the hypothalamus followed by irradiation.

## MEDICAL MANAGEMENT

Frequently, there are endocrine deficiencies before primary tumor treatment, and they are almost universal afterwards. Therefore, hormone replacement treatment (thyroxin, glucocorticoid, sex steroids, growth hormone, and vasopressin analog) is needed for hypopituitarism. Postoperative fluid balance must be maintained. Introduction of growth hormone treatment in children is often delayed until postoperative growth can be assessed. Some children grow quickly after operation even though they are deficient of growth hormone, and this phenomenon is usually associated with hyperphagia and weight gain. There is no evidence that hormonal replacement causes earlier or more frequent recurrence.

Educational and other assistance for the visually impaired may be needed. Cognitive dysfunction, especially memory impairment, needs to be addressed using psychological strategies.

## PROGNOSIS

Ten-year survival rates have been reported by several centers and vary from 60 to 93%. Although these rates may seem to be good, the underlying tumor is non-malignant and quality of life in survivors can be seriously impaired. Excluding endocrine dysfunction,

approximately two-thirds of childhood craniopharyngioma survivors have major morbidity, often visual. Hypothalamic damage in children is particularly debilitating. Extensive resection of large tumors is a bad prognostic factor for survival and for good quality of life.

Standardized mortality rates (SMRs) are increased 5-fold, with a higher risk for females (11-fold), and there is a 3-fold increase in cardiovascular and cerebrovascular mortality. When compared with other pituitary tumors causing hypopituitarism, those with craniopharyngioma had higher SMRs.

### See Also the Following Articles

Delayed Puberty and Hypogonadism, Female • Pineal Tumors • Pituitary Region, Non-Functioning Tumors of • Pituitary Tumors, Clonality • Pituitary Tumors, Molecular Pathogenesis • Pituitary Tumors, Surgery

### Further Reading

*Allgemeine Deutsche Biographie* 27, 352–355 (1888). Verlag von Duncker & Humblot, Leipzig.  
Anderson, C. A., Wilkening, G. N., Filley, C. M., Reardon, M. S., and Kleinschmidt-DeMasters, B. K. (1997). Neurobehavioral

outcome in pediatric craniopharyngioma. *Pediatr. Neurosurg.* 26, 255–260.

Banna, M. (1976). Craniopharyngioma: Based on 160 cases. *Br. J. Radiol.* 49, 206–223.

Bulow, B., Attewell, R., Hagmar, L., Malmstrom, P., Nordstrom, C.-H., and Erfurth, E. M. (1998). Postoperative prognosis in craniopharyngioma with respect to cardiovascular mortality, survival, and tumor recurrence. *J. Clin. Endocrinol. Metab.* 83, 3897–3904.

Cushing, H. (1932). Craniopharyngiomas. In "Intracranial Tumors," pp. 93–98. Bailliere, Tindall & Cox, London.

De Vile, C. J., Grant, D. B., Hayward, R. D., Kendall, B. E., Neville, B. G. R., and Stanhope, R. (1996). Obesity in childhood craniopharyngioma: Relation to post-operative hypothalamic damage shown by magnetic resonance imaging. *J. Clin. Endocrinol. Metab.* 81, 2734–2737.

Graham, P. H., Gattamaneni, H. R., and Birch, J. M. (1992). Paediatric craniopharyngiomas: A regional review. *Br. J. Neurosurg.* 6, 187–194.

Janzer, R. C., Burger, P. C., and Giangaspero, F. (1997). Craniopharyngioma. In "Pathology and Genetics of Tumors of the Nervous System" (P. Kleihaus and W. K. Cavenee, eds.). International Agency for Research on Cancer, Lyon, France.

Sorva, R. (1988). Children with craniopharyngioma. *Acta Paediatr. Scand.* 77, 587–592.

Tomlinson, J. W., Holden, N., Hills, R. K., Wheatley, K., Clayton, R. N., Bates, A. S., Sheppard, M. C., and Stewart, P. M. (2001). Association between premature mortality and hypopituitarism. West Midlands Prospective Hypopituitary Study Group. *Lancet* 357, 425–431.



Cryptorchidism

see Undescended Testes



# Cytokine Actions, Cellular Mechanism of

Michael David

University of California, San Diego, California, United States

## Glossary

**cytokines (also known as lymphokines and monokines)** Low-molecular-weight, pharmacologically active ligands with diverse functions secreted by one cell for the purpose of altering either its own functions (autocrine effect) or those of adjacent cells (paracrine effect).

**inflammation** A potentially harmful process involving diverse cell types and soluble mediators, including cytokines, by which an organism counteracts potentially injurious insults, such as bacteria, viruses, and other pathogens.

**signal transduction** The perpetuation of ligand-receptor interaction-derived instructions through intracellular second messenger molecules or enzyme cascades.

Cytokines represent a group of transiently expressed, low-molecular-weight molecules that play a vital role in intercellular communication and regulate numerous physiological and pathological events, such as lymphocyte development and activation, inflammatory processes, and programmed cell death.

## INTRODUCTION

Research during the past decade has led to the discovery of many new cytokines, their sources, and respective cell surface receptors. The elucidation of their signaling pathways has facilitated a better understanding of cytokine receptor-mediated gene induction. In addition to their complex biological effects, cytokine-induced transcriptional activation is the product of an intricate network of multiple signaling cascades.

This article presents an overview of cytokine-mediated transcriptional responses mediated through STAT proteins, NF- $\kappa$ B, and the MAPK/SAPK cascades, signaling pathways widely used among cytokine receptors.

## OVERVIEW OF CYTOKINE SIGNALING

Two fundamental concepts are crucial for understanding the complexity of cellular responses to cytokine stimulation. First, the generation of opposing signals through multiple pathways emanating from the same receptor is common in cytokine-initiated signal transduction. This concept is illustrated by the conversion of the antiproliferative actions of interferon- $\gamma$  (IFN- $\gamma$ ) into a mitogenic response in the absence of STAT1 or by the simultaneous activation of the caspase cascade as well as the antiapoptotic NF- $\kappa$ B pathway by the tumor necrosis factor- $\alpha$  (TNF- $\alpha$ ) receptor. Second, every cell encounters a myriad of stimuli simultaneously *in vivo*, and it is the integration of the signaling events initiated by each individual receptor that ultimately determines the final cellular response. Synergistic amplification or antagonistic suppression of the biological effects of two simultaneously acting cytokines are exemplified by the augmenting effect of IFN- $\gamma$  on the IFN- $\gamma$  response or attenuating consequences of IFN- $\gamma$  on TGF- $\alpha$  stimulation, respectively.

## THE JAK/STAT PATHWAY

The discovery of the Jak/STAT pathway was the result of investigations on the mechanism by which interferons activate immediate-early response genes. Although research during the past decade has demonstrated that the Jak/STAT pathway is activated by a diverse group of cytokine and growth factor receptors, it is predominantly members of the hematopoietin subfamily of cytokines acting through class I and class II receptors that use this signaling cascade.

### Jak/STAT Family Members

The first STAT proteins were identified as the regulatory cytoplasmic components of the IFN- $\gamma$ -induced transcription factor complex, interferon-stimulated response element-3 (ISGF3). Two splice variants of STAT1, STAT2 and the DNA-binding subunit p48,

Table I Activation of Jak/STAT Family Members by Various Cytokines

Receptor family	Tyk2	Jak1	Jak2	Jak3	STAT1	STAT2	STAT3	STAT4	STAT5	STAT6
<b>IFN<sup>a</sup></b>										
IFN- $\gamma$	+	+	-	-	+	+	+	+	+	+
IFN- $\alpha$	-	+	+	-	+	-	-	-	+	-
<b>gp130</b>										
LIF	+	+	+	-	+	-	+	-	-	-
G-CSF	-	+	-	-	+	-	+	-	+	-
IL-6	+	+	+	-	+	-	+	-	+	-
IL-11	-	+	-	-	+	-	+	-	-	-
IL-12	+	-	+	-	-	-	+	+	-	-
CNTF	+	+	+	-	+	-	+	-	-	-
OSM	+	+	+	-	+	-	+	-	-	-
<b>Common <math>\alpha</math> chain</b>										
IL-2	-	+	-	+	+	-	+	-	+	-
IL-4	-	+	-	+	+	-	+	-	-	+
IL-7	-	+	-	+	-	-	-	-	+	-
IL-9	-	+	-	+	-	-	-	-	+	-
IL-13	-	+	-	-	?	?	?	?	?	?
IL-15	-	+	-	+	-	-	-	-	+	-
<b>Common <math>\beta</math> chain</b>										
IL-3	-	-	+	-	-	-	-	-	+	-
IL-5	-	-	+	-	-	-	-	-	+	-
GM-CSF	-	-	+	-	-	-	-	-	+	-

<sup>a</sup>Abbreviations used; IFN, interferon; LIF; G-CSF, granulocyte colony-stimulating factor; IL, interleukin; CNTF; OSM; GM-CSF, granulocyte macrophage colony-stimulating factor.

were found to bind the ISRE enhancer as a multiprotein complex in response to IFN- $\gamma$ . Alternatively, STAT1 homodimers activated via the IFN- $\gamma$  receptor can bind a distinct element called GAS (IFN- $\gamma$ -activated sequence). Several other cytokines and growth factors were reported to activate STAT-like DNA-binding proteins (Table I), and subsequent identification of these proteins resulted in the cloning of five additional mammalian STAT genes (*STAT3*, *STAT4*, *STAT5A*, *STAT5B*, and *STAT6*). Several of the STAT proteins are represented by various splice variants; however, *STAT5A* and *STAT5B* are encoded by two distinct genes. In contrast to the more restricted expression of *STAT4* in spleen, heart, brain, peripheral blood cells, and testis, most STAT proteins are rather ubiquitous. Several conserved structural and functional domains characterize all seven mammalian STAT proteins as well as their *Drosophila* and *Dictyostelium* orthologs. A carboxy-terminal tyrosine residue, which serves as the target for ligand-induced phosphorylation, is crucial for dimerization, nuclear translocation, and DNA binding of STAT molecules. The src-homology 2 (SH2) domain of STAT proteins facilitates homo- or heterodimerization via reciprocal interaction with the

phosphorylated tyrosine residue. In addition, the SH2 domain facilitates the specific recruitment of STAT molecules to activated, tyrosine phosphorylated cytokine and growth factor receptors as well as Jak tyrosine kinases. Dimerization of STAT proteins allows the centrally located DNA-binding domains to interact with consensus palindromic sequences that differ only in their core nucleotide sequence. In several STAT proteins, the region C terminal to the tyrosine phosphorylation site harbors the transactivation domain (TAD), whose posttranslational modification through serine phosphorylation, presumably by members of the ERK/SAPK kinase family, accounts for the regulated interaction of the STAT-TAD domains with transcriptional coactivators such as CBP/p300 or MCM5.

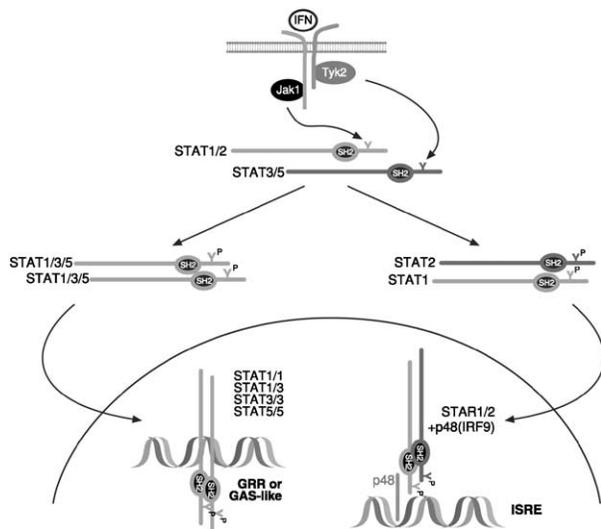
### Mechanism of Activation

It was again the paradigm of the interferon signaling cascade that led to the discovery of the participation of the Janus tyrosine kinases in STAT-mediated gene transcription. Disrupted expression of Tyk2 or Jak1 resulted in IFN- $\gamma$  unresponsive cells, whereas similar

abrogation of Jak1 and Jak2 expression caused the loss of IFN- $\gamma$  responsiveness. Contrary to the ubiquitous expression of these kinases, a fourth family member, Jak3, was identified whose expression is restricted to cells of hematopoietic origin. The most striking feature in the structure of Jak family members is the presence of a kinase-like domain and the absence of any SH2- or SH3-like structures that are commonly found in cytoplasmic tyrosine kinases. The relatively large (120–140 kDa) Jak kinases are constitutively associated with many cytokine and growth factor receptors. Nevertheless, this association can be increased after cytokine stimulation.

The general outline of the sequential events that led to STAT-mediated gene transcription has been derived mostly from evidence obtained from the interferon signaling paradigm (Fig. 1). It is believed that engagement of ligands to their respective receptors results in an increased local concentration of Jak proteins due to receptor aggregation and increased affinity of the receptors for Jak kinases. Subsequent transphosphorylation of the Jaks results in the activation of their kinase activity, such that they can phosphorylate tyrosine residues in the receptor chains. These phosphotyrosine moieties provide the docking sites for the STAT proteins via their SH2 domains, leading to their tyrosine phosphorylation, dimerization, and nuclear translocation. Parallel signaling

events, presumably involving ERK/SAPK family members, are responsible for further phosphorylation of the serine residues in the C terminus of the STAT proteins, which are required for efficient transcriptional response. For some receptors, the intrinsic tyrosine kinase activity of the receptor, and not the presence of Jak kinases, is required for the tyrosine phosphorylation of STAT proteins. In addition to these phosphorylation events, STAT arginine methylation has been determined to be a crucial requirement for IFN- $\gamma$ -induced transcriptional responses. Of equal importance to the activation of a signaling pathway is its spatially and temporally coordinated attenuation. Several independent mechanisms are responsible for the negative regulatory control over the Jak/STAT pathway. The tyrosine phosphatase SHP1 acts on activated receptors and the Jak kinases, whereas the tyrosine phosphatase TcPTP dephosphorylates several STAT proteins in the nucleus. The SH2 domain-containing SOCS proteins have been identified as a family of cytokine-inducible inhibitors of Jak kinase activity, thus acting in a classical negative feedback loop. In contrast, the constitutively expressed nuclear PIAS proteins do not prevent the phosphorylation of STAT proteins but exert their negative regulatory role by associating with tyrosine phosphorylated STAT dimers and preventing them from binding DNA. As is the case with other signaling pathways, much of our understanding of the diverse functions of Jak and STAT proteins and their regulation *in vivo* has been further clarified through the generation of transgenic or knockout animals.

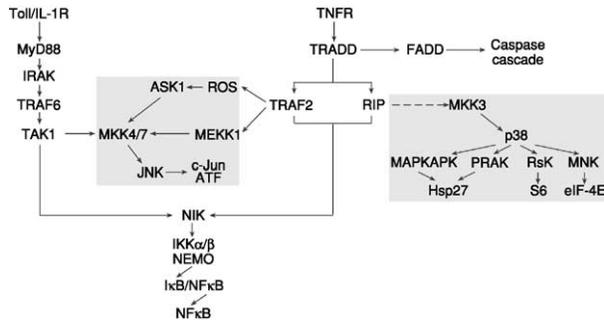


**Figure 1** The Jak/STAT pathway exemplified by the IFN signal transduction cascade: Ligand binding triggers the activation of the Jak kinases and subsequent phosphorylation of STAT proteins. STAT1/2 heterodimers use IRF9 as an adapter to bind to the ISRE, whereas STAT1/1, STAT3/3, and STAT5/5 homodimers and STAT1/3 heterodimers bind GRR/GAS-like enhancer elements directly.

## SIGNAL TRANSDUCTION VIA THE SAPK CASCADE AND NF- $\kappa$ B

Components or products of infectious microorganisms (e.g., LPS) evoke an inflammatory response characterized by the expression of proinflammatory cytokines, such as interleukin-1 (IL-1) and TNF- $\alpha$ . The binding of these cytokines to class III and class IV cytokine receptors rapidly induces a genetic program to respond to cellular stress and to generate inflammatory mediators (e.g., chemokines, acute phase proteins, proteases, adhesion molecules, and prostaglandins).

The signaling events triggered by the IL-1 receptors closely resemble those initiated by the Toll-like receptors (TLRs), which are responsible for the cellular recognition of pathogen-characteristic molecular patterns. Indeed, a region called the TIR (Toll/IL-1 receptor) domain is highly conserved between the TLRs and the IL-1 receptor. Although the



**Figure 2** Toll/IL-1R- and TNFR-mediated activation of the NF- $\kappa$ B and SAPK pathways: Signals originating from the Toll/IL-1R and TNFR converge on NIK to activate NF- $\kappa$ B. The pathways leading to the activation of the SAPKs p38 and JNK are shaded.

TNF receptors are structurally distinct from their TIR-containing counterparts, they share numerous signaling components in order to promote ligand-induced transcriptional responses (Fig. 2). Activation of the transcription factor NF- $\kappa$ B represents a point of convergence for the signaling cascades originating from TLR/IL-1R and TNFR. Although a key mediator in the inflammatory response, NF- $\kappa$ B was first isolated in B cells as a factor necessary for immunoglobulin kappa light-chain transcription. NF- $\kappa$ B binding sites have since been identified in a number of genes encoding cytokines and chemokines, adhesion molecules, acute phase proteins, antiapoptotic genes, or transcription factors. Thus, NF- $\kappa$ B was implicated in innate immunity as well as in the adaptive immune response. NF- $\kappa$ B exists as a homo- or heterodimer of any of the five isolated subunits: RelA (p65), RelB, c-Rel, p50, and p52. The p50 and p52 subunits are first generated as the longer p105 and p100 forms, respectively, which are then proteolytically processed into the shorter, active forms. All members contain a conserved rel homology domain (RHD) within their N terminus that contains the dimerization, I $\kappa$ B-binding, and nuclear localization signal (NLS) regions. Additionally, RelA, RelB, and c-Rel contain a C-terminal transactivation domain that is absent in p50 and p52. Hence, p50 and p52 homodimers are transcriptionally repressive.

Activation of NF- $\kappa$ B occurs through the protein synthesis-independent degradation of inhibitors of NF- $\kappa$ B (I $\kappa$ B), freeing NF- $\kappa$ B proteins from the cytosolic tether and exposing the NLS. Once in the nucleus, NF- $\kappa$ B negatively regulates its own activity by inducing transcription of I $\kappa$ B, which enters the nucleus and chaperones NF- $\kappa$ B molecules out through an active export pathway involving the I $\kappa$ B nuclear export sequence.

The I $\kappa$ B family members (I $\kappa$ B $\alpha$ , I $\kappa$ B $\beta$ , I $\kappa$ B $\epsilon$ , I $\kappa$ B $\gamma$ , and Bcl-3) contain ankyrin repeats that bind the NF- $\kappa$ B RHD, an N-terminal regulatory domain, and a C-terminal PEST motif for proteolytic degradation. Activators of the NF- $\kappa$ B pathway trigger serine phosphorylation of I $\kappa$ B proteins that targets them for ubiquitin-mediated destruction by the 26S proteasome. Two closely related proteins with I $\kappa$ B kinase (IKK) activity, IKK $\alpha$  and IKK $\beta$ , and a scaffolding subunit IKK $\gamma$  (NEMO) assemble into a large-molecular-weight complex capable of phosphorylating I $\kappa$ B proteins. Upon stimulation by TNF and IL-1, the IKK subunits are activated by phosphorylation on serine residues by the MAPKKK family member NIK.

A number of adapter proteins assemble into a signaling scaffold and link the Toll/IL-1 and TNF receptors to IKK activation and consequently NF- $\kappa$ B activation. The adapter protein MyD88 (myeloid differentiation factor 88) binds to TLR and IL-1R. Interacting through its death domain, MyD88 binds the serine/threonine kinase IRAK (IL-1R-associated kinase), which in turn recruits TRAF6 to the complex. Additionally, the MAPKKK TAK1 (TGF- $\beta$ -activated kinase) associates with TRAF6 and activates NIK in an IL-1-dependent manner. Similarly, the death domain containing TNFR promotes the formation of a large signaling complex by interacting with the death domain of TRADD. TRADD links the TNFR to FADD, thereby initiating proapoptotic signals via the caspase cascade, but also induces the antiapoptotic NF- $\kappa$ B pathway by recruiting TRAF2. Although TRAF2 is sufficient to recruit the IKK complex to the TNFR, the RIP kinase is required for full activation of IKK kinase activity.

The nonapoptotic TRAF2-dependent cascade also leads to the activation of the SAPKs p38 and JNK via two distinct pathways. Activation of MKK3 in a RIP-dependent manner triggers p38 phosphorylation, whereas the pathway leading to JNK activation involves MEKK1 and MKK4/7. Furthermore, a role for ASK1 in supporting sustained JNK activity was observed under apoptotic conditions. In the case of IL-1R, the events leading to ERK, p38, and JNK activation are less well documented. Nevertheless, it appears that the pathways leading to p38 and JNK activation diverge from the NF- $\kappa$ B activation cascade at the level of TAK1, which facilitates activation of MKK4/7 in addition to NIK.

## CONCLUSION

The discovery of the Jak/STAT pathway and the identification of numerous crucial elements in the NF- $\kappa$ B

and SAPK activation cascade have provided new momentum for research on cytokine responses. It is important to note, however, that many additional signaling molecules contribute to the cellular responses toward cytokine exposure. The raf–MEK–ERK cascade, the PI3 kinase/Akt pathway, the small GTP-binding proteins Rac/Rho/Cdc42, and the caspase cascade are significant contributors to many cytokine responses. The continuous exploration of the multitude of specific, overlapping, and redundant activities of individual cytokines within a larger signaling network remains an extraordinary challenge for the future.

### See Also the Following Articles

Adipocytokines • Cytokine Receptors • Cytokines, Constitutive Secretion • Cytokines, Evolutionary Aspects and Functions • Cytokines, Extracellular Transport and Processing • Janus Kinases and Cytokine Receptors

### Further Reading

- Akira, S. (1999). Functional roles of STAT family proteins: Lessons from knockout mice. *Stem Cells* **17**(3), 138–146.
- Dunne, A., and O'Neill, L. A. J. (2003). The interleukin-1 receptor/Toll-like receptor superfamily: Signal transduction during inflammation and host defense. *Science STKE* **171**, re3.
- Haddad, J. J. (2002). Cytokines and related receptor-mediated signaling pathways. *Biochem. Biophys. Res. Commun.* **297**, 700–713.
- Ishihara, K., and Hirano, T. (2002). Molecular basis of the specificity of cytokine action. *Biochim. Biophys. Acta.* **1592**, 281–296.
- Martin, M. U., and Wesche, H. (2002). Summary and comparison of the signaling mechanisms of the Toll/interleukin-1 receptor family. *Biochim. Biophys. Acta.* **1592**, 265–280.
- Schindler, C., and Darnell, J. E., Jr. (1995). Transcriptional responses to polypeptide ligands: The JAK–STAT pathway. *Annu. Rev. Biochem.* **64**, 621–651.
- Stark, G. R., Kerr, I. M., Williams, B. R., Silverman, R. H., and Schreiber, R. D. (1998). How cells respond to interferons. *Annu. Rev. Biochem.* **67**, 227–264.
- Wajant, H., Pfizenmaier, K., and Scheurich, P. (2003). Tumor necrosis factor signaling. *Cell Death Differ.* **10**, 45–65.

or interferon- $\gamma$  (IFN- $\gamma$ ), cytokine production is one of the many functions of the producing cells. Hormones are released in the circulation and act at distant sites. On the other hand, cytokine production is transient; they act over short distances and are not found in substantial amounts in the circulation. Polypeptide hormones are frequently restricted in their action to a limited set of cellular targets, with the notable exception of insulin. In contrast, many cytokines are pleiotropic, affecting a variety of cells and tissues. Also unlike hormones, structurally different cytokines have overlapping actions, as illustrated by IL-1 and tumor necrosis factor (TNF).

None of these distinguishing features are absolute. For instance, transforming growth factor- $\beta$  (TGF- $\beta$ ) and M-CSF are present in substantial amounts in the circulation under normal conditions, and IL-6 is produced in response to local inflammatory signals and acts distantly on the liver, contributing to the acute phase response.

## RECEPTORS OF THE IMMUNOGLOBULIN SUPERFAMILY

Receptors of the immunoglobulin (Ig) superfamily include those for the inflammatory cytokines IL-1 and IL-18 and a series of structurally similar orphan receptors whose function is only partly known.

IL-1 is a group of related cytokines including two agonist proteins (IL-1 $\alpha$  and IL-1 $\beta$ ) and an antagonist protein (IL-1ra). IL-1 plays a key role in the onset and development of the host reaction to invasion, being an important factor in the initiation of the inflammatory response and in the triggering of immune functions. IL-1 has a pivotal role in the regulation of the hypothalamic-pituitary-adrenal axis and in the modulation of thyroid functions, ovulation and pregnancy, and the brain-endocrine-immune axis in general. Receptors for IL-1 are expressed ubiquitously; thus, most of the cells and tissues of the body are responsive to IL-1 effects.

IL-18, a molecule structurally very similar to IL-1, is a very potent inducer of IFN- $\gamma$  production and activation of Th1 cells and natural killer (NK) cells. IL-18 is produced by several immune and nonimmune cells and tissues, including the zona reticularis and zona fasciculata of the adrenal cortex, that produce glucocorticoids. In Cushing's syndrome, the increased levels of IL-18 correlate with those of serum cortisol. IL-18 has been associated with other endocrine disorders, such as autoimmune thyroiditis of the obese chicken (a model of Hashimoto's thyroiditis) and

type 1 autoimmune diabetes. Physiologically, like IL-1, IL-18 is apparently involved in the regulation of ovary functions, ovulation, and pregnancy. Six other proteins belonging to the IL-1 family have been identified (IL-1F5-F10), whose receptor specificity and functional characteristics have yet to be determined.

Due to the high potency of its inflammatory effects, IL-1 activity is tightly regulated in the body by a complex network of control systems, which include two types of inhibitors—the receptor antagonist IL-1ra and the second type of IL-1 receptor (IL-1RII), which is a natural scavenger of IL-1 (decoy receptor). Regulation of IL-1 activity is attained by a strict hierarchy of binding affinity of the two receptors (the activating IL-1RI and the decoy IL-1RII) for the various members of the IL-1 family. The presence of a third receptor chain, the accessory protein IL-1RAcP, adds a further level of regulation since this chain is required for IL-1RI-mediated activation but it can be sequestered into an inactive complex by IL-1RII. A schematic representation of the mechanism of IL-1 binding to its receptor complex and of the IL-1-triggered pathway of signal transduction is shown in Fig. 1.

The IL-18 receptor system largely resembles that of IL-1. The IL-18 receptor is composed of two chains, IL-18R $\alpha$  and IL-18R $\beta$ , that are structurally very similar to IL-1RI and IL-1RAcP and initiate the

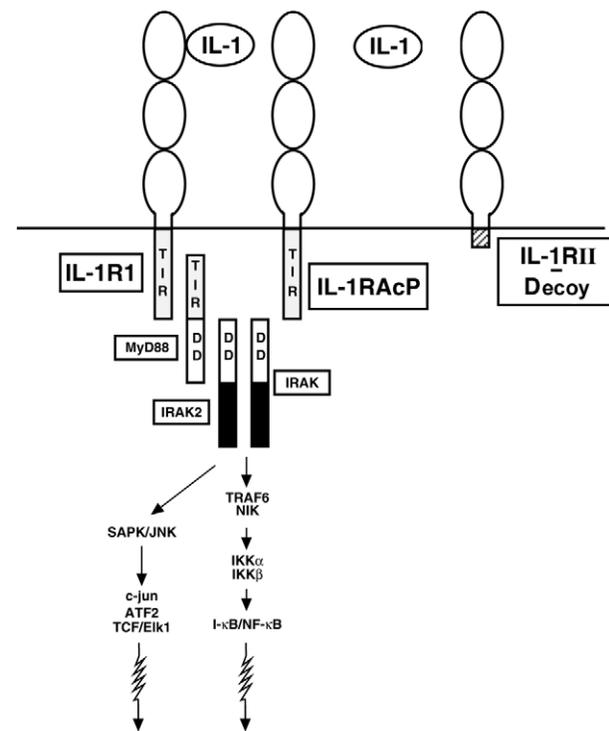


Figure 1 IL-1 receptor complex and signal transduction pathway.

Table I Ig-like Receptor Family: IL-1R/IL-18R Molecules

Receptor	Ligand	Function
IL-1RI	Agonist IL-1 $\alpha$ and IL-1 $\beta$ , antagonist IL-1ra, IL-1F10 (to soluble receptor form)	Initiates and amplifies the immune and inflammatory response activated by agonist ligands; inhibited upon binding the antagonist ligand
IL-1RII	IL-1 $\beta$ , less efficiently IL-1 $\alpha$ and IL-1ra	Decoy receptor, unable to initiate signal transduction
IL-1RAcP	Coreceptor for IL-1RI bound to agonist ligands IL-1 $\alpha$ and IL-1 $\beta$	Responsible for signaling together with ligand-bound IL-1RI; can form inactive complexes with IL-1RII bound to IL-1
IL-18R $\alpha$	IL-18, IL-1F7	Initiates IL-18-dependent activation of Th1 and NK cells; function upon binding of IL-1F7 unknown
IL-18R $\beta$	Coreceptor for IL-18R $\alpha$ bound to IL-18	Responsible for signaling together with ligand-bound IL-18R $\alpha$
T1/ST2	Orphan receptor	Unknown; possible role in the activation of Th2 responses
TIGIRR	Orphan receptor	Unknown
IL-1RAPL	Orphan receptor	Deletions in the gene are associated with X-linked mental retardation
IL-1Rrp2	Binds IL-1F5 (antagonist of IL-1F9) and IL-1F9	Unknown
SIGIRR	Orphan receptor	Unknown

same signal transduction pathway. In addition to the receptor chains for the inflammatory cytokines IL-1 and IL-18, the family of IL-1 receptors includes a series of IL-1R-like orphan receptors. The gene coding for one of these receptors (IL1RAPL, expressed in the hippocampus) is a strong candidate for X-linked nonsyndromic mental retardation locus, suggesting that molecules resembling IL-1 and IL-18 may play a role in the development or function of the central nervous system. The 10 known members of the IL-1R/IL-18R family are shown in [Table I](#).

## TNF RECEPTORS

The TNF receptor family includes more than 20 members that are involved in a variety of functions. [Figure 2](#) shows the two TNF receptors (RI and RII), which serve as paradigms for members of the family acting primarily as death receptors. Ligands of members of the TNF-R family generally have a long-chain  $\beta$ -sheet structure. When membrane-associated as a type II integral membrane protein or when released, TNF forms a trimer. TNF release involves a member of the matrix metalloprotease family (TNF- $\alpha$ -converting enzyme). Receptor engagement and oligomerization in the case of TNF-RI and related receptors (fas) lead to cell death via activation of the caspase cascade. Members of this group of TNF receptors have as first enzymatic activity a protease (caspase 8). TNF-RII and related receptors activate NF- $\kappa$ B, via adapter proteins and promote survival.

Receptors of the TNF family serve diverse functions in pathophysiology, ranging from induction of apoptosis (fas and TNF-RI) to ontogeny of the lymphoid system (TNF and lymphotoxin receptor), bone remodeling (RANK and its decoy receptor osteoprotegerin), innate immunity, and inflammation. Blocking TNF with soluble TNF-RII or antibodies has proven useful in human rheumatoid arthritis and Chron's disease.

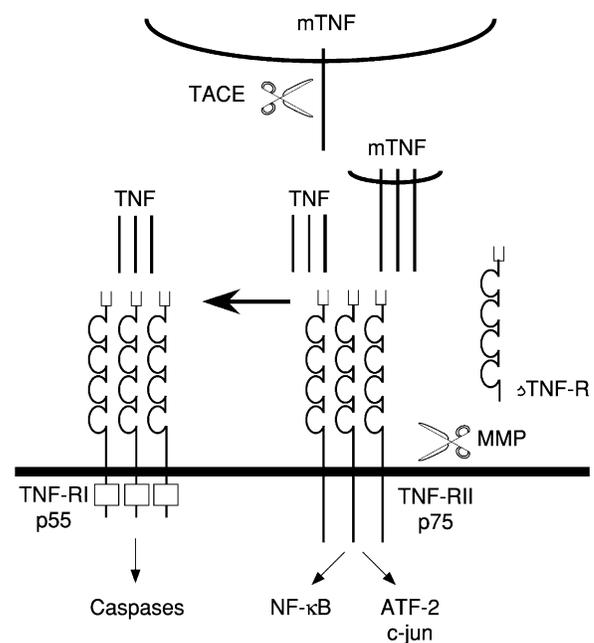


Figure 2 TNF receptors. TACE, TNF- $\alpha$ -converting enzyme.

## COMMON $\gamma$ -CHAIN RECEPTORS

The common cytokine receptor  $\gamma$ -chain ( $\gamma_c$ ) is the signaling chain in the receptor complexes of the cytokines IL-2, IL-4, IL-7, IL-9, and IL-15. Mutation in the gene that encodes  $\gamma_c$  results in severe combined immunodeficiency in humans and mice. Signals that are dependent on  $\gamma_c$  have been shown to be important for T cell differentiation, activation, expansion, and survival.

IL-2 is the prototypic T cell growth factor. The biologic effects of IL-2R signals influence T cell growth and can also promote cell survival, effector function, and apoptosis. IL-2 and IL-2R are also expressed by cells other than lymphocytes (e.g., pituitary and adrenal cells), suggesting a role for this cytokine in the autocrine and paracrine regulation of the hypothalamic–pituitary–adrenal axis. The IL-2 receptor complex is composed of three chains. IL-2R $\alpha$  binds IL-2 with low affinity. IL-2R $\beta$  increases the affinity of the IL-2/IL-2R $\alpha$  complex. The  $\gamma_c$  associates with the IL-2/IL-2R $\alpha$ /IL-2R $\beta$  complex and initiates signaling.

IL-7, a product of stromal cells, induces potent B cell lymphopoiesis, followed by bone loss, resembling estrogen deficiency. The IL-7 receptor is composed of two chains, IL-7R $\alpha$  and  $\gamma_c$ . Following receptor cross-linking, rapid activation of several classes of kinases occurs, including members of the Janus and Src families and PI3-kinase. A number of transcription factors are subsequently activated, including STATs, c-myc, NFAT, and AP-1.

IL-9 is a Th2 cytokine with major involvement in malignant proliferation (e.g., T cell lymphomas and leukemias and Hodgkin's disease). The IL-9 receptor is heterodimeric, formed by the specific IL-9-binding chain IL-9R $\alpha$  and by the signaling  $\gamma_c$ .

The heterotrimeric IL-15R complex is very similar to that for IL-2 and comprises the unique IL-15R $\alpha$  subunit, IL-2R $\beta$  (CD122), and the  $\gamma_c$ . The  $\gamma_c$  and  $\beta$  subunits are responsible for transducing intracellular signals and the  $\alpha$  subunit mediates specific, high-affinity binding of IL-15. The production of IL-15 and IL-2 by essentially different cell types suggests that these cytokines mediate similar functions in distinct sets of tissues and/or organs. IL-2R $\beta$  knockout mice, as well as IL-15R $\alpha$  knockout mice, are deficient in both NK cells and NK-T cells (i.e., cells implicated in so-called "innate immunity").

IL-4 is a Th2 cytokine that acts on both hematopoietic and nonhematopoietic cells through different types of receptor complexes. There are two different forms of IL-4R. Type I IL-4R is predominantly expressed in hematopoietic cells and consists of IL-4R

p140 (IL-4R $\alpha$ ) and the  $\gamma_c$ . The type II form of IL-4R is predominantly expressed in nonhematopoietic cells, which do not express the  $\gamma_c$ , and consists of IL-4R $\alpha$  and IL-13R $\alpha_1$  chains. This alternative form of IL-4R is also used as a functional component in the IL-13R complex.

## IFN RECEPTORS AND RELATED MOLECULES

Interferons are potently active proteins synthesized by somatic cells of all mammalian species. Broadly, IFNs are clustered into two subfamilies, type I and type II. The type I IFN family includes IFN- $\alpha$ , IFN- $\beta$ , IFN- $\omega$ , and IFN- $\tau$ . These molecules use the same cell surface receptors, IFN- $\alpha$ R1 and IFN- $\alpha$ R2 (belonging to the type 2 cytokine receptor family for  $\alpha$ -helical cytokines), and have similar functions (antiproliferative, immunomodulatory, and antiviral effects). Type II IFN is the cytokine IFN- $\gamma$ , which has lower antiviral activity than type I IFN but has a major immunomodulatory role. Cellular responses to IFN- $\gamma$  are mediated by its heterodimeric cell surface receptor (IFN- $\gamma$ R1/IFN- $\gamma$ R2).

The family of IL-10-related cytokines comprises a series of members, including IL-19, IL-20, IL-22, IL-24, and IL-26. Although the predicted helical structure of these homodimeric molecules is conserved, certain receptor-binding residues are variable and define the interaction with specific heterodimers of different members of type 2 cytokine receptors. This leads, through the activation of signal transducer and activator of transcription (STAT) factors, to diverse biological effects. For example, whereas IL-10 is a well-known immunomodulatory cytokine, IL-22 mediates acute phase response signals in hepatocytes and IL-20 induces the hyperproliferation of keratinocytes, one of the possible pathogenic mechanisms of psoriasis.

Receptors for INFs, IL-10, and its related cytokines are all members of the cytokine receptor family type 2 (CRF2). The family also includes tissue factor (TF), the membrane-binding chain for FVII/FVIIa. CRF2 members are present in the extracellular part of fibronectin III domains and have conserved intracellular regions involved in the interaction with downstream signaling molecules. As a prototype, the IL-10 receptor consists of a long chain (IL-10R1), the major signaling component, and an additional receptor chain (IL-10R2) with a short intracellular segment. Similarly, all the other receptor heterodimers of this family combine a long chain with a large intracellular

Table II CRF2 Receptors and Their Ligands

Receptor complex	Ligand	Function
IFN- $\alpha$ R1/IFN- $\alpha$ R2	IFN type I	Antiviral, antiproliferative, immunomodulatory effects
IFN- $\gamma$ R1/IFN- $\gamma$ R2	IFN- $\gamma$	Immunomodulatory, inflammatory, antiviral effects
IL-10R1/IL-10R2	IL-10	Down-regulation Th1 activities, antiinflammatory
IL-20R1/IL-20R2	IL-19	Unknown
IL-20R1/IL-20R2	IL-20	Keratinocyte activation, keratin expression, synergistic with IL-1
IL-20R1/IL-20R2	IL-24	Tumor growth inhibition
IL-22R/IL-10R2	IL-22	Induction of acute phase proteins in hepatic cells, activation of pancreatic acinar cells; no IL-10-like activity, no inhibition of IL-10 activity
IL-22R/IL-20R2	IL-20	Unknown
IL-22R/IL-20R2	IL-24	Tumor growth inhibition

domain and a smaller accessory chain. Signaling through the heterodimeric CRF2 complexes for the individual IL-10-like cytokines involves the activation of JAK kinases and the phosphorylation of STAT factors, which induce  $\gamma$ -activated sequence-dependent or STAT-dependent transcription of target genes. Because STAT3 seems to be a common major transcription factor mediating stimulatory effects, the tissue-specific surface expression of the specific combination of CRF2 subunits likely plays the key role in determining the function of the respective IFN/IL-10 family member. A summary of the CRF2 receptors and their ligands is provided in [Table II](#).

## CHEMOKINE RECEPTORS

Chemokines are a superfamily of small proteins vital for immune and inflammatory reactions and viral infection. Most chemokines cause chemotactic migration of leukocytes, but they also affect angiogenesis, collagen production, and the proliferation of hematopoietic precursors. Based on a cysteine motif, CXC, CC, C, and CX3C families have been identified. The chemokine scaffold consists of an N-terminal loop connected by Cys bonds to the more structured core of the molecule (three  $\beta$ -sheets) with a C-terminal  $\alpha$ -helix. Approximately 50 human chemokines have been identified.

Chemokines interact with seven-transmembrane domain, G protein-coupled receptors. Eight CC (CCR1–8), five CXC (CXCR1–5), and one CX3C (CX3CR1) receptor have been identified.

The chemokines' main function is chemotaxis for leukocytes. Chemokines are redundant in their action on target cells. No chemokine is active on only one leukocyte population, and usually a given leukocyte population has receptors for and responds to different

chemokine molecules. The chemokines' interactions with their receptors show considerable promiscuity. Most known receptors have been reported to interact with multiple ligands, and most ligands interact with more than one receptor. G protein-coupled receptors are a classic target in pharmacology. Likewise, given the role of chemokines in various human diseases, ranging from HIV infection to allergies, their receptors' pharmacology is a prime target for research. Although redundancy is a formidable problem, blocking one agonist or one receptor may be beneficial in certain disease models. Therefore, simple chemicals with chemokine antagonistic properties or with selective inhibitory activity on chemokine production are a holy grail in cytokine pharmacology.

## gp130 USING RECEPTORS

Receptors for IL-6 and related cytokines (oncostatin M, leukemia inhibitory factor, ciliary neurotrophic factor, and cardiotrophin) share a common gp130 chain, which activates a JAK/STAT signaling cascade. Ligands for these receptors share a four- $\alpha$ -helical bundle, long-chain structure. The IL-6R $\beta$  chain binds the ligand and forms a signaling receptor complex in membrane-bound and soluble form (trans-signaling). The liver is a prime target for IL-6, which activates the production of acute phase proteins (e.g., C-reactive protein). In addition, IL-6 amplifies inflammation in tissues.

## See Also the Following Articles

Adipocytokines • Chemokines • Cytokine Actions, Cellular Mechanism of • Cytokines, Constitutive Secretion • Cytokines, Evolutionary Aspects and Functions • Cytokines, Extracellular Transport and Processing • Janus Kinases and Cytokine Receptors

## Further Reading

- Kotenko, S. V. (2002). The family of IL-10-related cytokines and their receptors: Related but to what extent? *Cytokine Growth Factor Rev.* **13**, 223–240.
- Mantovani, A. (ed.) (2001). “Chemokines.” Karger, Basel.
- Mantovani, A., Dinarello, C. A., and Ghezzi, P. (2000). “Pharmacology of Cytokines.” Oxford Univ. Press, Oxford, UK.
- Mantovani, A., Locati, M., Vecchi, A., Sozzani, S., and Allavena, P. (2001). Decoy receptors: A strategy to regulate inflammatory cytokines and chemokines. *Trends Immunol.* **22**, 328–336.
- Sims, J. E. (2002). IL-1 and IL-18 receptors and their extended family. *Curr. Opin. Immunol.* **14**, 117–122.
- Waldmann, T., Tagaya, Y., and Bamford, R. (1998). Interleukin-2, interleukin-15, and their receptors. *Int. Rev. Immunol.* **16**, 205–226.



# Cytokines, Constitutive Secretion

Nobuyuki Tanaka and Kazuo Sugamura  
Tohoku University Graduate School of Medicine, Sendai, Japan

## Glossary

**human T-cell leukemia virus type I (HTLV-I)** A causative agent for adult T-cell leukemia.

**interleukin-2 (IL-2)** A cytokine that regulates lymphocyte proliferation via the IL-2 receptor (IL-2R).

**interleukin-4 (IL-4)** A cytokine that regulates B-lymphocyte activation and differentiation into antibody-secreting cells.

**tumor necrosis factor- $\alpha$  (TNF- $\alpha$ )** An inflammatory cytokine, secreted mainly by macrophages, that provokes increased vascular permeability, inflammation, and apoptosis.

Cytokines are potent protein messengers participating in cell-to-cell communication with a variety of biological activities. Their constitutive secretions may lead to autoimmunity, allergy, and cancers.

## INTRODUCTION

The identification of a number of cytokines has shed light on our understanding of the regulation of homeostasis. In the immune system, accumulating evidence has revealed roles of cytokines in cell fate determination such as development, proliferation, differentiation, and cell death. In addition to the “favorable” roles of cytokines to maintain homeostasis in a healthy condition, uncontrolled secretion of potent cytokines may provoke autoimmune diseases, allergies, and cancers. This article discusses some examples of human diseases as well as lessons learned from model animals, including transgenic mice.

## CYTOKINE CONSTITUTIVE SECRETION POSSIBLY INVOLVED IN LEUKEMOGENESIS

Interleukin-2 (IL-2) is a cytokine first identified as a T-cell growth factor. Constitutive expression of IL-2 and its receptor (IL-2R) has been suggested to be related

to development of adult T-cell leukemia (ATL). ATL is caused by infection with human T-cell leukemia virus type I (HTLV-I). HTLV-I-infected T cells derived from healthy individuals as well as ATL patients constitutively express IL-2R and are easily immortalized in the presence of IL-2. The functional IL-2R consists of three subunits:  $\alpha$ -,  $\beta$ -, and  $\gamma$ -chains. The  $\gamma$ -chain is constitutively expressed on every hematopoietic cell population, including T cells, but expression of the  $\alpha$ - and  $\beta$ -chains and production of IL-2 are inducible after antigen stimulation in normal T cells. HTLV-I-infected T cells constitutively express all three IL-2R subunits. Furthermore, the HTLV-I genome contains a pX region coding for a transactivator protein named p40<sup>tax</sup>, which has been demonstrated to transactivate transcription of at least IL-2 and IL-2R $\alpha$  genes. Hence, it can be speculated that HTLV-I-infected T cells are stimulated to proliferate constitutively by IL-2 in an autocrine or paracrine manner, and this may result in further malignant transformation of the cells. Such a mechanism of HTLV-I-induced leukemogenesis favors the notion that ATL cells are derived from CD4<sup>+</sup> T cells with potentiality for IL-2 production.

Another example of autocrine-related lymphoid malignancy is Hodgkin's lymphoma (HL). HL is a unique lymphoma characterized by the appearance of Reed-Sternberg cell and multiple cytokine secretion. One report suggests that most Hodgkin's lymphoma cells secrete IL-13 and express IL-13R $\alpha$ -chain, leading to their proliferation in an IL-13 autocrine manner. Because IL-13 shares the  $\gamma$ c-chain as a receptor subunit with IL-2, it is of note that constitutive signaling by the cytokines sharing the  $\gamma$ c-chain may relate to leukemogenesis.

## EXCESSIVE INFLAMMATORY CYTOKINE SECRETION AS A PATHOGENESIS FOR AUTOIMMUNE DISEASES

IL-6 was originally identified as a B-cell differentiation factor, and it is now known to be multifunctional in

regulation of the immune system, hematopoiesis, acute phase reaction, and inflammation. The first clue to suggest its role in autoimmunity came from the finding that cardiac myxoma cells produce IL-6 and patients suffering from this disease frequently manifest autoimmune symptoms such as hypergammaglobulinemia and production of autoreactive antibodies. Because surgical removal of the cardiac myxoma induces the disappearance of the autoimmune conditions, myxoma cells that contain a high level of IL-6 mRNA expression seemingly contribute to the autoimmune pathogenesis. Among the autoantibodies, anti-DNA and antichromatin antibodies are shown to be dependent on IL-6. Systemic lupus erythematosus (SLE), a systemic autoimmune disease, also has been suggested to be related to IL-6. Serum IL-6 levels of SLE patients are usually normal, but SLE patients with serositis or SLE patients during disease exacerbations show increased IL-6 levels in sera. In addition to IL-6, B cells in patients with active SLE express IL-6R as well as IL-2R, suggesting the competency of these cells toward IL-6-induced reaction. Alternatively, an anti-IL-6R mAb inhibits the spontaneous production of anti-DNA antibodies, indicating that B cells are activated by IL-6R-mediated signals in SLE. Transgenic mice expressing IL-6 develop a massive polyclonal plasmacytosis with autoantibodies and mesangial cell proliferative glomerulonephritis (MPGN). Dysregulated expression of IL-6 can trigger polyclonal plasmacyte proliferation, resulting in generation of a malignant monoclonal plasmacytoma.

Another possible pathogenesis in relation with IL-6 is an inflammatory bowel disease, Crohn's disease (CD). CD activity indexes correlate well with serum IL-6 levels, and steroid therapy reduces IL-6 levels to induce remission of CD. In fact, soluble IL-6R, which can signal in the absence of membrane-bound IL-6R, is significantly high in sera of CD patients.

Another inflammatory and autoimmune disease possibly related to IL-6 is rheumatoid arthritis (RA), which is a heterogeneous and chronic joint disease. In RA patients, affected joints are invaded by leukocytes and synoviocytes are activated; thus, destructive arthritis occurs, resulting in cartilage destruction, bone resorption, and deformity. Although proinflammatory cytokines produced by macrophages and dendritic cells, such as TNF- $\alpha$  and IL-1 $\beta$ , also are detectable in RA-affected synovial fluids, high levels of IL-6 affect B cells to produce autoantibodies such as rheumatoid factor (RF). Involvement of IL-6 is also suggested in juvenile chronic arthritis (JCA). Serum IL-6 levels of patients with JCA are increased during active status, correlating with a severity of the joint.

Because osteoclasts are stimulated by IL-6, bone destruction and resorption may also be activated by IL-6. Mice deficient in IL-6 result in delayed onset and reduced severity of an experimentally induced autoimmune disease called collagen-induced arthritis (CIA). These findings imply that excessive secretion of IL-6 is associated with the destructive phenotype of RA, although IL-6 is not the only factor responsible for the pathogenesis.

In addition to IL-6, autoimmune RA is associated with constitutive secretion of TNF- $\alpha$ . High-affinity anti-TNF- $\alpha$  mAbs, neutralizing its function, have been used successfully for clinical therapy in moderate to severe RA patients with appreciable effects. In *in vitro* studies, anti-TNF- $\alpha$  mAb profoundly decreased IL-1 secretion from synovial cells derived from RA patients. As expected, IL-1R antagonist or soluble IL-1R type II is also under way for clinical trials, either alone or in combination with anti-TNF- $\alpha$  mAb.

## SKEWED CYTOKINE SECRETION FOR THE PATHOGENESIS OF ALLERGY

Allergic inflammatory reactions are induced by T-helper 2 (Th2) cells secreting the "Th2 cytokines" such as IL-4, IL-5, and IL-13. IL-4, first identified as a B-cell-stimulating factor, activates and differentiates B cells and, therefore, enhances antibody production. Immunoglobulin E (IgE)-class antibodies secreted by B cells are bound to the Fc receptor on the surface of the mast cells, which react to the specific antigens to release chemical mediators such as histamine. In mice, such Th2 reactions are surprisingly different between strains. For example, on antigen challenge, CD4+ T cells from BALB/c mice overproduce IL-4, whereas those from C57/BL6 mice produce much less IL-4. On pathogen challenge such as *Leishmania major*, BALB/c mice show a lethal phenotype, whereas C57/BL6 mice may survive with the same pathogenic burden. Because IL-4 knockout mice manifest reduced Th2 response, IL-4 should play a pivotal role in the pathogenesis of allergic disorders, although a certain genetic background definitely contributes to the susceptibility.

## EXCESSIVE CYTOKINE SECRETION DURING INFECTION

Toll-like receptors (TLRs) are now well known for pathogen recognition during the early phase of infection. For example, lipopolysaccharides (LPSs) are recognized by TLR4 along with other association

molecules, resulting in macrophage and dendritic cell activation. An immediate outcome of LPS-induced activation is a secretion of proinflammatory cytokines such as TNF- $\alpha$ , IL-1 $\beta$ , and IL-6. These cytokines favor elimination of the pathogen; they work locally at the site of infection to initiate inflammation with intravascular mononuclear cell recruitment, leading to protection against the pathogens. However, if massive and systemic secretion occurs, the host will face systemic and fatal reaction with circulation insufficiency, referred to as endotoxin shock or sepsis. Oversecretion of TNF- $\alpha$  is considered to be paramount in the development of shock given that injection of TNF- $\alpha$  into mice leads to septic-like shock, whereas TNFR type II knockout mice do not fall into septic shock. Because an initial response to the pathogen is important to activate acquired immunity, proinflammatory cytokines may be necessary to activate lymphocytes as well as to provoke inflammation.

Another example of excessive innate immune response is familial Mediterranean fever (FMF), an inherited disease characterized by recurrent episodes of fever and inflammation. Most patients with FMF carry missense mutations in the C-terminal region of PYRIN protein, and this induces a hyperreaction to a small amount of LPS. Recurrent and continuous secretion of IL-1 $\beta$  then results in recurrent fever in patients.

## CONCLUSION

Over the past two decades, great progress has been achieved in our knowledge of cytokines both *in vitro* and *in vivo*. Identification of a causative gene for

immune-deficient diseases and inflammatory diseases has facilitated further understanding of the roles of cytokines themselves as well as of their signaling mechanisms. In addition, animal models, including transgenic and gene-targeted mice, have facilitated and continue to facilitate our insight into the pathogenesis of diseases as well as development of new clinical interventions. Although we need to pay attention to their side effects, anti-cytokine therapies would be an important measure for controlling immune disease, infection, and cancers.

## See Also the Following Articles

Adipocytokines • Chemokines • Cytokine Actions, Cellular Mechanism of • Cytokine Receptors • Cytokines, Evolutionary Aspects and Functions • Cytokines, Extracellular Transport and Processing • Interleukin-2 • Janus Kinases and Cytokine Receptors • Tumor Necrosis Factor (TNF)

## Further Reading

- Firestein, G. S. (2003). Evolving concepts of rheumatoid arthritis. *Nature* **423**, 356–361.
- Murphy, K. M., and Reiner, S. L. (2002). The lineage decisions of helper T cells. *Nat. Rev. Immunol.* **2**, 933–944.
- Sugamura, K., and Hinuma, Y. (1993). The retroviridae. In “Human Retroviruses: HTLV-I and HTLV-II” (J. A. Levy, ed.), vol. 2, pp. 399–435. Plenum Press, New York.
- Sugamura, K., Asao, H., Kondo, M., Tanaka, N., Ishii, N., Nakamura, M., and Takeshita, T. (1995). The common gamma-chain for multiple cytokine receptors. *Adv. Immunol.* **59**, 225–277.
- Wynn, T. A. (2003). IL-13 effector functions. *Annu. Rev. Immunol.* **21**, 425–456.

immunological innate response, recognition of “foreignness” is one of the first responses of phagocytes (monocytes or macrophages). One of the most important roles of the phagocyte is to alert the rest of the immune system of an attack. The phagocyte has many secreted proteins in its arsenal, some of the most important being the cytokines.

## CYTOKINES

Cytokines are soluble low-molecular-weight proteins ( $\approx 20$  kDa) secreted by immune cells and they are the intracellular messenger molecules of the immune system. Cytokines determine the scope and type of an immune response that is generated after infection. Cytokine is a word that is derived from *cyto-*, meaning cell, and *-kinin*, a combining form used in naming hormones, usually peptide hormones (e.g., bradykinin). The term tends to be used as generic shorthand for several related signaling molecules (Table I). Cytokine is an imprecise term, but it is in very common usage. The categorization of cytokines has always been a problem because they were originally named for the activity that they described. Generally, growth factors are not classified as cytokines.

The cytokine pathway is akin to the endocrine system, with the cytokines being analogous to hormones in that cytokines serve as the intracellular messengers of the immune system. All the actions of cytokines are receptor mediated with binding to high-affinity receptors on target cells so they are active at picomolar ( $10^{-12}$  M) concentrations. Cytokines are not stored as preformed molecules within the cell; their synthesis is initiated by new gene transcription. This transcriptional activity is usually brief, ensuring that cytokine synthesis and release are transient. Cytokine actions can be autocrine (binding to its receptor and then activating the same cell that secreted it), paracrine

(binding to its receptor on a target cell in close proximity to the cell that released it), or systemic, which is also sometimes called endocrine (binding to its receptors on a target cell in a distant part of the body). Cytokines are not used exclusively by the immune system, as cells of the endocrine and nervous system will respond to the actions of cytokines. The attributes of cytokines include pleiotropy (one cytokine with different biological effects on different target cell types), redundancy (two or more cytokines with the same biological effects on a target cell), synergy (combined effect of two cytokines on a target cell), and antagonism (one cytokine blocking the effect of another cytokine on a given target cell).

Cytokines mediate many of the fundamental functions of the innate immune response. They are also involved in all aspects of the inflammatory response (e.g., fever, tissue remodeling). Table II lists just a few of the myriad actions of cytokines. Just as with innate immunity, many of the critical activities of the effector cells involved in acquired immunity are mediated by cytokines. For example, during the activation phase of specific immune responses, cytokines are involved in the growth and differentiation of lymphocytes.

In general terms, the database of cytokine sequences continues to be confined to human cytokines, laboratory model species, and economically significant domestic animals. The accumulation of primary sequence data on a wider range of species will be a valuable addition to researchers' knowledge of this medically and scientifically important group of molecules.

## CYTOKINE EVOLUTION

More than 200 different cytokines, mostly from humans and mice, have been described. Evidence suggesting that protein molecules with cytokine-like activities could be isolated from other vertebrate

Table I Nomenclature of Cytokines

Synonym	Simple definition
Interleukins	Macrophage-derived stimulatory/inhibitory molecules
Monokines	Stimulatory/inhibitory molecules secreted by mononuclear phagocytes
Lymphokines	Stimulatory/inhibitory molecules secreted by activated T lymphocytes
Colony-stimulating factors	Stimulatory/inhibitory molecules involved in the growth and differentiation of immature leukocytes in the bone marrow
Interferons	Stimulatory/inhibitory molecules that generally protect against viral infections
Chemokines	A family of molecules, related to interleukin-8, that function mainly as leukocyte chemoattractants
Tumor necrosis factor	A potent mediator of inflammatory and immune functions, released by leukocytes, that is selectively cytotoxic for many transformed cells

Table II Some Examples of the Actions of Cytokines

T cell proliferation	B cell activation and proliferation
Cell growth and proliferation	Production of fever
Stimulation of neutrophil development	Acute-phase protein production
Stimulation of eosinophil growth and differentiation	Stimulation of mast cells
Lymph node development	Stimulation of apoptosis
Stimulation of cytokine production	Stimulation of B cell class-switching
Natural killer cell activation	Mononuclear cell chemoattractant
Increases in major histocompatibility complex class I expression	Cytotoxicity of transformed cells
Stimulation of intestinal epithelium	Macrophage activation
Inhibition of macrophage migration	Stimulation of bone resorption
Stimulation of immunoglobulin secretion	Endothelial cell activation
Stimulation of bone marrow cell progenitors	Suppression of macrophage functions
Stimulation of serum cation concentrations	

animals has accumulated. This biological activity was based on cross-reactivity with antibodies and substitution of active molecules in mammalian-based assay systems. Primary sequence data from numerous vertebrate species have been reported. Most of these cytokines have been cloned using polymerase chain reaction-based techniques based on primers derived from existing sequence homologies principally between humans and mice.

Interleukin-1 (IL-1) is recognized in vertebrates as being a major cytokine of innate immune responses. In addition to IL-1, IL-6 and tumor necrosis factor  $\alpha$  (TNF $\alpha$ ) make up a collection of cytokines with overlapping biological properties; these proteins are known as the pro-inflammatory cytokines. The majority of cytokines isolated and sequenced from non-mammalian vertebrates have been pro-inflammatory cytokines.

In 1985, two distinct, but distantly related cDNAs encoding proteins sharing human IL-1 biological activity (termed IL-1 $\alpha$  and IL-1 $\beta$ ) were isolated from a macrophage cDNA library. A third member of the IL-1 family is an antagonist protein that shares sequence similarity with IL-1 $\alpha$  and IL-1 $\beta$ , the IL-1 receptor antagonist (IL-1ra). With the sequencing of the human and mouse genomes and subsequent database searches, other mammalian members of the interleukin-1 family (IL-1F) have been identified. One of those molecules is a homologue of IL-1ra. IL-1F5 (IL-1 $\delta$ ) shows 52% amino acid homology to IL-1ra. Even though IL-1F5 is similar to IL-1ra, it is not yet known whether it behaves like IL-1ra in regulating the response of IL-1 $\alpha$  and IL-1 $\beta$  via competitive inhibition. Along with IL-1F5, three other molecules were identified and cloned: IL-1F6 (IL-1 $\epsilon$ ), IL-1F7 (IL-1 $\zeta$ ), and IL-1F8 (IL-1 $\eta$ ). The novel genes demonstrate

significant sequence similarity to IL-1 $\alpha$ , IL-1 $\beta$ , IL-1ra, and IL-18, and in addition they maintain a conserved intron-exon composition that is shared with the other members of the family. One of the newest members of the family is IL-1 $\theta$  (IL-1F10), which may also be an IL-1ra homologue.

Cytokines have been isolated from other vertebrates as well. The following paragraphs give a quick summation of some studies, all of which would be too numerous to describe here. For example, IL-1 $\alpha$ , IL-1 $\beta$ , IL-4, IL-5, IL-6, IL-8, IL-10, IL-15, granulocyte/macrophage colony-stimulating factor, interferon- $\gamma$  (IFN- $\gamma$ ), and TNF $\alpha$  have all been cloned and characterized from nonhuman primates [e.g., rhesus monkey (*Macaca mulatta*) and pigtailed monkey (*Macaca nemestrina*)]. These cytokines share 93 to 99% homology at the nucleic acid and protein levels with their human equivalents.

IL-1 $\beta$ , IL-10, TNF $\alpha$ , and TNF $\beta$  have been isolated from marsupials [Virginia opossum (*Trichosurus vulpecula*) and wallaby (*Macropus eugenii*)]. When compared to human and mouse amino acid sequences, identities average > 60% for both species.

IL-2, a cytokine produced by activated T lymphocytes, which is a T cell-stimulatory factor, has been cloned from beluga whale (*Delphinapterus leucas*) and gray seal (*Halichoerus grypus*) T cells. The deduced amino acid sequence of IL-2 from beluga whale shared 61–93% identity with the sequences from 12 other mammalian species and that from gray seal shared 59–88% identity with the mammalian sequences. In addition, IL-6 was cloned and sequenced from harbor seal (*Phoca vitulina*), killer whale (*Orcinus orca*), and Southern sea otter (*Enhydra lutris nereis*). When these amino acid sequences were compared with those from 10 other mammalian species, the

sequence identities ranged from 41 to 84%, 42 to 80%, and 37 to 89%, respectively.

Cytokines have been found in lower vertebrates as well. The frog (*Xenopus laevis*) IL-1 $\beta$  amino acid sequence shares >35% identity with the human IL-1 $\beta$  sequence. An important negative regulatory cytokine, transforming growth factor- $\beta$  (TGF- $\beta$ ), has also been cloned from *Xenopus*. The *Xenopus* TGF- $\beta$  amino acid sequence is >75% identical to its human counterpart.

In studies of lower vertebrate cytokines, the cytokines from birds and fish have been among the best characterized. The potential of avian cytokines as natural mediators and regulators of the immune response is thought to offer exciting alternatives to conventional antibiotic therapeutics. IL-1 $\beta$ , IL-2, IL-15, IFN- $\alpha$ , IFN- $\beta$ , and IFN- $\gamma$  have all been cloned from chickens (*Gallus gallus domesticus*), turkeys (*Meleagris gallopava*), or ducks (*Anas platyrhynchos domesticus*). The chicken IL-1 $\beta$  amino acid sequence is 25% identical to the human sequence. Surprisingly, analysis of the chicken IL-2 amino acid sequence shows that it is ancestral to both mammalian IL-2 and IL-15. This finding suggests that the chicken gene duplicated and mutated before the evolutionary separation of mammals and birds.

Fish cytokines have been extensively studied and molecularly characterized. The therapeutic use of cytokines as adjuvants in aquaculture has been driving the research. TNF $\alpha$  from flounder (*Paralichthys olivaceus*) and rainbow trout (*Oncorhynchus mykiss*) has been cloned. The amino acid sequence of the flounder cytokine displayed 20–35% identity with mammalian cytokines, whereas the trout sequence displayed 31–36% similarity to mammalian TNF $\alpha$ . Rainbow trout, carp (*Cyprinus carpio*), seabass (*Dicentrarchus labrax*), seabream (*Sparus aurata*), turbot (*Scophthalmus maximus*), and dogfish (*Scyliorhinus caniculus*) IL-1 $\beta$  genes have all been characterized. In general, they all share approximately 25% amino acid sequence identity with the human gene. A second distinct IL-1 $\beta$  gene has been identified in rainbow trout and carp. The significance and function of this second IL-1 $\beta$  gene in fish immune responses have yet to be determined. Numerous IL-8-like molecules have also been sequenced in fish.

## FUNCTIONS OF CYTOKINES IN VERTEBRATE SPECIES

Cytokines are involved in orchestrating the immune and inflammatory responses in all vertebrates

(Table II). The biological activity of an individual cytokine is almost indistinguishable when its activity is observed in any vertebrate's immune system. For example, the function of IL-1 in the human immune system is the same as it is in fish, chicken, or frog. Similarly, levels of IL-2 in stressed seals are reported to be elevated, whereas levels of IL-6 in seals with inflammatory disorders are also elevated, just as would be observed in birds, fish, or any other vertebrate animal. Cytokines are important, ancient, and functionally conserved molecules.

## INVERTEBRATE CYTOKINE-LIKE MOLECULES?

Since invertebrate defense is only innate in nature, it is reasonable to assume that an IL-1-like molecule (or other pro-inflammatory cytokines) could be a part of the invertebrate immune response. Evidence of this possibility came in 1986 when an IL-1-like activity was isolated and characterized in an invertebrate (the echinoderm *Asterias forbesi*). The echinoderm IL-1-like molecule stimulated the proliferation of murine thymocytes and fibroblasts, which are standard assay systems for mammalian IL-1. In addition, an antibody to human IL-1 inhibited the activity of the echinoderm IL-1-like molecule in these assays. This finding gave credence to the hypothesis that there is an evolutionary continuity of IL-1. Other studies have shown that this result was not an isolated finding. Proteins with IL-1-like properties have been found in other invertebrates, such as tunicates, insects, and other echinoderms.

Other cytokine-like molecules have been identified in numerous invertebrates ranging from mussels (*Mytilus galloprovincialis* Lmk) to snails (*Biomphalaria glabrata*). These molecules include TNF $\alpha$ , IL-2, IL-6, and IL-8. All of these studies relied on functional assays that indicated a likeness to vertebrate cytokine activities. The significance and rationale of searching for cytokines that are involved in immunity in invertebrates is the assumption that these molecules will serve a similar function in the invertebrate immune response. There may be a central role for such molecules. Even though several invertebrate genomes have been sequenced, no cytokine-like proteins have been identified. The search for the gene sequences associated with these molecules continues.

Another line of evidence suggests that cytokine molecules could be found in invertebrates. The Toll receptor (TLR) proteins in *Drosophila* are involved in establishing the dorsal–ventral axis during

embryogenesis and also participate in the innate immune response. The molecule shows striking similarities in plants, invertebrates, and vertebrates. Twelve mammalian/human TLRs have thus far been described. The TLRs respond to different microbial products through a shared IL-1 receptor (IL-1R) signaling pathway. The TLRs share the IL-1R cytoplasmic signaling cascade. Both type I IL-1 receptor and TLRs are highly homologous in their cytosolic domains. This complex signaling pathway activates both the IL-1R and the Toll-like receptors in vertebrates.

### See Also the Following Articles

ACTH,  $\alpha$ -MSH, and POMC, Evolution of • Adipocytokines • Corticotropin-Releasing Hormone (CRH) and Inflammation • Cytokine Actions, Cellular Mechanism of • Cytokine Receptors • Cytokines, Constitutive Secretion • Cytokines, Extracellular Transport and Processing • Immune System, Hormonal Effects on • Interleukin-6 • Janus Kinases and Cytokine Receptors • Steroid Receptors, Evolution of • Tumor Necrosis Factor (TNF)

### Further Reading

- Beck, G. (1998). Macrokinines: Invertebrate cytokine-like molecules? *Front. Biosci.* **3**, 559–569.
- Beck, G., and Habicht, G. S. (1996). Immunity and the invertebrates. *Sci. Am.* **275**, 60–66.
- Beschino, A., Bilej, M., Torreele, E., and De Baetselier, P. (2001). On the existence of cytokines in invertebrates. *Cell. Mol. Life Sci.* **58**, 801–814.
- Dunn, E., Sims, J. E., Nicklin, M. J. H., and O'Neill, L. A. J. (2001). Annotating genes with potential roles in the immune system: Six new members of the IL-1 family. *Trends Immunol.* **22**, 533–536.
- Hoffmann, J. A., Kafatos, F. C., Janeway, C. A., and Ezekowitz, R. A. B. (1999). Phylogenetic perspectives in innate immunity. *Science* **284**, 1313–1318.
- Lowenthal, J. W., Larribrecht, B., van den Berg, T. P., Andrew, M. E., Strom, A. D. G., and Bean, A. G. D. (2000). Avian cytokines—The natural approach to therapeutics. *Dev. Comp. Immunol.* **24**, 355–365.
- O'Neill, L. (2000). The Toll/interleukin-1 receptor domain: A molecular switch for inflammation and host defense. *Biochem. Soc. Trans.* **28**, 557–563.
- Scapigliati, G., Bird, S., and Secombes, C. J. (2000). Invertebrate and fish cytokines. *Eur. Cytokine Network* **11**, 354–361.
- Scheerfinck, J. P. Y. (1999). Functional and structural comparison of cytokines in different species. *Vet. Immunol. Immunopathol.* **72**, 39–44.
- Secombes, C. J., Bird, S., Hong, S., Laing, K. J., and Zou, J. (2001). Phylogeny of vertebrate cytokines. In “Phylogenetic Perspectives on the Vertebrate Immune System” (G. Beck, M. Sugumaran, and E. L. Cooper, eds.), pp. 89–94. Kluwer Academic/Plenum, New York.
- Secombes, C. J., Wang, T., Hong, S., Peddie, S., Crampe, M., Laing, K. J., Cunningham, C., and Zou, J. (2001). Cytokines and innate immunity of fish. *Dev. Comp. Immunol.* **25**, 713–723.
- Siatskas, C., and Boyd, R. (2000). Regulation of chicken haemopoiesis by cytokines. *Dev. Comp. Immunol.* **24**, 37–59.

Sheddases are also known to process cytokine receptors such as TNF family receptors, IL-1 receptors, and IL-6 receptor  $\alpha$  chain. Theoretically, the importance of the regulation of cytokine activity by shedded cytokine receptors is twofold. First, sheddase inactivates receptors by removing the ligand-binding domain from the cell surface. Because this is a very quick process to inactivate receptors, this mechanism may be critical for shutting off cytokine signals. Second, soluble ligand-binding domains of receptors bind to their ligands and sequester the ligands from functioning. It has also been reported that some soluble cytokine receptors can stabilize cytokines when they are at low concentration. However, the physiological or pathological importance of the shedded cytokine receptors *in vivo* has to be elucidated further. Interestingly, the soluble binding protein for the inflammatory cytokine IL-18, IL-18-binding protein (IL-18BP), was identified. This protein is encoded by a unique gene and functions as an inhibitor for IL-18. Besides soluble receptor molecules, soluble cytokine-binding proteins can be another mechanism to regulate cytokine activities. This article summarizes the most extensively studied cytokine-processing proteases of the ADAM family and the caspase family.

## CYTOKINE PROCESSING BY THE ADAM FAMILY OF PROTEASES

### Sheddase for TNF Family

#### Identification of Sheddase of TNF Family

Phorbol ester, a well-known activator of protein kinase C, induces shedding of various cytokines, such as the TNF family, and cytokine receptors, such as TNF receptors, IL-1 receptors, and IL-2 and IL-6 receptor  $\alpha$  chains, from the plasma membrane. Other stimuli, such as bacterial toxins and C-reactive proteins, can induce shedding of some of these proteins as well. The shedding of these proteins by PMA is inhibited by hydroxamate compounds known as inhibitors for metalloproteases. TNF $\alpha$  is first produced as a 26-kD, membrane-bound form. TNF $\alpha$ -converting enzyme (TACE) sheds membrane-bound TNF $\alpha$  to yield a 17-kD mature soluble form. The activity of TACE is up-regulated by PMA and inhibited by hydroxamate compounds such as TNF $\alpha$  protease inhibitor (TAPI), indicating that TACE belongs to the metalloprotease family. TACE was biochemically purified, and the gene encoding TACE was cloned. It encodes 824 amino acids. TACE turned out to be one of the ADAM family proteases and was designated as ADAM17.

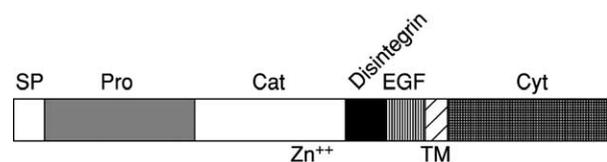
### Structure of ADAM

ADAM family proteases are also known as metalloprotease disintegrin cysteine rich (MDC). The first ADAM family member was identified as protease expressed on the sperm membrane and is crucial for the sperm-egg fusion. So far, there are 34 identified members in the mammalian ADAM family. In addition to the shedding of ectodomain of cytokines and cytokine receptors, some of the family members are involved in sperm-egg fusion, neural development, myoblast fusion, and many other developmental processes. ADAM family protease is a multidomain, type I transmembrane protein (Fig. 1). It consists of an N-terminal signal sequence, a pro-domain, a catalytic domain of metalloprotease that harbors a Zn<sup>2+</sup> binding site, a disintegrin cysteine-rich region, an epidermal growth factor (EGF) repeats cysteine-rich region, a transmembrane region, and a cytoplasmic tail of variable length. The domain structures are highly preserved; however, the function of each domain needs to be elucidated.

The disintegrin and EGF repeats domains of ADAMs play some roles in cell-cell interaction. These cysteine-rich domains of ADAM1 are required for the sperm-egg binding. The same region of ADAM12 mediates adhesion and cell spreading by binding to syndecans. The disintegrin domain of sheddase, ADAM17, is partly required for the susceptibility of the metalloprotease activity to hydroxamate.

Pro-domain is processed at the amino acid sequence "RVKR" by fulin-like convertases in the trans-Golgi network. Because the pro-domain is blocking the catalytically active site, the removal of the pro-domain is necessary for ADAMs to become active metalloproteases.

The catalytic domain of some ADAMs contains a consensus sequence "HEXXHXXGXXH(D)" that is highly conserved in Zn<sup>2+</sup>-dependent metalloproteases. Three histidine residues in the motif are involved in binding the catalytically essential Zn ion.



**Figure 1** Domain structure of the ADAM family. ADAM family protein consists of a signal peptide (SP), a pro domain (Pro), a metalloprotease catalytic domain with a Zn<sup>2+</sup> binding site (Cat), a disintegrin cysteine-rich region (Disintegrin), an EGF repeats cysteine-rich region (EGF), a transmembrane domain (TM), and a cytoplasmic domain (Cyt).

Anchoring ADAM17 to the membrane is essential for this enzyme to shed various cell surface proteins. The function of the cytoplasmic tail is unknown. The cytoplasmic domain of ADAM17 is not required for shedding various substrates, including TNF $\alpha$ .

#### **ADAM17-Disrupted Mice**

R. A. Black and colleagues created mice harboring the deletion of the Zn<sup>2+</sup>-binding domain of ADAM17, thereby inactivating the metalloprotease activity. They also established immortalized fibroblast cell lines from these mutant mice. By using the mutant mice as well as the cell lines, they showed that ADAM17 is the major sheddase not only for TNF $\alpha$  but also for p75 tumor necrosis factor receptor (p75TNFR), IL-1 receptor-II (IL-1R-II), p66TNFR, tumor growth factor- $\alpha$  (TGF $\alpha$ ), and colony-stimulating factor-1(CSF-1). The essential role of ADAM17 in development was also revealed. The majority of ADAM17 mutant mice die between embryonic day 17.5 and the first day after birth. The mutant mice show failure of eyelid fusion as well as defects in epithelial cell maturation and organization observed in skin, intestine, lung, stomach, and the airway branching. The phenotypes of eye, hair, and skin are similar to those of mice disrupted with EGFR or EGFR ligands such as EGF and TGF $\alpha$ , suggesting that shedding of these cytokines by ADAM17 is critical for their function in the development of skin. There are six members of the EGFR ligand: EGF, heparin-binding EGF (HB-EGF), TGF $\alpha$ , amphiregulin, betacellulin, and epiregulin. Presumably, ADAM17 is required for the shedding of all of those EGFR ligand proteins.

#### **G Protein-Coupled Receptor and HB-EGF Shedding**

It is well known that activators of G protein-coupled receptor (GPCR) induces transient low-level signaling through EGFR, a process termed "transactivation." Transactivation is observed in various signaling pathways of different types of cells. In the heart, vasoactive molecules such as phenylephrine (PE), angiotensin II (AngII), and endothelin-1 (ET-1) use GPCR and induce cardiac hypertrophy. The signal through GPCR activates sheddase that sheds HB-EGF, a member of the EGFR ligand family. Among the various metalloprotease inhibitors, KB-R7785 showed selectivity to HB-EGF shedding. In two-hybrid screening of the human heart cDNA library by protein kinase C- $\delta$  (PKC $\delta$ ) as a bait,

ADAM12 was isolated. KB-R7785 inhibited pressure overload-induced or vasoactive molecule-induced hypertrophy in mice. Furthermore, in ADAM12-deficient mice, the shedding of HB-EGF from mouse embryonic fibroblasts was severely impaired. These results suggest that ADAM12 is a specific sheddase for HB-EGF, especially for cardiomyocyte. HB-EGF shedding activity of ADAM9 was shown by the over-expression of the wild type and mutant form of this ADAM family member *in vitro*. However, ADAM9-deficient mice did not show the impairment of the shedding of HB-EGF. The target specificity of the ADAM family *in vivo* might be regulated not only by the substrate specificity of ADAM itself but also by the association with other molecules required for the interaction with the substrates.

### **REGULATION OF INFLAMMATORY CYTOKINES, IL-1 $\beta$ AND IL-18**

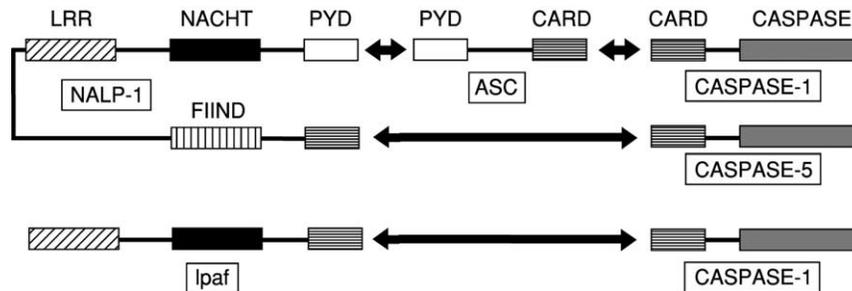
#### **Regulatory Mechanism of the Processing of IL-1 $\beta$ and IL-18**

##### **Biological Activities of IL-1 $\beta$ and IL-18**

IL-1 $\beta$  and IL-18 are two important pro-inflammatory cytokines produced by various cell types, including macrophages and dendritic cells. IL-1 $\beta$  stimulates inflammatory responses such as fever, hypotension, lymphocyte activation, recruitment of leukocytes to the site of inflammation, anorexia, hypoglycemia, and induction of acute phase proteins by activating the production of pro-inflammatory cytokines, including itself, IL-6, IL-8, and TNF $\alpha$ . IL-18 plays an important role in inflammatory responses, especially in the induction of pro-inflammatory cytokine IFN $\gamma$  and in the enhancement of T-helper subset Th1 differentiation with IL-12.

##### **Processing of IL-1 $\beta$ and IL-18**

Both cytokines are produced as pro-form (31 kDa for IL-1 $\beta$  and 24 kDa for IL-18) that have very weak biological activities and no signal peptides. Therefore, they are not produced in the regular secretory pathway but rather are produced in cytoplasm. Pro-form of these cytokines is converted to active form on the processing by caspase-1, also known as the IL-1 $\beta$ -converting enzyme (ICE), one of the members of inflammatory caspases. Human caspases-1,-4, and-5 and mouse caspases-1,-11, and-12 are grouped as inflammatory caspases, whereas other caspases play roles in apoptosis. Caspase-11 collaborates with caspase-1 for the processing of these cytokines in mice. This has been clearly shown by the defect of the active



**Figure 2** Structure of the large protein complex of inflammatory caspases caspase-1/ASC/NALP-1/caspase-5 and caspase-1/Ipaf. CARD domain, ■; caspase catalytic domain, ■; PYD domain, □; NACHT domain, ■; LRR domain, ▨; FIIND domain, ▨.

form of IL-1 $\beta$  or IL-18 production in mice deficient in either caspase-1 or caspase-11. The human counterpart of mouse caspase-11 is presumably caspase-5. These caspases are produced with N-terminal pro-domain that needs to be removed for the protease activity and caspase recruitment domain (CARD) in addition to the C-terminal protease domain.

#### Regulation of IL-1 $\beta$ and IL-18 Activities

Both IL-1 $\beta$  and IL-18 are regulated at the transcriptional level. These cytokine genes are induced by the signaling pathway from IL-1Rs, TNFRs, toll-like receptors (TLR), or other inflammatory cytokine receptors. Transcription factors such as NF- $\kappa$ B, IFN consensus sequence-binding protein (ICSBP), Pu.1, and AP-1 have been shown to be important for the transcriptional activation of these cytokine genes. Another mechanism for the regulation of this cytokine activity is the processing of pro-domain. Stimuli such as bacterial toxins, hypotonic stress, nigericin, antimicrobial peptides, and LPS can facilitate the processing of IL-1 $\beta$ . Pro-IL-18 has been shown to be processed on bacterial or viral infection. In addition, the majority of these stimuli can induce apoptosis of the cytokine-producing cells.

The mechanism that regulates the activity of inflammatory caspases is unknown; however, those caspases are shown to form large-molecular-weight complexes (approximately 1 mDa) of multiple adaptor proteins involved in inflammatory responses (Fig. 2). These complexes of various inflammatory proteins are termed “inflammasome.” One inflammasome consists of adaptor proteins, NALP-1 (nacht, LRR, and PYD containing protein 1) and ASC (apoptosis-associated, speck-like protein containing a CARD), and two inflammatory caspases, caspase-1 and caspase-5. Both caspase-1 and caspase-5 harbor the CARD domain to interact with the CARD domain of ASC and

NALP-1, respectively, in homophilic fashion. ASC associated with caspase-1 binds to NALP-1 through the homophilic binding of the PYD (pyrin) domain. NALP-1 self-associates and presumably forms a multisubunit complex with caspase-1 and caspase-5. Caspase-1 associates with another adaptor protein, Ipaf (ICE protease-activating factor/CARD12), by the homophilic interaction of the CARD domain. The activation mechanism of pro-apoptotic caspase, caspase-9, has been proposed. Cytochrome C released from mitochondria induces the assembly of adaptor protein, Apaf-1 (apoptotic protease-activating factor-1), to form apoptosome. Then, pro-caspase-9 binds to each Apaf-1 through the homophilic interaction of the CARD domain, resulting in the multimerization of pro-caspase-9. This assembly of pro-caspase-9 leads to the intermolecular processing and activation. This model might be applicable to the caspase-1 combined with Ipaf or NALP-1, pycard, and caspase-5. However, the ligand equivalent to cytochrome C for the inflammasome has not yet been identified.

Proteins carrying only the CARD domain, pseudo-ICE or ICEBER, inhibit caspase-1 activation by blocking the assembly of caspases with adaptor protein complex through the CARD–CARD interaction. In addition, activated inflammatory caspases are released from cells rapidly. This could be the mechanism to cease continuous activation of the inflammatory cytokines. Activation of inflammatory cytokines could be very harmful to the host; thus, the activity must be strictly regulated.

#### IL-18-Binding Protein

Besides the shedded cytokine receptors, cytokine-binding protein has been reported. IL-18-binding protein (IL-18BP) was identified from both mouse and human. IL-18BP carries signal peptide but no

transmembrane domain. Its molecular weight is approximately 40 kDa. It binds to IL-18 and prevents it from binding to the receptor. Thus, this protein can inhibit the IL-18-dependent IFN $\gamma$  production both *in vitro* and *in vivo*. Administration of IL-18BP into mice ameliorates colitis in the mouse model. However in patients with Crohn's disease, both IL-18 and IL-18BP are increased in the intestinal tissue. Its physiological and pathological functions need to be further elucidated. A creation of gene-targeted mice is especially necessary. A search for such binding proteins for other cytokines is also required.

### See Also the Following Articles

Adipocytokines • Cytokine Actions, Cellular Mechanism of • Cytokine Receptors • Cytokines, Constitutive Secretion • Cytokines, Evolutionary Aspects and Functions • EGF and Related Growth Factors • Janus Kinases and Cytokine Receptors • Tumor Necrosis Factor (TNF)

### Further Reading

- Asakura, M., Kitakaze, M., Takashima, S., Liao, Y., Ishikura, F., Yoshinaga, T., Ohmoto, H., Node, K., Yoshino, K., Ishiguro, H., Asanuma, H., Sanada, S., Matsumura, Y., Takeda, H., Beppu, S., Tada, M., Hori, M., and Higashiyama, S. (2002). Cardiac hypertrophy is inhibited by antagonism of ADAM12 processing of HB-EGF: Metalloprotease inhibitors as a new therapy. *Nature Med.* **8**, 35–40.
- Burns, K., Martinon, F., and Tschopp, J. (2003). New insights into the mechanism of IL-1 $\beta$  maturation. *Curr. Opin. Immunol.* **15**, 26–30.
- Corbaz, A., ten Hove, T., Herren, S., Graber, P., Schwartzburd, B., Belzer, I., Harrison, J., Plitz, T., Kosco-Vilbois, M. H., Kim, S. H., Dinarello, C. A., Novick, D., van Deventer, S., and Chvatchko, Y. (2002). IL-18-binding protein expression by endothelial cells and macrophages is up-regulated during active Crohn's disease. *J. Immunol.* **168**, 3608–3616.
- Kheradmand, F., and Werb, Z. (2002). Shedding light on sheddases: Role in growth and development. *BioEssays* **24**, 8–12.
- Mullberg, J., Althoff, K., Jostock, T., and Rose-John, S. (1999). The importance of shedding of membrane proteins for cytokine biology. *Eur. Cytokine Network* **11**, 27–38.

or developmental defects of the hypothalamus or by destructive lesions such as tumors, inflammatory processes, vascular lesion, and trauma. Usually, patients with isolated gonadotropin deficiency have normal height for their age during the prepubertal period, in contrast to patients with constitutional delay, who are short. In hypogonadotropic hypogonadism, gonadotropin responses to GnRH stimulation may be subnormal, but because of the functional hypogonadotropism in constitutional delay, the differential diagnosis between these two conditions may be difficult.

### Temporary Gonadotropin Deficiencies

#### **Anorexia Nervosa**

Anorexia nervosa is usually associated with severe or even fatal weight loss, which is due to distorted body image, obsessive fear of obesity, and avoidance of food. Virtually all patients have primary or secondary amenorrhea. Functional hypogonadotropic hypogonadism is at least partly due to severe weight loss, but amenorrhea may also precede the onset of weight loss. The underlying pathophysiology of amenorrhea is due to GnRH deficiency because the luteinizing hormone (LH) secretory pattern in pubertal-aged girls with anorexia is similar to that seen in girls during prepuberty: low or absent LH pulses and blunted LH response to exogenous GnRH. Pulsatile administration of GnRH has been shown to restore a pubertal pattern of LH secretion, confirming the hypothalamic location of the defect. Recovery of normal weight will normalize most endocrine and metabolic functions, but amenorrhea may persist for years.

#### **Athletic Training**

Overly intensive exercise may inhibit GnRH secretion, arrest pubertal development, and cause amenorrhea. These disorders are common especially among long-distance runners, gymnasts, and ballerinas. Hypogonadotropic hypogonadism may develop even when the female athletes have normal weight but have less fat and more muscle than do nonathletic girls. In female athletes with delayed or arrested pubertal development, adrenarche usually takes place at the normal age. The mechanism of delayed puberty is unclear, but interruption of intensive training advances puberty and menarche before any change in body composition or weight, suggesting a direct effect physical activity on GnRH secretion.

#### **Malnutrition and Chronic Diseases**

In malnutrition and chronic diseases, weight loss below the level of 80% of ideal body weight can cause delayed or arrested pubertal development. Nutrition plays an important yet uncharacterized role in the control of GnRH secretions. For example, in regional enteritis, gonadotropin secretion remains normal if nutrition is optimally balanced, but a non-optimal nutritional status will result in a hypogonadotropic state and arrested pubertal maturation. Chronic renal insufficiency delays pubertal development, but after successful renal transplantation, gonadotropin secretion is usually restored.

### Central Nervous System Tumors

Tumors causing delayed puberty most commonly interfere with GnRH synthesis or secretion. Deficiency of other pituitary hormones is common. Associated posterior pituitary hormone deficiencies are often manifested by diabetes insipidus.

#### **Craniopharyngioma**

The most common neoplasm causing hypothalamic-pituitary dysfunction and hypogonadotropic hypogonadism is craniopharyngioma. It is actually a congenital tumor that most commonly becomes symptomatic between 6 and 14 years of age. At presentation, the most common symptoms are headache, visual disturbances, short stature, delayed puberty, polyuria, and polydipsia. Skull radiographs may show suprasellar or intrasellar calcification or an abnormality in sella turcica. CT scans may reveal fine calcifications that are not apparent on routine skull radiographs. The structure of the tumor varies from solid to cystic. Treatment consists of surgery and radiotherapy, but the recurrence rate is high even when complete surgical removal is attempted.

#### **Langerhans' Cell Histiocytosis**

Langerhans' cell histiocytosis (LCH), also called Hand-Schüller-Christian disease or histiocytosis X, is characterized by infiltration of lipid-containing histiocytic cells in the skin, bone, and viscera. Cyst-like areas can be found in flat bones of the skull, pelvis, dorsolumbar spine, scapula, and long bones of the arms and legs by X ray. CNS involvement and, in particular, hypothalamic-pituitary involvement are well-described features of LCH. The precise incidence of CNS-LCH disease is unknown, and the natural history is poorly understood. Diabetes insipidus (DI) is reported to be the most common and

well-described manifestation of hypothalamic–pituitary involvement (up to 50%). Anterior pituitary dysfunction has been reported in up to 20% of patients with LCH and occurs almost exclusively concurrently with DI. Although histiocytosis is not a tumor, it can be treated with chemotherapeutic agents, especially vinblastine. However, the natural course of the disease is fluctuating, making evaluation of treatment effect difficult. Endocrine function is not improved following medical treatment of LCH with chemotherapy and glucocorticoids. All LCH patients should undergo a thorough endocrine evaluation periodically.

### **Germinomas**

Germinomas are the extrasellar tumors that often cause delayed puberty, although these tumors are a rarity among primary CNS tumors. Polydipsia, polyuria, and visual disturbances are the most common symptoms associated with these tumors, followed by arrested growth and delayed puberty. Germinomas are commonly located in the pituitary stalk, in the suprasellar region of the hypothalamus, or in the proximity of the pineal gland. Seeding of the tumor to the cerebrospinal fluid is common, and this can also be used in the diagnosis, where examination of tumor markers (hCG $\beta$  and alphafetoprotein) or germ cells (with positive placental alkaline phosphatase staining) in the cerebrospinal fluid may be helpful. These laboratory findings, together with clinical features and an excellent response to radiation therapy, are so characteristic that surgery is rarely indicated except for biopsy to establish histological diagnosis.

## **Other Central Nervous System Disorders**

### **Defects in Development**

Various malformations affecting the development of the prosencephalon may cause hypogonadotropic hypogonadism combined with deficiency of any or all other pituitary hormones. Midline malformations are often associated with optic dysplasia, and absent septum pellucidum is often found by imaging techniques (septo-optic dysplasia). Other congenital midline defects, which may range from holoprosencephaly to cleft lip and palate, may also be associated with variable hypothalamic–pituitary dysfunction.

Genetic defects affect development of the anterior pituitary cause hypopituitarism, including hypogonadotropic hypogonadism in some cases. During fetal development of the anterior pituitary gland, a number

of sequential processes occur that affect cell differentiation and proliferation. Recent advances in molecular biology have revealed several steps that are required for pituitary cell line specification, and several genes have been identified to play a role in control of these steps. Mutations in the DAX-1 gene cause X-linked adrenal hypoplasia congenita and mutations in the DAX-1-related orphan nuclear receptor, steroidogenic factor-1, and leptin and prohormone convertase-1 may influence GnRH release and processing of the GnRH receptor. The pituitary transcription factors, HESX-1, LHX-3, and PROP-1, are important for the development of gonadotropin-secreting cells. Depending on the condition, different approaches for counseling are needed. Despite recent advances, the pathophysiological basis of hypogonadotropic hypogonadism in the majority of individuals remains unclear. Recently, compound heterozygote mutations in the GnRH receptor gene were described in both males and females, and hormonal resistance was confirmed *in vitro*.

Because of different causes, there is a wide spectrum of phenotypes, ranging from complete hypogonadotropic hypogonadism with lack of pubertal development to a partial hypogonadism with an arrest of pubertal development. In complete GnRH resistance, endogenous LH secretory patterns are abnormal, either apulsatile or characterized by a low to normal pulse frequency with small pulses or erratic pulses of low amplitude. In patients with partial resistance, basal LH plasma concentrations are low, but follicle-stimulating hormone (FSH) levels are in the normal range.

### **Kallmann Syndrome**

Kallmann syndrome is the most common form of isolated hypogonadotropic hypogonadism. Hypogonadism is due to GnRH deficiency, and the other components of the syndrome include anosmia or hyposmia due to hypoplasia of the olfactory lobes and occasionally cleft lip and palate, unilateral renal agenesis, short metacarpals, sensorineural hearing loss, and color-blindness. About half of Kallmann syndrome patients have mutations in the KAL gene on chromosome Xp22.3. This gene encodes an extracellular matrix protein that regulates axonal path finding and cellular adhesion. Defect in this gene causes failure of fetal GnRH neurosecretory neurons to migrate from the olfactory placode to the medio–basal hypothalamus, causing hypoplasia of the olfactory sulci. Autosomal disorders (dominant or recessive) may also cause Kallmann syndrome, but the gene defects of the forms have not yet been characterized.

Because about half of the patients have X-linked disorder, the syndrome is more common in boys than in girls.

### **Iatrogenic Gonadotropin Deficiencies**

Treatment of CNS tumors, leukemia, or neoplasms with cranial irradiation may result in gradual development of hypothalamic–pituitary failure. GH deficiency is the most common component of the radiation-induced hormone disorder, but gonadotropin deficiency also occurs when the radiation dose is high enough. Development of radiation-induced hypothalamic–pituitary failure usually takes from 1 year to several years to develop.

## **HYPERGONADOTROPIC HYPOGONADISM**

### **Gonadal Dysgenesis Syndromes**

#### **Turner Syndrome**

The syndrome of gonadal dysgenesis (Turner syndrome) is the most common form of hypergonadotropic hypogonadism, affecting about 1 in 2500 live-born girls. About half of girls with Turner syndrome have the 45,X karyotype, but about 99% of fetuses with this karyotype abort spontaneously, and in 1 of 15 spontaneous abortions the fetus has the 45,X karyotype. Turner syndrome may be regarded as a continuum ranging from the typical 45,X phenotype to a normal male or female phenotype. Chromosomal mosaicism and structural abnormalities of the sex chromosomes modify the clinical features. Typical features include short stature (which may be apparent already at birth), lymphedema of the extremities and loose posterior cervical skinfolds during the newborn period, low-set or deformed ears, epicanthal folds, ptosis, micrognathia, high arched palate, dental abnormalities, wide-spaced nipples caused by shield-like chest, hypoplastic areolae, short neck with low hairline, and cubitus valgus. Abnormalities of the left side of the heart include coarctation of the aorta, aortic stenosis, and bicuspid aortic valves. Renal anomalies include abnormal position or alignment (horseshoe kidney) and various anomalies on the collecting system. Patients with Turner syndrome have increased incidence of inflammatory bowel disease, autoimmune thyroiditis, Graves' disease, and insulin resistance. Intelligence is usually normal, but spatiotemporal processing, visuomotor coordination, and mathematical skills performance may be impaired.

Ovarian insufficiency is also apparent at birth, as evidenced by high gonadotropin concentrations during the neonatal period. During childhood, with the development of the CNS-mediated inhibition of GnRH secretion, gonadotropin levels decrease to near normal levels, but by 10 years of age they usually are elevated again. The Müllerian structure (uterus and fallopian tubes) is present but remains infantile if the ovarian failure is not adequately treated with hormone replacement therapy. Histologically, the ovaries are streaks of connective tissue, with a decreased number of primordial follicles and oocytes for age. Spontaneous oocyte death is accelerated, resulting in complete loss of the oocyte pool in most cases.

Sexual infantilism is one of the most common clinical findings in girls with Turner syndrome. More than 90% have gonadal failure. It is important to remember, however, that up to 30% of girls will undergo spontaneous pubertal development and that 2 to 5% will have spontaneous menses and may have the potential to achieve pregnancy without medical intervention. Pubertal development may be delayed and, in most patients, is followed by progressive ovarian failure.

Most patients are small for their gestational age at birth, and the slow growth rate is apparent after 3 years of age. Most girls fail to have a pubertal growth spurt due to insufficient estrogen production in the ovaries. The mean adult height is approximately 143 to 146 cm, depending on both parental heights and the overall height of the same genetic population.

#### **Pure Gonadal Dysgenesis**

The term “pure gonadal dysgenesis” refers to phenotypic females with no pubertal development and the 46,XX or 46,XY karyotype without detectable chromosomal abnormalities. Patients with 46,XX gonadal dysgenesis have normal stature, bilateral streak gonads, normal female internal and external genitalia, and (sometimes) sensorineural deafness. Malignant transformation of the streak gonad is rare. Most cases are sporadic, but autosomal-recessive form has also been described. Patients with familial or sporadic 46,XY gonadal dysgenesis have normal female internal and external genitalia with occasional clitoral enlargement due to increased testosterone production by the gonad, bilateral streak gonads, and normal or tall stature with eunuchoid body proportions. The dysgenetic gonads may undergo neoplastic transformation, so gonadectomy is indicated.

## Other Causes of Primary Ovarian Failure

### **Irradiation and Chemotherapies**

Damage to the gonads by irradiation or chemotherapy depends on the patient's gender, age at the time of treatment, radiation dose and fractionation schedule, and total dose and nature of chemotherapy delivered. Most chemotherapy protocols use multiple agents whose effects may be synergistic. Biochemical detection of gonadal damage is rarely possible before puberty, so treatment-induced gonadal damage during childhood may present with infertility or premature menopause during adulthood. Abdominal, pelvic, and total body irradiation may result in ovarian and uterine damage. The human oocyte is sensitive to radiation, with an estimated LD50 of less than 4 Gy. Less than 2% of children receiving total body irradiation subsequently became pregnant, although there may be some protection of ovarian function in prepubertal girls. Uterine radiation increases the incidence of nulliparity, fetal loss, and small-for-dates infants, and it reduces the success of assisted reproduction. Suppression of the pituitary–gonadal axis with gonadal steroids or GnRH agonists does not protect the ovary from the damage induced by irradiation or chemotherapy.

### **Hypergonadotropic Ovarian Failure**

Hypergonadotropic ovarian failure, also termed the “resistant ovary syndrome,” is a heterogeneous disorder with unknown etiology in most cases. Approximately 40% of patients have a point mutation, causing a single amino acid substitution in the extracellular domain of the FSH receptor and subsequent inactivation of the receptor function. Patients with and without the receptor mutation have amenorrhea with variable development of secondary sex characteristics and have high serum levels of gonadotropins. Histological examinations of ovary biopsies show the presence of follicles in all patients with the receptor defect, whereas only 1 in 4 of those with unknown etiology has follicles. Hence, whereas the receptor defect causes a specific arrest in follicular maturation, most patients with hypergonadotropic ovarian failure have true ovarian dysgenesis. The ovarian phenotype in patients with inactivating FSH receptor mutation is informative in regard to the role of FSH in the regulation of follicular development: the early phases of follicular maturation (up to the preantral stage) are independent of FSH, but for the final maturation of the follicle, this gonadotropin is absolutely necessary.

### **Autoimmune Ovarian Failure**

Autoimmune ovarian failure is often one of the components of autoimmune polyendocrinopathies. Autoimmune polyglandular syndrome type I is an autosomal recessive disorder caused by mutation in the autoimmune regulator (AIRE) gene, which maps to 21q22.3. It is characterized by two of the three major clinical symptoms that may be present: Addison disease, hypoparathyroidism, and chronic mucocutaneous candidiasis. Moniliasis often precedes symptoms and signs of endocrinopathy. Furthermore, hypoparathyroidism usually reveals itself before adrenal insufficiency.

Ahonen and colleagues reported data from a 10-month to 31-year follow-up of 68 patients from 54 families, ages 10 months to 53 years at the time of report. Hypoplasia of the dental enamel and keratopathy were frequent and were not attributable to hypoparathyroidism. Some of the manifestations of the disorder did not appear until the fifth decade. Thus, all patients need lifelong follow-up for the detection of new components of the disease. Candidiasis was the initial manifestation in 60% of the patients and was present in all patients at some time. Hypoparathyroidism was present in 79%, adrenocortical failure in 72%, and gonadal failure in 60% of female patients over 13 years of age.

## DIAGNOSIS

The cutoff age for identifying girls who need evaluation for delayed puberty may vary in different ethnic groups, but in most populations early signs of secondary sexual development should be present by 13 years of age.

A thorough history should note the signs of anorexia, the intensity of athletic training, and the timing of puberty of both parents (Fig. 1). In constitutional delay of puberty, often either one of the parents developed late. A history of chronic illnesses, such as celiac disease and inflammatory bowel disease, will suggest a temporary or secondary delay of puberty. Stature and height velocity should be evaluated using appropriate growth charts. Height velocity is usually slow in patients with constitutional delay and is usually normal in patients with hypogonadotropic hypogonadism, but especially hypogonadotropic states cannot be ruled out by short stature and slow growth rate. Likewise, bone age (X-ray film of left hand and wrist read according to standards such as Greulich and Pyle) delay provides useful information

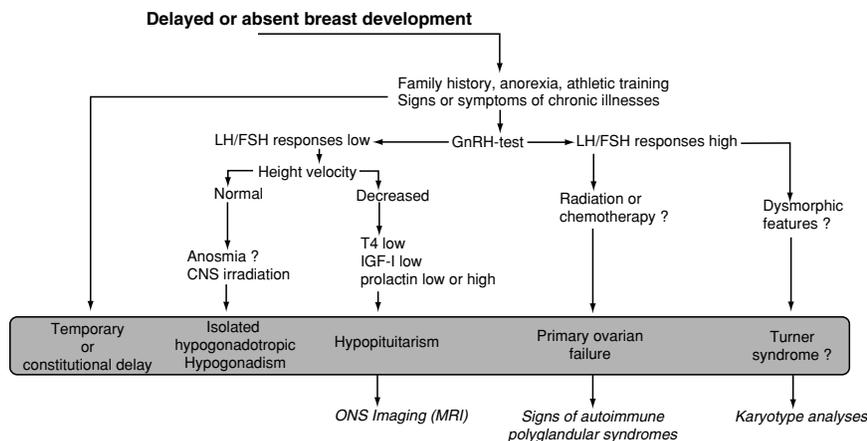


Figure 1 Differential diagnosis of delayed puberty.

in the growth analysis, but it contributes little to the differential diagnosis.

Gonadotropin levels assessed by a single basal LH and FSH determination are often increased in primary ovarian failure or in Turner syndrome, but the basal gonadotropin values are not useful in the differential diagnosis of constitutional delay and hypogonadotropic hypogonadism. Dynamic testing, such as administration of synthetic GnRH, may provide information for the differential diagnosis. In some girls with constitutional delay, a pubertal pattern of response (post-GnRH maximum LH higher than maximum FSH) may be observed, but a low prepubertal response to GnRH can be found in some girls with constitutional delay and typically in hypogonadotropic hypogonadism. Follow-up is often warranted before a definitive diagnosis can be made.

## MANAGEMENT OF DELAYED PUBERTY

When estrogen therapy is required to induce pubertal development, the dosing and timing should be aimed at mimicking normal pubertal development, taking account of the individual's desire to begin puberty and also of the family history of age at onset of puberty. Doses should be adjusted to the responses of individual patients, who may be monitored in terms of the development of secondary sex characteristics, bone maturation, and/or uterine volume.

In hypopituitary girls and in girls with Turner syndrome, estrogen therapy should be coordinated with the use of GH. This should be individualized for each patient to optimize both growth and pubertal development. When growth promotion is a priority,

consideration should be given to delaying estrogen therapy to avoid compromising final height.

Before initiation of estrogen therapy in girls with Turner syndrome, serum gonadotropin levels should be determined to exclude the possibility of delayed spontaneous pubertal development. If gonadotropin levels are normal, a sonographic examination should be undertaken to determine the status of the gonads. Hormonal induction of feminization should be initiated and carried out in a manner that simulates the normal growth and development of secondary sex characteristics as closely as possible. Estrogen therapy needs to be initiated and adjusted according to the needs and priorities of the individual. Thus, if growth promotion is a priority, estrogen therapy should not be initiated before 12 years of age unless height has already been maximized. In girls with Turner syndrome, estrogen therapy should ideally be started by 15 years of age. Estrogen therapy should be initiated at a low dose (one-sixth to one-quarter of the adult dose) and increased gradually (at intervals of 3–6 months). Doses can then be adjusted to the response (Tanner stage, bone age, or uterine growth), with the aim of completing feminization gradually over a period of 2 to 3 years.

A progestin such as medroxyprogesterone should be added either when vaginal bleeding first occurs or after 12 to 24 months of estrogen therapy to establish monthly menstrual cycles.

Individuals with Turner syndrome who have functioning ovaries and who progress through puberty spontaneously should receive contraceptive and genetic counseling. However, ovulatory function should be documented (FSH and LH measurements) because a perimenopausal pattern of anovulation can lead to endometrial hyperplasia.

## See Also the Following Articles

Constitutional Delay of Growth and Puberty (CDGP) • Craniopharyngiomas • Delayed Puberty and Hypogonadism, Male • Delayed Puberty, Male • Precocious Puberty, Central (Female) • Pseudoprecocious Puberty, Female • Puberty: Physical Activity and Growth

## Further Reading

- Achermann, J. C., and Jameson, J. L. (2001). Advances in the molecular genetics of hypogonadotropic hypogonadism. *J. Pediatr. Endocrinol.* **14**, 3–15.
- Ahonen, P., Myllarniemi, S., Sipila, I., and Perheentupa, J. (1990). Clinical variation of autoimmune polyendocrinopathy–candidiasis–ectodermal dystrophy (APECED) in a series of 68 patients. *N. Engl. J. Med.* **322**, 1829–1836.
- Aittomaki, K., Lucena, J. L., Pakarinen, P., Sistonen, P., Tapanainen, J., Gromoll, J., Kaskikari, R., Sankila, E-M., Leivaslaiho, H., and Engel, A. R. (1995). Mutation in the follicle-stimulating hormone receptor gene causes hereditary hypergonadotropic ovarian failure. *Cell* **82**, 959–968.
- Blaschke, R. J., and Rappold, G. A. (2000). SHOX: Growth, Leri-Weill, and Turner syndromes. *Trends Endocrinol. Metab.* **11**, 227–230.
- Saenger, P., Wikland, K. A., Conway, G. S., Davenport, M., Gravholt, C. H., Hintz, R., Hovatta, O., Hultcrantz, M., Landin-Wilhelmsen, K., Lin, A., Lippe, B., Pasquino, A. M., Ranke, M. B., Rosenfeld, R., and Silberbach, M. (2001). Fifth International Symposium on Turner Syndrome: Recommendations for the diagnosis and management of Turner syndrome. *J. Clin. Endocrinol. Metab.* **86**, 3061–3069.
- Seminara, S. B., Hayes, F. J., and Crowley, W. F., Jr. (1998). Gonadotropin-releasing hormone deficiency in the human (idiopathic hypogonadotropic hypogonadism and Kallmann's syndrome): Pathophysiological and genetic considerations. *Endocrine Rev.* **19**, 521–539.
- Wallace, W. H. B., Shalet, S. M., and Hendry, J. H. (1989). Ovarian failure following abdominal irradiation in childhood: The radio-sensitivity of the human oocyte. *Brit. J. Radiol.* **62**, 995–998.

Table I The Classification of Delayed Puberty and Hypogonadism in the Male

<b>Idiopathic (constitutional) delay in growth and puberty</b> <b>Hypogonadotropic hypogonadism: Sexual infantilism related to gonadotropin deficiency</b>	<b>Idiopathic and genetic forms of multiple pituitary hormone deficiencies including Prop-1 mutation</b>
<p><i>CNS disorders</i></p> <p>Tumors</p> <ul style="list-style-type: none"> <li>Craniopharyngiomas</li> <li>Germinomas</li> <li>Other germ cell tumors</li> <li>Hypothalamic and optic gliomas</li> <li>Astrocytomas</li> <li>Pituitary tumors (including MEN1, prolactinoma)</li> </ul> <p>Other causes</p> <ul style="list-style-type: none"> <li>Langerhans' histiocytosis</li> <li>Postinfectious lesions of the CNS</li> <li>Vascular abnormalities of the CNS</li> <li>Radiation therapy</li> <li>Congenital malformations especially associated with craniofacial anomalies</li> <li>Head trauma</li> <li>Lymphocytic hypophysitis</li> </ul> <p><i>Isolated gonadotropin deficiency</i></p> <ul style="list-style-type: none"> <li>Kallmanns syndrome <ul style="list-style-type: none"> <li>With hyposmia or anosmia</li> <li>Without anosmia</li> </ul> </li> <li>LHRH receptor mutation</li> <li>Congenital adrenal hypoplasia (DAX1 mutation)</li> <li>Isolated LH deficiency</li> <li>Isolated FSH deficiency</li> <li>Prohormone convertase 1 deficiency</li> </ul>	<p><i>Miscellaneous disorders</i></p> <ul style="list-style-type: none"> <li>Prader-Willi syndrome</li> <li>Laurence-Moon syndrome</li> <li>Bardet-Biedl syndrome</li> <li>Functional gonadotropin deficiency <ul style="list-style-type: none"> <li>Chronic systemic disease and malnutrition <ul style="list-style-type: none"> <li>Sickle cell disease</li> <li>Cystic fibrosis</li> </ul> </li> <li>Acquired immunodeficiency syndrome</li> <li>Chronic gastroenteric disease</li> <li>Chronic renal disease</li> </ul> </li> <li>Malnutrition</li> <li>Anorexia nervosa</li> <li>Bulimia</li> <li>Impaired puberty in male athletes</li> <li>Hypothyroidism</li> <li>Diabetes mellitus</li> <li>Cushing's disease</li> <li>Hyperprolactinemia</li> <li>Marijuana use</li> <li>Gaucher's disease</li> </ul> <p><b>Hypergonadotropic hypogonadism</b></p> <ul style="list-style-type: none"> <li>Seminiferous tubular dysgenesis syndrome and its variants (Klinefelter's syndrome)</li> <li>Other forms of primary testicular failure <ul style="list-style-type: none"> <li>Chemotherapy</li> <li>Radiation therapy</li> <li>Testicular steroid biosynthetic defects</li> <li>Sertoli-only syndrome</li> <li>LH receptor mutation</li> <li>Anorchia and cryptorchidism</li> <li>Trauma/surgery</li> </ul> </li> </ul>

*Source.* Modified from Grumbach, M. M. and Styne, D. M. (2003). Puberty: Ontogeny, neuroendocrinology, physiology, and disorders. *In* "Williams' Textbook of Endocrinology" (P. R. K. H. M. Larsen, S. Melmed, and K. S. Polonsky, eds.), pp. 1115–1286. W. B. Saunders, Philadelphia PA, with permission.

reflecting maturation of the hypothalamic–pituitary system. Alternatively, an 8 AM serum testosterone value of 0.7 nmol/liter (20 ng/dl) heralds the development of puberty in boys within 12–15 months according to one study. However, although these indicators are helpful, no single test reliably distinguishes a constitutional delay in growth and puberty from hypogonadotropic hypogonadism. All indices in constitutional delay before the onset of puberty are identical to endocrine measurements in a child of the same delayed skeletal age, which are, in turn, similar to those found in hypogonadotropic hypogonadism.

Growth rate before the physical changes of the onset of puberty in these boys is often suboptimal for chronological age but growth velocity usually increases to normal levels after puberty begins. Growth hormone (GH) release to GH secretagogues may be decreased in children with constitutional delay before puberty begins although the amplitude of GH secretion and the GH response to GH-releasing hormone increase after the administration of exogenous aromatizable androgens or estrogens or after the onset of puberty. Thus, affected boys have functional, temporary GH insufficiency for chronological age but not for bone age; this does not constitute a rationale for

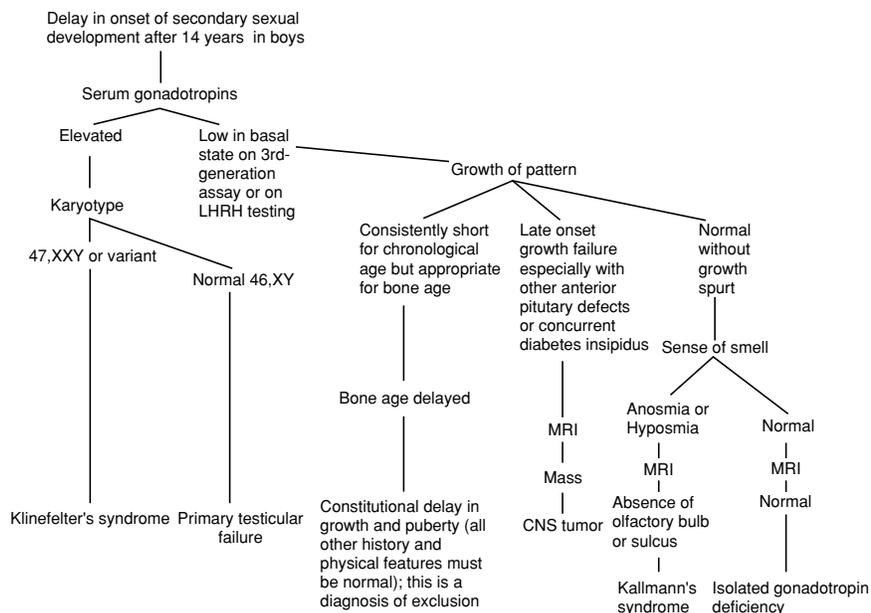


Figure 1 The diagnosis of delayed puberty. Modified from Grumbach, M. M., and Styne, D. M. (2003).

treatment of constitutional delay with growth hormone, as growth hormone therapy does not improve adult height in customary constitutional delay although it may increase growth rate.

The discovery of the critical role of estradiol in skeletal maturation in boys as well as girls led to the suggestion that in boys with constitutional delayed growth and puberty, treatment with an aromatase inhibitor would improve adult height by inhibiting skeletal maturation. In a double-blind, randomized, placebo-controlled study, this proved to be true, suggesting future possibilities for such a plan.

Normal volumetric (but not areal) bone density is described in young men previously affected by constitutional delay in contrast to the decreased peak bone mass due to hypergonadotropic or hypogonadotropic hypogonadism.

## HYPOGONADOTROPIC HYPOGONADISM

Gonadotropin deficiency is due to insufficient pulsatile secretion of LHRH and the resulting follicle-stimulating hormone (FSH) and LH deficiencies lead to delayed sexual maturation. The magnitude of the LHRH deficiency may be quantitative—either absolute or relative—or qualitative; it may involve abnormalities in the amplitude or frequency of LHRH pulses or in both components and hence the phenotype

can vary from severe sexual infantilism to instances in which the separation from constitutional delay of puberty is difficult. LHRH deficiency may be congenital or acquired. With associated GH deficiency, decreased growth velocity, especially during the expected pubertal growth spurt, and short stature result. Patients with isolated FSH and LH deficiencies are usually of normal height for age by the middle adolescent years, in contrast to boys with constitutional delay in puberty who are usually short for chronological age throughout their growing period.

## Central Nervous System Disorders

Central nervous system (CNS) disorders, including tumors that cause delayed puberty, are usually extra-sellar masses that interfere with LHRH synthesis, secretion, or stimulation of pituitary gonadotropes. Most of these classes of patients with gonadotropin deficiency also have a deficiency of one or more additional pituitary hormones (or an increased concentration of plasma prolactin) and those with GH deficiency have a late onset of growth failure compared with congenital hypopituitarism where there is growth failure early in life. The presence of both anterior and posterior pituitary deficiencies developing after infancy suggests an expanding lesion, whereas this combination manifesting in infancy suggests a midline developmental defect.

**Craniopharyngioma**

Craniopharyngioma is the most common brain neoplasm associated with hypothalamic–pituitary dysfunction and sexual infantilism; symptoms usually arise before the age of 20 years and the peak incidence occurs between the ages of 6 and 14 years. This tumor of Rathke's pouch originates from epithelial rests along the pituitary stalk that extend superiorly to the hypothalamus; craniopharyngiomas stay within the sella turcica or, more rarely, may be found in the nasopharynx and CNS. Signs develop as the tumor encroaches on surrounding structures. Symptoms include headache, visual disturbances, short stature, polyuria and polydipsia of diabetes insipidus, and weakness of one or more limbs. Signs include visual defects (including bilateral temporal field deficits), optic atrophy or papilledema, and the poor growth of GH deficiency, delayed puberty, and hypothyroidism. Most subjects are already below the mean in height and height velocity at the time of diagnosis. Laboratory findings include deficiencies in one or more anterior and posterior pituitary hormones. The plasma concentration of prolactin may be normal or increased. Bone age is usually delayed. Approximately 70% of patients with craniopharyngioma have an abnormal sella as well as suprasellar or intrasellar calcification and less than 1% of normal individuals have such calcification. Computed tomographic (CT) scans [but not magnetic resonance imaging (MRI) scans] reveal fine calcifications that are not apparent on routine roentgenograms, and CT or MRI scans with contrast (the diagnostic procedure of choice) can determine whether the tumor is cystic or solid and indicate the presence of hydrocephalus.

Smaller, intrasellar craniopharyngiomas can be resected or decompressed by transsphenoidal microsurgery, but larger or suprasellar masses usually require craniotomy. The recurrence rate is high for complete surgical removal. The combination of limited tumor removal and radiation therapy leads to at least as satisfactory a neurological prognosis and better cognitive outcome as attempts at complete surgical extirpation and to a better endocrinological outcome. Virtually all subjects with radical removal of craniopharyngioma require replacement with gonadal steroids and growth hormone. Postoperative hyperphagia and obesity may develop and correlate with the magnitude of the hypothalamic damage, probably due to injury to the hypothalamic ventromedial nuclei (associated with increased parasympathetic activity and hyperinsulinemia) and/or the paraventricular nuclei.

**Rathke-Cleft Cyst**

A Rathke-cleft cyst can produce symptoms and signs indistinguishable from those of a craniopharyngioma but the usual treatment for this condition is surgical drainage and excision of the cyst wall.

**Other Extrasellar Tumors**

Other extrasellar tumors may cause other extrasellar tumors that arise in or encroach on the hypothalamus.

**Germinomas**

Germinomas (previously termed pinealomas, ectopic pinealomas, atypical teratomas, or dysgerminomas) or other germ cell tumors of the CNS are the extrasellar tumors that most commonly cause sexual infantilism, although these are rare considering all primary CNS tumors. They usually present during the second decade and polydipsia and polyuria are among the most common symptoms, followed by visual difficulties and abnormalities in growth and puberty. The most common endocrine deficiencies involve vasopressin and GH, but other anterior pituitary hormone deficiencies (including gonadotropin deficiency) and elevated serum prolactin levels are frequent. These germ cell tumors secrete human chorionic gonadotropin (hCG), as well as  $\alpha$ -fetoprotein, into spinal fluid and serum. Germ cell tumors in boys may delay puberty or may cause isosexual precocity by secretion of hCG, which stimulates the Leydig cells of the testes.

A germ cell tumor may arise in the suprasellar hypothalamic region, in the pineal region, or in another area of the CNS. Subependymal spread along the lining of the third ventricle is common and seeding may lead to involvement of the lower spinal cord and *corda equina*. MRI scans with contrast enhancement are useful in the diagnosis of tumors more than 0.5 cm in diameter and for the detection of isolated enlargement of the pituitary stalk, an early finding on MRI scans, which must periodically be monitored by MRI for further development of a tumor. Although the pituitary gland increases 100% in size between year 1 and 15, the pineal gland does not change after age 1; any enlargement after 1 year should indicate suspicion of a mass lesion. Pure germ cell tumors (germinomas) are radiosensitive and radiation is the preferred treatment; the clinical features and the response to radiation therapy are so characteristic that surgery is usually indicated only for biopsy to establish a tissue diagnosis. However, a mixed germ cell tumor may require both radiation therapy and chemotherapy.

*Hypothalamic and Optic Gliomas or Astrocytomas*

Hypothalamic and optic gliomas or astrocytomas may cause sexual infantilism as a component of neurofibromatosis (von Recklinghausen's disease) or independently.

**Pituitary Tumors**

Only 2–6% of all pituitary adenomas occur in childhood and adolescence. Most pituitary adenomas occurring before adulthood are prolactinomas, with GH-secreting adenomas and chromophobe adenomas being the next most common. Hyperprolactinemia due to micro- or macroprolactinomas of the pituitary is uncommon in childhood and adolescence and is a rare cause of delayed puberty in both boys and girls, although removal or diminution of the tumor allows resumption of pubertal development. Galactorrhea may be absent by history but it is often demonstrable by manual manipulation of the nipples (serum prolactin rises after manipulation of the nipples, so samples should be obtained before examination or many hours later). Transsphenoidal resection of microprolactinomas in children and adolescents is an effective treatment. The dopamine agonist bromergocryptine can decrease serum prolactin concentrations and decrease the size of the tumors; this may be used when resection of the adenoma is incomplete or to reduce the size of the tumor in large macroprolactinomas before attempted surgical removal.

**Other CNS Disorders Leading to Delayed Puberty***Langerhans' Cell Histiocytosis*

Langerhans' cell histiocytosis (Hand-Schüller-Christian disease, or histiocytosis X) is a clonal proliferative disorder of Langerhans' histiocytes or their precursors and is characterized by the infiltration of lipid-laden histiocytic cells or foam cells in the skin, viscera, and bone. Diabetes insipidus usually results from infiltration of the hypothalamus and/or the pituitary stalk and is the most common endocrine manifestation, although GH deficiency and delayed puberty may occur. There may be visceral involvement including the lung, liver, and spleen and cyst-like areas in flat bones of the skull, the ribs, the pelvis, and the scapula, in the long bones of the arms and legs, and in the dorsolumbar spine. Lesions of the mandible lead to the radiographic impression of "floating teeth" within rarefied bone and the clinical finding of absent or loose teeth. Infiltration of the orbit may lead to exophthalmos, and mastoid or temporal bone involvement may lead to chronic otitis media. Treatment with glucocorticoids, anti-neoplas-

tic agents, and radiation is promising in terms of survival but more than 50% of patients have late sequelae or progression and the natural waxing and waning course of this disease makes evaluation of therapy difficult.

*Postinfectious Inflammatory Lesions of the CNS, Vascular Abnormalities, and Head Trauma*

Postinfectious inflammatory lesions of the CNS, vascular abnormalities, and head trauma rarely cause hypogonadotropic hypogonadism. Tuberculous or sarcoid granulomas of the CNS are the rare causes. When hydrocephalus or a subarachnoid cyst causes delayed puberty, decompression will often reverse this result.

*Radiation of the Head*

Treatment of CNS tumors, leukemia, or neoplasms of the head and the face may result in the gradual onset of hypothalamic–pituitary failure. GH deficiency is the most common but gonadotropin deficiency also occurs. Decreased growth due to GH deficiency with precocious puberty can lead to a decrease in the final height of children with acute lymphocytic leukemia treated with CNS radiation.

*Developmental Defects*

Developmental defects, mainly midline malformations of the head and the CNS, cause a variety of endocrine deficiencies. Septo-optic dysplasia or optic dysplasia is due to abnormal development of the prosencephalon with the optic nerve usually affected, leading to small, dysplastic, pale optic discs, pendular (evenly moving side to side) nystagmus, and possibly blindness. There may be GH deficiency and diabetes insipidus as well as adrenocorticotrophic hormone (ACTH), thyroid-stimulating hormone (TSH), and gonadotropin deficiencies. Short stature and delayed puberty may result, although true precocious puberty is an alternative outcome. The septum pellucidum is often absent in association with optic hypoplasia or dysplasia, leading to septo-optic dysplasia. The pituitary gland may be hypoplastic and some patients may have an ectopic location of the neurohypophysis on MRI. A mutation in the HESX1 is a rare cause of septo-optic dysplasia.

Other developmental defects of the anterior pituitary gland are caused by mutations in encoding transcription factors involved in the early aspects of pituitary development. Mutations in the LHX3 and PROP-1 homeobox genes may cause autosomal-recessive hypogonadotropic hypogonadism and other pituitary hormone deficiencies. PROP-1 mutations cause GH and TSH deficiencies but may be associated

with delayed puberty or the late onset of secondary hypogonadism in adulthood and may also cause ACTH deficiency. Other congenital midline defects ranging from complete dysraphism and holoprosencephaly to cleft palate or lip are also associated with hypothalamic–pituitary dysfunction. Homozygous mutations in the *LHX3* gene are associated with multiple pituitary hormone deficiencies including LH and FSH and severe restriction of head rotation.

Individuals with myelomeningocele (myelodysplasia) have an increased frequency of endocrine abnormalities, including hypothalamic hypothyroidism, hyperprolactinemia, and elevated gonadotropin concentrations, whereas some patients demonstrate true precocious puberty.

### Isolated Gonadotropin Deficiency

Isolated gonadotropin deficiency is due to a defect in the hypothalamus involving the LHRH pulse generator and/or the pituitary gonadotropes without an anatomic lesion. Puberty is delayed or pubertal maturation is incomplete. The testes, if palpable, are small and the concentration of serum inhibin B, an estimate of seminiferous tubule function, is low. Micropenis or undescended testes or both signs are evidence of a fetal testosterone deficiency.

Isolated gonadotropin deficiency may be sporadic or occur in families in an autosomal-dominant, autosomal-recessive, or X-linked recessive pattern. Whereas patients with CNS tumors usually have associated GH deficiency and growth failure, and those with constitutional delay in growth and adolescence are short for chronological age, patients with isolated gonadotropin deficiency are usually of appropriate height for their age. Due to delayed bone age and epiphyseal fusion, these patients develop increased arm span for height and decreased upper/lower ratios (eunuchoid body proportions) and usually become tall adults.

### **Kallmann's Syndrome**

Kallmann's syndrome is the most common form of isolated hypogonadotropic hypogonadism with delayed puberty, in which anosmia or hyposmia resulting from agenesis or hypoplasia of the olfactory lobes and/or sulci is associated with LHRH deficiency. Affected individuals may not notice impaired olfaction; testing with graded dilutions of pure scents is useful to discriminate the magnitude of the deficit. Undescended testes and gynecomastia are common

in this and all types of hypogonadotropic hypogonadism in boys. Approximately one-half of males with Kallmann's syndrome are born with a micropenis. Associated defects inconstantly present are cleft lip and/or palate, imperfect facial fusion, seizure disorders, short metacarpals, pes cavus, neurosensory hearing loss, cerebellar ataxia and nystagmus, ocular motor abnormalities, and, limited to the X-linked form, unilateral, or rarely bilateral, renal aplasia or dysplasia and mirror movements of the upper extremities (synkinesia). All of these structures and organs are sites of expression of the *KAL* gene in the human fetus.

There may be bilateral agenesis of the olfactory bulbs on coronal and axial cranial MRI scans and absent or abnormal olfactory sulci bilaterally. Over 90% of those with Kallmann's syndrome have an abnormal cranial MRI. Kallmann's syndrome can be transmitted as an X-linked, autosomal-dominant, or autosomal-recessive trait. The majority of cases of Kallmann's syndrome do not have mutations in the *KAL* gene on the X chromosome, but these cases are more likely to have a complete absence of gonadotropin secretion pulses and have an absence of migration of GnRH neurons to the hypothalamus.

The Xp22.3 locus is the site of the *KAL-1* gene, an X-linked gene that escapes X inactivation and encodes a glycoprotein, termed anosmin-1, with characteristics of an extracellular neural adhesion molecule, which could function as a pathfinder in the guidance of LHRH neurons from the olfactory placode of the fetus to the medial basal hypothalamus. A variety of deletions and mutations of the *KAL* gene have been described including large and small (exon) deletions, point mutations, and a variety of nonsense mutations leading to frameshift and premature stop codons. Contiguous gene deletions in this region of the X chromosome can lead to an association of Kallmann's syndrome with X-linked ichthyosis caused by steroid sulfatase deficiency, mental retardation, and chondroplasia punctata. The fetal LHRH-containing cells and neurites are arrested in their migration to the brain and end in a tangle around the cribiform plate and in the dural layers adjacent to the meninges beneath the forebrain.

Hypogonadotropic hypogonadism may be transmitted by autosomal-recessive inheritance with none of the other features of Kallmann's syndrome. Males with cerebellar ataxia and deficient gonadotropin production have been reported in kindreds with X-linked inheritance (possibly a variant form of Kallmann's syndrome) and hypogonadotropic hypogonadism may be associated with the multiple lentigenes and basal cell nevus syndromes.

### **LHRH Receptor Mutations**

LHRH receptor mutations are reported in familial and sporadic patients and lead to various degrees of hypogonadotropic hypogonadism with normosmia. In the autosomal-recessive form, there are heterozygous or homozygous mutations in the LHRH receptor. The patients may present with severe features of isolated hypogonadotropic hypogonadism, sexual infantilism, ranging to delayed puberty, or relatively mild hypogonadism and infertility. The clinical variants include the fertile eunuch variant. In all types of congenital gonadotropin deficiency, male patients are likely to manifest micropenis due to a lack of fetal gonadotropin stimulation of fetal testes during the last half of gestation. Micropenis may also occur in boys with congenital growth hormone deficiency.

No mutations of the LHRH genes have been reported.

### **X-linked Congenital Adrenal Hypoplasia and Hypogonadotropic Hypogonadism**

X-linked congenital adrenal hypoplasia and hypogonadotropic hypogonadism are due to a deletion or mutation in the DAX-1 gene (dosage-sensitive sex reversal-A adrenal hypoplasia congenita gene on the X chromosome gene 1). The DAX-1 gene, a member of the nuclear receptor superfamily, encodes an orphan receptor that is a putative transcriptional repressor. The DAX-1 locus undergoes X inactivation. It maps to the DSS (dosage-sensitive sex reversal) locus (Xp21); a double dose of DAX-1 is associated with a female phenotype or ambiguous genitalia in 46,XY males. DAX1 has a steroidogenic factor 1 (SF1) response element in the 5' promoter region. SF1 is another orphan member of the nuclear hormone receptor superfamily; both DAX-1 and SF1 are expressed in the adrenal glands, gonads, pituitary, and hypothalamus, which raises the possibility of an important interaction between these two genes and their products. Abnormalities of the DAX-1 gene are characterized by severe glucocorticoid deficiency, mineralocorticoid deficiency, and, at puberty, androgen deficiency. A mature adrenal cortex is lacking and the abnormal structure of the adrenal cortex resembles that of the fetal zone made up of disorganized vacuolated cytomegalic cells. In the majority of affected boys, the severe primary adrenal insufficiency with hyponatremia, hyperkalemia, acidosis, and hypoglycemia (failure to thrive, vomiting, poor feeding, dehydration, circulatory collapse, increased pigmentation) is lethal if untreated early in life. The concentration of plasma ACTH is high and plasma

cortisol and aldosterone levels are low. The testes are undescended in less than one-half of the patients; micropenis is rare, but occasionally urogenital abnormalities and hearing loss are present. Most commonly, signs of sexual maturation at the age of puberty, including the absence of pubic and axillary hair and testicular enlargement, are lacking; the concentrations of serum FSH, LH, and testosterone are low.

Contiguous gene syndromes are not uncommon in association with X-linked congenital adrenal hypoplasia, as the gene maps to Xp21, distal to the glycerol kinase gene and the Duchenne muscular dystrophy gene and proximal to the gene associated with mental retardation. Rare patients have defects in all of these areas.

### **Isolated LH Deficiency**

Isolated LH deficiency or the fertile eunuch syndrome is associated with deficient testosterone production (which responds to administration of hCG) in the presence of variable spermatogenesis. In most instances, it is an incomplete form of isolated gonadotropin deficiency; the disorder may be idiopathic or secondary to a hypothalamic pituitary neoplasm.

### **Isolated FSH Deficiency**

Isolated FSH deficiency is rarely due to mutations in the FSH  $\beta$ -subunit, either homozygous or compound heterozygous. Females are mostly affected but two affected men have been described; both had azoospermia, small, soft testes, and an absence of serum FSH; some patients have delayed puberty.

### **Idiopathic Hypopituitary Dwarfism**

Idiopathic hypopituitary dwarfism is usually caused by a deficiency of hypothalamic releasing factors. In the untreated state, patients with deficient LHRH have delayed or absent puberty and, in contrast, patients with isolated GH deficiency ultimately undergo spontaneous, if delayed, pubertal development, without exogenous gonadal steroids, when the bone age reaches the pubertal stage of 11 to 13 years, usually during GH therapy. Common to many patients with idiopathic hypopituitary dwarfism is early onset of growth failure; late onset of diminished growth suggests the presence of a CNS tumor or other serious problems. Breech delivery, especially in males, perinatal distress, idiopathic hypopituitarism, and malformations of the pituitary stalk demonstrable by MRI are common in such patients. Familial forms of multiple pituitary hormone deficiencies with either

autosomal-recessive or X-linked inheritance are less common.

Treatment of prepubertal children with isolated growth hormone deficiency with growth hormone can increase the rate of pubertal development. Alternatively, if GH treatment is instituted in children already in puberty who have a limited height potential, limitation of the amount of growth attained with GH treatment often results; in these instances, the use of LHRH agonists to suppress pubertal development in addition to the use of GH is useful in increasing final height. The judicious use of low-dose testosterone in affected boys of pubertal age with associated gonadotropin deficiency does not seem to impair growth achieved by growth hormone replacement.

## Miscellaneous Conditions

### **Prader-Willi Syndrome**

Prader-Willi syndrome is a syndrome of early onset childhood hyperphagia, pathological obesity and carbohydrate intolerance, infantile central hypotonia and lethargy, delayed onset and poor fetal activity, a tendency for intrauterine growth retardation, short stature by 15 years of age, small hands and feet, mild to moderate mental retardation, emotional instability including perseveration, obsessions, and compulsions, and characteristic facies with almond-shaped eyes, triangular mouth, and narrow bifrontal diameter. Delayed puberty is usually related to hypogonadotropic hypogonadism caused by hypothalamic dysfunction. Affected boys often have micropenis and cryptorchidism. The frequency is approximately 1 in 20,000. It is very rarely familial and is caused by abnormalities involving the long arm of chromosome 15 in the region q11–q13. Most patients with Prader-Willi syndrome have a paternal deletion of 15q11–q13; a minority of cases have maternal uniparental disomy (either isodisomy or heterodisomy), where both chromosomes 15 are derived from the mother, possibly by non-disjunction during maternal meiosis, and represent a striking example of genomic imprinting. Rarely, an imprinting center defect has been detected. One candidate-imprinted gene that maps to this region, SNRPN (small nuclear ribonucleoprotein-associated polypeptide SmN), implicated in splicing pre-mRNA, is expressed in the brain and hypothalamus and is one explanation of the syndrome.

Plasma GH responses to provocative stimuli and to sleep are usually low due to obesity but may be normal. In June 2000, the U.S. Food and Drug

Administration approved rhGH treatment as an indication for children with Prader-Willi syndrome without a requirement for assessing growth hormone secretion; however, genetic testing is required to confirm the clinical diagnosis. GH treatment decreases body fat, fat utilization, lean body mass, linear growth, energy expenditure in the syndrome with a possible improvement in physical strength, and motor development.

### **Laurence-Moon and the Bardet-Biedl Syndromes**

The Laurence-Moon and Bardet-Biedl syndromes are rare autosomal-recessive traits that both combine retinitis pigmentosa and hypogonadism of various etiologies; many Bardet-Biedl patients have developmental delay as do all of the Laurence-Moon patients. Laurence-Moon syndrome, however, demonstrates spastic paraplegia, whereas the Bardet-Biedl syndrome has postaxial polydactyly, onset of obesity usually in early infancy, renal dysplasia, and a relatively high prevalence among the Bedouin of the Middle East. Similar findings are present in Biemond's syndrome II with iris coloboma, hypogenitalism, obesity, polydactyly, and developmental delay, but it too is a distinct entity. The genetically and phenotypically heterogeneous Bardet-Biedl syndrome is linked to six loci that map to chromosomes 2, 3, 11, 15, 16, and 20; in most cases, three mutant genes are required for the phenotype to develop.

## Functional Gonadotropin Deficiencies

Severe systemic and chronic disorders and malnutrition are associated with delayed puberty or failure to progress through puberty. In general, weight loss of any cause to less than 80% of ideal weight for height can lead to gonadotropin deficiency and low serum leptin levels; weight regain usually restores hypothalamic-pituitary gonadal function over a variable period. If adequate nutrition and body weight are maintained in patients with regional enteritis or chronic pulmonary disease, gonadotropin secretion is usually adequate. The weight loss of anorexia nervosa is classically found in girls, but boys may develop this psychiatric disease as well.

Cystic fibrosis also delays puberty, in large part through malnutrition, but even with normal pubertal progression, boys with cystic fibrosis almost universally have oligospermia caused by obstruction of the spermatic ducts unrelated to their nutritional status. The greater prevalence of reproductive difficulties in male patients with cystic fibrosis compared

with female patients may be due to the greater prevalence of the cystic fibrosis transmembrane regulator in male reproductive tissues, such as the epididymis and vas deferens, and as a consequence more viscid luminal contents, which ultimately damage the testes. Alternatively, the epididymides may be deficient and the vas differentia absent.

Sickle-cell disease may cause a delay of the pubertal growth spurt and peak height velocity although adult height may be normal; Leydig cell function is often impaired due to ischemia of the testes, gonadotropin deficiency, or both factors. Thalassemia major leads to abnormal sexual maturation in a majority of boys and many experience growth failure due to hypothyroidism or gonadotropin deficiency. Thalassemia carries the risk of hemochromatosis, which causes multiple endocrine abnormalities, and the desferioxamine therapy itself may cause skeletal dysplasia and compromise pubertal growth; GH secretion is usually not affected.

Chronic gastrointestinal diseases, such as Crohn's or celiac diseases, are often accompanied by delayed puberty; successful therapy to restore nutrition enables puberty to progress. Chronic renal disease has been associated with delayed pubertal development and decreased pulsatile gonadotropin secretion; successful renal transplantation usually restores gonadotropin secretion. Patients with nephrotic syndrome or glomerulonephritis treated with alternate-day glucocorticoid therapy have poor pubertal growth, poor secondary sexual development, and deficient gonadotropin secretion in a pattern resembling constitutional delay in puberty.

Advances in the treatment of leukemia have improved the prognosis; children with early onset and long-term remission experience puberty at an appropriate age or with only slight delay, whereas patients with initial symptoms of leukemia in late childhood may have considerable delay of pubertal development. Further CNS radiation may cause hypogonadotropic hypogonadism and/or growth hormone deficiency, and radiation to the abdomen or pelvis and certain types of chemotherapy, especially if administered during puberty, may impair gonadal function and cause primary hypogonadism. The adult height in boys may be reduced less with a dose of 1800 cGy rather than 2400 cGy.

If hypothyroidism delays the onset of puberty, treatment with levothyroxine will allow progression. Poorly controlled diabetes mellitus can lead to poor growth, fatty infiltration of the liver, and sexual infantilism (Mauriac's syndrome), probably related to poor nutritional status. Prepubertal children with

diabetes mellitus are most vulnerable to poor glyce-mic control, whereas pubertal subjects exhibit normal growth unless severe hyperglycemia occurs; adolescents with even moderately poor control frequently have some degree of growth impairment and delayed puberty. Cushing's disease can cause delayed onset or arrest of gonadarche, which is corrected by removal of the ACTH-secreting pituitary adenoma.

Although men are less often affected than women, males may also be affected by rigorous physical training, with decreased LH response to LHRH and decreased spontaneous LH pulse frequency and amplitude resulting; the serum testosterone is normal or low.

### HYPERGONADOTROPIC HYPOGONADISM: SEXUAL INFANTILISM CAUSED BY PRIMARY GONADAL DISORDERS

Primary gonadal failure and the impaired secretion of gonadal steroids lead to decreased negative feedback and elevated LH and FSH levels. The most common forms in boys are associated with Klinefelter's syndrome and its variants.

#### Klinefelter's Syndrome

In Klinefelter's syndrome (syndrome of seminiferous tubular dysgenesis), elevated gonadotropin levels are found postpubertally; before the age of 12 years, gonadotropin concentrations are in the prepubertal range. Rarely, low gonadotropin concentrations occur when hypogonadotropic hypogonadism is associated with 47,XXY Klinefelter's syndrome. Prepubertally, patients can be detected by the disproportionate length of the extremities and decreased upper/lower body ratio without an increase in arm span. There is variation in Leydig cell function, but the plasma concentration of testosterone tends to be in the normal range until approximately the age of 14 years, after which it may fail to rise to normal adult levels. Thus, the onset of puberty may not be delayed, but impaired Leydig cell reserve and low testosterone levels may cause slow progression or arrest of pubertal changes. Serum estradiol/testosterone ratios and TeBG levels are higher than those in normal males, which indicates an increased estrogen effect and a decreased testosterone effect. These factors probably account, at least in part, for the gynecomastia characteristic of Klinefelter's syndrome during adolescence.

Neurobehavioral abnormalities, primarily in language and frontal executive functions, are frequent,

and some say universal, in Klinefelter's syndrome. These problems may be severe enough to lead to evaluation in childhood and the prepubertal diagnosis. The global IQ in unselected populations of Klinefelter's syndrome is normal or near normal but verbal IQ is usually lower than performance IQ. Patients with clinical or psychological problems are referred more often for evaluation, so some studies suggest more significant deficits than found in an unselected population of XXY individuals.

Approximately 20% of mediastinal germ cell tumors are associated with Klinefelter's syndrome and they occur at a younger age than the mediastinal germ cell tumors that are not associated with Klinefelter's syndrome (average ages of 16 years versus 27 years in one study). These germ cell tumors, which may be located in the midline anywhere from the CNS to the pelvis, secrete hCG and can induce sexual precocity in young boys. Klinefelter's syndrome must be considered in boys with hCG-secreting germ cell tumors, especially if the tumor is located in the mediastinum or CNS.

### Other Forms of Primary Testicular Failure

Much cancer therapy affects testicular function, and as more children survive with effective therapy for cancer, adult infertility may result. Chemotherapeutic agents used in the treatment of nephrotic syndrome, leukemia, or Hodgkin's disease, such as chlorambucil, vinblastin, Mustargen, Oncovin, and procarbazine, have led to Sertoli cell, Leydig cell, and germ cell damage (sometimes reversible) in prepubertal patients. Even if spontaneous progression through puberty occurs, serum FSH and LH concentrations may be elevated and the inhibin B concentrations decreased during puberty, indicative of gonadal damage. COPP/MOPP therapy for Hodgkin's disease can cause severe damage to germinal cells apparently without much effect on Leydig cells even if therapy occurred in the prepubertal period. In addition, adriamycin, bleomycin, vinblastine, and dacarbazine also can cause germ cell depletion. Gonadal damage can occur earlier as a result of therapy in the prepubertal period but may not be demonstrable until the age of puberty.

Radiation to the gonads can cause primary testicular failure, usually resulting in azoospermia, although normal testosterone secretion may be associated with elevated LH and FSH values (compensated Leydig cell failure); the gonads must be shielded from the treatment, if possible. Although doses of 0.35 Gy to the testes may lead to temporary aspermia, doses over

2 Gy lead to permanent aspermia and over 15 Gy may cause Leydig cell dysfunction.

LH resistance caused by Leydig cell LH receptor abnormalities are reported in rare boys with a male phenotype and lack of pubertal development along with gynecomastia, elevated plasma LH levels, and early pubertal plasma testosterone concentrations that did not increase after hCG administration. Several different mutations may be at fault.

### DIAGNOSIS OF DELAYED PUBERTY AND SEXUAL INFANTILISM

When prepubertal boys present at age 14, the physician must make a clinical judgment as to which are variants of the norm and which require extensive evaluation and treatment. Lack of progression through the stages of puberty, even if the age at onset is normal, may also require evaluation; a boy who has not completed secondary sexual maturation within 4.5 years after onset of puberty may have a hypothalamic, pituitary, or gonadal disorder. The diagnosis of hypergonadotropic hypogonadism is readily established by elevation of random plasma LH and FSH concentrations. However, the differential diagnosis of hypogonadotropic hypogonadism versus constitutional delay in growth and adolescence is more difficult because of the overlap in physical and laboratory findings in the two conditions and no single test consistently makes this distinction.

The medical history must elicit all symptoms of chronic or intermittent illnesses and all details pertaining to growth and development as well as the patient's sense of smell and nutritional status. Disorders of pregnancy and abnormalities of labor and delivery, if present, suggest that a congenital or neonatal event may be related to the delay in puberty. Poor linear growth and poor nutritional status during the neonatal period and childhood may reflect longstanding abnormalities of development. A growth chart is plotted to represent graphically the increase in stature and to assess growth velocity from birth. Late onset of growth failure usually indicates a serious condition. Family history may reveal disorders of puberty or infertility, anosmia, or hyposmia in relatives as well as delay in the age at onset of puberty in parents or siblings. Recalled age of pubertal onset is relatively reliable in women but less often accurate in men. A history of consanguinity is important in the detection of autosomal-recessive disorders.

The physical examination includes determination of height and weight; the upper/lower segment ratio

or sitting height is calculated and the arm span is measured and compared with the height to determine whether there is a lack of spine growth from, e.g., radiation, or whether eunuchoid proportions are present. The height velocity should be documented over a period of at least 3 months, preferably 6 months. The signs of puberty are noted and the stage of secondary sexual development is determined according to the "Tanner" standards. The length and width of the testes are recorded or the volume is assessed using an orchidometer. The length and diameter of the stretched penis are determined. Obese boys often appear to have a small penis because of excessive adipose tissue surrounding the phallus; only when the fat is retracted can the full extent of phallic development be assessed. (This feature is among the most common causes of inappropriate referral for hypogonadism.) The extent of axillary hair is noted, as is the degree of acne. The possibility of cryptorchidism, anorchia, or retractile testes should be differentiated if no testes are palpated in the scrotum. Neurological examination, including examination of the optic discs and visual fields by frontal confrontation perimetry and evaluation of olfaction, may reveal findings suggesting the presence of a CNS neoplasm or a developmental defect (Kallmann's syndrome). The small testes and gynecomastia of Klinefelter's syndrome may suggest this diagnosis. Complete physical examination is important in the search for a chronic disorder that may delay puberty.

Laboratory studies include determination of serum LH and FSH concentrations by "third-generation" assays, measurement of the rise in LH levels after LHRH administration, determination of testosterone concentrations, and measurements of thyroxine and prolactin concentrations if the clinical features warrant. It is important to use one of the few national endocrine laboratories for the determinations of the hormones of puberty as most local laboratories are interested only in differentiating normal, higher, adult values from inappropriately low levels and not the low concentrations characteristic of the early stage of pubertal development. Radiographic examination includes bone age determination and, if the diagnosis is at all consistent with a CNS lesion, an MRI of the brain with specific attention to the pituitary and hypothalamic area using contrast; only advanced pituitary tumors or significantly calcified craniopharyngiomas will appear on lateral skull films and, though a positive result is useful, a negative radiograph cannot rule out a CNS defect. CT scanning can detect calcification, in contrast to MRI scans. Assessment of chromosomal karyotype should be considered in boys with

suspected Klinefelter's stigmata or behavior. A presumptive diagnosis of constitutional delay in growth and adolescence is made if the history is normal and the growth chart reveals a history of short stature but consistent growth rate for skeletal age (and no signs or symptoms of hypothalamic lesions), if the family history includes parents or siblings with delayed puberty, if the physical examination (including assessment of the olfactory threshold) is normal, if optic discs and visual fields are normal, or if the bone age is significantly delayed. In classic cases, an MRI of the hypothalamic-pituitary region may not be necessary but constitutional delay is a diagnosis of exclusion. The rate of growth in these patients is usually appropriate for bone age. A decrease in growth velocity occurs in some normal children just before the appearance of secondary sexual characteristics and may awaken concerns if such a pattern occurs in these subjects; if the decrease is significant, another diagnosis may be warranted.

Measurement of 8 AM serum testosterone is proposed to be an accurate indication of impending pubertal development; a value of greater than 0.7 nmol/liter (20 ng/dl) predicts enlargement of the testes to greater than 4 ml by 12 months in 77% of cases and in 15 months in 100% of cases, whereas in those with a value less than 0.7 nmol/liter only 12% entered puberty in 12 months and only 25% entered puberty in 15 months. Failing progression into puberty, there does not appear to be a practical and reliable endocrine test for indisputably differentiating between constitutional delay in growth and adolescence and hypogonadotropic hypogonadism. Watchful waiting remains the procedure of choice.

Patients with deficiency of gonadotropins combined with deficiency of other pituitary hormones require careful evaluation for a CNS neoplasm. Visual field or optic disc abnormalities support the diagnosis of a CNS tumor; even if these tests are normal, a cranial MRI should be performed to evaluate the pituitary gland and stalk and the hypothalamic region. CT scans, but especially MRI of the head, are valuable in detecting mass lesions and developmental abnormalities of the hypothalamic-pituitary region. Late onset of hypothalamic-pituitary defect is an ominous sign requiring evaluation.

## TREATMENT OF DELAYED PUBERTY AND SEXUAL INFANTILISM

Patients with constitutional delay in growth and adolescence ultimately have spontaneous onset and

progression through puberty so that reassurance and continued observation are sufficient. However, psychological stress may occur due to the immature appearance and school work may suffer due to the child's poor self-image. Thus, for psychological reasons, prepubertal boys of age 14 years or older may receive a 3- to 6-month course of testosterone enanthate; cyprionate or cyclopropionate may be helpful or transdermal testosterone may be applied according to the doses and provisos given below. If spontaneous puberty does not ensue or the concentrations of plasma gonadotropins and plasma testosterone in boys do not increase, the treatment may be repeated once. When treatment is discontinued after bone age has advanced to 13 or 14 years in boys, patients with constitutional delay usually continue pubertal development on their own, whereas those with gonadotropin deficiency do not progress and may, in fact, regress.

A fourth-generation aromatase inhibitor, letrozole, administered along with testosterone in a randomized controlled trial in boys with constitutional delay in puberty and growth decreased the advancement in bone age, an effect that will presumably lead to a greater adult height by strikingly decreasing the synthesis of estradiol from testosterone. This treatment is experimental; it may improve the decreased adult height in some boys with constitutional delay compared to their predicted genetic potential, a decrease that testosterone cannot overcome.

Functional hypogonadotropic hypogonadism associated with chronic disease is treated by alleviating the underlying problem or improving nutrition.

Congenital or acquired gonadotropin deficiency as a result of a lesion or surgery requires replacement therapy with gonadal steroids at an age approximating the normal age of onset of puberty. If untreated GH deficiency and shorts stature coexist with gonadotropin deficiency, testosterone treatment may be delayed; if bone age advancement and epiphyseal fusion are brought about by testosterone replacement before therapy with GH causes adequate linear growth, adult height will be compromised. Alternatively, if puberty is not initiated early enough, the patient may suffer psychological damage. Generally, puberty in such patients can be initiated with low-dose gonadal steroids by the age of 14 years in boys, regardless of the definitive diagnosis of gonadotropin deficiency. Isolated growth hormone-deficient patients may have a delayed onset of puberty; with growth hormone administration, puberty usually occurs at an appropriate age but may progress faster than in normal individuals. Clinical trials are in progress to

determine the effects of artificially delaying puberty with an LHRH analogue to attempt to achieve a greater final height in patients with isolated GH deficiency treated with hGH.

Klinefelter's syndrome is compatible with varying degrees of masculinization at puberty, but some patients require testosterone replacement. The concentration of plasma testosterone and LH should be monitored every 6 months during puberty and yearly thereafter. If the LH level rises more than 2.5 SD above the mean value or the testosterone level decreases below the normal range for age, testosterone replacement therapy is indicated. There is a growing feeling among parents that testosterone treatment in the early pubertal period improves language skills, reading ability, and behavior in boys with Klinefelter's syndrome, and well-controlled studies supporting this contention are available.

Hypogonadotropic hypogonadism or hypergonadotropic hypogonadism are both treated with gonadal steroid replacement. Alkylated testosterone preparations are not used because of the risk of peliosis hepatitis (hemorrhagic liver cysts); progression to liver failure can occur. Boys may receive testosterone enanthate, propionate, or cyprionate, 50–100 mg every 4 weeks intramuscularly at the start; later the dosage is gradually increased to 200 to 300 mg every 2 to 3 weeks. Testosterone may also be administered by a cutaneous patch on scrotal skin or on nonsexual skin or by new testosterone gel preparations, usually rubbed onto the forearms; both methods have been approved for adults but have not yet been approved for boys under the age of 16 years.

### See Also the Following Articles

Agonadism, Male and Female • Bardet-Biedl Syndrome • Constitutional Delay of Growth and Puberty (CDGP) • Delayed Puberty and Hypogonadism, Female • Genes and Gene Defects Affecting Gonadal Development and Sex Determination • Germ Cell Differentiation Signaling Events, Male • Hypergonadotropic Hypogonadism • Kallmann's Syndrome and Idiopathic Hypogonadotropic Hypogonadism • Klinefelter's Syndrome • Prader-Willi Syndrome • Testes, Embryology of • Undescended Testes

### Further Reading

Beuschlein, F., Keegan, C. E., Bavers, D. L., Mutch, C., Hutz, J. E., Shah, S., *et al.* (2002). SF-1, DAX-1, and acd: Molecular determinants of adrenocortical growth and steroidogenesis. *Endocr. Res.* **28**, 597–607.

- Boone, K. B., Swerdloff, R. S., Miller, B. L., Geschwind, D. H., Razani, J., Lee, A., *et al.* (2001). Neuropsychological profiles of adults with Klinefelter syndrome. *J. Int. Neuropsychol. Soc.* **7**, 446–456.
- Grumbach, M. M., Hughes, I. A., and Conte, F. A. (2003). Disorders of sex differentiation. In “Williams’ Textbook of Endocrinology” (P. R. K. H. M. Larsen, S. Melmed, and K. S. Polonsky, eds.), pp. 842–1002. W. B. Saunders, Philadelphia, PA.
- Grumbach, M. M., and Styne, D. M. (2003). Puberty: Ontogeny, neuroendocrinology, physiology, and disorders. In “Williams’ Textbook of Endocrinology” (P. R. K. H. M. Larsen, S. Melmed, and K. S. Polonsky, eds.), pp. 1115–1286. W. B. Saunders, Philadelphia, PA.
- Harlan, W. R., Grillo, G. P., Cornoni-Huntley, J., and Leaverton, P. E. (1979). Secondary sex characteristics of boys 12 to 17 years of age: The U. S. Health Examination Survey. *J. Pediatr.* **95**, 293–297.
- Layman, L. C. (2002). Human gene mutations causing infertility. *J. Med. Genet.* **39**, 153–161.
- Roche, A. F., Wellens, R., Attie, K. M., and Siervogel, R. M. (1995). The timing of sexual maturation in a group of U. S. white youths. *J. Pediatr. Endocrinol.* **8**, 11–18.
- Sedlmeyer, I. L., and Palmert, M. R. (2002). Delayed puberty: Analysis of a large case series from an academic center. *J. Clin. Endocrinol. Metab.* **87**, 1613–1620.
- Seminara, S. B., Oliveira, L. M., Beranova, M., Hayes, F. J., and Crowley, W. F., Jr. (2000). Genetics of hypogonadotropic hypogonadism. *J. Endocrinol. Invest.* **23**, 560–565.
- Styne, D. M. (2002). The testes. In “Pediatric Endocrinology” (M. A. Sperling, ed.), pp. 562–629. W. B. Saunders, Philadelphia, PA.
- Wu, F. C., Brown, D. C., Butler, G. E., Stirling, H. F., and Kelnar, C. J. (1993). Early morning plasma testosterone is an accurate predictor of imminent pubertal development in prepubertal boys. *J. Clin. Endocrinol. Metab.* **76**, 26–31.

## Secondary Sexual Characteristics

The timing and tempo of pubertal development are linked to alterations in the HPG and GH/IGF-I axes. There may be great variability in the onset of pubertal development, but once entrained, there is less variability in progress through the pubertal stages. Still, there is significant variability in the progression of puberty (tempo) that may be very relevant to the boy with delayed adolescence. The external signs of pubertal development have been summarized into relatively easily discernible stages. In general, scrotal thinning and reddening occur concomitantly with an increase in testicular volume to 4 ml or greater. As testis size increases and the levels of the male hormone testosterone increase, there is lengthening of the phallus followed by an increase in its circumference as it further lengthens. Pubic hair increases in area from just above the phallus to growth in all directions, including the medial thigh. The quality of the hair becomes more coarse and curly. With delayed pubertal development, the chronological age differs from the “biological” age. The various stages of growth noted previously—take-off and PHV—and the endocrinological changes noted in what follows are more tightly correlated to the biological age than to the chronological age.

## THE ENDOCRINE SYSTEM AND PUBERTAL DEVELOPMENT

### Gonadotropins

After peaks of activity during fetal life and within the first few months of extrauterine life, the HPG axis becomes quiescent, only to reawaken at puberty. Gonadotropin release is pulsatile at all ages. Puberty is anticipated by an increase in the amplitude of luteinizing hormone (LH) and follicle-stimulating hormone (FSH) secretion that may be detected several years before the external signs of pubertal development are present. Initially, biologically relevant surges (“pulses”) of LH occur only after the onset of deep sleep. These permit the testis to increase the production of the potent androgen testosterone. However, the hypothalamic-pituitary axis for LH remains exquisitely sensitive to the negative feedback actions of testosterone or one of its metabolites, estradiol, such that the minor increases in testosterone are able to dampen LH release. As puberty unfolds, the small increments in testosterone levels are no longer able to maintain suppressed levels of LH. The increases in LH continue for longer and longer periods

during the day, pushing testosterone levels even higher. Concomitantly, the feedback sensitivity of the HPG axis diminishes and the ambient levels of testosterone are no longer sufficient to suppress the surges of LH. At the end of the pubertal process, the pulsatile release of LH occurs every 1 or 2 h during the day and night.

### Gonadal Steroid Hormones

Nighttime increases in circulating levels of testosterone are often detectable even before the external signs of puberty develop. The daytime levels rise later as the testis volume increases. The circulating levels of testosterone are substrate for at least two important enzymes, 5  $\alpha$ -reductase (converts testosterone to dihydrotestosterone, DHT) and aromatase (converts testosterone to estradiol). The effects on the lean body mass are in part due directly to testosterone and indirectly to estradiol because the latter leads to the marked increases in GH and IGF-I due to an action of estradiol on the hypothalamus and pituitary. Circulating estradiol levels cause maturation and eventually closure of the epiphyses of the long bones, signaling the virtual cessation of linear growth. Diminished bone mass is present in boys with delayed pubertal development, including severe constitutional delay of growth and puberty (CDGP).

### Growth Hormone and Insulin-like Growth Factor-I

Growth hormone levels increase throughout the first stages of pubertal development to reach a maximum at the time of PHV. The levels of IGF-I also rise, denoting a switch to positive feedback or at least an absence of negative feedback action of IGF-I on GH release. Although serum testosterone levels increase markedly by PHV, it is likely that the conversion to estradiol is responsible for the alteration in pulsatile GH release.

### Summary and Integration

During pubertal development, there is an intricate interplay between the increased activation of the HPG and the GH/IGF-I axes. Both contribute to the pubertal spurt in linear growth and to the marked alterations in body composition and the regional distribution of body fat. They have independent and at least additive, if not synergistic, actions at puberty.

## CAUSES OF DELAYED PUBERTAL DEVELOPMENT IN BOYS

Delayed pubertal development has many causes, both physiological (functional) and pathological. It usually occurs because of inadequate secretion of the gonadal steroid hormones, especially that of testosterone in boys. The causes may be conveniently separated into hypogonadotropic (central) and hypergonadotropic causes, depending on primary hypothalamic–pituitary dysfunction or primary gonadal failure.

Of these categories, the physiological causes are by far the most common, especially in 14- to 16-year-old boys. As the second decade of life ends, the likelihood of finding a pathological cause rises sharply (see Table I).

## EVALUATION OF DELAYED PUBERTY IN BOYS

As with any evaluation that includes growth and development, one begins with the growth curve. Normal

**Table I Causes of Delayed Pubertal Development in Boys**

Hypergonadotropic (primary) hypogonadism
Congenital
Chromosomal abnormality
Klinefelter syndrome 47 XXY and variants
Anorchia (vanishing testis syndrome)
Acquired
Trauma (or surgery)
Chemo or radiation therapy
Post infectious (e.g., mumps orchitis)
Hypogonadotropic (secondary) hypogonadism
Congenital
Various forms of hypothalamic–pituitary disease
Gonadotropin-releasing hormone deficiency
Multiple pituitary hormone deficiencies
Congenital malformations as part of holoprosencephalic anomalies
Acquired
Tumors, particularly craniopharyngioma
Pituitary apoplexy
Malnutrition and acute and chronic systemic disease
Endocrine deficiencies
Hypothyroidism
Hyperprolactinemia
Head trauma
Physiological
Constitutional delay of growth and puberty

growth over an extended period of time speaks to the overall general good health of the individual. The childhood growth rate is relatively constant, usually between 5 and 7 cm per year. In boys who undergo puberty at the usual time, there is an almost imperceptible slowing of growth (the take-off point). This is followed by rapid acceleration to PHV and then deceleration toward zero velocity as the epiphyses of the long bones close. The greater the delay in pubertal development, the greater the slowing of growth before take off—often called the preadolescent “dip.” This occurs at the time that one’s peers are undergoing their most rapid growth, highlighting the differences among likely normal individuals. However, this knowledge does not mitigate the distress of the smaller, and often slighter, individual.

The evaluation is no different from any other medical evaluation in that the history can often give clues as to what to look for in the physical examination or what laboratory tests to order to integrate the information to produce a differential diagnosis. Of particular importance in the history is pattern of growth, not only of the child but also of his siblings and parents. The pubertal development of the siblings and parents can be of help as well because there is some familial patterning.

Although it is not often emphasized, one should take a detailed dietary history because delayed pubertal development may be related to a total energy deficit—caloric deprivation and/or excessive use. Prior medical conditions and some of the medications used to treat them can delay growth and pubertal development. There are many individual organ systems that may be disordered and cause delayed growth and pubertal development. Many are related to the gastrointestinal (GI) system—decreased intake and/or excessive caloric losses. The other major organ systems that can be affected are the endocrine system and central nervous system (CNS). Disorders of the HPG, GH/IGF-I, thyroid, and adrenal axes can lead to delayed growth and sexual maturation. CNS disease may be heralded by headache, visual disturbance, seizures, and nausea and vomiting.

The general physical examination might not be too rewarding unless there are congenital (or acquired) anomalies. The typical boy with delayed pubertal development has a relatively normal physical examination—but for a younger child. The key elements are the measurements of height and weight and the genitalia, including pubic hair. There are convenient stages to mark pubertal progression; noting the Tanner stages can integrate not only the timing but also the tempo of pubertal development.

Laboratory testing that may help to define the diagnosis includes the hemogram and tests of liver, kidney, and GI function. From the endocrine point of view, one seeks evidence for low IGF-I and IGF binding proteins (BP)-3, low thyroxine level, low testosterone level, or high prolactin or cortisol level. The detailed endocrine evaluation is beyond the scope of this article.

Imaging studies may help; an X ray of the left hand and wrist can indicate a bone age (biological age). That is critical as one tries to determine whether the delay is significantly outside of the normal variation. Other imaging, whether CNS or GI or the like, should follow from clues gathered in the history, physical examination, and laboratory evaluations. A karyotype is indicated if Klinefelter syndrome is considered. Boys with this condition often have delayed tempo of puberty but are of normal or slightly increased stature.

The younger the child during his teenage years, the more likely that physiologically delayed puberty (CDGP) is the correct "diagnosis." Treatment depends on the diagnosis.

## TREATMENT OF DELAYED PUBERTAL DEVELOPMENT IN BOYS

For most causes of delayed puberty, testosterone administration will form part of the therapeutic plan. It may be exogenous (usual) or may involve merely waiting for the endogenous testosterone to be secreted. By the time most boys with CDGP arrive in the specialist's office, testosterone therapy has already been considered. It is critical to start slowly and increase the dose slowly so as not to cause too rapid epiphyseal closure and shortened adult height. Testosterone therapy can be halted as the testes increase in size, in concert with increasing endogenous testosterone secretion.

For pathological conditions, the treatment of the individual obviously depends on the diagnosis. It may be as simple as adding calories or dampening the activity of inflammatory bowel disease, or it may be as complex as removing a CNS tumor or correcting significant renal failure. Many of the endocrine deficits can be replaced (e.g., GH, thyroxine, testosterone), and some of the excesses can be dampened (e.g., hyperprolactinemia, Cushing syndrome).

Whatever the cause, it is critical to follow the growth and development status over time. The amount of testosterone given should escalate over time to mimic the natural evolution of puberty. Physiologically delayed puberty is very common in 12- to 15-year-old boys. As they become older adolescents, it is less likely that a physiological cause for delayed pubertal development will be found. Often, it is the disease and distress that are the most prominent symptoms, and discussions with the patient and his parents along with short-term therapy with testosterone may be all that is required to help get the boy through pubertal development and toward young adulthood. In addition to growth and genital development, there are remarkable alterations in body composition, including bone mass and fat distribution. Dissatisfaction with his body image is often a cardinal feature of the adolescent boy's distress, and therapy with testosterone has major salutary effects.

## See Also the Following Articles

Delayed Puberty and Hypogonadism, Female • Hypergonadotropic Hypogonadism • Precocious Puberty, Central (Male) • Pseudoprecocious Puberty, Male • Puberty, Male: Mechanisms of Onset and Progression

## Further Reading

- Abbassi, V. (1998). Growth and normal puberty. *Pediatrics* **102**, 507-511.
- Bourguignon, J. P. (1988). Linear growth as a function of age at onset of puberty and sex steroid dosage: Therapeutic implications. *Endocrinol. Rev.* **9**, 467-488.
- Crowne, E. C., Shalet, S. M., Wallace, W. H., Eminson, D. M., and Price, D. A. (1990). Final height in boys with untreated constitutional delay in growth and puberty. *Arch. Dis. Child.* **65**, 1109-1112.
- Houchin, L. D., and Rogol, A. D. (1998). Androgen replacement in children with constitutional delay of puberty: The case for aggressive therapy. *Baillieres Clin. Endocrinol. Metab.* **123**, 427-440.
- Kerrigan, J. R., Martha, P. M., Jr., Blizzard, R. M., Christie, C. M., and Rogol, A. D. (1990). Variations of pulsatile growth hormone release in healthy, short prepubertal boys. *Pediatr. Res.* **28**, 11-14.
- Rogol, A. D., and Roemmich, J. N. (2000). Puberty as a growth promoted event. In "The Onset of Puberty in Perspective" (J. P. Bourguignon and T. M. Plant, eds.), pp. 97-111. Elsevier Science, Amsterdam.
- Tanner, J. M. (1962). "Growth at Adolescence." Charles C Thomas, Springfield, IL.

have depression.  $T_3$  has been reported to influence the effect of antidepressants either by hastening the onset of response or by converting a nonresponder into a responder. Although it has been given as adjuvant therapy for refractory depression for more than two decades, efficacy has not been definitively established in double-blind control trials and its use for this purpose remains controversial.

## CLINICAL MANIFESTATIONS OF DEPRESSION AND THYROID DISEASE

At a given point in time, major depression may affect up to 15% of the general population. The diagnosis is made in those who present with a depressed or irritable mood and who have a lack of interest or pleasure for at least 2 weeks in combination with other symptoms (see Table I). The initial episode of depression most frequently occurs during the fourth or fifth decade of life but may occur at any age. If left untreated, the duration of an episode can vary greatly, from a few months to 1 or more years. Occasionally, it will persist as chronic depression. A family history of mood disorders is common. Recurrence, which can be seen in half of such patients, generally occurs within 2 years of the first episode.

As in thyroid disease, depression occurs more often in women than in men. Although the peak incidence in women occurs between 35 and 45 years of age, the diagnosis is now made more frequently in older women than was the case previously. The prevalence of autoimmune thyroid disease is highest among women more than 40 years of age. Most individuals presenting with depression do not have biochemical evidence of thyroid dysfunction; despite this, depression and thyroid dysfunction share many clinical features (see Table II). In those cases where patients present with depression and findings that suggest thyroid dysfunction, it may be difficult to distinguish clinically which entity is the cause of the symptoms.

**Table I** Symptoms of Major Depression

- |  |
|--|
| <ul style="list-style-type: none"> <li>• Change in appetite and weight</li> <li>• Insomnia or hypersomnia</li> <li>• Fatigue or loss of energy</li> <li>• Motor agitation or retardation</li> <li>• Feelings of worthlessness or guilt</li> <li>• Decreased ability to concentrate and make decisions</li> <li>• Recurrent thoughts of death or suicide</li> </ul> |
|--|

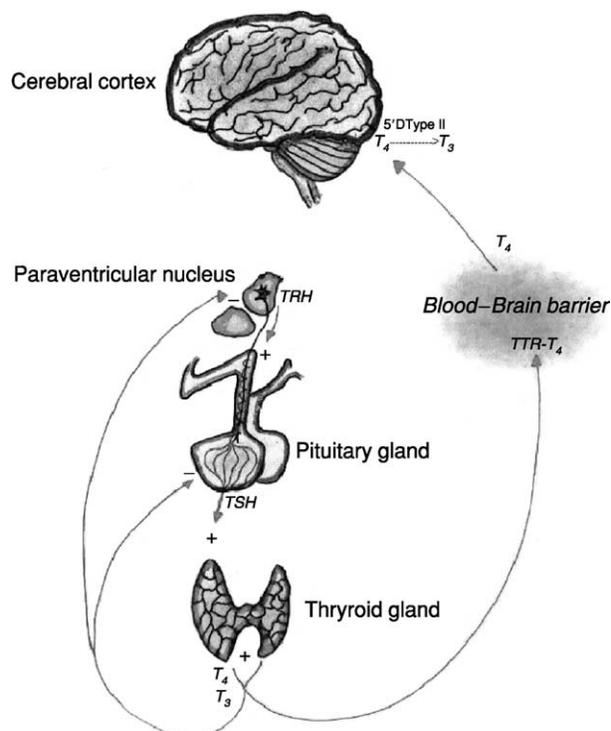
**Table II** Clinical Findings Common to Depression and Hypothyroidism

- |   |
|---|
| <ul style="list-style-type: none"> <li>• Anorexia</li> <li>• Apathy</li> <li>• Anhedonia</li> <li>• Constipation</li> <li>• Lethargy</li> <li>• Anergia</li> <li>• Decreased libido</li> <li>• Dysphoria</li> <li>• Fatigue</li> <li>• Somnolence</li> <li>• Melancholia</li> <li>• Weight gain or weight loss</li> </ul> |
|---|

## REGULATION OF THE HYPOTHALAMIC–PITUITARY–THYROID AXIS

The fundamental actions of the hypothalamus–pituitary–thyroid axis (HPTA) are based on stimulation of the pituitary thyrotrophs balanced by a negative feedback inhibition. The hypothalamus regulates the synthesis and release of thyrotropin (TSH) through the secretion of thyrotropin-releasing hormone (TRH). TRH is produced in the paraventricular nucleus (PVN) of the hypothalamus and then is transported to the specialized nerve terminals in the median eminence before being liberated into the hypophyseal portal blood. The basophilic cells of the anterior pituitary produce TSH in response to TRH. Then, TSH is liberated in the circulation and stimulates the thyrocytes to produce thyroid hormones via the thyrotropin receptor (TSHr). Once thyroxine ( $T_4$ ) and  $T_3$  are liberated, they exert an inhibitory effect on TSH and TRH secretion (see Fig. 1). Apart from thyroid hormones, there are other inhibitors of TSH secretion. This group includes somatostatin, dopamine, glucocorticoids, and certain cytokines (see Table III). On the other hand,  $\alpha$ -adrenergic agonists stimulate the thyrotrophs.

The most abundant thyroid hormone produced by the thyroid gland is  $T_4$ . The monodeiodination of  $T_4$  in extrathyroidal tissues gives rise to 80% of the  $T_3$  in the peripheral circulation, with the remainder being derived from direct thyroid gland secretion. Within the CNS, local deiodination activity in the brain is of great importance because it enables the CNS to maintain an optimal  $T_3$  content. Type II 5'-deiodinase and 5-deiodinase are the major mediators of this process. The former, which is responsible for the conversion of  $T_4$  to  $T_3$ , is found in the cerebral



**Figure 1** The hypothalamic–pituitary–thyroid axis (HPTA). The fundamental action of the HPTA is based on stimulation of the pituitary thyrotropin by thyrotropin-releasing hormone (TRH). In response to this signal, thyrotropin (TSH) is liberated into the circulation and reaches the thyroid gland, which secretes thyroxine ( $T_4$ ) and triiodothyronine ( $T_3$ ), with the former being the most abundant hormone produced by the gland. Through a negative feedback response, both TRH secretion and TSH secretion are inhibited by the thyroid hormone. Other substances can also inhibit TSH secretion to a certain degree. Once  $T_4$  is in the circulation, it crosses the blood–brain barrier with the aid of transthyretin (TTR). Within the cerebral cortex,  $T_4$  is converted into  $T_3$  by 5′-deiodinase type II (5DType II). Finally,  $T_3$  is inactivated by 5-deiodinase (5D) into diiodothyronine ( $T_2$ ). Triiodothyronine is the active thyroid hormone at the cellular level.

cortex, pituitary, and hypothalamus. The latter is also found in the cerebral cortex but not in the anterior pituitary. It results in degradation of  $T_3$  and inactivates  $T_4$ , yielding reverse triiodothyronine ( $rT_3$ ). In addition, 5-deiodinase activity predominates in the euthyroid state, and the  $rT_3$  produced plays a role in the regulation of local deiodination (see Fig. 1). The distribution of these enzymes within the various regions of the nervous system varies. The cortex has the highest level of activity, followed by the midbrain, pons, hypothalamus, and brainstem (in that order).

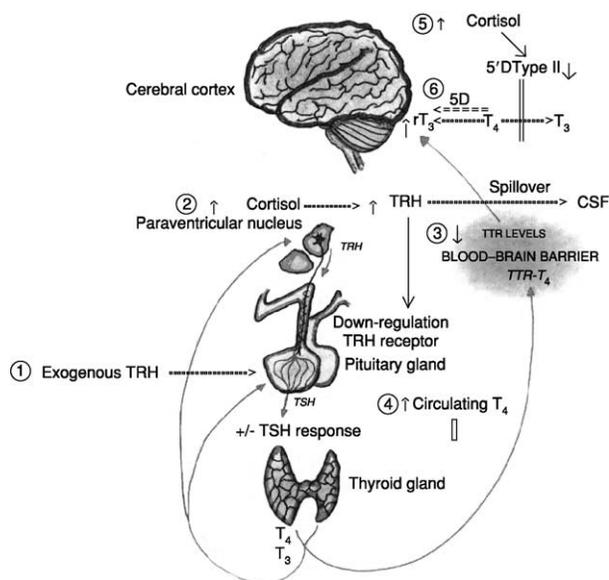
**Table III** Conditions with a Blunted TSH Response to Exogenous TRH

• Depression
• Elderly males
• Acromegaly
• Cushing's syndrome
• Alcoholism
• Medications
Glucocorticoids
Dopamine
Phenytoin
Somatostatin

### ALTERATIONS OF THE HYPOTHALAMIC–PITUITARY–THYROID AXIS IN DEPRESSION

Patients with depression can present many abnormalities of the HPTA (see Fig. 2). The most widely recognized disturbance is blunting of the TSH response to TRH stimulation. Defined as a TSH rise of less than  $5 \mu\text{IU/ml}$ , this phenomenon occurs in 25 to 30% of depressed individuals. Normalization of this response is noted once clinical recovery ensues. It differs from thyrotoxicosis in which the TSH response is flat and the circulating  $T_3$  levels are elevated. A blunted TSH response is not unique to depression and may occur in other clinical settings (see Table III). The mechanism of this blunted response is not entirely clear, but one hypothesis proposes that there is TRH hypersecretion with down-regulation of the TRH receptors on the thyrotrophs.

Another common abnormality of the HPTA is held to be an increase in total and/or free thyroxine ( $\text{FT}_4$ ) levels, although still within the conventional normal range, that regress after successful treatment of depression. It has been reported that some patients admitted with acute psychosis, including depression, manifest a transient elevation above normal levels of  $T_4$  and/or  $\text{FT}_4$  and occasionally TSH. These findings usually resolve spontaneously within 2 weeks. In a fetal rat hypothalamic culture system, we showed an increase in TRH gene expression on glucocorticoid exposure. This effect is notable because human depression is characterized by hypercortisolemia, which we believe leads to activation of the TRH neuron and, consequently, thyroid function. The hypercortisolemia of depression probably results from impaired function of the hippocampus, a locus for glucocorticoid negative



**Figure 2** Alterations of the hypothalamic–pituitary–thyroid axis (HPTA) in depression. In depression, various alterations of the HPTA have been described. (1) Blunted (+/–) thyrotropin (TSH) response to exogenous thyrotropin-releasing hormone (TRH) administration. (2) Elevated cortisol levels activate the TRH neuron, which can lead to down-regulation of the TRH receptor, providing one possible explanation for blunting of the TSH response to exogenous TRH. A spillover of TRH into the cerebrospinal fluid (CSF) can occur. (3) Decreased levels of transthyretin (TTR), a thyroxine ( $T_4$ ) transport protein, have been found. (4) Circulating  $T_4$  levels have been found to be elevated. (5) Inhibition of 5'-deiodinase type II (5DType II) occurs, possibly due to elevation of cortisol, resulting in decreased triiodothyronine ( $T_3$ ). (6)  $T_4$  is converted into reverse  $T_3$  ( $rT_3$ ), resulting in higher  $rT_3$  levels in the CSF.

feedback of the hypothalamic–pituitary–adrenocortical axis. Although glucocorticoids generally inhibit the thyroid axis in humans and rats *in vivo*, a “functional” disconnection of the hypothalamus from the rest of the brain (as is postulated to occur in depression) would remove this inhibitory influence from the hypothalamus. Indeed, a fornical lesion that severs the hypothalamus from hippocampal regulation will increase thyroid function. In many ways, the hypothalamus in culture is analogous to a deafferented hypothalamus *in vivo*. We hypothesize that the direct stimulation of the TRH neuron by glucocorticoids seen *in vitro* is overridden *in vivo* by an inhibitory influence emanating from the hippocampus in the normal human or rat but not in some persons with clinical depression. In these individuals, activation of the TRH neuron could lead to increased hypothalamic TRH secretion with down-regulation of the TRH receptors. This could result in a blunted TSH response to exogenous TRH and spillover of TRH into the cerebrospinal fluid

(CSF), increased levels of which have been reported. In addition, we have explored a direct effect of antidepressants on the TRH neuron to explain the reversal of hyperthyroxinemia with successful treatment of clinical depression. The results showed that the selective serotonin reuptake inhibitors (SSRIs) and the tricyclic antidepressants (TCAs) inhibit TRH secretion. These studies suggest that the fall in circulating  $T_4$  levels seen with antidepressant medication might reflect a direct effect on the TRH neuron and consequent reduced activation of the thyroid axis. Antidepressants may also be clinically efficacious by enhancing  $T_4$ -to- $T_3$  conversion in the CNS.

As mentioned previously,  $T_3$  is the active thyroid hormone in the brain. It has been hypothesized that depression leads to inhibition of type II deiodinase, possibly due to the elevated cortisol levels in this disorder. Therefore,  $T_4$  is converted to  $rT_3$  by 5-deiodinase. This would explain the elevated levels of  $rT_3$  in the CSF of individuals with unipolar depression.

Transthyretin (TTR), a  $T_4$  transport protein, is synthesized by the choroid plexus and accounts for up to 25% of the protein in the CSF. It has a relative binding affinity of 39.3% for  $T_4$  and of only 1.4% for  $T_3$ ; thus, it is unlikely that TTR plays a significant role in the transport of  $T_3$  across the blood–brain barrier. Interestingly, in a study of eight patients with refractory depression, CSF levels of TTR were much lower when compared with nine patients with neurological disease but without depression. The authors of this study proposed that low levels of TTR could give rise to “brain hypothyroidism” with normal peripheral thyroid hormone concentrations in depression.

Autoimmune thyroiditis is found in at least 15% of depressed patients and is associated with an exaggerated response to TRH stimulation. In rapid cycling bipolar disease (four or more episodes of manic depression per year), the prevalence of autoimmune thyroid disease is even higher, reaching up to 50% in one series. Furthermore, women with postpartum thyroiditis often suffer from depression. In a study of 145 women with presence of antithyroid antibodies, 47% had significant depressive symptoms. When these women were compared with a control group, with negative titers to antithyroid antibodies, only 32% presented with depressive symptoms regardless of the thyroid function abnormalities. Among the antibody-positive individuals, 62 had episodes of thyroid dysfunction; a total of 27 had signs of hypothyroidism, 11 had signs of hyperthyroidism, and 24 showed biochemical evidence of hyperthyroidism followed by hypothyroidism.

Alterations in the circadian rhythm of TSH have been described, including absence of the normal nocturnal TSH surge that may result in an overall diminution of thyroid hormone secretion. This suggests that there may be a degree of central hypothyroidism in some patients with depression. Furthermore, sleep deprivation, which has an antidepressant effect, leads to restoration of the nocturnal TSH rise and an elevation in the levels of  $T_4$  and  $T_3$ .

## DEPRESSION IN THYROID DISEASE

### Hypothyroidism

Hypothyroidism indicates any degree of thyroid hormone deficiency. The Wickham survey found an incidence of 2% among 2800 individuals. The mean age at diagnosis was 57 years. Female predominance was noted, although in debilitated geriatric populations the prevalence does not have a gender difference. Subclinical hypothyroidism (SCH) is characterized by a mild elevation of the serum TSH, normal circulating levels of thyroid hormones, and absence of the typical symptoms of overt hypothyroidism. The prevalence of SCH is 2 to 8%. It is higher in women than in men (7.5 vs. 2.8%), particularly if the individual is over 60 years of age. In that instance, the prevalence can be as high as 16%. Progression to overt hypothyroidism occurs in 20 to 50% of individuals within a period of 4 to 8 years. Those patients with elevated titers of antithyroid antibodies have a higher incidence of progression (close to 80% within 4 years).

Studies in adult patients with hypothyroidism have demonstrated decreased cerebral blood flow and consumption of oxygen and glucose (27% below normal). Results from positron emission tomography (PET) studies indicate that TSH levels correlate inversely with global and regional blood flow as well as cerebral glucose metabolism. In addition, when PET studies were performed in individuals with severe hypothyroidism of short duration, the brain activity was globally reduced without the regional modifications observed in primary depression. As noted previously,  $T_3$  in the brain is derived from deiodination of  $T_4$  by type II deiodinase. The activity of this enzyme is increased in hypothyroidism. Intracerebral generation of  $T_3$  increases as serum concentrations of  $T_4$  decline. It seems that intracellular  $T_3$  concentrations will remain fairly stable until serum  $T_4$  has been depleted. Depression has been described in all grades of hypothyroidism. The occurrence of depression is believed to be higher in those who have a history of

a first-degree relative with depression. Up to 10% of patients admitted for treatment of depression are found to have subclinical or frank hypothyroidism. Clinical manifestations are indistinguishable from non-thyroid-related depression. Response is usually refractory to antidepressants alone; on certain occasions, it has also been resistant to thyroid hormone replacement.

### Hyperthyroidism

Hyperthyroidism can be seen in up to 3% of individuals over 60 years of age. Typically, Graves' disease is still the most common cause of thyrotoxicosis in the elderly, but toxic multinodular goiter or adenomas are more frequent than they are in young people. The classic findings of hyperthyroidism are not seen in the elderly. The term "apathetic thyrotoxicosis" has been used to address such a condition, which may be associated with depressive symptoms (see Table IV), that are relieved by antithyroid treatment. The laboratory data may show a suppressed TSH with or without frank elevation of  $T_4$  and  $T_3$ . In studies of patients with thyrotoxicosis, depression has been documented in up to 30 to 60% of the cases.

## THYROID HORMONE SUPPLEMENTATION IN DEPRESSION

In the evaluation of thyroid hormone as adjuvant therapy for depression, there is a need to exclude patients with borderline TSH elevation as well as those with detectable antithyroid antibodies. This population may respond favorably to thyroid hormone supplementation simply because of underlying hypothyroidism. In addition, when parameters of thyroid function are evaluated, the antithyroid peroxidase (anti-TPO) antibody should be included. It is conceivable that patients with evidence of autoimmune thyroid disease might benefit from treatment

Table IV Manifestations of Apathetic Thyrotoxicosis

- 
- Anorexia
  - Apathy
  - Depression
  - Confusion/Slow mentation
  - Weight loss
  - Constipation
  - Atrial fibrillation
  - Angina exacerbation
  - Muscle atrophy
-

with levothyroxine because the treatment corrects a cryptic underlying decrease in thyroid function. Patients with rapid cycling bipolar disease may benefit from pharmacological doses of levothyroxine. Such cases may reflect an aberrant expression of hypothyroidism. In studies involving thyroid hormone as adjuvant therapy in depression,  $T_3$  appears to be superior to  $T_4$  and so is used much more often. Cooke and colleagues reported that  $T_3$  augmented the response to antidepressant therapy in  $T_4$ -replaced hypothyroid patients in a randomized controlled trial during a 3-week period. Joffe found  $T_3$  to be significantly more effective than  $T_4$  in patients with depression who did not respond to TCAs. However, because  $T_4$  equilibrates in tissues more slowly than does  $T_3$ , 6 to 8 weeks of  $T_4$  therapy may be required for adequate comparison of its efficacy with the more rapidly acting  $T_3$ .

### Thyroid Hormone Use to Hasten Antidepressant Response

The use of  $T_3$  has been proposed to hasten antidepressant effect. The therapeutic response of TCAs is delayed up to 1 month. Prange and colleagues studied 20 patients with retarded depression, and 10 patients received  $T_3$  sodium (25  $\mu$ g daily) in addition to imipramine (150 mg daily). The remaining patients were given imipramine, at equivalent doses, as well as placebo. After 10 days of treatment, they reported a significant improvement in the Hamilton Rating Scale (HRS) for depression. The scores improved by 50%, when compared with the initial value, in those taking the combination of imipramine and  $T_3$ . The patients treated only with imipramine reached the same reduction in the HRS, but not until after the third week of treatment. The benefit of the combination was limited to women. Wilson and colleagues reported a similar benefit from adjuvant  $T_3$  in 20 individuals with nonretarded depression.

Wheatley studied a double-blind comparison of amitriptyline alone and in combination with two doses of  $T_3$  (40 and 20  $\mu$ g daily). These patients had been admitted for neurotic depression with minimal anxiety symptoms. A total of 52 patients were enrolled and then divided into three groups. All received amitriptyline (100 mg daily). There was a steady and significant improvement in depression scores for those who received  $T_3$  (40  $\mu$ g daily) after 14 days of therapy. Interestingly, the patients who received the lower dose of  $T_3$  did not seem to be that different from the placebo group when the self-rating scores were

reported. Once adjustments were made for gender, there was a significant response in females but not in males, confirming prior findings. There was no significant difference in responses once results were adjusted to age and anxiety score.

Some studies have not found any benefit from the use of thyroid hormone as an accelerator. One of these was a randomized double-blind study of 49 patients with primary depression in which patients were treated with imipramine (200 mg daily) alone or in combination with  $T_3$  (25  $\mu$ g daily for 10 days). The researchers were unable to demonstrate at the end of the study period (22 days) that there was a significant difference between the two arms. The authors of the study questioned the efficacy of  $T_3$  in enhancing the antidepressive activity of imipramine.

As Joffe and colleagues pointed out, these and other studies performed more than two decades ago suffered from the major limitations of small sample size and what are currently judged to be inadequate doses of antidepressants. Therefore,  $T_3$  has not been established as a means of accelerating the onset of response to TCAs.

### Thyroid Hormone Use to Convert Nonresponders to Antidepressants into Responders

Approximately 25% of patients with depression fail to improve with TCAs alone. As a consequence, there has been much interest in the possibility that augmentation therapy with  $T_3$  could convert a nonresponder into a responder. While investigating the use of thyroid hormone as an augmentation strategy, Goodwin and colleagues treated 12 patients who had failed treatment with TCAs for at least 4 weeks and added  $T_3$  (25–50  $\mu$ g daily). After 28 days of treatment, 8 patients experienced a significant improvement in the rating (score <5) that was sufficient to allow discharge from the hospital, whereas 4 reached a significant reduction in the HRS score (decrease of >3 points). Another study, one of the largest published to date, took 51 nonresponders to TCAs and added either  $T_3$ , lithium, or placebo to the antidepressant. There was a significant response in half of those included in the  $T_3$  and lithium groups when compared with those in the placebo group, although no superiority was seen in the use of either medication for augmentation. Thase and colleagues evaluated the efficacy of augmentation with  $T_3$  in a series of 20 outpatients with recurrent unipolar depression who had failed an extended and closely monitored course

(>12 weeks) of imipramine and interpersonal psychotherapy. In their open clinical trial of adjunctive  $T_3$  (25  $\mu\text{g}$  daily) for 4 weeks, the overall response did not differ significantly from that of a matched historical comparison group managed with imipramine alone. Similar findings were noted by Steiner and colleagues, who used a double-blind study to compare the addition of  $T_3$  to imipramine with imipramine and placebo and with electroconvulsive therapy (ECT). At the end of the 5 weeks, 3 of the 4 women with recurrent unipolar endogenous depression in each group responded to the specific regimen administered.

Currently, the most frequent prescribed antidepressants are the SSRIs. There are no controlled trials of  $T_3$  augmentation with these agents. However, we can find case reports of beneficial effects from  $T_3$  after failure to respond to them. An example is a 30-year-old woman with major depression who did not respond to a 10-week trial of fluoxetine (20 mg daily for 4 weeks and 40 mg daily for the remainder). Just 6 days after the addition of  $T_3$  (25  $\mu\text{g}$  daily), she reported improvement in her mood. The HRS score had dropped from 23 to 10. By 4 weeks of treatment, the score was 8, her suicidal ideations had disappeared, and her energy level and motivation had improved.

Moreover, recent studies have focused on whether a relationship can be detected between the swiftness in reaching a clinical remission or the risk of recurrence and the baseline thyroid hormone levels. Cole and colleagues observed that approximately one-third of 65 patients who presented in the depressive phase of bipolar I disorder had a thyroid profile consisting of a normal free thyroxine index and TSH values, although the levels were above the median in the latter. The remaining patients did not present with this profile and had a median time to remission of 1 year, whereas the group just mentioned experienced remission 4 months faster. Thus, the authors questioned whether adjuvant therapy with thyroid hormone may optimize the management of patients presenting in the depressive phase of bipolar disorders. Joffe and colleagues looked at the relationship between basal thyroid hormone levels and the life course of depressive illness in 75 outpatients with unipolar major depressive disorder. The significant positive predictors of recurrence were comorbid anxiety, number of previous episodes of depression, prolonged course of episodes, and level of  $T_3$ . An increase in  $T_3$  was associated with a 22% decrease in the risk of recurrence.

Although  $T_3$  adjuvant therapy may help 25% of TCA nonresponders as some researchers claim, ob-

servers must be cautious about this conclusion for many reasons. First, we do not know with certainty which clinical and/or biochemical parameters determine the subset of depressed patients who will benefit from adjuvant  $T_3$ . Second, the clarification of the thyroid status was not universally reported in the studies that are available, and unrecognized "subclinical" hypothyroidism or autoimmune thyroid disease could have led researchers to overestimate the therapeutic response to  $T_3$ . Third, the relationship between  $T_3$  dose and clinical response is unclear. Fourth, because the longest duration of  $T_3$  administration in studies has been only a few weeks, the extent of treatment remains imprecise, although it has been recommended by some investigators that thyroid augmentation be discontinued 8 to 12 weeks after a response and then reinstated if symptoms recur. Fifth, studies regarding long-term side effects from use of  $T_3$  in depressed patients are not available, and when instituting this form of therapy, the potential effect of thyroid hormone overreplacement on the heart and bones has to be strongly taken into account. Finally, longer and larger randomized, double-blind, placebo-controlled trials are needed, especially with SSRIs given that these are the contemporary antidepressants.

### USE OF THYROTROPIN-RELEASING HORMONE IN REFRACTORY DEPRESSION

The therapeutic use of TRH in mood disorders is based on the fact that this hormone has direct effects on the CNS independent of pituitary and thyroid stimulation. High-affinity receptors to TRH are found throughout the brain especially, in the amygdala and the hippocampus, and modulate the effects of serotonin and dopamine. In addition, in some studies the concentrations of TRH in the CSF of untreated individuals with mood disorders were elevated. It remains unclear whether this corresponds to a natural compensatory mechanism or to a pathological occurrence. This elevation could render an explanation for the blunted TSH response after exogenous administration of TRH because if the relatively high endogenous levels of TRH are causing receptor down-regulation, the pituitary might be less responsive to exogenous stimulation. Some treatments for mood disorders, such as carbamazepine, increase endogenous production of TRH, further elevating the CSF levels and favoring the hypothesis that this is a compensatory mechanism possibly trying to overcome

the brain hypothyroidism. Marangell and colleagues administered 500  $\mu\text{g}$  of TRH intrathecally to eight drug-free patients with refractory depression. Identical sham spinal punctures were performed a week apart. The study was carried out in a double-blind fashion. Five patients had a clinically significant, rapid, and robust but short-lived improvement in mood after TRH infusion, but not after the sham puncture. Systemic thyroid function was unaltered. None of the patients showed a worsening of depression after infusing TRH, suggesting that the elevation of TRH in the CSF is likely due to a physiological response. Intrathecal TRH might be a positive modulator of mood and might have a role in the treatment of refractory depression.

## CONCLUSIONS

Hypothyroidism predisposes patients to depression that may be wholly reversed by thyroxine replacement. When first seen for depression, most patients have normal circulating TSH,  $\text{T}_3$ , and  $\text{T}_4$ . However, 15% may have subclinical hypothyroidism and have a TSH hyper-response to TRH stimulation. Another 10 to 15% may show evidence of autoimmune disease. In this setting, thyroxine may alleviate depressive symptoms alone or in combination with antidepressants. Various thyroid axis abnormalities in depression have been described, including a blunted TSH response to exogenous TRH, an increase in circulating  $\text{T}_4$ , and elevated TRH levels in the CSF. It is hypothesized that there is enhanced hypothalamic TRH secretion that is triggered by dysfunction in the hypothalamic–pituitary–adrenal axis. The clinical features of depression often resemble those of hypothyroidism and may reflect brain hypothyroidism despite systemic euthyroidism. The mechanism may be related to impaired conversion of  $\text{T}_4$  to  $\text{T}_3$  in the brain due to inhibition of type II 5'-deiodinase, possibly by increased levels of circulating cortisol and/or reduced transport of  $\text{T}_4$  across the blood–brain barrier. Adjuvant therapy with  $\text{T}_3$  for patients receiving antidepressant medication has been postulated to augment responsiveness to therapy in about a quarter of cases of refractory depression, possibly by correcting brain hypothyroidism. TCAs and SSRIs have been reported to inhibit TRH gene expression *in vitro*, a fact that might explain the normalization of serum  $\text{T}_4$  levels following treatment; TCAs also enhance the

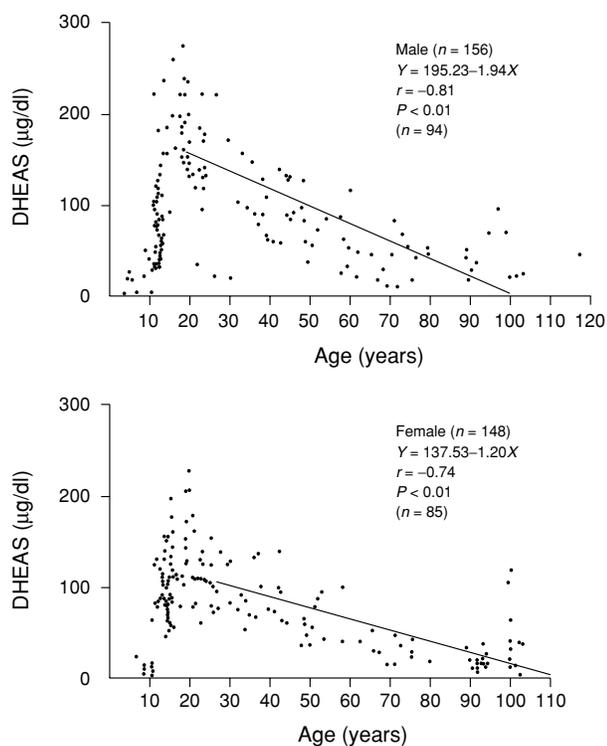
conversion of  $\text{T}_4$  to  $\text{T}_3$  in the brain, providing another potential explanation for their therapeutic efficacy in depression. In some studies, the concentrations of TRH in the CSF of untreated individuals with mood disorders have been found to be elevated, and it is possible that this corresponds to a natural compensatory mechanism. Intrathecal TRH might be a positive modulator of mood, and its administration in refractory depression might be a reasonable strategy in managing the depression, but further studies are needed.

## See Also the Following Articles

Graves' Disease • Hypothalamus–Pituitary–Thyroid Axis • Hypothyroidism, Subclinical • Hypothyroidism, Treatment of • Thyroid Autoimmunity • Thyrotoxicosis, Overview of Causes • Thyrotoxicosis, Systemic Manifestations • Thyrotropin-Releasing Hormone (TRH)

## Further Reading

- Braverman, L. E., and Utiger, R. D. (2000). "Werner & Ingbar's The Thyroid: A Fundamental and Clinical Text," 8th ed. Lippincott Williams & Wilkins, Philadelphia.
- Cole, D. P., Thase, M. E., Mallinger, A. G., Soares, J. C., Luther, J. F., Kupfer, D. J., and Frank, E. (2002). Slower treatment response in bipolar depression predicted by lower pretreatment thyroid function. *Am. J. Psychiatry* **159**, 116–121.
- Constant, E. L., De Volder, A. G., Ivanoiu, A., Bol, A., Labar, D., Seghers, A., Cosnard, G., Melin, J., and Daumerie, C. (2001). Cerebral blood flow and glucose metabolism in hypothyroidism: A positron emission tomography study. *J. Clin. Endocrinol. Metab.* **86**, 3864–3870.
- Hennessey, J. V., and Jackson, I. M. D. (1996). The interface between thyroid hormones and psychiatry. *The Endocrinologist* **6**, 214–223.
- Jackson, I. M. D., and Asamoah, E. O. (1999). Thyroid function in clinical depression: Insights and uncertainties. *Thyroid Today* **22**(2), 1–11.
- Joffe, R. T. (1998). The use of thyroid supplements to augment antidepressant medication. *J. Clin. Psychiatry* **59**(Suppl. 5), 26–29.
- Joffe, R. T., and Marriott, M. (2000). Thyroid hormone levels and recurrence of major depression. *Am. J. Psychiatry* **157**, 1689–1691.
- Marangell, L. B., and Callahan, A. M. (1998). Mood disorders and the thyroid axis. *Curr. Opin. Psychiatry* **11**, 67–70.
- Marangell, L. B., George, M. S., Callahan, A. M., Ketter, T. A., Pazzaglia, P. J., L'Herrou, T. A., Leverich, G. S., and Post, R. M. (1997). Effects of intrathecal thyrotropin-releasing hormone (Protirelin) in refractory depressed patients. *Arch. Gen. Psychiatry* **54**, 214–222.
- Nierenberg, A. A. (2000). Depression: Recent developments and innovative treatments—Antidepressant augmentation and combinations. *Psychiatr. Clin. North America* **23**, 743–755.

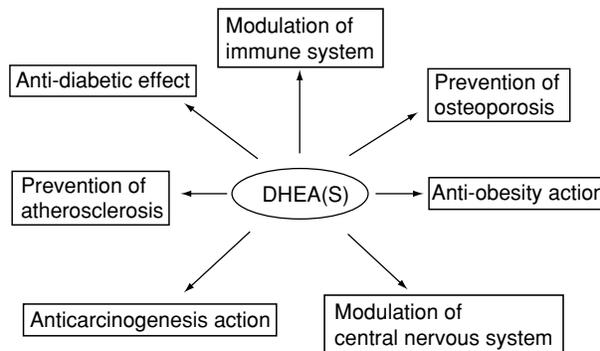


**Figure 1** Age-related change in plasma DHEAS in the male and female.

## BIOLOGICAL EFFECTS AND MECHANISM OF ACTION OF DHEA

The mechanism of action of DHEA(S) remains unclear, but three possible mechanisms have been hypothesized. The first possible mechanism is a direct action through the specific receptor. Although specific binding sites for DHEA have been reported in various cells and tissues, the receptor has not been successfully cloned as yet. The second possible mechanism is an indirect action, in which DHEA is converted to testosterone, or is further converted to estrogen by aromatase, in peripheral tissues and acts through the androgen receptor or the estrogen receptor. Thus, DHEA can be converted to active sex steroids and function in the same cell. This mechanism is called "intracrine action." The third possibility is that the hydrophobic DHEA molecule may alter cell function by interacting with certain macromolecules, such as enzymes.

Experimental studies have demonstrated that DHEA(S) has various beneficial effects including an anti-diabetic effect, prevention of atherosclerosis,



**Figure 2** Beneficial effects of DHEA(S).

anti-obesity action, prevention of osteoporosis, and modulation of the immune system (Fig. 2).

### Anti-diabetic Effect

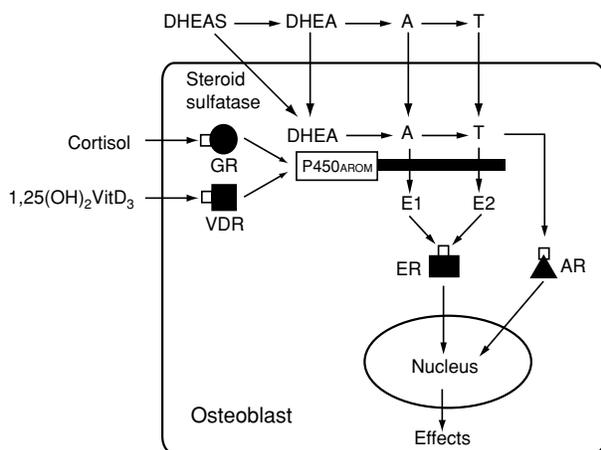
It is reported that administration of DHEA improved insulin resistance by suppression of the increased activities of glucose-6-phosphatase and fructose-1,6-bisphosphatase in a non-insulin-dependent diabetes model, db/db mice. DHEA is also reported to improve insulin resistance observed in rats and decrease the serum level of tumor necrosis factor- $\alpha$  in obese Zucker rats. DHEA increases glucose uptake *in vitro* through the activation of protein kinase C or phosphatidylinositol 3-kinase. In human studies, the serum DHEAS level is reported to be lower in non-insulin-dependent diabetic patients than in age-matched normal male subjects.

### Anti-obesity Action

An anti-obesity effect of DHEA has been observed in animal studies and may be related to several mechanisms including decreased lipogenesis and increased thermogenesis. In humans, the plasma concentration of DHEA decreases inversely with body mass index.

### Prevention of Atherosclerosis

There are many retrospective and cross-sectional human studies demonstrating reciprocal relationships between the serum concentration of DHEA(S) and the incidence of ischemic heart disease. Animal studies also showed that the administration of DHEA



**Figure 3** Schematic representation of intracrine action of DHEA in the osteoblast. GR, glucocorticoid receptor; VDR, vitamin D receptor; ER, estrogen receptor; AR, androgen receptor; A, androstenedione; T, testosterone; E1, estrone; E2, estradiol; 1,25(OH)<sub>2</sub>VitD<sub>3</sub>, 1,25-dihydroxyvitamin D<sub>3</sub>; P450<sub>AROM</sub>, cytochrome P450 aromatase.

suppressed atherosclerosis induced by intimal injury or a high-cholesterol diet. It has been reported that DHEA suppressed the accumulation of cholesterol esters induced by the addition of acetylated low-density lipoprotein in cultured macrophage cells.

### Prevention of Osteoporosis

Postmenopausal osteoporosis is a widespread clinical problem that correlates with low circulating levels of sex steroid hormones. A significant positive correlation between bone mineral density and the serum DHEAS level, but not the serum estradiol level, was observed in postmenopausal women, suggesting a possible role for DHEA(S) in preventing osteoporosis. It has been demonstrated that primary cultured human osteoblasts have aromatase activity, which is enhanced by dexamethasone and 1,25-hydroxyvitamin D<sub>3</sub> (Fig. 3). These results suggest that osteoblasts may utilize the circulating DHEA(S) and convert it to estrogen by aromatase to maintain bone mass after menopause.

### REPLACEMENT THERAPY OF DHEA IN HUMANS

The replacement of DHEA has attracted attention in view of both the prevention of age-related diseases and the control of aging. DHEA is commercially available as a food supplement in the United States,

but the purity and content of that DHEA are not clearly specified. Clinical intervention trials of DHEA replacement are essential for establishing the significance of that therapy. Results from three randomized placebo-controlled studies support the concept that oral replacement of DHEA has beneficial effects.

Six months of daily replacement of DHEA (50 mg/day) in healthy adults (13 men and 17 women) restored the level of DHEA and DHEAS to the high levels observed in young people and caused an increase in the serum concentrations of androgen and insulin-like growth factor-I. This result is associated with a remarkable increase in perceived physical and psychological well-being in both sexes without an effect on libido.

The administration of DHEA (50 mg/day) in 24 female patients with adrenal insufficiency normalized serum concentrations of DHEA, DHEAS, androstenedione, and testosterone. DHEA significantly improved overall well-being as well as scores for depression and anxiety, the frequency of sexual thoughts, sexual interests, and satisfaction of both the mental and physical aspects of sexuality.

Two hundred eighty healthy men and women (60–79 years old) were given DHEA (50 mg/day) or placebo orally for 1 year in a double-blind, placebo-controlled study (DHEAge Study). DHEA administration reestablished a “young” concentration of DHEAS and a small increase in testosterone and estradiol with no adverse side effects. A significant increase in most libido parameters was observed in older women. An improvement in skin status in terms of hydration, epidermal thickness, sebum production, and pigmentation was also noted, particularly in women. Bone turnover was improved in women over 70 years of age as assessed by bone mineral density and bone resorption markers.

### CLINICAL TRIAL FOR THE TREATMENT OF VARIOUS DISEASES

In a study of 14 postmenopausal women, 12 months of daily administration of DHEA via 10% cream caused a significant increase in bone mineral density accompanied by a 2.1-fold increase in the plasma osteocalcin level with no adverse side effects. Oral replacement of DHEA (50 mg/day) in 18 elderly people also increased bone mineral density and decreased fat mass.

The intravenous administration of DHEAS (200 mg/day for 8 weeks) in 11 patients with myotonic dystrophy led to an improvement in the activities of

daily living, an increase in muscle strength, and a decrease in myotonia.

Oral administration of DHEA (50–200 mg/day for 1 year) in 50 female patients with systemic lupus erythematosus (SLE) induced a decrease in the activity of SLE and led to a decrease in the requisite dose of prednisolone.

## CONCLUSION

DHEA is theorized to act as an anti-aging hormone against age-related diseases. However, it is not known whether or not the increase in sex steroid levels in long-term DHEA replacement is safe with regard to the development of steroid-dependent cancers, such as ovarian and prostate cancers. It is necessary to establish the clinical evidence for DHEA replacement therapy in the elderly by conducting prospective long-term randomized studies.

## See Also the Following Articles

Adrenal Androgens • Aging and Longevity of Human Populations • Aging: Muscle • Alzheimer's Disease and Hormones

• Neuroendocrine System and Aging • Osteoporosis in Older Men • Osteoporosis in Older Women

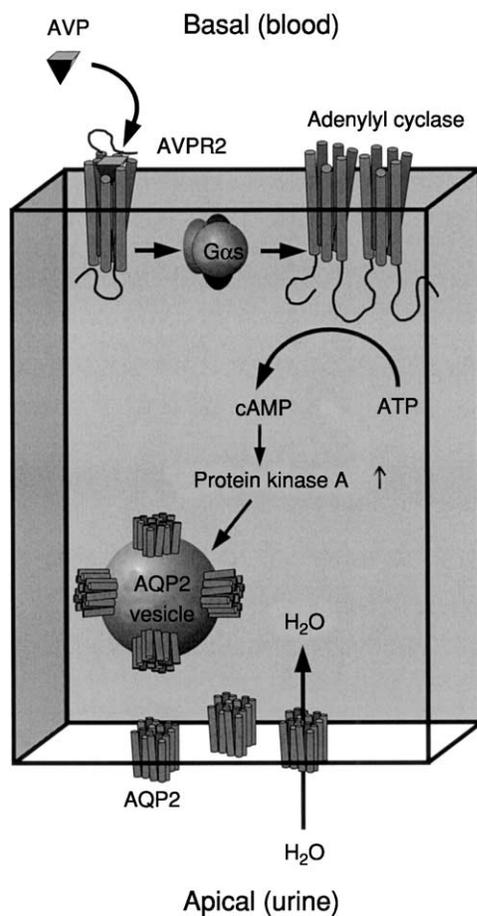
## Further Reading

- Baulieu, E.-E., Thomas, G., Legrain, S., *et al.* (2000). Dehydroepiandrosterone (DHEA), DHEA sulfate, and aging: Contribution of the DHEAge study to a sociobiomedical issue. *Proc. Natl. Acad. Sci. USA* **97**, 4279–4284.
- Becker, K. L. (2001). "Principles and Practice of Endocrinology and Metabolism," 3rd ed. Lippincott Williams & Wilkins, Philadelphia, PA.
- Larsen, P. R., Kronenberg, H. M., Melmed, S., and Polonsky, K. S. (2003). "Williams Textbook of Endocrinology," 10th ed. W. B. Saunders, Philadelphia, PA.
- Nawata, H., Yanase, T., Goto, K., *et al.* (2002). Mechanism of action of anti-aging DHEA-S and the replacement of DHEA-S. *Mech. Ageing Dev.* **123**, 1101–1106.
- Roth, G. S., Lane, M. A., Ingram, D. K., *et al.* (2002). Biomarkers of caloric restriction may predict longevity in humans. *Science* **297**, 811.
- Takayanagi, R., Goto, K., Suzuki, S., *et al.* (2002). Dehydroepiandrosterone (DHEA) as a possible source for estrogen formation in bone cells: Correlation between bone mineral density and serum DHEA-sulfate concentration in postmenopausal women, and the presence of aromatase to be enhanced by 1,25-dihydroxyvitamin D3 in human osteoblasts. *Mech. Ageing Dev.* **123**, 1107–1114.

responsible for acquired NDI. In addition to drug treatment, acquired NDI can also occur temporarily or permanently as a result of certain diseases, such as chronic kidney failure, sickle cell disease, amyloidosis, sarcoidosis, and Sjögren syndrome, and infections.

## MOLECULAR MECHANISMS OF INHERITED NDI

Water reabsorption by the collecting duct is under hormonal control of AVP. The hormone binds to *AVPR2s* of collecting duct cells and initiates a series of intracellular events (Fig. 1). *AVPR2* is a peptide



**Figure 1** Mechanism of AVP-induced increase in water permeability in collecting duct cells. The first step in the antidiuretic action of vasopressin (AVP) is its binding to the V2 vasopressin receptor (*AVPR2*), which is located on the basolateral membrane of collecting duct cells. The receptor is linked to a heterotrimeric Gs protein, which activates adenylyl cyclases. The cyclases increase the production of cyclic AMP (cAMP). By an unknown process that involves protein kinase A, *aquaporin-2* (*AQP2*)-containing vesicles are translocated to the apical plasma membrane, those increasing water permeability of the collecting duct. Mutation of the *AVPR2* or the *AQP2* gene can cause inherited NDI.

receptor that belongs to the large superfamily of G protein-coupled receptors (GPCRs). The receptor activates adenylyl cyclases by interacting with the heterotrimeric G $\alpha$ s protein. Adenylyl cyclase activation increases the intracellular concentration of cyclic AMP (cAMP). cAMP induces the translocation of *AQP2*-laden vesicles to the apical plasma membrane. Ultimately, an increase in the number of *AQP2* water channels incorporated into the luminal membrane of collecting duct principal cells elevates cellular water permeability. The molecular mechanisms underlying docking and fusion of *AQP2*-bearing vesicles are currently unknown.

Following the identification of the DNA sequences and genomic structures of the *AVPR2* and *AQP2* genes, mutations in both genes were determined to be responsible for inherited NDI forms. Autosomal NDI accounts for approximately 10% and X-linked NDI is responsible for approximately 90% of cases of congenital NDI. However, X-linked NDI is still a rare disorder and occurs with a frequency of 4–8 per 1 million male live births. Identification of the molecular defect underlying inherited NDI is of clinical significance because early diagnosis and treatment of affected infants can prevent physical and mental retardation resulting from repeated episodes of dehydration. Probably because of dehydration and hypernatremia in early infancy, only 11% of untreated patients with inherited NDI have normal intelligence. Therefore, immediate genetic testing and identification of inactivating mutations within the candidate genes followed by adequate treatment of inherited NDI result in almost normal physical and mental development.

### Mutations in the *AVPR2* Gene

The *AVPR2* gene is located at the genomic locus Xq28, and more than 170 disease-causing *AVPR2* mutations have been identified in X-linked NDI families. A large variety of genomic alterations have been identified in the *AVPR2* gene, yielding approximately 51% missense mutations, 23% small deletions, 10% nonsense mutations, 8% small insertions, and 6% large or complex deletions with no preference for a distinct receptor portion. The remaining 2% affect proper mRNA splicing.

Mutations within the *AVPR2* gene can affect receptor function at different levels. At the genomic level, partial or complete gene deletion can interfere with gene transcription. Small genomic alterations can lead to nonsense mutation-mediated mRNA decay and splicing errors (mRNA level). Mutation-induced

structural changes in the receptor protein can interfere with proper receptor folding and trafficking (post-translational level), thus promoting protein degradation in endosomes. Finally, mutations can also abolish receptor signaling by disturbing the ligand-binding domain or G-protein interaction sites (functional level). After a GPCR has been synthesized, the receptor polypeptide adopts a structure that enables the receptor to pass through the endoplasmic reticulum (ER) quality control machinery. More than 80% of all missense mutations lead to incompletely folded receptor proteins that are retained intracellularly by the ER quality control system. Some clinically relevant examples of *AVPR2* mutations show that loss of receptor function is mainly caused by reduced affinity for the natural ligand (e.g., F105V and V290G). The impaired ligand binding is indicated by a concentration-response curve that is shifted toward higher vasopressin concentrations. To clearly differentiate between trafficking defects and abolished agonist binding and/or functional coupling, second messenger and radioligand binding assays and immunologic methods are applied in *in vitro* test systems.

### Mutations in the *AQP2* Gene

In less than 10% of the NDI families studied, congenital NDI has an autosomal recessive mode of inheritance. Approximately two dozen disease-causing mutations have been identified in the *AQP2* gene, which is located in chromosome region 12q13. Sixty-five percent are missense mutations, 23% frameshift mutations caused by small nucleotide deletions or insertions, 8% nonsense mutations, and 4% splice site mutations. Similar to many *AVPR2* mutations, the majority of mutant *AQP2* proteins underlying autosomal recessive NDI were found to be intracellularly trapped and not expressed at the luminal plasma membrane.

Mutations in the *AQP2* gene are also the basis of the autosomal dominant NDI. Aquaporins form homotetramers containing four water channel molecules. It has been demonstrated that a mutation in only one allele of the *AQP2* gene (e.g., E258K) has a dominant negative effect on the wild-type *AQP2* by interfering with proper membrane targeting of the *AQP2* homotetramer.

### Animal Models of NDI

A nonsense mutation known to cause X-linked NDI in humans (E242X) was introduced into the mouse

genome. The mutant male mice died within the first week after birth, apparently due to hypernatremia and dehydration caused by the inability of these animals to concentrate urine. The body weight and the urine osmolality levels of the *AVPR2*-deficient mice were significantly reduced. Interestingly, heterozygous adult female mice showed clear symptoms of NDI, including reduced urine-concentrating ability, polyuria, and polydipsia.

A mouse knock-in model of autosomal NDI (T126M mutation) was generated by a targeted *AQP2* gene replacement strategy. The mutant mice died within 6 days after birth unless given supplemental fluid. Urine/serum analysis showed a urinary concentrating defect with serum hyperosmolality and low urine osmolality that could not be corrected by a *AVPR2* agonist. Immunohistologic and histologic analyses revealed kidney collecting duct dilatation and papillary atrophy.

### CLINICAL SYMPTOMS AND DIAGNOSTIC

The primary symptoms of NDI are polyuria and polydipsia. Inherited NDI may appear in the first weeks of life with symptoms that include fever, irritability, constipation, failure to thrive, lack of appetite, vomiting, and high blood levels of sodium chloride. Due to recurrent electrolyte and fluid imbalances, NDI patients can develop mental retardation. In some cases, secondary symptoms of NDI are observed, such as an enlarged urinary bladder, dilated ureters, or nonobstructive hydronephrosis.

If a patient shows the classic symptoms of DI (polyuria and polydipsia), testing includes measurements of urine osmolality and serum/plasma parameters, such as sodium, potassium, chloride, osmolality, and AVP level. To confirm an NDI diagnosis and to help determine the type of NDI, the patient is given a water-deprivation test. Measurement of the patient's response to a synthetic analog of AVP, 1-desamino-8-D-arginine vasopressin (desmopressin, DDAVP) is also required. If the patient shows almost normal urine concentration abilities in response to DDAVP, a pituitary DI is more likely. It has been reported that *AQP2* is detectable in the urine in both soluble and membrane-bound forms. In normal subjects, an infusion of DDAVP increased the urinary excretion of *AQP2*. In patients with central DI, administration of vasopressin in the same form likewise increased urinary excretion of *AQP2*, but there was no increase in patients with X-linked or autosomal NDI.

The expression of AVPR2 is not only restricted to the kidney but also found in extrarenal tissue, such as endothelium cells. Thus, DDAVP stimulates the release of factor VIII-related antigen from vascular epithelium and factor VIII coagulant activity from liver and other unidentified sites in normal subjects, in addition to increasing plasma renin activity and stimulating the release of von Willebrand factor. DDAVP administration also exerts a vasodilatory action, manifested by facial flushing, a decrease in diastolic blood pressure, and an increase in pulse rate. These effects of DDAVP administration were absent in NDI patients and reduced in obligatory carriers. In autosomal recessive NDI patients with a defect in *AQP2*, the defect is limited to the kidney and the extrarenal effects of DDAVP are normal. Extrarenal parameters may be useful to differentiate clinically between the *AVPR2* and *AQP2* gene defects.

Some patients may have a partial form (partial resistance to AVP) of DI. In these cases, diagnosis is more challenging. A significant number of patients who have either the acquired or the inherited forms of NDI can concentrate their urine during standard fluid deprivation and/or DDAVP tests because they have only partial resistance to the antidiuretic effects of vasopressin.

## TREATMENT

There is no causal cure for inherited NDI. Affected males should be treated immediately with abundant water intake day and night, a low-sodium diet, and thiazide diuretics. Temperature, urine volume, water intake, appetite, and growth should be monitored closely during the first months. The urinary output is decreased by only 30%, and a normal growth curve is difficult to attain during the first 2 or 3 years of life, despite employing the previously discussed treatments and providing intensive attention.

Thiazide diuretics can reduce polyuria of NDI patients, but they can also deplete the body's potassium stores. To maintain sufficient potassium in the body, combining a thiazide (e.g., hydrochlorothiazide) with a potassium-sparing diuretic (e.g., amiloride) may be more effective than using a thiazide alone. Thiazides can reduce the degree to which the kidney can excrete lithium and should be used with care in

cases of lithium-induced NDI because of a toxic buildup of lithium plasma levels. Amiloride is more widely used in these cases because it inhibits the accumulation of lithium. Sometimes, thiazides are used in combination with indomethacin. Indomethacin inhibits the synthesis of prostaglandin in the kidney, which leads to a reduction of polyuria in cases of both inherited and acquired NDI. Low-sodium diets (300–500 mg sodium per day) are recommended because the improvement of water absorption is negated by a diet heavy in salt.

There are some experimental approaches to rescue the function of mutant *AQP2* and *AVPR2* molecules. It has been demonstrated that small molecules such as glycerol (so-called chemical chaperones) can assist mutant *AQP2s* to exit from the ER and ultimately be transported to the plasma membrane. Similarly, non-peptide antagonists have been successfully used to improve plasma membrane trafficking and function of intracellularly retained AVPRs harboring missense mutations. Based on findings that all *GPCRs* are likely composed of multiple folding units, it was demonstrated that truncated *AVPR2s* can be functionally reconstituted by coexpressing the missing receptor fragments. In the future, these new approaches may provide alternatives for restoring kidney concentration abilities. The therapeutic feasibility of these potential strategies needs to be tested in mouse models for NDI.

## See Also the Following Articles

Diabetes Insipidus, Neurogenic • Diabetes, Type 1 • Diabetes, Type 2 • Kidney Disease in Diabetes

## Further Reading

- Bichet, D. G. (1998). Nephrogenic diabetes insipidus. *Am. J. Med.* **105**, 431–442.
- Morello, J. P., and Bichet, D. G. (2001). Nephrogenic diabetes insipidus. *Annu. Rev. Physiol.* **63**, 607–630.
- Yang, B., Gillespie, A., Carlson, E. J., Epstein, C. J., and Verkman, A. S. (2001). Neonatal mortality in an aquaporin-2 knock-in mouse model of recessive nephrogenic diabetes insipidus. *J. Biol. Chem.* **276**, 2775–2779.
- Yun, J., Schöneberg, T., Liu, J., Schulz, A., Ecelbarger, C. A., Promeneur, D., Nielsen, S., Sheng, H., Grinberg, A., Deng, C., and Wess, J. (2000). Generation of a mouse model for X-linked nephrogenic diabetes insipidus. *J. Clin. Invest.* **106**, 1361–1371.

components. The AVP is then transported within these vesicles down the neuronal axon to the posterior pituitary, where it awaits an excitatory stimulus for release into the circulatory system.

### Secretion

In healthy adults, AVP secretion is regulated primarily by the effective osmotic pressure of extracellular fluid. This pressure is determined by the concentration of solutes that do not readily cross cell membranes, and it results largely from the concentration of serum sodium and its associated anions. Total osmolarity may be determined by direct plasma measurement or estimated by doubling the serum sodium, with minor additions for blood urea nitrogen (BUN) and serum glucose (Glu), as follows:

$$\text{Calculated serum osmolarity} = 2(\text{Na}) + \text{BUN}/2.8 + \text{Glu}/18$$

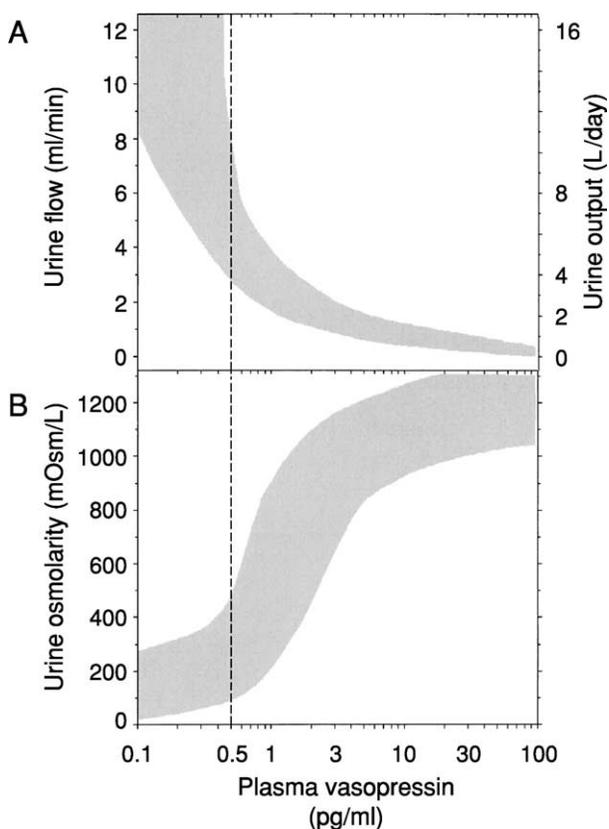
The influence of osmolarity on AVP secretion is mediated by specialized sensory cells known as osmoreceptors that are thought to be located in the front of the hypothalamus near the supraoptic nucleus. They are capable of detecting minute changes in the plasma concentration of sodium and certain other solutes. The level at which AVP secretion begins is commonly known as the osmotic threshold or setpoint. Below this level of osmolarity, AVP secretion is suppressed to low or undetectable levels. Above it, AVP secretion rises steeply in direct proportion to plasma osmolarity. In healthy adults, the osmotic setpoint for AVP release averages approximately 280 mOsm/liter, corresponding to a serum sodium concentration of approximately 135 mEq/liter, although it may vary appreciably due to genetic influences, pregnancy, and even menstruation. The sensitivity of osmoregulation also varies widely between individuals.

In addition to osmoregulation, nonosmotic stimuli may play a major role in AVP release under pathologic conditions. Major changes in blood pressure or volume may trigger specialized sensors known as baroreceptors, which are located in the atria, aorta, and carotid vessels. This pathway undoubtedly plays an important role under conditions such as hemorrhage or dehydration, but it is not believed to play a major role under normal conditions since major changes on the order of 10–20% are necessary to appreciably increase AVP release. Other nonosmotic stimuli, such as nausea, cortisol deficiency, and low blood sugar, may also effect an increase in AVP release. Notably, nausea is one of the most powerful

stimuli, resulting in a 50- to 100-fold increase in basal AVP levels.

### Action

In humans, the primary biologic action of plasma AVP is to fine-tune the urinary excretion of water by selectively increasing its reabsorption from the reduced volume of glomerular filtrate that reaches the collecting ducts. This antidiuretic effect is mediated by binding of AVP to receptors located on cells lining the collecting duct, thereby increasing the number of water channels in these cells. The net result is an increase in water permeability, allowing water to passively diffuse from the tubular fluid into the hyper-tonic milieu of the renal medulla. Resultant urine excretion is marked by low volume and high concentration. Urine concentration, measured as osmolarity or specific gravity, varies as a positive exponential function of plasma AVP concentration (Fig. 1B) and has a curvilinear inverse relationship to urine flow (Fig. 1A). This relationship may be altered in various



**Figure 1** Typical response in normal individuals: (A) inverse curvilinear relation of urine volume and flow rate to plasma vasopressin levels and (B) direct exponential response of urine concentration (expressed in osmolarity) to plasma vasopressin levels.

ways by disorders of the kidney or variations in the amount of sodium, chloride, urea, or other solute necessary for excretion.

### Thirst and Satiation

In humans, the regulation of water intake is as important as the control of urine output for maintaining normal water balance. Thirst, defined as a consciously perceived desire for water, is the principal mechanism for ensuring that water intake is always sufficient to replenish losses. It is regulated by hypothalamic osmoreceptors in much the same way as AVP. The only difference seems to be that the osmotic threshold for thirst is “set” slightly higher (5–15 mOsm/liter) than that for AVP release. Thus, in a healthy adult, thirst and polydipsia usually do not begin until plasma osmolarity rises to a level at which AVP secretion is sufficient to produce a maximum antidiuresis. This arrangement ensures that the antidiuretic mechanism will always be utilized initially in the defense against hypertonic dehydration, with provision for an essential backup protection against water losses that threaten to overwhelm the other compensatory mechanisms. Some kind of satiety mechanism for inhibiting water intake may also exist in healthy adults since fixing antidiuresis near the maximum for weeks does not result in overhydration because basal, *ad libitum* fluid intake decreases in proportion to urine output. Thus, the mechanisms for regulating fluid intake not only provide a necessary function but also are able to do so with admirable precision in maintaining a normal water balance under various adverse conditions.

### CAUSES

Neurogenic diabetes insipidus (DI) is considered an uncommon disorder, although there are no reliable data regarding incidence or prevalence, and diagnosis may frequently be missed. Approximately 50% of cases can be attributed to destruction of the neurohypophysis by an identifiable genetic, congenital, or acquired disease, including trauma, neoplastic infiltration from either primary or metastatic disease, granulomatous disease, and infection (Table I). The remainder are classified as idiopathic. Of the genetic forms, the most common is a completely penetrant, autosomal dominant form caused by mutations of the AVP gene.

### PATHOPHYSIOLOGY

Typically, irreversible destruction of 80–85% of the neurohypophysis is necessary to elicit neurogenic DI.

**Table I Causes of Neurogenic Diabetes Insipidus**

Acquired
Head trauma (closed and penetrating)
Neoplasms
Primary
Craniopharyngioma, adenoma, dysgerminoma, meningioma
Metastatic (lung, breast)
Hematologic (lymphoma, leukemia)
Granulomas
Sarcoidosis, histiocytosis
Infectious
Chronic meningitis, viral encephalitis, toxoplasmosis
Inflammatory
Lymphocytic infundibuloneurohypophysitis, Wegener's granulomatosis, lupus erythematosus, scleroderma
Vascular
Sheehan's syndrome, aneurysm (internal carotid), infarction
Pregnancy (associated with vasopressinase)
Idiopathic
Congenital malformations (present at birth, regardless of causation)
Septo-optic dysplasia
Midline craniofacial defects
Holoprosencephaly
Hypogenesis, ectopia of pituitary
Genetic
Autosomal dominant
Autosomal recessive
X-linked recessive

In adults with normal thirst and access to water, overt signs of dehydration are absent since the polyuria produces only a 1 to 2% decrease in total body water prior to commensurate increases in thirst and fluid consumption. Only when thirst is deficient and/or access to water is limited are patients unable to compensate for the urinary loss, and they can develop significant elevations in serum sodium (hypernatremia) and hyperosmolarity. The AVP deficiency can vary in severity from patient to patient, resulting in large individual differences in the urinary response to fluid deprivation. Many patients with partial AVP deficiency retain the ability to concentrate urine above 300 mOsm/liter.

### DIAGNOSIS

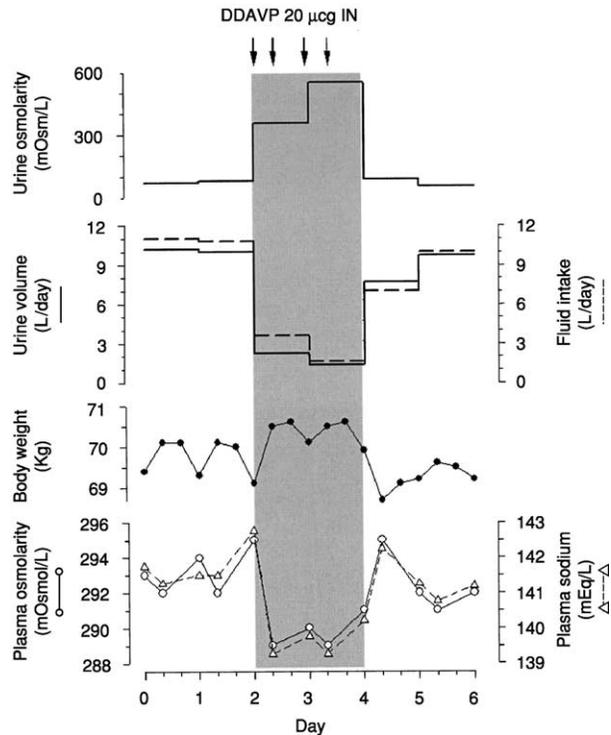
The initial step in the evaluation is to establish the presence of DI by obtaining a 24-h urine collection to document that the volume is abnormally high (typically exceeding 45 ml/kg per 24 hours) and the concentration or osmolarity is abnormally low (typically less

than 300 mOsm/liter). A urine creatinine should be obtained simultaneously to confirm an adequate collection.

Once a diagnosis of DI is established, the next step is to determine the type of DI. This is essential because treatment strategies vary markedly and, if misapplied, can have dire consequences. Measurements of basal plasma osmolarity or sodium are rarely, if ever, helpful because they are almost always normal and vary over the same range in all types of DI. Therefore, except in the rare patient with hypertonic dehydration (i.e., hypernatremia) under basal conditions, the first step in the differential diagnosis is to determine whether the patient can concentrate urine in response to fluid restriction. Close monitoring is necessary, with serial determinations of body weight, fluid intake and output, and periodic measurements of plasma and urine for both osmolarity and sodium, in order to define the patient's response. Failure to concentrate urine above 300 mOsm/liter prior to either a 5% decrease in body weight or an increase in plasma osmolarity or sodium above the upper limit of normal excludes partial forms of DI and indicates a severe form of nephrogenic or neurogenic DI. In this instance, further differentiation between nephrogenic and neurogenic DI can be accomplished by determining the response in urinary concentration to antidiuretic hormone by injection of desmopressin acetate in a typical treatment dose. A precipitous decline in urine volume accompanied by an increase in excess of 50% in urinary osmolarity following administration of desmopressin indicates severe neurogenic DI (Fig. 2). However, an absent or attenuated response indicates nephrogenic DI.

If the patient concentrates urine in response to fluid deprivation, severe forms of nephrogenic or neurogenic DI are excluded, but the patient may still have primary polydipsia or partial forms of nephrogenic or neurogenic DI. Further differentiation among these disorders may be performed by measuring plasma AVP when plasma osmolarity and/or sodium are above the normal range (Fig. 3). This requisite of plasma hypertonicity may be difficult to achieve rapidly by fluid deprivation alone in this subset of patients who retain some degree of urinary concentrating ability, and cautious intravenous infusion of hypertonic (3%) saline may be required.

Magnetic resonance imaging (MRI) of the brain may also be helpful in the differential diagnosis of DI as well as in identifying the pathology responsible for the neurogenic and dipsogenic forms. On T1-weighted midsagittal images, the posterior pituitary emits a hyperintense signal (or bright spot) in 85–90%

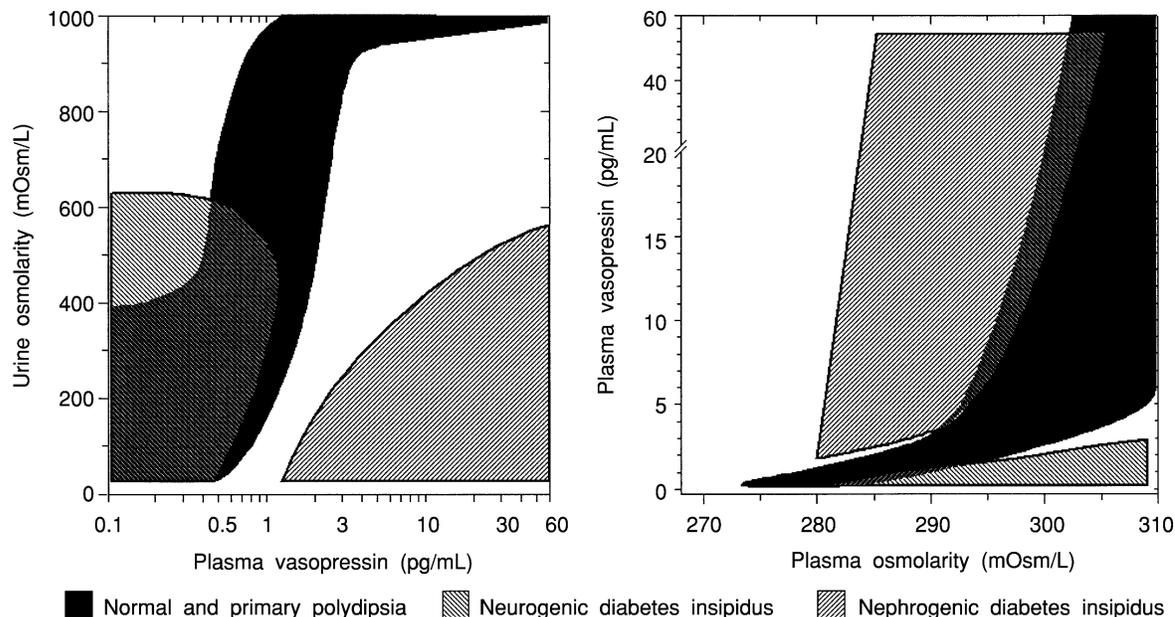


**Figure 2** Characteristic pattern of selected variables observed in individuals with severe neurogenic DI under basal conditions and in response to desmopressin (DDAVP) administration. Note that treatment primarily restores urinary concentration and rapidly reduces thirst and fluid consumption, with normalization of urine volume.

of healthy individuals. This bright spot is almost always present in patients with primary polydipsia but is invariably absent or small in patients with neurogenic DI. However, it is also usually absent in cases of nephrogenic DI. Therefore, a normal posterior pituitary bright spot excludes neurogenic DI, mitigates against nephrogenic DI, and strongly favors dipsogenic DI. The absence of the bright spot is less meaningful but is a strong indication against primary polydipsia and favors the diagnosis of neurogenic or nephrogenic DI.

## MANAGEMENT

There are two major aspects in the treatment of neurogenic DI. The first is the elimination of the polyuria, thirst, and polydipsia by treating with desmopressin acetate, a synthetic analogue of the natural hormone that is more resistant to degradation and devoid of smooth muscle effects. It may be given orally (p.o.), intranasally (i.n.), or parenterally [subcutaneously (s.q.) or intravenously (i.v.)], with relative doses between the different routes approximating



**Figure 3** Relationship of plasma vasopressin to urine osmolarity (left) and plasma osmolarity (right). The wide range of values for both plasma osmolarity and AVP levels was generated by fluid deprivation, with hypertonic saline administration as needed. The shaded areas illustrating neurogenic and nephrogenic DI represent both complete and partial forms of each syndrome.

a logarithmic scale. Typical replacement doses for an adult are 50–200  $\mu\text{g}$  p.o., 5–20  $\mu\text{g}$  i.n., or 1 or 2  $\mu\text{g}$  i.v. or s.q. given two or three times per day as needed to maintain the antidiuretic effect. Individual requirements vary widely and must be determined empirically by assessing relief of symptoms and changes in 24-h urine volume.

If the patient with neurogenic DI has a normal thirst mechanism and drinks only when thirsty, administration of desmopressin in doses sufficient to completely normalize urine output will not result in water intoxication because spontaneous fluid intake decreases in exact proportion to urine output and insensible loss. However, if the patient has primary polydipsia or an associated hyperdipsia, the same treatment will rapidly produce water intoxication because fluid intake is not reduced as much as urine output and the excess cannot be excreted due to the antidiuretic effect of desmopressin. This can have a variety of consequences, ranging from mild asymptomatic hyponatremia to mental status changes, seizure, coma, and even death in extreme circumstances. Therefore, in the first 3–7 days following the initiation of desmopressin or a change in dose, patients should be monitored to verify that their thirst and polydipsia cease and serum sodium remains within the normal range. Furthermore, they should be counseled regarding the symptoms and risk of

water intoxication and the paramount importance of drinking only enough to satisfy thirst.

An alternative to desmopressin in the treatment of neurogenic DI is chlorpropamide. This drug, classified as a first-generation sulfonylurea, has typically been used for treatment of diabetes mellitus, although it has been found to be effective in controlling or at least ameliorating signs and symptoms of neurogenic DI. The mechanism of its antidiuretic effect is uncertain, but it is thought to be due to either potentiation of existing plasma AVP or direct receptor activation. Regardless, for the patient with neurogenic DI, the net effect is similar to that achieved with desmopressin, although antidiuresis is typically slower in onset and its maximal effect is less.

The second aspect of treatment is the search for the underlying cause of neurogenic DI. At a minimum, this involves obtaining an MRI of the brain to search for a space-occupying lesion, such as a neoplastic, infectious, or granulomatous process, that might be treated medically or surgically. If the MRI does not show a likely cause and/or there is a history of childhood onset or other affected family members, the possibility of a genetic cause should be investigated by direct genetic testing. It should be remembered, however, that despite a thorough investigation, no cause is found in up to 50% of cases, resulting in a diagnosis of idiopathic neurogenic DI.

### See Also the Following Articles

Diabetes Insipidus, Nephrogenic • Diabetes, Type 1 • Diabetes, Type 2 • Kidney Disease in Diabetes • Obesity and Diabetes, Regulation of Food Intake • Thirst and Dehydration in the Elderly

### Further Reading

Bayliss, P. (2001). In "Endocrinology" (L. J. DeGroot and J. L. Jameson, eds.), 4th ed., pp. 363–376. Saunders, Philadelphia.

Drincic, A., and Robertson, G. L. (1999). Treatment of diabetes insipidus in adults. In "Hormone Replacement Therapy for Contemporary Endocrinology" (A. W. Mickle, ed.), pp. 21–38. Humana Press, Totowa, NJ.

Robertson, G. L. (1995). Diabetes insipidus. *Endocrinol. Metab. Clin. North Am.* **24**(3), 549–572.

Robertson, G. L. (2001). Disorders of the neurohypophysis. In "Harrison's Principles of Internal Medicine" (Braunwald, Fauci, Kasper, Hauser, Longo, and J. L. Jameson, eds.), 15th ed., pp. 2052–2060. McGraw-Hill, New York.

Robertson, G. L. (2001). Antidiuretic hormone: Normal and disordered function. *Endocrinol. Metab. Clin. North Am.* **30**(3), 671–694.

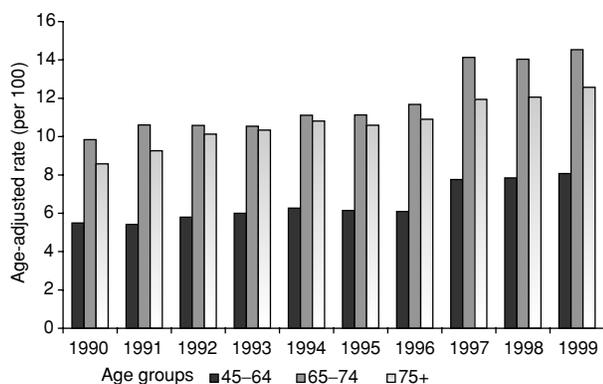
## PREVALENCE OF DIABETES MELLITUS

The prevalence of diabetes is increasing among the U.S. population across age groups. According to data from the Centers for Disease Control (CDC), diabetes affects 6.2% of the population, which represents approximately 17 million people. In addition, undiagnosed diabetes occurs in more than one-third of this population group, or approximately 5.9 million people. Estimates of the prevalence of the disease among older adults in 2000 indicated that 7 million people, or 20% of all people older than age 65, had diabetes. The prevalence of diabetes was 22 and 19% among elders in the age categories 65–74 and 75+, respectively.

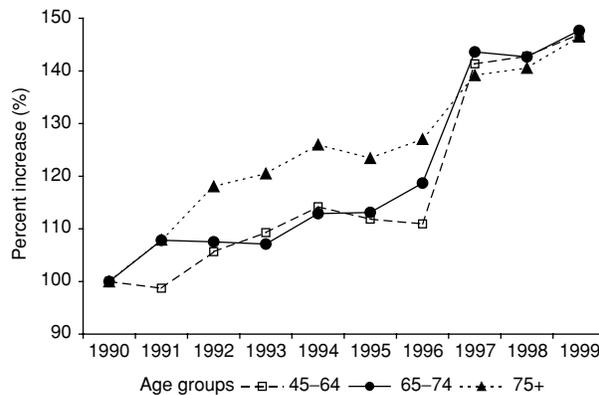
Analyses of annual trends of the prevalence of diabetes among the elderly indicate that in addition to being high, it is also on the rise. In 1990, the prevalence of diagnosed diabetes among the older adult population (aged 45+) ranged from 5.5% among those 45–64 years of age to approximately 10% among people aged 65–74 (Fig. 1). By 1999, the prevalence of diabetes was 8.1 and 14.5% for the same age groups, respectively, with increases from 8.6 to 12.6% among people 75 years of age or older.

Between 1990 and 1999, the largest increase in the prevalence of diabetes was observed among people 65–74 years of age, for whom diagnosed diabetes increased 48% (from 10 to 15 per 100 people) (Fig. 2). Increases of 47 and 46.5% were also observed for older adults in the 45–64 and 75+ age groups, respectively.

People with diabetes are approximately two times more likely to die than those free of the condition, with the most frequent cause of death being cardiovascular disease. In 1999, diabetes was the sixth leading cause of death based on CDC data. Furthermore,



**Figure 1** Age-adjusted prevalence of diabetes in the U.S. older adult population (45 years or older) by age group, per 100 population, 1990–1999.



**Figure 2** Increases in the prevalence of diabetes in the U.S. older adult population (45 years or older) by age group, 1990–1999.

diabetes imposes a major economic burden on the health sector of most countries. According to estimates from the American Diabetes Association (ADA), in 1997 the total costs in the United States associated with diabetes were \$98 billion, of which 45% represented direct medical costs and the remaining 55% (\$54 billion) was attributed to indirect costs, including disability, work loss, and premature mortality due to the disease.

## RISK FACTORS ASSOCIATED WITH DIABETES

There is uncertainty about the specific factors that trigger the development of type 2 diabetes. However, it is known that certain high-risk groups of people are more likely to develop the condition than other groups. Table I summarizes some of the main risk factors for diabetes. Age is an important risk factor, with a positive association beginning after age 30. As previously noted, more than 20% of the U.S. population aged 65 or older has diabetes. It has been reported that Hispanics, African Americans, Native Americans, Pacific Islanders, and other ethnic and racial groups are at a higher risk for diabetes than the general population. Also, evidence shows that approximately 80% of adults with type 2 diabetes are overweight and physically inactive. Those with a family history of diabetes are also at increased risk for the disease. Women who have had gestational diabetes or who have given birth to a baby weighing more than 9 pounds are also at higher risk for developing diabetes later in their lives. Furthermore, there is elevated risk for diabetes among women with polycystic ovary syndrome who are also overweight. The risk is also high for those known to have impaired glucose tolerance (IGT) or impaired

**Table I Risk Factors for Type 2 Diabetes**

Older than 45 years of age
Obesity
Family history of diabetes
Ethnicity
History of gestational diabetes
Hypertension
Dyslipidemia
Impaired fasting glucose
Impaired glucose tolerance

fasting glycemia (IFG) as well as for those who have high blood pressure or signs of heart disease or poor circulation.

### SYMPTOMS OF DIABETES IN THE ELDERLY

Type 2 diabetes has a gradual onset, with signs of the disease developing over years. The classic symptoms, listed in [Table II](#), are usually subdued or even absent in older individuals, and the disease is often diagnosed several years after its onset, when complications are already present. Regular screening for the disease after age 45 is usually the best way to detect early type 2 diabetes and avoid complications associated with the disease.

### DIAGNOSIS

The current criteria for diagnosis of diabetes mellitus are based on the 1997 recommendations of an international expert committee from the ADA and the World Health Organization. Based on these recommendations, and according to the 2003 guidelines from the ADA, three diagnosis criteria may be used, with confirmation on a subsequent day. [Table III](#) lists these criteria.

Hyperglycemia between normal and diabetes levels is classified as prediabetes and includes two categories: IFG, which is present with fasting plasma glucose

**Table II Symptoms of Type 2 Diabetes**

Polyuria (frequent urination)
Polydipsia (excessive thirst)
Weight loss
Severe fatigue and/or nausea
Blurred vision
Frequent and recurrent infections
Slow healing of wounds or sores

**Table III Diagnosis of Diabetes Mellitus<sup>a</sup>**

Parameter	Criteria for diagnosis
Fasting plasma glucose	7.0 mmol/liter (126 mg/dl)
Random plasma glucose <sup>b</sup> and symptoms of diabetes	11.1 mmol/liter (200 mg/dl) (see <a href="#">Table II</a> )
Oral glucose tolerance test	11.1 mmol/liter (200 mg/dl) 2 h after a glucose load 75 g

<sup>a</sup>According to the American Diabetes Association (2003a).

<sup>b</sup>Random or casual refers to any time, independent of meal times.

between 6.1 and 6.9 mmol/liter (110–126 mg/dl), and IGT, which is diagnosed with an oral glucose tolerance test that results in 2-h plasma of 7.8–11.0 mmol/liter (140–199 mg/dl).

Detection of diabetes among older individuals generally occurs late. It is estimated that onset of type 2 diabetes precedes clinical diagnosis by 10–12 years. It has been reported that among people with undiagnosed type 2 diabetes, 10–29% have retinopathy, 10–37% have proteinuria, and 9% have neuropathy at the time of diagnosis. These complications associated with uncontrolled diabetes take 10 or more years to develop in the presence of the disease.

### COMPLICATIONS OF TYPE 2 DIABETES

Since the metabolic problems associated with diabetes affect the entire body, elderly people with this condition face multiple potential complications affecting many different organ systems. If the management of the disease is poorly controlled, if the disease is left untreated, or if the diagnosis is made late, serious complications can occur. Moreover, poor glyce-mic control among people with diabetes is a major risk factor for end-stage complications of this condition.

Cardiovascular disease (e.g., coronary heart disease and stroke) is the leading cause of morbidity and mortality among persons with diabetes. Cardiovascular disease is two to four times more common in adults with diabetes. The rate of stroke is 2.5 times higher in this population, and high blood pressure affects 60–65% of those with diabetes.

Diabetes is the leading cause of blindness and visual impairment in adults. Nerve damage combined with peripheral vascular disease make diabetes the most common cause of lower extremity amputation. Between 60 and 70% of persons with diabetes have mild to severe forms of diabetic nerve damage, and more than half of lower limb amputations in the United States occur among diabetics.

There is conclusive evidence that good control of blood glucose levels can substantially reduce the risk of developing complications and slow their progression in all types of diabetes. Results from the Diabetes Control and Complications Trial Research on type 1 diabetes, the U.K. Prospective Diabetes Study on type 2 diabetes, and the Kumamoto study, also on type 2 diabetes, confirm that better glycemic control is associated with reductions in both the incidence and the progression of retinopathy and nephropathy in subjects with type 1 and type 2 diabetes mellitus. Available evidence also links macrovascular disease (cardiovascular, peripheral vascular, and cerebrovascular disease) to glycemic status. Furthermore, subjects with good long-term glycemic control have better survival rates than those with average fasting blood glucose.

## CLINICAL MANAGEMENT OF DIABETES IN THE ELDERLY

Initial clinical management of diabetes in the elderly usually consists of medical nutrition therapy and increases in physical activity. However, few elderly people can adhere strictly to the required changes in diet and physical activity, and controlling hyperglycemia becomes a challenge. Therefore, in the vast majority of cases, drug therapy is needed soon after the disease is detected. Currently, drug options include second-generation oral sulfonylureas, metformin, troglitazone, acarbose, insulin, insulin analogs, and combinations of these drugs. These drugs have demonstrated efficacy in increasing insulin availability, decreasing insulin requirements, or both. Careful planning and adherence to the clinical management designed for the elderly with diabetes can provide good glycemic control. Self-monitoring of blood glucose by the elder diabetic, along with clinical diabetes monitoring of hemoglobin A1c, which informs the medical team about glucose control during the past 3 or 4 months, are good parameters for monitoring the disease. A hemoglobin A1c level lower than 7% is usually the goal for achieving good control.

Evidence is lacking about the benefits of tight glycemic control in older adults (older than 65 years of age). However, elderly diabetic patients with a prognosis for extended life and who are able to manage their disease should have the same goals for glycemic control as younger adults (Table IV). For elders with advanced diabetes, disease complications, advanced cognitive or physical impairment, or comorbidities that may reduce their life expectancy, less intensive goals can be pursued. In any case, it is advisable to

Table IV Goals for Glycemic Control of Adults<sup>a</sup>

Goal	Criteria
<b>For glycemic control</b>	
Hemoglobin A1c	<7%
Preprandial plasma glucose	5.0–7.2 mmol/liter (90–130 mg/dl)
Postprandial plasma glucose	<10.0 mmol/liter (180 mg/dl)
<b>For cardiovascular health control</b>	
<b>Blood pressure</b>	Systolic <130 mmHg and/or diastolic <80 mmHg
<b>Lipid profile</b>	
High-density lipoprotein cholesterol	>1.1 mmol/liter (40 mg/dl)
Low-density lipoprotein cholesterol	<2.6 mmol/liter (100 mg/dl)
Triglycerides	<1.7 mmol/liter (150 mg/dl)

<sup>a</sup>Based on the American Diabetes Association (2003b).

individualize the clinical management of diabetes among elderly individuals.

Proper clinical and self-management of elderly people with diabetes is often difficult and costly due to circumstances associated with aging and comorbidities. Physical ailments, affective or cognitive disorders, psychosocial problems such as depression, functional impairment, and living situation can all interfere with proper management of diabetes in the elderly and potentially exacerbate problems associated with noncompliance with medications, poor eating, the absence of glucose monitoring, and lack of adequate physical activity.

Furthermore, the impact of cardiovascular risk factors is significant among people with diabetes. Therefore, the management of diabetes requires control of cardiovascular risk factors, including hyperglycemia, hypertension, and dyslipidemia. Diabetes needs to be viewed and treated as part of the metabolic syndrome, keeping in mind that the main goals of diabetes management, whenever the circumstances apply, are tight glycemic control and more aggressive management of risk factors associated with cardiovascular disease.

## CONCLUSION

The rapidly increasing rates of diabetes among older adults and the tremendous social and economic impact of the disease require that this disease be treated as a high-priority public health problem. Prevention of the disease among the growing elderly population is the best strategy to reduce the burden of diabetes. Results from the Diabetes Prevention Trial showed that lifestyle strategies (healthy eating

and increased physical activity) are more effective than drugs for preventing the development of diabetes, with a 58% reduction in the incidence rate of diabetes among people at high risk. These lifestyle changes were particularly successful for people older than age 60. From this evidence, a strong argument can be made for focusing on the prevention of diabetes in older adults through improvement of the favorable factors associated with diabetes and the metabolic syndrome, particularly lifestyle factors.

### Acknowledgments

This material is based on work partly supported by the U.S. Department of Agriculture under agreement No. 581950-9-001. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author and do not necessarily reflect the views of the U.S. Department of Agriculture.

### See Also the Following Articles

Cardiovascular Disease in Diabetes • Diabetes, Type 1 • Diabetes, Type 2 • Eye Disease in Diabetes • Foot Disease in Diabetes • Hypertension and Diabetes • Hypoglycemia • Lipid Disorders in the Elderly • Obesity and Diabetes, Regulation of Food Intake

### Further Reading

American Diabetes Association (2003a). Report of the Expert Committee on Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care* 26(Suppl. 1), S5–S20.

- American Diabetes Association (2003b). Standards of medical care for patients with diabetes mellitus. *Diabetes Care* 26(Suppl. 1), S33–S50.
- Centers for Disease Control and Prevention (2002, March 27). National diabetes fact sheet: National estimates on diabetes. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Atlanta, GA.
- Diabetes Control and Complications Trial Research Group (1993). The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin dependent diabetes mellitus. *N. Engl. J. Med.* 329, 977–986.
- Diabetes Prevention Program Research Group (2002). The Diabetes Prevention Program (DPP): Description of lifestyle intervention. *Diabetes Care* 25(12), 2165–2171.
- National Institute of Diabetes and Digestive and Kidney Diseases (2003, March 24). *Diabetes Diagnosis*, National Diabetes Information Clearinghouse, NIH Publication No. 00-4642. U.S. Department of Health and Human Services, National Institutes of Health, Bethesda, MD.
- Ohkubo, Y., Kishikawa, H., Araki, E., Miyata, T., Isami, S., Motoyoshi, S., *et al.* (1995). Intensive insulin therapy prevents the progression of diabetic microvascular complications in Japanese patients with non-insulin-dependent diabetes mellitus: A randomized prospective 6-year study. *Diabetes Res. Clin. Practice* 28(2), 103–117.
- Pietropaolo, M., Barinas-Mitchell, E., Pietropaolo, S. L., Kuller, L. H., and Trucco, M. (2000). Evidence of islet cell autoimmunity in elderly patients with type 2 diabetes. *Diabetes* 49(1), 32–38.
- U.K. Prospective Diabetes Study (UKPDS) Group (1998). Intensive blood-glucose control with sulphonylurea or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet* 352, 837–853.
- World Health Organization (1999). *Definition, Diagnosis and Classification of Diabetes Mellitus and Its Complications. Report of a WHO Consultation*. World Health Organization, Department of Noncommunicable Disease Surveillance, Geneva.

## DIAGNOSIS

The diagnosis of diabetes is made by the presence of the classical symptoms of diabetes (polyuria, polydipsia, and unexplained weight loss) and unequivocal hyperglycemia documented on two occasions. Hyperglycemia is defined as a random plasma glucose  $\geq 200$  mg/dl plus symptoms of diabetes, a fasting plasma glucose  $\geq 126$  mg/dl, or a 2-h plasma glucose  $\geq 200$  mg/dl following a 75-g oral glucose challenge in adults (1.75 g/kg in children). Between 10 and 40% of patients with new-onset type 1 diabetes have ketoacidosis at presentation. Unlike patients with type 2 diabetes, type 1 patients are typically lean, have had weight loss, and are frankly symptomatic. When the diagnosis is in doubt, type 1 can be confirmed by the finding of islet-related autoantibodies [islet cell antibodies (ICA), antibodies to glutamic acid decarboxylase (GADA), insulin autoantibodies (IAA), and insulinoma-associated antigen-2 (IA-2)]. ICA and GADA are found in approximately 70% of new-onset cases, and if all four are measured, in excess of 90% of new-onset type 1 diabetes patients are positive.

## MANAGEMENT

In 1921, Frederick Banting and Charles Best discovered what was thought to be the cure for diabetes—insulin. Prior to the discovery of insulin, type 1 diabetes was a slow but sure death sentence for which the only treatment was a diet low in carbohydrates and sugar and high in fats and protein. Of course, insulin is not the cure for diabetes but is, rather, the vital tool for managing this disease and controlling complications. The Diabetes Control and Complications Trial demonstrated a strong relationship between intensive management and good metabolic control and the rate and progression of complications. Intensive therapy, however, is neither workable nor appropriate for all patients. What is evident, however, is that improved overall blood glucose control should be sought at all ages while avoiding severe hypoglycemia, weight gain, or emotional problems for the patient and the family.

Type 1 diabetes management in children (and adults) is complex, requiring a multidisciplinary medical team, including a diabetologist, a nurse who is a certified diabetes educator, a nutritionist, a psychologist, and, if possible, a social worker. The disease should be approached as affecting not just the individual with diabetes but also the entire family. The team works together to help teach family members and involve them in the decision-making process for the

patient. A successful management program must be consistent and flexible. A balance must be sought between the needs of the medical establishment (insulin, nutrition, exercise, blood glucose testing, HbA1c tests, lipids, and complication surveillance) and the lifestyle demands and desires of the patient. Each patient needs individualized therapy based on age, schedule, social environment, and capabilities. A health care team that knows the patient and the family well will be in a better position to make recommendations that work.

Advances in diabetes management, including the development of newer long- and short-acting insulins and improved blood glucose monitoring and delivery systems, have facilitated the daily management of patients with the disease.

The following are the overall goals of diabetes management:

- Set realistic goals for each child and family: Consider the patient's age, family involvement and social situation, economic factors, and history of hypoglycemia.
- Near normalization of blood glucose levels and HbA1c measurements or, when not possible, improvement on subsequent follow-up.
- Prevention of diabetic ketoacidosis.
- Avoidance of severe hypoglycemia.
- Maintenance of normal quality of life.
- Achievement of normal growth, development, and maturation in children.
- Multidisciplinary support, including nutritional education and psychological support.
- Close surveillance and prevention of microvascular, macrovascular, and neuropathic complications.

A variety of insulin regimens are available for use that must be individualized depending on patient compliance, the patient's level of social support, and various other factors, including dietary and exercise regimens. Possible insulin regimens include subcutaneous injections of a basal (long-acting) insulin with bolus insulin therapy before meals, mixed injections of fast- and intermediate-acting insulin several times per day, and the insulin pump. Inhaled insulin, the insulin patch, and buccal/intranasal insulins are under investigation. In addition to improved insulin therapy, glucose monitoring devices have become simpler to use, allowing for blood testing from sites other than the finger that are less painful, and they have also become faster. Minimally invasive near continuous glucose monitoring devices have been applied to the clinical setting.

## GENETICS OF TYPE 1 DIABETES

In the United States, the Caucasian population has a 0.3% risk of developing type 1 diabetes, whereas first-degree relatives of patients with the disease are 15 times more likely to develop the disease. Monozygotic twins have a concordance rate of approximately 50%, whereas dizygotic twins have a reduced concordance rate of 6–10%. The high rate of discordance even in identical twins suggests that environmental factors, not just genetic susceptibility, play a major role in the development of this disease.

Approximately 20 susceptibility gene intervals have been identified for type 1 diabetes. The major susceptibility gene region is the human leukocyte antigen (HLA) region located in the major histocompatibility complex (MHC) on chromosome 6p21, which accounts for up to 65% of genetic susceptibility. The MHC complex spans a 3.5-megabase region of chromosome 6 and is subdivided into three regions (classes I–III). The most important genetic factors in type 1 diabetes are the HLA class II genes. Evidence from animal and human studies has shown that certain loci of class II HLA alleles—HLA-DQA1, HLA-DQB1, and HLA-DRB1—are involved in susceptibility to type 1 diabetes. The majority of patients with type 1 diabetes carry either the HLA-DR3 or HLA-DR4 class II alleles, with approximately 30% being DR3/DR4 heterozygous. In Caucasians, the DR3/DR4 genotype confers the highest risk for development of type 1 diabetes, followed by DR4 and DR3 homozygosity, respectively. Additionally, HLA-DQ alleles have been shown to be a key susceptibility factor in type 1 diabetes. In Caucasians, type 1 diabetes is strongly associated with two combinations of DQA1 and DQB1 alleles: DQA1\*0501–DQB\*0201, which encodes the DQ2 molecule, and DQA1\*0301–DQB1\*0302, which encodes the DQ8 molecule. Individuals who are heterozygous for the DQ2 and DQ8 molecules have the highest risk of developing type 1 diabetes, with approximately 30% of individuals affected with type 1 diabetes being heterozygous for DQ2/DQ8. It has been proposed that the degree of susceptibility/protection from the HLA genotype correlates with specific amino acid positions of the DQB1 gene. Position 57 is thought to play a crucial role in determining susceptibility to type 1 diabetes because the residue at position 57 contributes to the shaping of the antigen presenting pocket of the molecule, suggesting that the development of diabetes may depend on events surrounding antigen presentation. However, class II genes do not explain all of the HLA association with type 1 diabetes. Class I genes

have been postulated to influence susceptibility and clinical aspects of the disease, such as age of onset and the rate of beta cell destruction. Other non-HLA susceptibility genes have been mapped, include the IDDM2 insulin gene VNTR (variable number of tandem repeats) locus on chromosome 11p15.5.

## ENVIRONMENTAL FACTORS

Environmental factors, such as diet, immunizations, and viruses, may potentially trigger the onset of autoimmunity or hasten progression to disease in patients who already have evidence of autoimmune “prediabetes.” A case for early exposure to viruses in the pathogenesis of type 1 diabetes can be made for congenital rubella. Approximately 30% of children with congenital rubella develop autoimmune diabetes. Viruses could trigger an autoimmune response either by directly damaging the pancreatic beta cells or by invoking molecular mimicry. At least 14 different viruses have been reported to be associated with the development of type 1 diabetes in human or animal models. Viruses suspected to induce autoimmunity to the beta cells via molecular mimicry include retrovirus, mumps, rubella, cytomegalovirus, and Epstein-Barr virus, whereas viruses that are suspected to induce autoimmunity via direct cytolytic damage to the beta cells include coxsackie B and other enteroviruses. The enteroviruses have received attention as a possible trigger of the development of diabetes because observational studies have noted an increase in the incidence of enteroviral infections in individuals with type 1 diabetes preceding the development of the disease, and some studies have demonstrated the presence of the viral genome in diabetic patients.

Immunizations have also been implicated in the pathogenesis of type 1 diabetes. An increase in the incidence of type 1 diabetes in Finland was correlated with the initiation of a mandatory immunization program against diphtheria–pertussis–tetanus. There were also reports of an association between vaccination for hemophilus influenza and an increased incidence of type 1 diabetes. However, in 1998 an expert committee from the National Institutes of Health found no evidence to support these assertions.

Of the putative dietary triggers for diabetes, cow's milk has been the most extensively studied. Epidemiologic, ecologic, and immunological studies suggest a possible link between the early introduction of cow's milk into an infant's diet and/or a decreased period of breast-feeding and the subsequent development of type 1 diabetes. However, this issue is debated

and will likely only be settled with the completion of the Trial to Prevent Diabetes in the Genetically At-Risk study, in which the introduction of cow's milk to high-risk relatives is delayed.

## AUTOIMMUNITY

As previously mentioned, islet autoantibodies (ICA, GADA, IAA, and IA-2A) distinguish type 1 diabetes from other types of diabetes. None of the islet cell antibodies have been found to play a role in the pathologic destruction of the beta cells.

Combinations of these autoantibodies have been found to denote significant risk for the subsequent development of diabetes, both in unaffected relatives of individuals with type 1 diabetes and in the general population.

## IDENTIFICATION OF AT-RISK SUBJECTS

There are two approaches to identifying individuals genetically susceptible to the development of type 1 diabetes prior to clinical manifestation of the disease: (i) antibody testing followed by quantification of risk by further antibody, genetic, and metabolic testing and (ii) primary genetic screening of newborns with determination of high-risk HLA genes and ensuing quantification of risk by further autoantibody, genetic, and metabolic testing. Primary antibody testing is based on the finding that 60–80% of patients with new-onset type 1 diabetes have islet autoantibodies. Studies have shown that the presence of a single autoantibody imparts an approximate 10–20% risk of subsequently progressing to disease, whereas combinations of autoantibodies carry a 5-year risk of  $\geq 50\%$ . The presence of metabolic derangements in addition to serologic markers increases the predictability of screening. Persistent loss of first-phase insulin response during an intravenous glucose tolerance test carries a 50–70% 5-year risk in ICA-positive individuals. Changes in glucose tolerance measured during an oral glucose tolerance test generally occur later in the disease process, indicating significant beta cell destruction. Identification of at-risk individuals has led to a multitude of clinical trials in an attempt to prevent the development of this disease.

## PREVENTION TRIALS

The ability to identify individuals who will subsequently develop the disease, together with the tremendous and increasing burden of the disease, has

facilitated studies aimed at the prevention of type 1 diabetes. After several pilot studies, three large-scale, adequately powered trials were commenced in the mid-1990s.

## Diabetes Prevention Trial—Type 1

The objective of the U.S. Diabetes Prevention Trial—Type 1 (DPT-1) was to determine whether antigen-based (insulin) therapy (either parenteral or oral) in at-risk individuals could prevent or delay the onset of type 1 diabetes. Studies of animal models of type 1 diabetes, as well as encouraging pilot studies of prediabetic relatives of probands, provided strong rationale for this trial. In the DPT-1 trial, more than 100,000 first- and second-degree relatives of patients with diabetes underwent genetic, immunologic, and metabolic assessment to quantify their risk. Patients with a projected 5-year risk  $\geq 50\%$  were randomized either to an annual 4- or 5-day intravenous infusion of regular insulin followed by the administration of subcutaneous ultralente insulin twice daily or to a close observation arm. Those with a risk of 25–50% were randomized to oral insulin or placebo. Unfortunately, in neither arm was insulin successful in preventing or delaying the onset of diabetes.

## European Nicotinamide Diabetes Intervention Trial

Nicotinamide, the water-soluble amide of nicotinic acid and an oxygen free radical scavenger, was shown to prevent the development of type 1 diabetes in mouse models and showed encouraging results in small studies of at-risk human subjects. A double-blind study was conducted in several European countries in which either nicotinamide or placebo was administered to islet cell antibody-positive, first-degree relatives of patients with type 1 diabetes. Unfortunately, there was no difference in the incidence of diabetes between the two groups after 5 years of follow-up.

## Trial to Prevent Diabetes in Genetically At-Risk

The Trial to Prevent Diabetes in Genetically At-Risk was designed to evaluate whether avoidance of cow's milk for at least the first 6 months of life could prevent or delay the onset of type 1 diabetes in high-risk infant relatives of type 1 probands. Mothers of infants enrolled on the basis of high-risk HLA genes breast-feed their infants for at least 6 months. When weaning

occurs, the infants are randomized to either a standard cow's milk-based formula or a casein hydrolyzed formula. The study is ongoing in Scandinavia and North America. The creation of a worldwide multicenter network (TrialNet) will facilitate the conduct of multifaceted approaches to prevent the disease.

## SEARCHING FOR A CURE

Advances in both pancreatic and especially islet cell transplantation have led to both enhanced patient and graft survival. Unfortunately, less than 10,000 transplants have been conducted worldwide, and in the United States alone, more than 1 million people have type 1 diabetes. Thus, there is a need to derive an unlimited supply of insulin-producing cells. In addition, patients require lifelong immunosuppression to prevent rejection (alloimmunity) and possibly prevent recurrence (autoimmunity). Potential sources include deriving beta cells from stem cells (pancreatic ducts, bone marrow, liver, cord blood, and embryonic stem cells) and gene therapy approaches. Encouraging results have been obtained in several animal models, although no gene therapy studies have been conducted in humans with type 1 diabetes. It is hoped that the successes in the laboratory and animal models will be emulated in humans.

## CONCLUSION

The incidence of type 1 diabetes continues to increase, as does the amount of health care dollars spent on the day-to-day management and long-term complications of this disease. Despite therapeutic advances, the burden to patients in terms of cost to the individual and cost to society remains high. The quest for prevention and cure remains the priority.

## Acknowledgments

This work was supported by Grant 2MO1RR0082 from the National Institutes of Health.

## See Also the Following Articles

Carbohydrate Metabolism: Diabetes Mellitus, Genomic Aberrations • Cardiovascular Disease in Diabetes • Diabetes Mellitus, Diagnosis and Treatment in the Elderly • Diabetes, Type 2 • Hypertension and Diabetes • Hypoglycemia • Obesity and Diabetes, Regulation of Food Intake

## Further Reading

- Akerblom, H. K., Vaarala, O., Hyoty, H., Ilonen, J., and Knip, M. (2002). Environmental factors in the etiology of type 1 diabetes. *Am. J. Med. Genet.* **115**(1), 18–29.
- Chaudhari, M., and Schatz, D. A. (2001). Management of diabetes mellitus in children and adolescents. In "Conn's Textbook of Therapy" (R. E. Rakel and E. T. Bope, eds.), pp. 580–586. Sanders, Philadelphia.
- Diabetes Prevention Trial—Type 1 Diabetes Study Group (2002). Effects of insulin in relatives of patients with type 1 diabetes mellitus. *N. Engl. J. Med.* **346**(22), 1685–1691.
- Emerich, D. F. (2002). Islet transplantation for diabetes: Current status and future prospects. *Expert Opin. Biol. Ther.* **2**(8), 793–803.
- Expert Committee on the Diagnosis and Classification of Diabetes Mellitus (2002). Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care* **25**(Suppl.), S5–S20.
- Morales, A., She, J.-X., and Schatz, D. A. (2001). Prediction and prevention of type 1 diabetes. *Curr. Diabetes Rep.* **1**, 28–32.
- Morales, A., She, J.-X., and Schatz, D. A. (2003). Genetics of type 1 diabetes. In "Pediatric Endocrinology: Mechanisms, Manifestations, and Management" (O. H. Pescovitz and E. Eugster, eds.). Lippincott Williams & Wilkins, Baltimore.
- Petrovsky, N., and Schatz, D. A. (2002). Immunology of insulin-dependent diabetes mellitus. In "Textbook of Diabetes" (G. Williams and J. Pickup, eds.), pp. 18.1–18.14. Blackpool, Liverpool, UK.
- White, N. H., Cleary, P. A., Dahms, W., Goldstein, D., Malone, J., and Tamborlane, W. V. (2001). Diabetes Control and Complications Trial (DCCT), Epidemiology of Diabetes Interventions and Complications (EDIC) Research Group. Beneficial effects of intensive therapy of diabetes during adolescence: Outcomes after the conclusion of the Diabetes Control and Complications Trial (DCCT). *J. Pediatr.* **139**(6), 804–812.
- Winter, W. E., and Schatz, D. A. (2003). Prevention strategies for type 1 diabetes mellitus: Current status and future directions. *BioDrugs* **17**, 39–64.
- Winter, W. E., Harris, N., and Schatz, D. A. (2002). Type 1 diabetes islet autoantibody markers. *Diabetes Tech. Ther.* **4**(6), 817–839.

Table 1 1997 ADA Diagnostic Criteria for Dysglycemic Disorders

Criteria for diabetes mellitus	Criteria for prediabetic states (i.e., impaired fasting glucose and impaired glucose tolerance)
Symptoms of hyperglycemia with random plasma glucose $\geq 200$ mg/dl Fasting plasma glucose $\geq 126$ mg/dl on at least two occasions 2 h plasma glucose $\geq 200$ mg/dl after 75 g oral glucose challenge	Impaired fasting glucose: Fasting plasma glucose $\geq 110$ mg/dl and $< 126$ mg/dl Impaired glucose tolerance: 2 h glucose $\geq 140$ mg/dl but $< 200$ mg/dl after 75 g oral glucose challenge

deficiency secondary to autoimmune destruction of beta cells or genetic defects.

According to the 1997 ADA proposal, several criteria may be used independently to establish the diagnosis of diabetes: (1) a FPG  $\geq 126$  mg/dl on more than one occasion; (2) a random plasma glucose of 200 mg/dl or more with typical symptoms of diabetes; or (3) a 2 h value  $\geq 200$  mg/dl during a 75 g OGTT.

The recommendations of the Expert Committee of the ADA, however, discouraged the routine use of the OGTT in clinical practice. A new entity, impaired fasting glucose (IFG), was created and defined as a FPG between 110 and 125 mg/dl as a more sensitive indicator of impaired glucose homeostasis based on the fasting glucose. The diagnosis of IGT remained essentially unchanged and was defined as a FPG of  $< 126$  mg/dl with a 2 h postchallenge blood glucose between 140 and 199 mg/dl during an OGTT.

In 1999, the WHO revised their previous (1985) diagnostic criteria for the diagnosis and classification of diabetes mellitus and took into account the 1997 ADA criteria. The principal difference between the 1997 ADA and the 1999 WHO criteria is that the ADA discourages the use of the OGTT as a routine diagnostic tool but the WHO does not.

## EPIDEMIOLOGY OF TYPE 2 DIABETES MELLITUS

The data from the U.S. Third National Health and Nutrition Examination Survey (NHANES III), conducted between 1988 and 1994, involved 18,825 subjects  $\geq 20$  years of age and provided a determination of prevalence for T2DM, IFG, and IGT. Using only the FPG of  $\geq 126$  mg/dl as the diagnostic criterion for T2DM (i.e., the 1997 ADA criteria), diabetes (diagnosed and undiagnosed) affected approximately 7.8% of adults  $\geq 20$  years of age in the United States. Rates increased with age, reaching 18.8% in individuals  $\geq 60$  years old, and U.S. minorities were more frequently affected by diabetes. A higher prevalence of undiagnosed T2DM was found using the 2 h OGTT

criterion since 48% of individuals with a post-challenge glucose  $\geq 200$  mg/dl have FPG levels  $< 126$  mg/dl. IFG was found in 6.9% of those  $\geq 20$  years of age, whereas IGT was diagnosable in 15.6% of adults between 40 and 74 years of age. The lower prevalence of IFG compared with IGT was due to the fact that 70% of individuals with 2 h plasma glucose 140–199 mg/dl have a FPG  $< 110$  mg/dl.

Irrespective of diagnostic criteria, prevalence rates for T2DM have been increasing in developed countries over the past 3–4 decades and there has been an alarming increase in developing countries over the past 10 years. Thus, the problem of T2DM, which is already a major cause of patient suffering, mortality, and health care costs, is expected to worsen as a worldwide public health burden.

## GENETICS OF TYPE 2 DIABETES MELLITUS

T2DM aggregates within families and concordance rates in monozygotic twin pairs (60–90%) are greater than in dizygotic twins (30–40%). This pattern of transmission is consistent with a polygenic form of inheritance. Causal mutations have been identified in rare monogenic forms of diabetes, for example, mutations in the insulin receptor gene in patients with type A severe insulin resistance and acanthosis nigricans and mutations in glucokinase and certain hepatic nuclear factor genes in some families with maturity onset diabetes of the young. Patients with single gene mutations conform to a particular subphenotype of the disease.

The polygenes that contribute to the common form of T2DM have not yet been elucidated. It is likely that T2DM is heterogeneous in that multiple polygenic forms of the disease share many characteristics of a common phenotype. Environmental determinants that predispose to the disease include physical inactivity, obesity, and high-calorie, high-fat diets. The interaction between genes and environment is evident when groups of Africans, Australian Aborigines, or Amerindians, who have a low prevalence of T2DM in

their native cultures, adopt a Western lifestyle and develop diabetes at higher rates than other racial/ethnic groups in the same environment.

## PATHOPHYSIOLOGY OF TYPE 2 DIABETES MELLITUS

Although type 2 diabetes may be heterogeneous, several major metabolic defects consistently contribute to hyperglycemia. These major defects include the following: (1) peripheral insulin resistance due to decreased glucose transport activity in skeletal muscle; (2) impaired insulin secretion in response to glucose; and (3) elevated rates of hepatic glucose production. All three of these metabolic defects are present in overt diabetes and contribute significantly to the diabetic state.

Insulin resistance (IR) is considered by some investigators to be a fundamental defect in patients with T2DM. The 1997 recommendations of the Expert Committee of the ADA considered IR to be a necessary aspect of the diagnosis of T2DM. Hence, by definition, all patients with T2DM would be characterized by IR. However, proof for this contention is problematic for two reasons. First, rigorous measurement of insulin resistance requires research methods such as the hyperinsulinemic glucose clamp or the frequently sampled intravenous glucose tolerance test (fsIVGTT). Second, since insulin sensitivity is a continuous variable, there is no established cutoff value below which individuals would be considered to be insulin resistant. In 1997, Haffner and colleagues, using the fsIVGTT to estimate insulin sensitivity in 479 T2DM patients, determined that the majority of subjects were insulin resistant, regardless of race or ethnicity. Other studies have found lower prevalence rates, of approximately 80%, for IR among T2DM patients. Therefore, the prevalence of IR in T2DM depends on the definition and methods employed, ranging from 80 to 100%. In fact, the 1999 WHO proposal for the definition and classification of T2DM, which otherwise agrees with most of the 1997 ADA criteria, did not consider the presence of IR to be a requirement for the diagnosis of T2DM.

Inadequate beta-cell function is also usually considered an essential component of T2DM pathogenesis. The most striking functional defect is a specific loss of first-phase or acute glucose-induced insulin secretion, whereas responses to other secretagogues, such as the amino acid arginine, can be relatively preserved. Defective insulin secretion becomes progressively worse over the course of the disease and this is

associated with the need for combination medical therapy and exogenous insulin for glycemic control. Chronic hyperglycemia contributes to progressive beta-cell failure in a process sometimes known as “glucotoxicity.” Both insulin resistance and defective glucose-mediated insulin secretion can be detected in prediabetic individuals; however, an insulin-resistant individual who maintains full insulin secretory responses will generally not develop T2DM.

## INSULIN RESISTANCE SYNDROME

There is a wide variation in insulin sensitivity in normoglycemic individuals. Relative insulin resistance is associated with additional metabolic traits that in combination place these individuals at risk for the future development of T2DM and/or cardiovascular disease. These traits can include hyperinsulinemia; some degree of glucose intolerance; upper body fat distribution or abdominal obesity; dyslipidemia characterized by high triglycerides, low high-density lipoprotein (HDL) cholesterol, and small dense low-density lipoprotein (LDL) particles; elevated blood pressure, elevated c-reactive protein; high plasminogen activator inhibitor-1 levels; and a positive family history for T2DM. The clustering of several of these traits has been linked to insulin resistance in a single individual and referred to as insulin-resistant syndrome (IRS) or metabolic syndrome. IRS/metabolic syndrome represents an important new concept that is critical for improved identification and management of metabolic and cardiovascular risk. For example, institution of lifestyle changes in these individuals resulting in mild weight loss (5–7%) and adherence to a regular exercise program can cut rates of progression to overt T2DM by over 50%, as demonstrated by the Diabetes Prevention Program. Furthermore, patients with IRS/metabolic syndrome should be targeted for more aggressive control of dyslipidemia and blood pressure.

One problem has been the lack of consensus on how to identify or diagnose IRS/metabolic syndrome. To address this problem, the National Cholesterol Education Program Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol (Adult Treatment Panel III) proposed in 2001 that the metabolic syndrome could be defined in individuals who met at least three of five criteria [waist circumference >102 cm in men and >88 cm in women; fasting glucose 110–125 mg/dl; triglycerides  $\geq$ 150 mg/dl; HDL cholesterol <40 mg/dl in men and <50 mg/dl in women; blood pressure >130/85].

Additional research is required to determine the adequacy of this scheme in identifying patients at risk.

## TREATMENT OF TYPE 2 DIABETES MELLITUS

The most important reason to treat T2DM aggressively is to prevent its chronic complications, both microvascular and macrovascular. It is well known that the degree of chronic hyperglycemia, best reflected in glycosylated hemoglobin measurements, is directly related to microvascular complications (neuropathy, nephropathy, and retinopathy). In addition, the treatment of hyperglycemia, aimed at obtaining near-normoglycemia, prevents or at least slows the progression of these complications independently of the glucose-lowering agent used.

Whereas microvascular complications cause significant morbidity and cost, it is the macrovascular complications (coronary artery disease, cerebrovascular disease, and peripheral vascular disease) that cause most of the excess mortality in patients with T2DM when compared to the rest of the population. The prevention of these macrovascular complications, particularly coronary artery disease, is not as dependent on simple glycemic control as that of the microvascular complications, although there is some evidence suggesting that strict glycemic control may have a beneficial effect in terms of reduction of cardiovascular events. Consequently, the prevention of the macrovascular complications requires aggressive management of the modifiable traditional risk factors for cardiovascular disease frequently present in these patients, namely, dyslipidemia, hypertension, obesity, and smoking. Therefore, the care of T2DM patients must focus on achieving aggressive targets for blood pressure (<130/80) and serum lipids (LDL cholesterol <100 mg/dl, triglycerides <150 mg/dl; HDL cholesterol >40 mg/dl) with drugs such as HMG (3-hydroxy-methyl-glutaryl) coenzyme A reductase inhibitors, fibric acid derivatives, niacin, and blood pressure-lowering preparations.

The cornerstones of anti-diabetic therapy are as follows:

### Diet

A very important component of the management of all patients with T2DM is diet. A trained dietician should be involved in assessing dietary history and lifestyle, reviewing dietary principles, and developing and executing the meal plan on an ongoing basis. The diet plan should be tailored to the individual and account

for personal preferences, age, gender, current nutritional and clinical status, cultural practices, degree of obesity, medications, and activity level. Although the recommended caloric distribution has fluctuated over time, the general approach is to limit fat calories, reduce cholesterol and saturated fatty acids, and allow modest protein and complex carbohydrates, with a dietary composition that approximates 50% carbohydrates, 30% fat, and 20% protein.

Weight loss can have a dramatic effect to reduce (or sometimes eliminate) the need for diabetes medications and can also ameliorate dyslipidemia and hypertension, in overweight patients. This is often best accomplished by a moderate caloric reduction of 250–500 calories below the weight maintenance in the context of a nutritionally sound meal plan, proper caloric distribution, and an exercise regimen. Alternatively, a very-low-calorie diet (500–800 kcal/day) can be used in select patients. With either moderate reduction or very-low-calorie diets, weight loss is frequently achieved but is difficult to sustain. However, significant metabolic and cardiovascular benefits can be realized without achieving ideal body weight. Hypoglycemia will improve within a few days, even before measurable weight loss, during the hypocaloric phase. However, at least a 5% weight reduction is required after weight stabilization. A further weight reduction of 10% will significantly improve glycemia, insulin sensitivity, and serum lipid profile if sustained.

### Exercise

The vast majority of diabetic patients should be prescribed some form of regular exercise. Exercise is an important adjunct in glycemic control, weight loss, maintenance of weight reduction, management of cardiovascular risk factors, and sense of well-being. Exercise training raises HDL cholesterol, lowers blood pressure, and leads to a 20–40% increase in insulin sensitivity by enhancing insulin action in skeletal muscle. There is evidence that regular physical activity of even modest intensity (e.g., walking three to five times per week) significantly improves insulin sensitivity in individuals at high risk for diabetes.

### Pharmacologic Therapy

The treatment of type 2 diabetes has been revolutionized in the past several years by the availability of several new classes of oral hypoglycemic agents, listed in [Table II](#). These agents influence biochemical processes in different organs and differentially ameliorate multiple metabolic defects that cause hyperglycemia.

Sulfonylureas and the two new nonsulfonylurea agents, repaglinide and nateglinide, interact with the sulfonylurea receptor/potassium channel complex on pancreatic beta cells and increase insulin secretion. Metformin's principal site of action is the liver and acts to reduce hepatic glucose production. The new oral agents in the thiazolidinedione class, rosiglitazone and pioglitazone, act as insulin sensitizers. They enhance insulin's ability to stimulate glucose uptake in the skeletal muscle and adipose tissue and also cause a decrease in the levels of circulating free fatty acids, which are usually elevated and are thought to play a pathogenic role in T2DM. Therefore, these classes of oral agents ameliorate deficient insulin

secretion, high hepatic glucose output, or insulin resistance, which are the three major defects that combine to produce hyperglycemia in nearly all patients with type 2 diabetes. In addition,  $\alpha$ -glucosidase inhibitors have been introduced, which impede carbohydrate digestion and absorption in the gut, thus reducing the postprandial glycemic rise.

Since multiple classes of effective drugs have become available, oral agents have become the preferred choice for initiating medical therapy in most patients. However, since T2DM is a progressive disease, secondary failure to the first oral agent is the rule. When this occurs, it is generally advantageous to add a second drug that acts on a different metabolic defect (i.e., insulin secretion, hepatic glucose output, or insulin resistance) than the first drug, since the two drugs will interact synergistically to help control glycemia.

However, exogenous insulin remains a mainstay of therapy, especially in patients with longer disease duration who have failed one or more of the oral agents. This often involves the addition of bedtime insulin in combination with an oral agent and, later, multiple daily insulin doses. Although both the liver and skeletal muscle are insulin resistant in T2DM, dose-response differences dictate that, at serum insulin levels produced by exogenous insulin therapy in many patients, hepatic glucose production can be partially or completely inhibited, whereas lesser degrees of glucose uptake stimulation are observed in skeletal muscle. This ability to restrain hepatic glucose production is a key property whereby insulin helps achieve targets for lower glycosylated hemoglobin (e.g., HbA1c) values either as a single agent or in combination with oral agents.

**Table II Pharmacologic Classification of Agents Used in the Treatment of Diabetes Mellitus**

<b>Oral insulin secretagogues</b>
<i>First-generation sulfonylureas</i>
Tolbutamide
Chlorpropamide
Tolazamide
Acetohexamide
<i>Second-generation sulfonylureas</i>
Glyburide
Glipizide
Glimepiride
<i>Nonsulfonylurea insulin secretagogues</i>
Repaglinide
Nateglinide
<b>Inhibitors of carbohydrate absorption</b>
<i><math>\alpha</math>-Glucosidase inhibitors</i>
Acarbose
Miglitol
<b>Inhibitors of hepatic glucose output</b>
<i>Biguanides</i>
Metformin
<b>Insulin sensitizers</b>
<i>Thiazolidinediones</i>
Pioglitazone
Rosiglitazone
<b>Exogenous insulin</b>
<i>Human insulin</i>
Regular insulin
NPH insulin
Lente insulin
Ultralente insulin
<i>Human insulin analogues</i>
Lispro insulin
Aspart insulin
Glargine insulin

## See Also the Following Articles

Carbohydrate Metabolism: Diabetes Mellitus, Genomic Aberrations • Diabetes Insipidus, Nephrogenic • Diabetes Insipidus, Neurogenic • Diabetes Mellitus, Diagnosis and Treatment in the Elderly • Diabetes, Type 1 • Glucose, Impaired Tolerance • Glucose Physiology, Normal • Hypertension and Diabetes • Insulin-Resistant States, Role of Free Fatty Acids (FFA) • Insulin Secretion: Functional and Biochemical Aspects • Kidney Disease in Diabetes • Obesity and Diabetes, Regulation of Food Intake

## Further Reading

Expert Committee on the Diagnosis and Classification of Diabetes Mellitus (1997). Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care* 20, 1183–1197.

- Garvey, W. T., and Birnbaum, M. J. (1997). Cellular insulin action and insulin resistance. *Bailliere Clin. Endocrinol. Metab.* **7**, 785–873.
- Haffner, S. M., Howard, G., Mayer, E., Bergman, R. N., Savage, P. J., Rewers, M., Mykkanen, L., Karter, A. J., Hamman, R., and Saad, M. F. (1997). Insulin sensitivity and acute insulin response in African-Americans, non-Hispanic whites, and Hispanics with NIDDM: The Insulin Resistance Atherosclerosis Study. *Diabetes* **46**, 63–69.
- Harris, M. I., Eastman, R. C., Flegal, K. M., Cowie, C. C., and Eberhardt, M. S. (1997). Comparison of diabetes diagnostic categories in the U. S. population according to 1997 ADA and 1985 WHO diagnostic criteria. *Diabetes Care* **20**, 1859–1862.
- Harris, M. I., Flegal, K. M., Cowie, C. C., Eberhardt, M. S., Goldstein, D. E., Little, R. R., Wiedmeyer, H. M., and Byrd-Holt, D. D. (1998). Prevalence of diabetes, impaired fasting glucose, and impaired glucose tolerance in U. S. adults. The Third National Health and Nutrition Examination Survey, 1988–1994. *Diabetes Care* **21**, 518–524.
- National Diabetes Data Group. (1979). Classification and diagnosis of diabetes mellitus and other categories of glucose intolerance. *Diabetes* **28**, 1039–1057.
- World Health Organization (WHO). (1999). Definition, Diagnosis and Classification of Diabetes Mellitus and Its Complications. Report of a WHO Consultation. Part 1. Diagnosis and Classification of Diabetes Mellitus WHO, Geneva, Switzerland.
- World Health Organization (WHO). (1985). Diabetes Mellitus: Report of a WHO Study Group. Technical Report Series 727, WHO, Geneva, Switzerland.

in reciprocal decreases of myoinositol and taurine. Depletion of these metabolites becomes rate limiting for essential intracellular metabolism. The vascular theory was once considered separate from the metabolic theory. It states that reduced nerve blood flow results in hypoxia/ischemia and causes the development of DPN. It is now clear that the metabolic and vascular theories are linked on many levels. Glucose flux through the polyol pathway causes the depletion of NADPH and NAD<sup>+</sup>. Without sufficient NADPH, cells cannot regenerate the glutathione required to neutralize reactive oxygen species. Consequently, the peripheral nerve sustains oxidative damage that decreases endoneurial blood flow and promotes ischemia. The activity of the potent vasodilator, nitric oxide (NO), is also tightly linked to NADPH. NO, synthesized by NO synthase, is a NADPH-dependent reaction. Therefore, depletion of NADPH limits NO synthesis and in turn causes vasoconstriction and ischemia, which contribute to nerve conduction slowing.

## DIAGNOSIS AND TREATMENT

Three treatment strategies are clinically available. Of most importance is the early diagnosis of DPN. Early diagnosis allows for the implementation of the second strategy: good glycemic control and foot care. The third approach is focused on the treatment of painful DPN and is added to good glycemic control and foot care.

The early diagnosis of DPN is imperative. It makes early intervention possible, thereby significantly decreasing patient morbidity. It is important that other causes of neuropathy are excluded prior to DPN being related to diabetes. This is especially important if there are unusual features of the neuropathy, such as rapid progression, marked asymmetry, or more motor than sensory deficits.

In 1988, the San Antonio Consensus Panel recommended five quantifiable measures for the accurate diagnosis of DPN: a symptom questionnaire, a standardized clinical examination, quantitative sensory testing, nerve conduction studies, and autonomic function testing. Patients are classified as having stage I (no symptoms) or stage II (symptoms) neuropathy. Each stage is further divided into grades from A to C depending on the number of positive test results and the severity of clinical impairment. These criteria are being used to study the Rochester Diabetic Cohort. Modified criteria have been implemented in several clinical trials, including the DCCT.

Simpler screening instruments for DPN are also available. These instruments were developed because

patient and physician resources are frequently limited, making completion of the San Antonio criteria difficult. In the United Kingdom, DPN is diagnosed if patients present with mild signs and moderate symptoms or moderate signs alone, in the absence of symptoms. The Michigan Neuropathy Screening Instrument includes inspecting the feet for dry skin, callus, fissure, or ulceration; assessment of vibratory sensation in the great toes; and testing for ankle reflexes. A score of >2 indicates neuropathy with a high sensitivity (80%) and specificity (95%). Because of its simplicity, the Michigan Neuropathy Screening Instrument is also highly reproducible. Another simple method for screening patients for the presence of DPN is the use of a 10-g nylon monofilament. The filament is pressed against the sole of the foot until the filament buckles, indicating a known force has been applied. If a patient is unable to perceive the filament, he or she is at increased risk for complications of DPN.

## Glycemic Control and Foot Care

The DCCT clearly demonstrates that improved glycemic control decreases the frequency of DPN in patients with type I diabetes mellitus. Glycemic control requires thorough patient education. Patients are instructed on the importance of diet and regular glucose monitoring. This is achieved with a team consisting of the patient, a diabetes nurse educator, a dietician, and a physician.

Good foot care is also important. Patients are instructed to inspect their feet every night for evidence of dry skin, cracking, or fissuring of the skin. The importance of footwear is also well documented. Shoes must cushion the points of contact between the foot and the shoe and must accommodate any inherent or acquired foot deformities. For patients with mild neuropathy, cushioned socks and high-quality athletic shoes with room for the forefoot and toes are helpful. In severe cases, patients may require customized inserts or molded shoes.

## Acute and Chronic Painful DPN

A stepwise treatment protocol is used for patients with painful DPN. Nonsteroidal anti-inflammatory drugs provide relief in many patients with chronic painful DPN. In a double-blind, placebo-controlled trial, both ibuprofen (600 mg four times per day) and sulindac (200 mg two times per day) effectively decreased pain associated with DPN. However, this class of drugs cannot be used in patients with renal impairment.

The tricyclic antidepressants are the best studied class of drugs for the treatment of painful DPN. In double-blind, placebo-controlled trials, amitriptyline, nortriptyline, and imipramine were each effective in the treatment of painful DPN. If a patient continues to experience disabling pain, a second drug, gabapentin, is added to the therapeutic regimen. Patients may also be started on gabapentin as a first-line therapy, with a tricyclic added if a second medication is needed. Carbamazepine is also a good drug to add to either a tricyclic antidepressant or a serotonin uptake inhibitor. Double-blind, placebo-controlled trials have found that it provides symptomatic relief to a large percentage of afflicted patients.

For patients who are on two medications and still experiencing significant discomfort, capsaicin cream may be used. Capsaicin cream is a topical therapy that inhibits the uptake of substance P at sensory endings. Patients are instructed to apply capsaicin cream (0.075%) four times per day, the regimen that was successful in a double-blind, placebo-controlled trial.

If these therapeutic strategies fail, the second drug is discontinued and a new drug is instituted. The first choice is the cardiac antiarrhythmic drug mexiletine, which is effective in patients who otherwise are refractory to treatment. The antiepileptic topiramate can also be used in refractory patients. If a patient remains resistant to these treatment strategies, he or she should be referred to a comprehensive pain clinic. Here, patients frequently receive local nerve blocks, a transcutaneous electrical nerve stimulation (TENS) unit, or, in certain cases, acupuncture. Unfortunately, the prognosis for good pain relief in patients who require a pain clinic referral is low.

## CONCLUSION

A systematic, stepwise approach to patients with DPN ensures optimal patient care and decreases the risks

for short- and long-term disability. Future therapies hold promise, particularly those targeted toward ameliorating the metabolic and vascular abnormalities that develop in the peripheral nervous system as a result of continued hyperglycemia.

## Acknowledgments

The authors thank Ms. Judy Boldt for expert secretarial assistance. This work was supported by National Institutes of Health Grant R01 NS38849 and grants from the Juvenile Diabetes Research Foundation.

## See Also the Following Articles

Cardiovascular Disease in Diabetes • Eye Disease in Diabetes • Foot Disease in Diabetes • Kidney Disease in Diabetes • Neurological Disease and Diabetes, Autonomic

## Further Reading

- Feldman, E. L., Stevens, M. J., and Greene, D. A. (1998). Clinical management of diabetic neuropathy. *In* "Clinical Management of Diabetic Neuropathy" (A. Veves, ed.), pp. 89–105. Humana Press, Clifton, NJ.
- Feldman, E. L., Stevens, M. J., Russell, J. W., and Greene, D. A. (1999). Diabetic neuropathy. *In* "Current Review of Diabetes" (S. Taylor, ed.), pp. 71–83. Current Medicine, Inc., Philadelphia, PA.
- Feldman, E. L., Stevens, M. J., Russell, J. W., and Greene, D. A. (2001). Diabetic neuropathy. *In* "Principles and Practice of Endocrinology and Metabolism" (K. L. Becker, ed.), 3rd ed., pp. 1391–1399. Lippincott Williams & Wilkins, Baltimore.
- Greene, D. A., Stevens, M. J., and Feldman, E. L. (1998). Glycemic control. *In* "Diabetic Neuropathy" (P. J. Dyck and P. K. Thomas, eds.), 2nd ed., pp. 297–315. Saunders, Philadelphia.
- Stevens, M. J., Feldman, E. L., Thomas, T., and Greene, D. A. (1998). Pathogenesis of diabetic neuropathy. *In* "Clinical Management of Diabetic Neuropathy" (A. Veves, ed.), pp. 13–48. Humana Press, Clifton, NJ.
- Windebank, A. J., and Feldman, E. L. (2001). Diabetes and the nervous system. *In* "Neurology and General Medicine" (M. J. Aminoff, ed.), 3rd ed., pp. 341–364. Churchill Livingstone, London.



## Down Syndrome

see Short Stature and Chromosomal Abnormalities

thyroid-related effects of cytokines, monoclonal antibodies, heparin, phenytoin, amiodarone, and inhibitors of hormone binding to plasma proteins.

Some of these drug effects can be put to therapeutic use, for example, the use of lithium, cholestyramine, cholecystographic contrast agents, iodide, or glucocorticoids in unusual or complicated cases of thyrotoxicosis. It should also be noted that abnormal thyroid function, either thyrotoxicosis or hypothyroidism, can influence the effect and potential toxicity of a wide range of therapeutic agents, predominantly by altering their metabolism or clearance.

### ALTERED SECRETION OF THYROID-STIMULATING HORMONE

Glucocorticoids, either endogenous or exogenous, are potent inhibitors of thyroid-stimulating hormone (TSH) secretion. Dopaminergic drugs, even at doses that do not influence blood pressure, cause profound inhibition of TSH secretion; after cessation of dopamine infusion, there is a rapid reversal of TSH suppression within a few hours.

Cytokines directly affect TSH secretion. For example, infusion of interferon- $\alpha$  in normal volunteers results in a 60–70% decrease in serum TSH within 8–12 h, prior to any significant change in serum thyroid hormone concentrations. In addition, cytokines can induce or aggravate autoimmune thyroid dysfunction.

During long-term amiodarone therapy, TSH tends to be high in relation to serum free thyroxine (T<sub>4</sub>), because this drug, or its active metabolite, desethylamiodarone, acts as an antagonist at thyroid hormone receptors. Heavy amphetamine abuse can result in higher than normal levels of serum TSH and may cause TSH-induced hyperthyroxinemia. In general, agents that increase or decrease the basal level of serum TSH have a similar effect on the TSH response to thyrotropin-releasing hormone, whenever the two effects have been directly compared.

### ALTERED RELEASE OF THYROID HORMONE FROM THE GLAND

Iodide at a high dosage acutely decreases the release of thyroid hormone from the gland, apparently by inhibition of proteolysis. Glucocorticoids exert a similar effect that was initially attributed to TSH suppression, but the effect also occurs when TSH is suppressed, as in Graves' disease. Lithium inhibits thyroid hormone

release as a result of decreased proteolysis of thyroglobulin. These agents are occasionally used as additional medications in the management of complicated thyrotoxicosis, especially when standard anti-thyroid drugs cannot be used.

### INHIBITION OF 5'-(OUTER RING) DEIODINATION OF T<sub>4</sub>

The outer ring deiodinase, or 5'-deiodinase, that is present predominantly in liver, kidney, thyroid, and heart, designated type 1, is a selenoprotein that catalyzes the peripheral conversion of T<sub>4</sub> to triiodothyronine (T<sub>3</sub>) and reverse T<sub>3</sub> (rT<sub>3</sub>) to 3,3'-T<sub>2</sub>. This activity is markedly diminished by caloric deprivation and in catabolic states. Activity of this enzyme is also inhibited by numerous drugs (Table I), resulting in decreased T<sub>3</sub> and increased rT<sub>3</sub> serum concentrations. Several iodinated compounds, including amiodarone and oral cholecystographic contrast agents, also inhibit the pituitary 5'-deiodinase type 2, an effect that results in decreased formation of T<sub>3</sub> in the pituitary, leading to a slight increase in serum TSH, generally within the normal reference range.

Beta-blockers differ from one another in their effect on 5'-deiodination. Propranolol in high doses diminishes the production of T<sub>3</sub>, an effect due to a quinidine-like membrane stabilizing effect, rather than specific beta-blockade. This T<sub>3</sub>-lowering effect is not seen with other beta-blockers, such as atenolol, metoprolol, or labetalol. The symptomatic benefit of beta-blockers in thyrotoxicosis appears to be independent of their influence on serum T<sub>3</sub>.

A subnormal serum T<sub>3</sub> concentration is a normal response to illness or caloric deprivation that is potentially beneficial; this acute response should not be interpreted as hormone deficiency. During non-thyroidal illness, or with the use of medications that can lower serum T<sub>3</sub>, the possibility of thyrotoxicosis should not be dismissed on the grounds of normal serum T<sub>3</sub>, if this diagnosis is supported by T<sub>4</sub> excess and suppression of TSH.

Because of its ability to rapidly lower serum T<sub>3</sub>, the oral cholecystographic contrast agent ipodate has been used as an additional agent in the management of thyrotoxicosis.

### DISCHARGE OF TISSUE POOLS OF THYROID HORMONE

Isotope studies demonstrate that acute administration of lipid-soluble oral cholecystographic contrast

**Table I Principal Medications That Influence Thyroid Hormone or TSH Levels**

<b>Inhibit pituitary TSH secretion</b>
Dopamine, dobutamine, glucocorticoids, octreotide
<b>Modify thyroid hormone synthesis and release</b>
Contrast agents, amiodarone, topical iodine preparations, lithium, glucocorticoids, aminoglutethimide?
<b>Inhibit T4-T3 5'-deiodination</b>
Amiodarone, glucocorticoids, beta-blockers, <sup>a</sup> oral cholecystographic contrast agents, e.g., iopanoic acid, ipodate
<b>Modify immune function</b>
Interferon- $\alpha$ , interferon- $\beta$ , monoclonal antibody therapy
<b>Alter binding of T4 and T3 to plasma proteins</b>
<i>Increase concentration of T4-binding globulin</i>
Estrogen, heroin, methadone, clofibrate, 5-fluorouracil, perphenazine, tamoxifen
<i>Decrease concentration of T4-binding globulin</i>
Glucocorticoids, androgens, L-asparaginase
<i>Displace T4 and T3 from binding proteins</i>
Furosemide, salicylates, phenytoin, carbamazepine, nonsteroidal anti-inflammatory agents, <sup>b</sup> heparin <sup>c</sup>
<b>Displace T4 from tissue pool</b>
Alkylating agents, oral cholecystographic agents
<b>Modify thyroid hormone action?</b>
Amiodarone, phenytoin
<b>Increase clearance of T4, T3</b>
Barbiturates, phenytoin, carbamazepine, rifampicin, Sertraline?, fluoxetine?, dothiepin?
<b>Impair absorption of ingested T4</b>
Aluminum hydroxide, ferrous sulfate, cholestyramine, calcium carbonate, colestipol, sucralfate, soya preparations, kayexalate

<sup>a</sup>Those with membrane-stabilizing effect, e.g., propranolol.

<sup>b</sup>Some members of the group.

<sup>c</sup>*In vitro* effect of *in vivo* heparin administration (see Fig. 3).

agents causes the acute release of tissue T4 stores, mainly from the liver, resulting in a transient increase in the serum T4 concentration of up to 30%. Reported acute transient increases in serum T4 of up to 60% following treatment with the alkylating agents cyclophosphamide and ifosfamide may also be due to hormone displacement from tissues.

## ALTERED CONCENTRATION OF PLASMA-BINDING PROTEINS

Estrogens, whether endogenous or exogenous, commonly affect tests of thyroid function by increasing total serum T4 and T3 due to an increase in the serum concentration of thyroxine-binding globulin (TBG), but free T4 and T3 remain normal. Estrogens increase the glycosylation of TBG, which slows its clearance, leading to a higher serum concentration,

or binding capacity, with normal binding affinity. Transdermal estrogens do not show this effect to the same extent, because they have less influence on hepatic proteins. In pregnancy, the mean total serum T4 concentration is increased by approximately 30%, with up to 40% of values falling outside the normal reference range in the second and third trimesters. It is not certain whether other drugs that alter the serum concentration of TBG (Table I) increase the synthesis or retard the clearance of TBG.

At least initially, until a new steady state is established, the dose requirement for T4 replacement can change in response to changes in binding protein. For example, a higher dosage of T4 may be required in women taking estrogens. Conversely, commencement of androgen treatment for breast cancer, which results in a 50% lowering of serum TBG, may result in acute features of thyrotoxicosis in some women maintained on stable replacement dosage of T4. An initial T4 dose reduction by 20–50% is recommended, although the need for permanent dose modification has not been established.

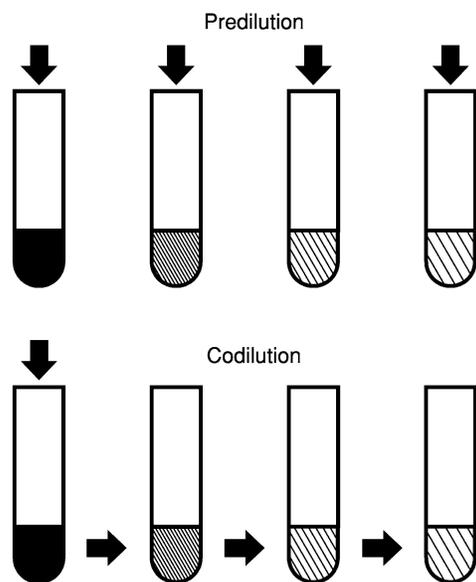
In response to an acute increase in the concentration of binding protein, there will be an initial tendency for the free hormone concentration to decrease, leading to a transient increase in serum TSH; an increase in total hormone concentration then restores the baseline free serum T4 concentration, but the free fraction of T4 is lower. Changes in T3 binding follow a similar pattern. Reverse changes follow a decrease in the concentration of binding protein. The turnover rate of hormone remains unchanged, but the extracellular pool size is altered. A similar sequence occurs if there is altered occupancy of the binding protein by a competitor.

## COMPETITION FOR THYROID HORMONE BINDING TO PLASMA PROTEINS

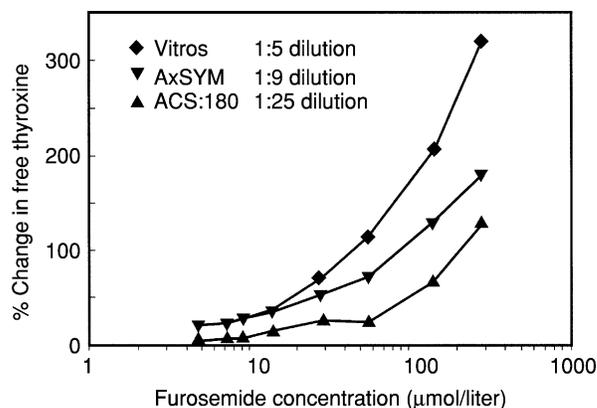
Thyroid hormone-binding proteins show wide cross-reaction with unrelated hydrophobic ligands, such as nonesterified fatty acids (NEFA) and various drugs (Table I). In contrast, the binding proteins for corticosteroids, vitamin D, and the sex hormones are more specific for a single family of ligands. Substances that interfere with T4 and T3 binding are themselves bound to albumin. Because of differences in relevant total therapeutic concentration and albumin binding, the hierarchy of drug competitor potency *in vivo* in undiluted serum differs markedly from the rank order of drug affinity for isolated TBG.

Serum dilution and albumin concentration need to be clearly defined in the interpretation of *in vitro* drug competition studies, because the concentrations of free ligand and competitor(s) and the number of unoccupied binding sites do not maintain the relationship that exists prior to serum dilution. The terms “predilution” and “codilution” describe two contrasting artifacts.

Predilution occurs when the concentration of albumin is decreased by dilution before a competitor is added (Fig. 1). At lower concentrations of albumin, there is a disproportionate increase in free competitor concentration. Hence, predilution before addition of competitor magnifies apparent competition. In contrast, an underestimate of competitor potency occurs with codilution, where total concentrations of binding proteins, hormone, and competitors diminish in parallel. Free ligand concentrations diverge widely with dilution; i.e., the free concentration of a drug with a free fraction in serum of 1:50 decreases markedly at a serum dilution of only 1:10, whereas the free concentration of T4, which has a free fraction of approximately 1:4000 before dilution, will be much better sustained with dilution. Hence, competitor effects will be lost with codilution and free T4 will be underestimated.



**Figure 1** Addition of a binding competitor after prior dilution of a serum sample will falsely accentuate hormone displacement, because dilution of serum albumin increases the free competitor concentration (predilution). In contrast, parallel dilution of all serum constituents tends to obscure displacement of T4 and T3 by a competitor that was present in the original serum sample (codilution), because the concentration of free competitor decreases more than the concentration of free hormone (see text).



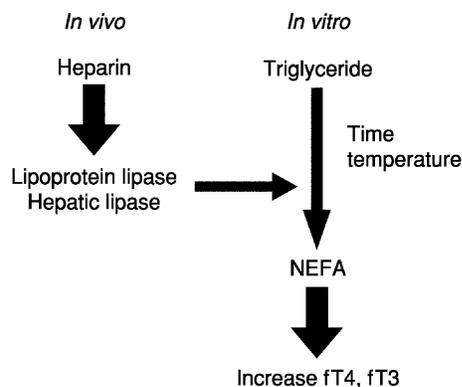
**Figure 2** Effect of furosemide on estimates of serum free T4 by three methods, showing that T4 displacement is progressively lost with increasing sample dilution. The least diluted serum sample most closely reflects the *in vivo* effect. Redrawn from Hawkins *et al.*, 1998.

This type of artifact was shown with therapeutic concentrations of phenytoin and carbamazepine, which increased the free fraction of T4 by 40–50% in undiluted serum, but produced no increase in free T4 when assayed in diluted serum. Similarly, the T4-displacing effect of furosemide is most obvious (Fig. 2) in methods with the least sample dilution.

Kinetics of the competitor will influence its effect on T4 and T3 *in vivo*. A competitor of long half-life will eventually result in a new steady state with normal free hormone and lowered total hormone concentration, i.e., an increased free fraction. In contrast, the effect of short half-life competitors, such as furosemide or salsalate, on serum free T4 and T3 will depend on the time between dosage and sampling.

Any substance that shares albumin-binding sites with a competitor can influence the free concentration of that competitor. Indoles and furans that accumulate in renal failure can displace drugs from albumin and may thus indirectly accentuate drug effects on hormone binding; i.e., the free hormone concentration can be influenced by substances that have little direct effect on hormone binding.

The effect of heparin on increasing the apparent free T4 concentration *in vitro* is an example of spurious competition. Heparin treatment causes the *in vivo* release of lipases, followed by generation of NEFA during sample storage or incubation *in vitro*, leading to much higher NEFA concentrations in the assay tube than were present *in vivo* (see Fig. 3). Heparin doses as low as 10 units can cause this effect after prolonged sample incubation, especially if serum triglyceride is increased. Low-molecular-weight heparin



**Figure 3** Heparin-induced changes can markedly increase the apparent concentration of serum free T4. Heparin *in vivo* (left) liberates lipoprotein lipase from vascular endothelium. Lipase acts *in vitro* (right) to increase the concentration of free fatty acids to  $>3$  mmol/liter, which displaces T4 and T3 from TBG. Generation of free fatty acids is increased by sample storage or incubation and by high serum triglyceride. The T4-displacing effect is accentuated at low albumin concentrations. Concept from Mendel *et al.*, 1987.

preparations have a similar effect. This artifact may account for reports of method-dependent increases in apparent free T4 concentrations in hospitalized subjects who are often exposed to heparin.

### IMPAIRED THYROXINE ABSORPTION

Numerous agents (Table I) can impair the absorption of ingested T4, probably by binding T4 in the bowel lumen. Such agents have little effect if the pituitary–thyroid axis is intact, but they may make fixed oral T4 dosage inadequate in hypothyroid patients who have no capacity to increase T4 production. This effect can generally be minimized by avoiding concurrent ingestion of T4 and potential inhibitors of absorption.

### ENHANCED THYROID HORMONE METABOLISM

There is usually little effect from medications that enhance thyroid hormone clearance if the pituitary–thyroid axis is normal, but effects may be profound when an individual is dependent on exogenous T4. Hepatic T4 and T3 metabolism is enhanced by numerous agents that stimulate the cytochrome P450 system. In addition to the known effects of rifampicin, phenytoin, carbamazepine, and barbiturates, other compounds may have this effect, for example, antibiotics, psychotropics, and nontherapeutic xenobiotics.

### DRUG INTERACTIONS

Tests of thyroid function may be especially affected when several agents are given together. For example, the concurrent use of high-dose furosemide with dopamine can result in profound hypothyroxinemia. Furosemide at high dosage displaces T4 from TBG, thereby accelerating its clearance, while concurrent dopamine infusion inhibits the normal TSH response to hypothyroxinemia. When dopamine treatment is concluded, serum TSH can rise transiently to supra-normal levels that can be misinterpreted as indicating primary hypothyroidism. The combination of rifampicin, which accelerates T4 clearance, with glucocorticoid-induced inhibition of TSH secretion may also cause hypothyroxinemia.

### AGENTS THAT MODIFY IMMUNE FUNCTION

Treatment of hepatitis C with interferon- $\alpha$  is frequently associated with thyrotoxicosis or hypothyroidism, sometimes in a sequence that suggests thyroiditis; the abnormalities are often transient, with resolution occurring over several months after treatment is ended. The prevalence of this effect has been estimated at 15–30%, with greater frequency in females and those with positive thyroid peroxidase antibody. In contrast, thyroid dysfunction is less common during interferon- $\beta$  treatment of neurological disorders, such as multiple sclerosis.

Monoclonal antibody treatment for multiple sclerosis may lead to persistent autoimmune thyroid dysfunction. Six to 30 months after treatment, one-third of previously euthyroid patients who received a 5-day course of the humanized anti-CD52 monoclonal antibody developed antibodies against the TSH receptor, associated with carbimazole-responsive thyrotoxicosis that tended to persist after withdrawal of carbimazole.

Transient primary hypothyroidism for 6 months to 2 years, perhaps of autoimmune origin, has been described in association with hypersensitivity reactions to phenytoin, carbamazepine, and sulfonamides; *in vitro* studies suggested an immune interaction with thyroid peroxidase.

### CLINICAL ISSUES WITH PARTICULAR DRUGS

#### Amiodarone

Amiodarone is the most complex and troublesome of the drugs that can affect thyroid status (Table II).

**Table II Medications with Multiple Effects on the Pituitary–Thyroid Axis**

<b>Glucocorticoids</b>
Inhibit TSH secretion
Impair thyroid hormone release from the gland
Inhibit 5'-deiodination of T4 and rT3
Decrease concentration of TBG
<b>Amiodarone</b>
Delivers iodine load
Iodine-induced thyrotoxicosis (Jod-Basedow)
Iodine-induced hypothyroidism
Causes unique form of thyrotoxicosis due to thyroiditis
Retards T4 clearance
Inhibits 5'-deiodination (type 1 and type 2) of T4 and rT3
Thyroid hormone receptor antagonist?
Augments TSH secretion
<b>Phenytoin</b>
Inhibits T4 and T3 binding to TBG
Accelerates T4 clearance and metabolism

Benign euthyroid hyperthyroxinemia, with high total and free T4, normal or subnormal serum T3, and increased serum rT3, occurs in up to one-third of amiodarone-treated subjects. This abnormality requires no treatment—the diagnosis of amiodarone-induced thyrotoxicosis should not be based on T4 excess alone.

In iodine-replete regions, the predominant amiodarone-induced thyroid problem is hypothyroidism, which is especially likely to occur on a background of thyroid autoimmunity. Routine replacement with T4 is effective, but therapy may need to be modified because of associated heart disease.

The most difficult abnormalities that result from amiodarone therapy are two unpredictable forms of thyrotoxicosis, one due to iodine excess and the other attributed to a unique form of persistent thyroiditis with specific intracellular inclusion bodies. Severe life-threatening thyrotoxicosis can occur rapidly, without premonitory abnormalities of thyroid function. Weight loss, deterioration of cardiac function, and severe myopathy are the important clues to this diagnosis. There can be poor correlation between circulating thyroid hormone levels and the clinical features of amiodarone-induced thyrotoxicosis, perhaps because of interaction of this drug, or its active metabolite desethylamiodarone, with thyroid hormone receptors.

A distinction between these two forms of amiodarone-induced thyrotoxicosis by color Doppler flow studies has been reported. If blood flow is increased, the thyrotoxicosis may be directly due to iodine

excess; standard anti-thyroid drugs, with the possible addition of potassium perchlorate, may be appropriate first-line therapy. In contrast, glucocorticoids are generally preferred in the thyroiditis variant that shows markedly decreased thyroid blood flow. Emergency thyroidectomy may be necessary in some cases of severe amiodarone-induced thyrotoxicosis that cannot be controlled medically.

## Lithium

This agent, widely used in the management of manic-depressive illness, has several effects on the pituitary–thyroid axis, the most important being the effect of inhibiting thyroglobulin hydrolysis and hormone release. Lithium exacerbates, or may possibly cause, autoimmune thyroid disease of the Hashimoto type, leading to eventual primary hypothyroidism, often with goiter. Women with positive anti-peroxidase antibodies are especially likely to be affected. Among 690 lithium-treated Scottish patients, 14% of women and 4.5% of men developed various grades of thyroid failure. There are also reports of lithium-induced thyrotoxicosis of probable autoimmune origin.

It is generally recommended that TSH, T4, and anti-peroxidase antibodies be assessed before commencement of lithium therapy, with serial evaluation of thyroid status every 6–12 months during treatment or if a goiter develops. Thyroxine replacement is recommended if there is TSH excess or progressive thyroid enlargement, even if TSH remains within the reference range. Doses of T4 sufficient to suppress TSH to subnormal levels probably retard goiter growth, but this point remains unproven.

## Phenytoin

This anti-epileptic commonly results in subnormal serum total T4, with an apparent lowering of free T4, not accompanied by the anticipated increase in TSH. Such findings are not easily distinguishable from central hypothyroidism due to pituitary deficiency. The effect on free T4 may be spurious, because the T4-displacing effect of phenytoin is poorly reflected by assays that use diluted serum, leading to an underestimate of the free T4 concentration (see above).

It remains difficult to make an accurate assessment of thyroid status in hypopituitary patients who are taking T4 together with phenytoin or carbamazepine; serum TSH is not useful and free T4 estimates can be misleading. Phenytoin accelerates T4 clearance by induction of cytochrome P450 enzymes, so that

the replacement dose may need to be increased. Treatment with phenytoin or carbamazepine can make previously optimal treatment of primary hypothyroidism inadequate or may unmask diminished thyroid reserve.

## EFFECTS OF THYROID STATUS ON DRUG EFFECTS

In general, thyrotoxicosis increases drug clearance, whereas hypothyroidism may markedly retard drug disposal. During thyrotoxicosis, standard drug dosage may be ineffective, as, for example, in the reputed “insensitivity” of thyrotoxic patients to digitalis preparations. A higher than normal dosage may be required, but as the thyrotoxicosis comes under control, digitalis toxicity can occur unless dosage is adjusted.

Severely hypothyroid subjects are abnormally sensitive to narcotics, sedatives, and analgesics, due to diminished clearance of these substances. This may be recognized as prolonged respiratory depression after anesthesia, when hypothyroidism has not been previously recognized.

Anticoagulant therapy is an exception to this rule. Consumption of coagulation factors tends to be more rapid in active thyrotoxicosis, with a tendency toward increased responsiveness and lower coumarin dose requirements. An increase in dosage may be required as control of thyrotoxicosis is achieved.

## See Also the Following Articles

Antithyroid Drugs • Hypothyroidism, Systemic Manifestations of • Lithium • Resistance to Thyroid Hormone (RTH) • Thyroid Function Tests • Thyroid Hormone Metabolism • Thyrotoxicosis: Diagnosis

## Further Reading

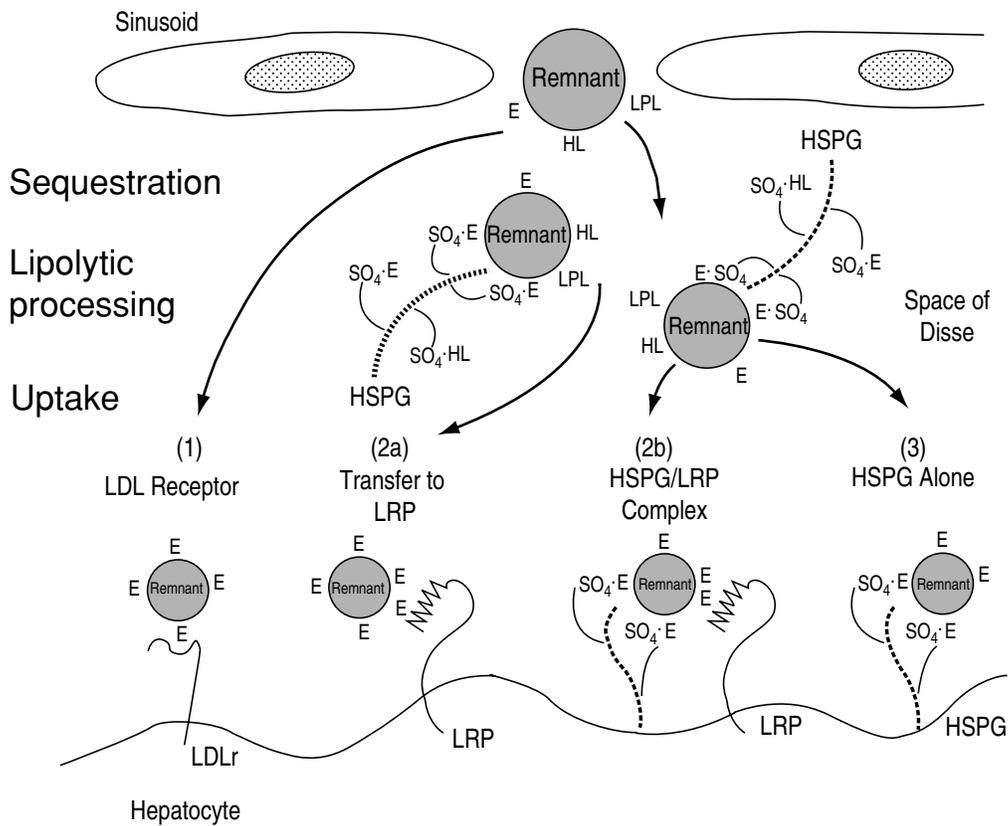
- Arafah, B. M. (1994). Decreased levothyroxine requirement in women with hypothyroidism during androgen therapy for breast cancer. *Ann. Intern. Med.* **121**, 247–251.
- Arafah, B. M. (2001). Increased need for thyroxine in women with hypothyroidism during estrogen therapy. *N. Engl. J. Med.* **344**, 1743–1749.
- Coles, A. J., Wing, M., Smith, S., *et al.* (1999). Pulsed monoclonal antibody treatment and autoimmune thyroid disease in multiple sclerosis. *Lancet* **354**, 1691–1695.
- Curran, P. G., and DeGroot, L. J. (1991). The effect of hepatic enzyme-inducing drugs on thyroid hormones and the thyroid gland. *Endocr. Rev.* **12**, 135–150.
- Durelli, L., Ferrero, B., Oggero, A., *et al.* (2001). Thyroid function and autoimmunity during interferon  $\beta$  treatment: A multicenter prospective study. *J. Clin. Endocrinol. Metab.* **86**, 3525–3532.
- Easton, S. E., Euinton, H. A., Newman, C. M., *et al.* (2002). Clinical experience with amiodarone over a 3 year period: Role of colour flow Doppler sonography. *Clin. Endocrinol.* **56**, 33–38.
- Harjai, K. J., and Licata, A. A. (1997). Effects of amiodarone on thyroid function. *Ann. Intern. Med.* **126**, 63–73.
- Hawkins, R. C. (1998). Furosemide interference in newer free thyroxine assays. *Clin. Chem.* **44**, 2550–2551.
- Hsieh, M. C., Chuang, W. L., Shin, S. J., *et al.* (2000). Virologic factors related to interferon- $\alpha$  induced thyroid dysfunction in patients with chronic hepatitis C. *Eur. J. Endocrinol.* **142**, 431–437.
- Kellett, H., Sawers, J., Boulton, F., *et al.* (1986). Problems of anticoagulation with warfarin in hyperthyroidism. *Q. J. Med.* **58**(225), 43–51.
- Lazarus, J. H. (1998). The effects of lithium therapy on thyroid and thyrotropin releasing hormone. *Thyroid* **8**, 909–913.
- Lim, C.-F., Bai, Y., Topliss, D. J., *et al.* (1988). Drug and fatty acid effects on serum thyroid hormone binding. *J. Clin. Endocrinol. Metab.* **67**, 682–688.
- Lim, C.-F., Stockigt, J. R., Curtis, A. J., *et al.* (1993). Influence of a naturally-occurring furanoid acid on the potency of drug competitors for specific thyroxine binding in serum. *Metabolism* **42**, 1468–1474.
- Mendel, C. M., Frost, P. H., Kunitake, S. T., and Cavalieri, R. R. (1987). Mechanism of the heparin-induced increase in the concentration of free thyroxine in plasma. *J. Clin. Endocrinol. Metab.* **65**, 1259–1264.
- Sapin, R., Schlienger, J.-L., Gasser, F., *et al.* (2000). Intermethod discordant free thyroxine measurements in bone marrow-transplanted patients. *Clin. Chem.* **46**, 418–422.
- Stockigt, J. R. (2001). Thyroid hormone binding. In “Endocrinology,” 5th ed. (L. J. DeGroot and L. Jameson, eds.). W. B. Saunders, Philadelphia, PA.
- Surks, M. I., and Defesi, C. R. (1996). Normal free thyroxine concentrations in patients treated with phenytoin or carbamazepine: A paradox resolved. *J. Am. Med. Assoc.* **275**, 1495–1498.
- Surks, M. I., and Sievert, R. (1995). Drugs and thyroid function. *N. Engl. J. Med.* **333**, 1688–1694.
- Van den Berghe, G., de Zegher, F., and Lauwers, P. (1994). Dopamine and the sick euthyroid syndrome in critical illness. *Clin. Endocrinol.* **41**, 731–737.
- Wang, R., Nelson, J. C., and Wilcox, R. B. (1998). Salsalate administration—A potential pharmacological model of the sick euthyroid syndrome. *J. Clin. Endocrinol. Metab.* **83**, 3095–3099.

**PATHOPHYSIOLOGY**

These diagnostic criteria for type III HLP reflect the prolonged circulation and accumulation of remnant lipoproteins in patients with this disorder. Remnants are lipoproteins at intermediate stages in the catabolism of TG-rich lipoproteins, both chylomicrons of dietary intestinal origin ( $d < 0.95$ ,  $S_f > 400$ , containing apo B-48 as the structural apolipoprotein) and VLDL of endogenous hepatic origin (containing apo B-100). Remnants are formed by the selective extraction of TG from the chylomicron or VLDL lipoprotein core, lipolysis mediated by lipoprotein lipase (LPL) that is bound by heparan sulfate proteoglycans (HSPG) to capillary endothelial cells (with apo C-II as an essential coenzyme). The remnant lipoproteins generated in this way are reduced in TG (and proportionally increased in cholesterol) and are smaller, denser, and of slower electrophoretic mobility than their precursors. They are normally

efficiently removed by the liver (Fig. 1) after sequestration in the space of Disse, where further lipolysis by LPL and hepatic lipase (HL) takes place prior to their uptake by hepatocytes. This uptake may take place via three alternative mechanisms: (1) the LDL receptor (LDL-r); (2) the LDL receptor-related protein (LRP), with or without bound HSPG; or (3) HSPG alone. Apo E on the remnant surface is critical in this process, serving as the principal ligand for the LDL-r and LRP. Of the various alternative removal mechanisms, the LDL-r is thought to be of primary regulatory importance, with the ontologically more primitive and pluripotential LRP serving in a "backup" role, apparent only when the LDL-r is deficient or defective (as in FH, in which remnants do not accumulate).

The capacity for remnant removal is normally sufficient to fully assimilate these lipoprotein intermediates during an 8- to 12-h overnight fast, even in a circumstance with a high flux of TG-rich lipoproteins



**Figure 1** Scheme of hepatic lipoprotein remnant processing in the space of Disse and uptake by hepatocytes, which can take place by three alternative mechanisms: via the LDL receptor (LDL-r) (1); transfer to the LDL receptor-related protein (LRP), either directly (2a) or as a complex with heparan sulfate proteoglycans (HSPG) (2b); or via hepatocyte-bound HSPG alone (3). The relative contributions of the three are affected by lipolytic processing via lipoprotein lipase (LPL) and hepatic lipase (HL) as well as by interactions between apo E and HSPG and the affinities of apo E and apo B for the LDL-r and LRP. Reproduced from *J. Lipid Res.* **40**, 1-16 (1999).

such as following an oral fat load or in the presence of high endogenous VLDL secretion rates (as in diabetes and the “metabolic syndrome” of insulin resistance). However, when remnant removal capacity is limited (as perhaps with hypothyroidism and certain dysglobulinemias), especially in a person whose apo E has reduced affinity for the LDL-r and LRP (i.e., with the apoE<sub>2,2</sub> genotype or who lacks apo E altogether on an inherited basis), and TG-rich lipoprotein flux is increased, remnants accumulate and acquire the properties of the cholesterol-rich,  $\beta$ -VLDL characteristic of type III HLP, including the persistence in fasting plasma of cholesterol-rich chylomicrons (at TG levels below 300 mg/dl).

The most common molecular basis of dysbetalipoproteinemia is the homozygous inheritance of isoapo E<sub>2</sub>, which differs from the wild-type E<sub>3</sub> in a single amino acid substitution (cysteine for arginine at residue 158). (E<sub>4</sub> has an additional arginine for cysteine substitution at residue 112 and, hence, is the most basic of the three.) The cys-arg substitution at residue 158 dramatically diminishes the affinity of E<sub>2</sub> for the LDL-r (by 98–99%) and for LRP (by 60–70%). Even in the hypolipidemic state, TG-rich lipoproteins of intermediate density with an increased C/TG ratio characteristic of dysbetalipoproteinemia predominate and  $\beta$ -VLDL can be detected in fasting plasma, albeit at very low levels. However, lacking a coexisting cause for hyperlipidemia, such persons are actually hypolipidemic, with low levels of LDL. Indeed, the apo E genotype is the most powerful genetic determinant of population plasma cholesterol levels; average LDL cholesterol levels (and cardiovascular risk) increase with increasing numerical apo E genotype: E<sub>2</sub>/E<sub>2</sub> < E<sub>2</sub>/E<sub>3</sub> < E<sub>3</sub>/E<sub>3</sub>  $\approx$  E<sub>2</sub>/E<sub>4</sub> < E<sub>3</sub>/E<sub>4</sub> < E<sub>4</sub>/E<sub>4</sub>. The gene frequencies of epsilon 2, 3, and 4 vary among populations; on average, the E<sub>2,2</sub> pattern is present in approximately 1%.

However, as noted earlier, when the E<sub>2,2</sub> genotype coexists with a cause for hyperlipidemia (genetic or acquired), this otherwise hypolipidemic phenotype is converted to one of the most hyperlipidemic (and atherogenic) of all (type III HLP). The required associated cause of hyperlipidemia is often a second inherited condition, most frequently FCHL (often considered synonymous with hyperapobetalipoproteinemia). FCHL may be present in as much as 1% of the general population (and in 10% of those with premature coronary disease); hence, type III HLP resulting from the coincidence of E<sub>2,2</sub> with FCHL may have a prevalence as high as 1 in 5000 to 10,000. Alternatively (or in addition to FCHL), insulin resistance may be the precipitating factor for type III HLP.

In a recent population-based Dutch study, the E<sub>2,2</sub> genotype was present in 0.6% and type III was present in 18% of these. In a referral sample from the same clinic, hyperinsulinemia was present in two-thirds of patients with type III HLP.

The pathophysiological hallmark of FCHL is apo B-100 (and VLDL) overproduction. Hence, in the presence of the E<sub>2,2</sub> genotype, VLDL apo B overproduction (with consequent increased remnant generation) appears to saturate a reduced remnant removal capacity, one that is attributable to impaired recognition of apo E<sub>2</sub> by the LDL-r and LRP. However, what mediates the efficient removal of E<sub>2</sub>-bearing remnants with diminished conversion to LDL in the absence of apo B overproduction remains enigmatic and subject to controversy. The prolonged plasma residence time of the remnants in type III HLP probably accounts for their progressive cholesterol enrichment (through exchange of lipids and apolipoproteins), oxidation, ultimate removal by scavenger receptors (as part of the atherogenic process), and deposition in other extravascular compartments, including the skin as xanthomas over areas subject to trauma.

A related area of concern in lipoprotein physiology is the general issue of the postprandial lipemia common to all forms of hypertriglyceridemia and the special atherogenicity of the remnants that circulate during prolonged alimentary lipemia. In this regard, even a partial limitation in apo E<sub>3</sub>- or E<sub>4</sub>-mediated remnant removal conferred by heterozygous E<sub>2</sub> status may contribute to accelerated atherogenesis in a person with apo B overproduction. ( $\beta$ -VLDL and premature ASCVD have been reported in persons of E<sub>2</sub>/E<sub>3</sub> phenotype in kindreds with FCHL.) Alternatively, diminished LDL-r capacity, whether on an inherited (FH) or acquired (hypothyroidism) basis, may also cause type III HLP in persons with apo E<sub>2</sub>/E<sub>2</sub>. However, in that circumstance, LDL cholesterol levels may actually be elevated (or at least not diminished) as a reflection of decreased LDL-r activity.

Type III HLP may also rarely be seen in patients lacking apo E altogether, as these patients also accumulate  $\beta$ -VLDL remnants and suffer accelerated atherogenesis. (Apo E knockout mice are favorite models for studying atherosclerosis.) Like their counterparts who are E<sub>2,2</sub> homozygotes, such persons without apo E demonstrate prolonged circulation of VLDL apo B and reduced conversion to LDL as manifestations of impaired remnant removal. (“Remnant removal disease” is a synonym for type III HLP.)

An even more severe limitation in remnant removal may occur in persons with rare apo E mutants that

produce type III HLP (with accelerated atherogenesis), in this instance genetically transmitted in an autosomal-dominant mode. This can occur from a single amino acid substitution (e.g., arg 136 → ser/cys) or the insertion of a tandem repeat of amino acids 121 to 127 (apo E-Leiden). Such apo E mutants appear to interfere with remnant removal mediated by HSPG, producing a blockade so complete that a coexisting cause of hyperlipidemia is not required for expression of type III HLP. An essential role for HL in remnant removal is also suggested by the presence of type III HLP in persons homozygously lacking HL on an autosomal-recessive genetic basis. However, in this circumstance, HDL levels are increased (and atherogenicity is relatively less than in the E<sub>2,2</sub> homozygote), presumably reflecting the role of HDL in reverse cholesterol transport (and the role of HL in accelerating HDL catabolism).

A final paradox remains to be resolved: the hypolipidemic effect of estrogen in type III HLP. Suggested clinically by the rare expression of the disorder in women prior to menopause in this autosomal-codominant condition, clinical studies (reproduced in animal models) have repeatedly demonstrated amelioration of type III HLP by estrogen therapy. TG levels decline (in contrast to their usual elevation with estrogen), as does VLDL cholesterol, and β-VLDL may disappear while HDL cholesterol rises; however, LDL levels do not rise, even though VLDL apo B is converted to LDL more readily. Thus, estrogen appears to facilitate remnant removal. Perhaps this may reflect enhanced clearance by LDL-r, LRP, or HSPG. However, it would seem unlikely that this effect would be mediated by enhanced TG lipolysis given that HL declines and LPL does not increase in women treated with estrogen.

Hence, at least a portion of the enigma of type III HLP remains. How does a genetic limitation in remnant removal result in decreased LDL? By what pathway are E<sub>2</sub>-bearing remnants removed from the circulation without down-regulating the LDL-r or becoming deposited in extravascular sites given that this would increase—not decrease—atherogenic risk? How does estrogen, which increases VLDL secretion, produce a net reduction in remnant accumulation in E<sub>2,2</sub> homozygotes with type III HLP?

## TREATMENT

As would be deduced from its pathophysiology, treatment of type III HLP begins with lifestyle modifications that reduce atherogenic risk in general and

diminish TG-rich lipoprotein production, especially via limitation of fat and cholesterol in an American Heart Association step II diet, coupled with caloric restriction/enhanced caloric expenditure to achieve ideal body weight. Drug therapy should proceed according to National Cholesterol Education Program-Adult Treatment Panel III (NCEP-ATP III) guidelines, although according to those based on non-HDL cholesterol levels rather than estimated LDL concentrations, which are inaccurate in type III HLP and besides which reduction of VLDL cholesterol is actually a primary therapeutic objective. TG-lowering (and HDL-raising) agents are especially effective (fibric acid derivatives and niacin), although HMG Co-A reductase inhibitors (statins) also appear to be effective and safe. Lipid lowering has been associated with regression of xanthomas in type III HLP, and coronary artery disease risk reduction would appear to be commensurate with non-HDL cholesterol reduction (although no specific study of individuals with type III HLP has been reported using cardiovascular disease end points). Estrogen replacement therapy has a special role in postmenopausal women, possibly alone but perhaps more effectively when combined with a fibrate or a statin.

## See Also the Following Articles

Abetalipoproteinemia • Atherogenesis • Familial Low Cholesterol Syndromes, Hypobetalipoproteinemia • Hypercholesterolemias, Familial Defective ApoB (FDB) and LDL Receptor Defects • Lipoprotein(a)

## Further Reading

- Chait, A., Brunzell, J. D., Albers, J. J., and Hazzard, W. R. (1977). Type-III hyperlipoproteinaemia ("remnant removal disease"): Insight into the pathogenetic mechanism. *Lancet* **1**, 1176-1178.
- Cooper, A. D. (1997). Hepatic uptake of chylomicron remnants. *J. Lipid Res.* **38**, 2173-2192.
- Davignon, J., Gregg, R. E., and Sing, C. F. (1988). Apolipoprotein E polymorphism and atherosclerosis. *Arteriosclerosis* **8**, 1-21.
- De Beer, F., Stalenhoef, A. F. H., Hoogerbrugge, N., Kastelein, J. J. P., Gevers Leuven, J. A., van Duijn, C. M., Havekes, L. M., and Smelt, A. H. M. (2002). Expression of type III hyperlipoproteinemia in apolipoprotein E2 (arg 158 → cys) homozygotes is associated with hyperinsulinemia. *Arterioscler. Thromb. Vasc. Biol.* **22**, 294-299.
- Hazzard, W. R., O'Donnell, T. F., and Lee, Y. L. (1975). Broad-beta disease (type III hyperlipoproteinemia) in a large kindred: Evidence for a monogenic mechanism. *Ann. Intern. Med.* **82**, 141-149.
- Hazzard, W. R., Warnick, G. R., Utermann, G., and Albers, J. J. (1981). Genetic transmission of isoapolipoprotein E phenotypes in a large kindred: Relationship to dysbetalipoproteinemia and hyperlipidemia. *Metabolism* **30**, 79-88.

- Hopkins, P. N., Wu, L. L., Schumacher, M. C., Emi, M., Hegele, R. M., Hunt, S. C., Lalouel, J.-M., and Williams, R. R. (1991). Type III dyslipoproteinemia in patients heterozygous for familial hypercholesterolemia and apolipoprotein E2: Evidence for a gene-gene interaction. *Arterioscler. Thromb.* **11**, 1137-1146.
- Huang, Y., Schwendner, S. W., Rall, S. C., Jr., Sanan, D. A., and Mahley, R. W. (1997). Apolipoprotein E2 transgenic rabbits: Modulation of the type III hyperlipoproteinemic phenotype by estrogen and occurrence of spontaneous atherosclerosis. *J. Biol. Chem.* **272**, 22685-22694.
- Hussain, M. M., Innerarity, T. L., Brecht, W. J., and Mahley, R. W. (1995). Chylomicron metabolism in normal, cholesterol-fed, and Watanabe heritable hyperlipidemic rabbits: Saturation of the sequestration step of the remnant clearance pathway. *J. Biol. Chem.* **270**, 8578-8587.
- Ji, Z.-S., Dichek, H. L., Miranda, R. D., and Mahley, R. W. (1997). Heparan sulfate proteoglycans participate in hepatic lipase- and apolipoprotein E-mediated binding and uptake of plasma lipoproteins, including high density lipoproteins. *J. Biol. Chem.* **272**, 31285-31292.
- Mahley, R. W., Huang, Y., and Rall, S. C., Jr. (1999). Pathogenesis of type III hyperlipoproteinemia (dysbetalipoproteinemia): Questions, quandaries, and paradoxes. *J. Lipid Res.* **40**, 1933-1949.
- Mahley, R. W., and Ji, Z.-S. (1999). Remnant lipoprotein metabolism: Key pathways involving cell-surface heparan sulfate proteoglycans and apolipoprotein E. *J. Lipid Res.* **40**, 1-16.
- Mahley, R. W., and Rall, S. C., Jr. (1995). Type III hyperlipoproteinemia (dysbetalipoproteinemia): The role of apolipoprotein E in normal and abnormal lipoprotein metabolism. In "The Metabolic and Molecular Bases of Inherited Disease" (C. R. Scriver, A. L. Beaudet, W. S. Sly, and D. Valle, eds.), 7th ed., pp. 1953-1980. McGraw-Hill, New York.
- Utermann, G., Hees, M., and Steinmetz, A. (1977). Polymorphism of apolipoprotein E and occurrence of dysbetalipoproteinemia in man. *Nature* **269**, 604-607.
- Utermann, G., Vogelberg, K. H., Steinmetz, A., Schoenborn, W., Pruin, N., Jaeschke, M., Hees, M., and Canzler, H. (1979). Polymorphism of apolipoprotein E. II: Genetics of hyperlipoproteinemia type III. *Clin. Genet.* **15**, 37-62.

## Bulimia

Persons with bulimia engage in cycles of bingeing and purging and/or other unusual means of controlling weight. These include using enemas and abusing laxatives and diuretics. Gorging cycles may also alternate with periods of severe food restriction and anorexics may also engage in bulimic behavior. In contrast to anorexia, weight often does not fall to critically low levels, although it may fluctuate. This, compounded with the secretive nature of the disorder, makes bulimia difficult to diagnose. Bulimia is most common in young women, especially those in high school and college, and often coincides with other impulsive behaviors and depression.

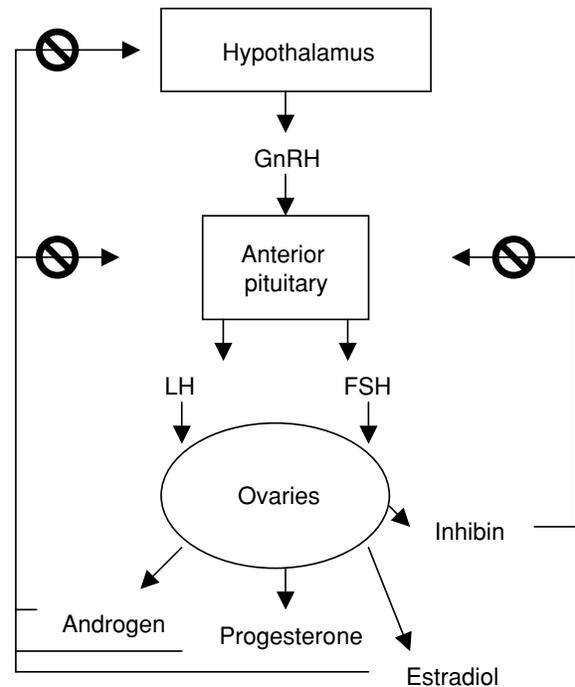
## NORMAL REPRODUCTIVE FUNCTION AND THE MENSTRUAL CYCLE

### Reproductive Hormonal Axis

To better understand the effects of eating disorders on the reproductive axis, the normal functioning of the female reproductive system should be considered. The mechanism of the normal menstrual cycle is the result of a number of different endocrine systems working in concert. These include the hypothalamus region of the brain and the adjacent anterior pituitary gland, as well as the ovaries and related organs (see Fig. 1). The hypothalamus secretes gonadotropin-releasing hormone (GnRH), signaling the pituitary gland to release luteinizing hormone (LH) and follicle-stimulating hormone (FSH). These two gonadotropins stimulate the ovaries to produce and secrete the sex hormones. The ovaries secrete primarily estrogens (mostly estradiol), progestins (mostly progesterone), and androgens.

### GnRH Pulsatility

The pulse generator in the arcuate nucleus of the hypothalamus controls the secretion of GnRH. Throughout the course of the female reproductive cycle, the hormone is released in varying spurts or pulses. The pulsatility of GnRH release causes subsequent pulses of gonadotropins, which affect the changes in the reproductive tract that characterize the menstrual cycle. GnRH release is regulated in response to changing levels of sex hormones, but the mechanism behind the pulse generator is not entirely understood. It is thought, however, to be highly



**Figure 1** Gonadotropin-releasing hormone (GnRH), secreted from the hypothalamus, induces the release of the gonadotropins [luteinizing hormone (LH) and follicle-stimulating hormone (FSH)] from the anterior pituitary gland. LH and FSH control the production and release of the sex hormones, estradiol, progesterone, and androgen, from the ovaries. These hormones, in turn, inhibit the secretion of GnRH, LH, and FSH. In addition, the ovaries produce inhibin, which exerts a negative feedback effect on the release of FSH.

sensitive to stress and metabolic factors. This may help explain the amenorrhea that often accompanies nutritional disturbances.

## INSULTS TO THE REPRODUCTIVE AXIS

### Eating Disorders and Reproductive Dysfunction

The physiologic effects of eating disorders on the female reproductive system are not completely understood, but appear to suppress ovulation and favor reproductive economy via a neuroendocrine mechanism.

### Anorexia Nervosa

In addition to the characteristic amenorrhea, anorexia produces other effects on the endocrine system, including changes in cortisol and thyroid hormones, which may affect the reproductive system.

**Bulimia**

Bulimics may also have menstrual irregularities, but reproductive dysfunction is generally less severe than in anorexics. Patients may be anovulatory with adequate estrogen secretion. Although amenorrheic, bulimics may maintain a normal or close to normal weight. However, bulimics have a wide variety of other medical problems that are superimposed on the anorectic syndrome. These include severe tooth decay, parotid enlargement, stomach rupture, metabolic alkalosis, carpopedal spasm, hypercarotenemia, and pancreatitis.

**Amenorrhea**

Amenorrhea, the lack of menstrual periods, is the primary effect of eating disorders on the reproductive system. It can be classified as primary or secondary, depending on the time of onset. Primary amenorrhea is a delay in menarche beyond the age of 16. Secondary amenorrhea occurs when periods are absent for more than 3 to 6 months in women who previously menstruated regularly. The dysfunctions that lead to amenorrhea are numerous and can result from an endocrine or physical disturbance at various steps along the reproductive axis. However, the type of amenorrhea generally seen in eating disorders is reversible hypothalamic amenorrhea. It can be either primary or secondary in nature, depending on whether or not the eating disorder began before puberty, thus delaying menarche.

The physiologic basis for hypothalamic amenorrhea is the disruption of the hypothalamus' pulsatile secretion of GnRH. A number of studies have shown a depression in GnRH pulsatility associated with starvation. This suggests that the GnRH pulse generator may be affected by metabolic fuels and/or is able to sense a drop in weight below a certain setpoint. When energy expenditure is greater than dietary intake, GnRH inhibition occurs, thereby lowering the release of LH and FSH from the anterior pituitary and shutting down or limiting ovarian stimulation and estradiol production.

Although actual levels of LH and FSH may not appear significantly altered, the 24h patterns of hormonal pulses more closely resemble prepubertal patterns than those of a mature female. Prepubertal LH and FSH patterns are characterized by decreased and low-amplitude pulsations. However, studies have shown that follicular maturation, and even menstruation, can be induced in many patients by the pulsatile administration of exogenous GnRH. Therefore, as is seen in patients who recover their weight, the

amenorrhea can be reversible when normal hormone pulsatility is restored.

**Other Types of Hypothalamic Amenorrhea****Athletic**

Hypothalamic amenorrhea is also prevalent among female athletes. In addition to the maintenance of low body weight, it is thought that a combination of other environmental factors, such as heavy exercise and mental stress, may contribute to the large number of amenorrheic women athletes. Some research suggests that women who sustain heavy exercise loads consistently over a period of time suffer an energy deficit that occurs because their daily energy expenditure exceeds their caloric intake. The body may compensate for this loss by decreasing resting metabolic rate and thereby disrupting the menstrual cycle. Because metabolic rate is depressed in these individuals, some are able to maintain normal weight despite a negative energy balance and amenorrhea.

**Functional Hypothalamic Amenorrhea**

Functional hypothalamic amenorrhea (FHA), a diagnosis of exclusion, occurs in normal weight, non-athletic women and has no apparent cause. Although previously attributed to stress, some evidence suggests that women with FHA may actually be incurring an energy deficit as a result of dieting. This energy deficit produces metabolic and endocrine effects similar to those seen in eating-disordered and athletic amenorrheics. Therefore, although anorexia and bulimia present extreme examples of the effect of nutritional aberrations on the reproductive system, menstrual irregularity and amenorrhea can occur from simple weight loss and dieting-associated weight loss. Understanding this effect may explain the significant number of recovered eating-disordered patients who do not regain menses with weight gain. By continuing disordered eating patterns, metabolic abnormalities may persist despite the return to normal weight.

The exact mechanism by which these adaptive metabolic changes influence the hormone cascade responsible for the maintenance of the menstrual cycle is unknown. However, researchers are pursuing an unknown metabolic signal or signals that may inform the GnRH pulse generator of the low weight and/or negative energy balance. Scientists are studying the roles of dehydroepiandrosterone, insulin-like growth factor, and cortisol, as well as leptin, a protein synthesized by the obesity gene, for possible clues.

## Leptin

Leptin is a small polypeptide hormone secreted primarily by adipocytes (fat cells). Leptin levels appear to be regulated by total energy intake and fat stores and correlate significantly with body mass index in humans. Because leptin is a regulator of the basal metabolic rate, it is thought to be a particularly important indicator of nutritional status. Studies have shown this effect in amenorrheic athletes and eating-disordered patients in whom leptin levels and metabolic rate tend to be low.

Studies have shown that leptin may also be a significant mediator of reproductive function. Research indicates that if leptin levels fall below a critical threshold, menstruation will not occur. Although the mechanism by which this occurs is unknown, leptin may affect changes in GnRH pulsatility as leptin receptors have been found on hypothalamic neurons thought to be involved in the control of the pulse generator.

Leptin is also connected to thyroid function and to the onset of puberty. When nutritional status is poor, altered leptin levels are directly correlated with changes in thyroid hormone. Thus, leptin may not only act as a regulator of metabolic rate, but also as a mediator of menstrual status by slowing metabolism and possibly returning a woman to a prepubertal-like state when challenged with starvation.

## OSTEOPOROSIS

One of the most permanent and devastating medical effects of hypothalamic amenorrhea in young women is significant osteopenia (reduced bone mass) and severe osteoporosis later in life. Studies suggest that almost half of an individual's bone mass is formed during adolescence and early adulthood, making this period especially critical to a woman's long-term health. Women who do not reach peak bone mass during this window are at a higher risk for osteoporosis once bone mass begins to decline.

Two homeostatic mechanisms act on bone simultaneously: hormones and mechanical stress. In normal circumstances, they maintain skeletal integrity and serum calcium levels. With aging or menstrual disturbance, factors such as diet, hormonal levels, and mechanical strain cause bones to become more vulnerable to fracture and osteoporosis. Studies have also shown that the normal mechanism that causes bones to strengthen in response to weight-bearing exercise is not functional in amenorrheic athletes. Although it

is unclear how this mechanism functions, these women have significantly lower bone mass density than their normally cycling counterparts.

Like amenorrheic athletes, anorexics have significantly decreased bone mass density. Although depressed estrogen levels may be partially responsible for this increased risk of osteoporosis, studies suggest that other mechanisms are also involved. The discovery of leptin receptors on bone may represent one such mechanism that could account for the low bone density and high stress fracture rates seen in women with hypothalamic amenorrhea.

## EFFECTS OF EATING DISORDERS ON FERTILITY

As a result of anovulation and hypoestrogenic amenorrhea, infertility may occur. Patients may need induction of ovulation with clomiphene citrate or gonadotropins. However, there is also a high miscarriage rate of 25 to 30% in this population, as well as a high frequency of low-birth-weight babies born to women who are underweight.

## TREATMENT

Though there are a variety of approaches used to treat eating disorder-induced and exercise-induced amenorrhea, all treatments are oriented toward a return to normal weight. Treatments include combinations of the following: psychotherapy, psychoanalysis, drug therapy, behavior modification, dietary therapy, and reduction of exercise load. Treatment with hormone replacement therapy or oral contraceptive therapy is controversial and does not appear to be uniformly effective.

## See Also the Following Articles

Anorexia Nervosa • Body Proportions • Constitutional Delay of Growth and Puberty (CDGP) • FSH (Follicle-Stimulating Hormone) • Gonadotropin-Releasing Hormone (GnRH) Actions • LH (Luteinizing Hormone) • Menstrual Cycle: An Integrative View • Obesity, Childhood and Adolescence • Obesity Regulation • Osteoporosis, Overview • Pregnancy Endocrinology

## Further Reading

Loucks, A. B., Laughlin, G. A., Mortola, J. F., Girton, L., Nelson, J. C., and Yan, S. S. C. (1992). Hypothalamic-pituitary-thyroidal

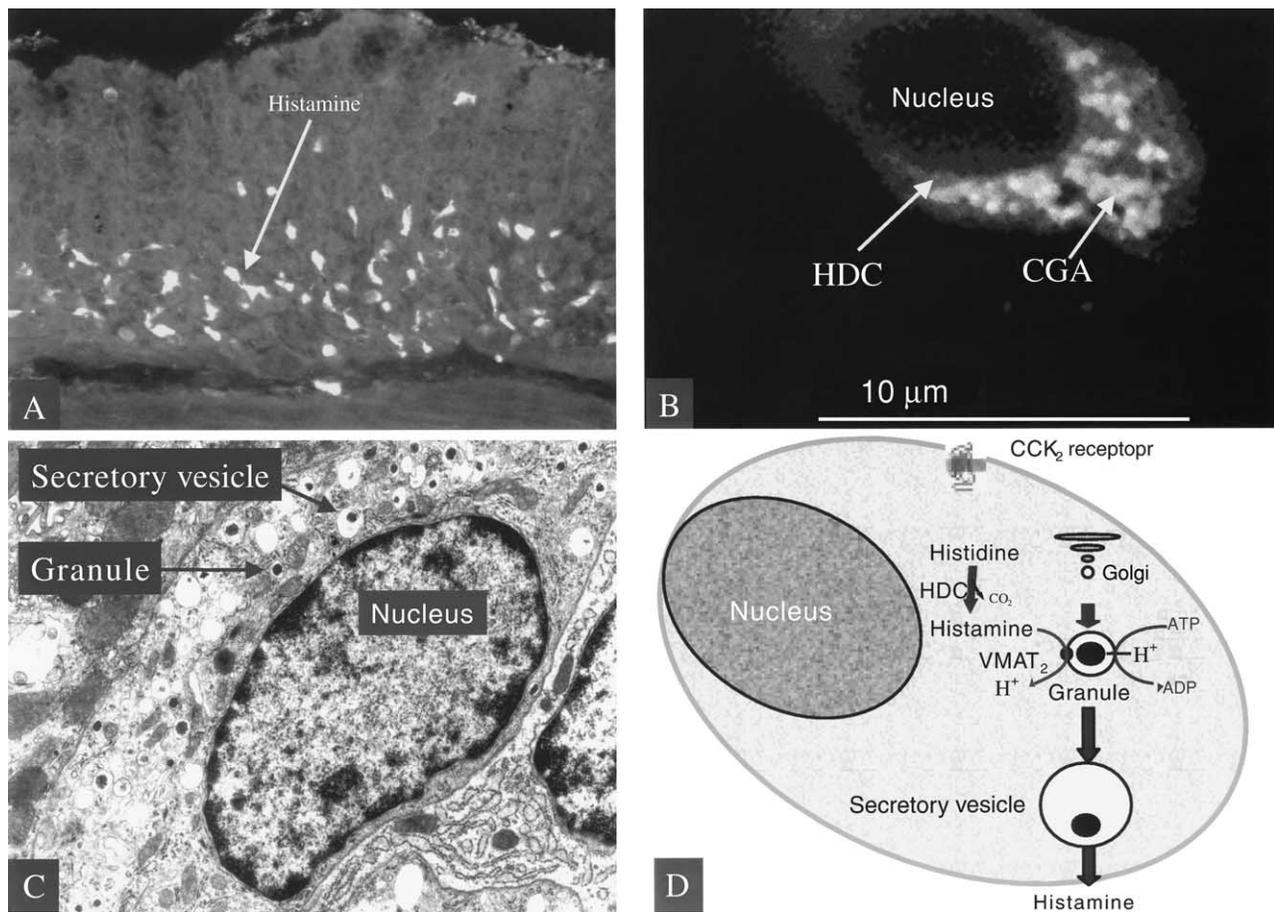
- function in eumenorrheic and amenorrheic athletes. *J. Clin. Endocrinol. Metab.* **75**, 514–518.
- Mehler, P. S. (2001). Diagnosis and care of patients with anorexia nervosa in primary care settings. *Ann. Intern. Med.* **134**, 1048–1059.
- Warren, M. P. (1997). Anorexia, bulimia, and exercise-induced amenorrhea: Medical approach. *Curr. Ther. Endocrinol. Metab.* **6**, 13–17.
- Warren, M. P., and Fried, J. L. (2001). Hypothalamic amenorrhea: The effects of environmental stresses on the reproductive system: A central effect of the central nervous system. *Endocrinol. Metab. Clin. N. Am.* **30**, 611–629.
- Warren, M. P., and Locke, R. J. (2001). Anorexia nervosa and other eating disorders. In “Principles and Practice of Endocrinology and Metabolism” (K. L. Becker, ed.), 3rd ed. pp. 1251–1257. Lippincott, Philadelphia, PA.

Table I Characteristic Features of ECL Cells

Properties in common with other peptide hormone-producing cells	Unique properties
Argyrophil staining	Histamine formation, storage, and secretion
Cytoplasmic secretory granules/vesicles	HDC
Chromogranins	Stimulation by gastrin (via CCK <sub>2</sub> receptors)
APUD properties	
Costorage of amines and peptides	
Proteolytic "processing" enzymes in granules/vesicles	
Exocytotic proteins	
PGP 9.5, Reg protein	

Note. APUD, amine precursor uptake and decarboxylation; CCK<sub>2</sub>, cholecystokinin-2; HDC, histidine decarboxylase; PGP 9.5, protein gene product, molecular weight 9.5 kDa; Reg, regenerating-gene.

that bud off the *trans*-Golgi network. These microvesicles are probably loaded with cargo, such as proteolytic enzymes and secretory peptide precursors (such as chromogranin A and prohormone). As a result of protein condensation (probably induced by low pH), the electron-lucent microvesicles are transformed into dense-cored granules while still in the *trans*-Golgi area. When the Golgi-associated granules embark on their journey toward the periphery of the cell, they start to take up histamine from the cytosol by means of the vesicle monoamine transporter type 2 (VMAT<sub>2</sub>), which is located in the granule/vesicle membrane. The continued accumulation of histamine is associated with transformation of the granule into a large electron-lucent (but dense-cored) secretory vesicle. As a consequence of stimulation of the cell



**Figure 1** (A) Immunofluorescence photograph of a transverse section of rat oxyntic mucosa after labeling with antiserum to histamine. Arrow points to one of the many ECL cells. (B) Immunofluorescence photograph of an ECL cell at high magnification after double-labeling with antisera to histidine decarboxylase (HDC) and chromogranin A (CGA). Confocal microscopy. Arrows indicate HDC in the cytosol and CGA in the secretory organelles. Bar, 10 μm. (C) Electron micrograph of an ECL cell. Arrows point to a typical secretory vesicle and a typical granule. (D) Cartoon illustrating the secretory pathway in the ECL cells from Golgi to exocytosis. HDC, histidine decarboxylase; VMAT<sub>2</sub>, vesicle monoamine transporter type 2.

(by, e.g., gastrin), the secretory vesicles will fuse with the cell membrane to release their contents by exocytosis (Fig. 1D). Small electron-lucent microvesicles in the periphery of the ECL cells are thought to be part of the regulated secretory pathway, arising from the process of membrane retrieval (endocytotic vesicles). In the case of exposure to sustained gastrin stimulation, the secretory vesicles in the cytoplasm will fuse with one another to form large vesicles, referred to as vacuoles. The development of vacuoles parallels the development of lipofuscin bodies, which probably arise from the oxidative conversion of native lipids and proteins to a polymerized, nondegradable, and chemically ill-defined material. Both vacuoles and lipofuscin bodies are crinophagic (and/or autophagic) organelles, which may incorporate or be associated with primary lysosomes to form part of the lysosomal compartment. The time-dependent accumulation of vacuoles and lipofuscin bodies is associated with a progressive impairment of ECL cell functions.

## FUNCTIONS OF ECL CELL PRODUCTS

Immunocytochemistry has revealed the presence of numerous potentially bioactive substances in the ECL cells (see Table I).

### Histidine Decarboxylase

Histidine decarboxylase (HDC) catalyzes the formation of histamine from histidine. Western blot analysis of oxyntic mucosal extracts has revealed the presence of at least three different isoforms of HDC, having molecular masses of approximately 74, 63, and 53 kDa; the 63 kDa form seems to be enzymatically active. ECL cells respond to gastrin with up-regulation of HDC at both the transcriptional and the translational/posttranslational levels and with down-regulation of its degradation.

### Histamine

ECL cell histamine plays an important role as mediator for gastrin-induced acid secretion. Gastrin stimulates the ECL cells by acting on cholecystokinin-2 (CCK<sub>2</sub>) receptors to mobilize histamine, which in turn stimulates the parietal cells via the histamine H<sub>2</sub> receptor to secrete acid. This pathway is referred to as the gastrin-ECL cell-parietal cell axis.

### Chromogranin A

Peptides of the chromogranin family (e.g., chromogranins A, B, and C) occur in endocrine cells through-

out the body. They probably function as chaperones in the packaging of secretory products in granules. It has also been suggested that they might be biologically active messenger molecules. The ECL cells are notably rich in pancreastatin (rat chromogranin A 264–314 amide) and pancreastatin-related products. In fact, the ECL cells contribute 70–90% of the circulating pancreastatin in rats.

## Exocytotic Proteins

Several exocytotic proteins and their isoforms, originally identified in the neurons, also occur in endocrine/neuroendocrine cells, including the ECL cells. These proteins include vesicle-associated membrane proteins (also called synaptobrevins), synaptotagmins, synaptosomal-associated protein of 25 kDa, syntaxin, Munc-18, and cysteine string protein. The process of exocytosis depends on these proteins. In the ECL cells, exocytosis is triggered by Ca<sup>2+</sup> entry, causing a rise in [Ca<sup>2+</sup>]<sub>i</sub>.

## Reg Protein

Reg protein is thought to be released from the ECL cells to act as a mediator of gastrin-induced growth of the oxyntic mucosa.

## REGULATION OF ECL CELL SECRETION

The ECL cells are controlled by a complex regulatory system. They carry the following receptors (preferred ligand is shown in parentheses): CCK<sub>2</sub> receptor (gastrin and CCK), PAC<sub>1</sub> receptor [pituitary adenylate cyclase-activating peptide (PACAP)], VPAC<sub>2</sub> receptor [vasoactive intestinal peptide (VIP)], somatostatin type 2 receptor, galanin receptor, β<sub>2</sub>-adrenergic receptor (adrenaline), EGF receptor, neuronal growth factor receptor type 1, fibroblast growth factor receptor type 1, and EP<sub>3</sub> receptor (prostaglandin E<sub>2</sub>). The nature of the ligands of these receptors is such that the regulation of ECL cell secretion can be categorized into endocrine, paracrine, and neural control. Briefly, gastrin, PACAP, and VIP stimulate ECL cell secretion, whereas somatostatin, galanin, and prostaglandin E<sub>2</sub> inhibit it. Interestingly, acetylcholine and histamine seem to have no effect on the ECL cells.

## REGULATION OF ECL CELL GROWTH

The turnover time of ECL cells in mice and rats is between 20 and 100 days. The ECL cells respond to sustained gastrin stimulation first with hypertrophy and then with general diffuse hyperplasia, which reflects a transient increase in the ECL cell self-replication rate. Potent anti-secretagogues, such as histamine H<sub>2</sub> receptor antagonists and proton pump inhibitors, are widely used clinically to treat acid-related disorders. An increased incidence of gastric carcinoids, later identified as ECL cell tumors or ECLomas, was noted in the murine stomach after long-term treatment with these acid inhibitors. In view of the known trophic effect of gastrin on ECL cells, the stimulus behind the development of ECL cell tumors is hypergastrinemia rather than achlorhydria. Thus, effective inhibition of gastric acid secretion abolishes the luminal acid feedback inhibition of the antral G cells and leads to hypergastrinemia, which in turn stimulates the ECL cells to divide, at first resulting in general diffuse ECL cell

hyperplasia, later resulting in focal hyperplasia with multiple micronodules, and finally resulting in frank tumors (carcinoids).

### See Also the Following Articles

CCK (Cholecystokinin) • Gastrin • GI Hormone Development (Families and Phylogeny)

### Further Reading

- Anderson, K., Chen, D., Mattsson, H., Sundler, F., and Håkanson, R. (1998). Physiological significance of ECL-cell histamine. *Yale J. Biol. Med.* **71**, 183–193.
- Chen, D., Zhao, C.-M., Lindström, E., and Håkanson, R. (1999). Rat stomach ECL cells: Update of biology and physiology. *Gen. Pharmacol.* **32**, 413–422.
- Håkanson, R., Chen, D., and Sundler, F. (1994). The ECL cells. In “Physiology of the Gastrointestinal Tract” (L. R. Johnson, ed.), 3rd ed., Vol. 2, pp. 1171–1184. Raven Press, New York.
- Prinz, C., Zanner, R., Gerhard, M., Mahr, S., Neumayer, N., Hohne-Zell, B., and Gratzl, M. (1999). The mechanism of histamine secretion from gastric enterochromaffin-like cells. *Am. J. Physiol.* **277**, C845–855.

these low-molecular-weight forms of the peptides are also the most readily available commercially (Fig. 1A).

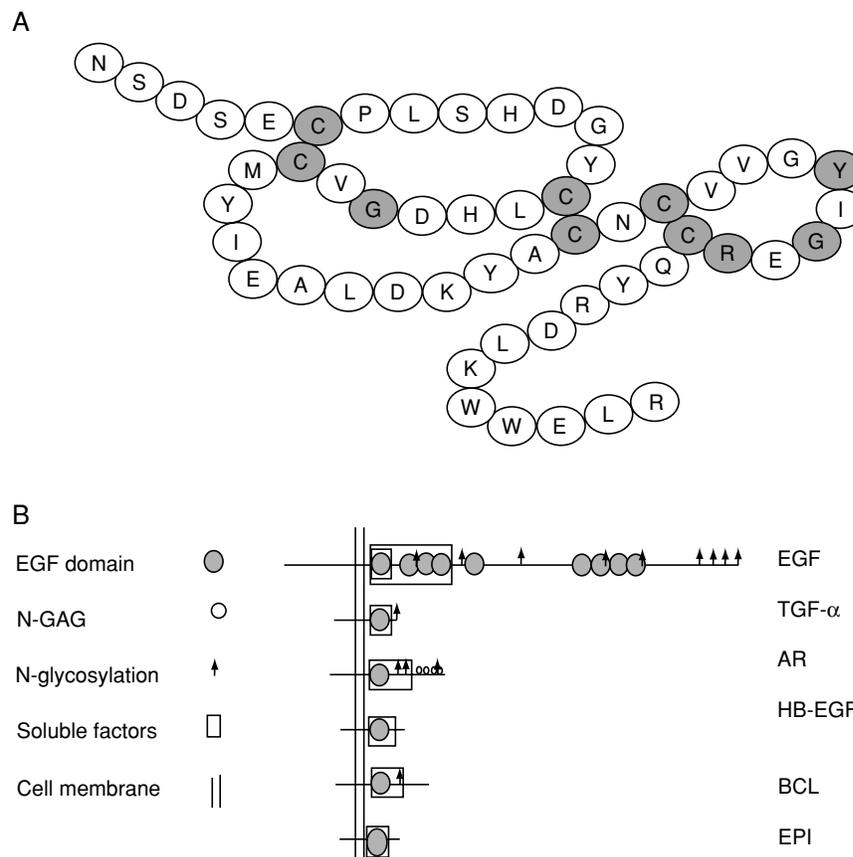
EGF and its related ligands are presented in Table I. The names of the ligands do not follow a systematic pattern, which can give rise to confusion. Transforming growth factor- $\alpha$  (TGF- $\alpha$ ), for example, is related to malignant growth no more than the other ligands. TGF- $\alpha$  was discovered together with TGF- $\beta$ , but the two ligands are not related. TGF- $\alpha$  belongs to the EGF family, whereas TGF- $\beta$  belongs to another group of growth factors. Heparin binding-EGF-like growth factor (HB-EGF) binds to heparin but so does amphiregulin.

The EGF motif is phylogenetically very old and is widely distributed. EGF-like peptides have been identified in primitive organisms such as *Caenorhabditis elegans* and the EGF motif is present in many types of proteins including receptors, enzymes, and matrix

proteins. Most likely, this motif is a multifunctional structure suitable for protein-protein interactions.

Even though the six EGF ligands are able to bind to the same receptor, only a few amino acids are conserved for all six peptides (Fig. 1A). The similarity between the individual peptides across species is considerably higher. For example, the mature 53-amino-acid EGF displays an identity of 64% among human, pig, rat, and mouse. Interestingly, other parts of the precursor molecule for EGF show an even higher degree of conservation. A stretch of 261 amino acids present in the EGF precursor shows an identity of 73% among human, pig, rat, and mouse, suggesting a conserved biological role also for this part of the molecule.

All of the ligands are synthesized as membrane-spanning peptides ranging from 162 to >1000 amino acids in length (Fig. 1B). Proteolytic cleavage



**Figure 1** The structure of epidermal growth factor and related ligands. (A) The amino acid sequence of human EGF comprises 53 amino acids and is characterized by three S-S bridges with a spacing that is almost identical for all of the EGF-related ligands. The few amino acids conserved among the six human ligands are indicated. The amino acids are indicated in single-letter code. (B) Precursor structure of the EGF-related ligands. The abbreviations employed are indicated in Table I. All the ligands are synthesized as membrane-spanning precursors. Both the membrane-bound forms and the forms released by enzymatic cleavage are biologically active. EGF is released in both low- and high-molecular-weight forms.

Table I Features of Human Epidermal Growth Factor and Related Growth Factors

Name of peptide	Abbreviation	Year of discovery and material in which it was isolated	No. of amino acids		Chromosomal location	Site where found	Receptor(s)	Function and relation to diseases other than cancer
			Precursor	Mature form				
Epidermal growth factor	EGF	1962, <sup>a</sup> /1975; mouse submandibular glands, human urine	1207	53	4q21-qter	Exocrine glands, kidney	HER1	Prevents gastrointestinal ulcers, <sup>a</sup> used to treat corneal ulcers
Transforming growth factor- $\alpha$	TGF- $\alpha$	1980; tumor cell lines	160	50	2p11-p13	Skin, brain	HER1	Growth of hair, eye development
Amphiregulin	AR	1988; MCF-7 breast carcinoma cells	252	84	4q13-q21	During implantation	HER1	Over expressed in psoriasis
Heparin-binding epidermal growth factor	HB-EGF	1991; macrophage-like U-937 cells	208	87	5q23	Lung, skeletal muscle, brain, and heart	HER1 and HER4	Diphtheria toxin receptor
Betacellulin	BCL	1993; breast adenocarcinoma	178	80	4q13-q21 <sup>b</sup>	Liver, kidney, weak in heart, lung, and intestine	HER1 and HER4	Promotes maturation of pancreatic beta cells
Epiregulin	EPI	1995 <sup>a</sup> ; human peptide identified by cloning	162	46		Macrophages, placenta	HER1 and HER4	

is needed to release the low-molecular-weight and soluble forms of the peptides. Though it is well known that, for example, EGF can be released as 6, 45, and 200 kDa molecules, little is known concerning the various forms of the EGF ligands and even less is known concerning the biological diversity found in molecular forms derived from the same precursor molecule of the ligand.

It has been questioned whether the membrane-spanning precursors could possibly play a role as receptors themselves. This is most obvious for the EGF precursor, both because it is the largest of all the precursors because of its well-conserved sequence and because it has structural features in common with the low-density lipoprotein (LDL) receptor. Thus far, however, a receptor function has been described only for the precursor of HB-EGF. In human and monkey, the HB-EGF precursor acts as the cellular receptor for diphtheria toxin.

One or more of the EGF-related ligands are present in all the tissues of the body (Table I). EGF has been the most thoroughly examined and its distribution is restricted to a few sites. It is produced in exocrine glands and in the kidney and it is secreted into urine and secretions such as tears and saliva. In contrast, other members of the EGF system, such as TGF- $\alpha$ , are more widely expressed.

The protein level of the peptides was initially determined from rather insensitive receptor-binding assays. Antibodies against the low-molecular-weight forms of all the peptides except for epiregulin have become commercially available and thus it is possible to perform both immunohistochemical studies and measurements employing, e.g., enzyme-linked immunosorbent assay methodologies. A major problem concerning such studies is that relatively little is known concerning the ability of the antibodies to recognize the precursor forms of the EGF ligands.

By and large, the ligands function close to the place of their synthesis. They act as paracrine or even autocrine growth factors and it is doubtful that they have a function as classical hormones produced in one place and transported through the bloodstream to act at other organs.

Many of the ligands are active in both their membrane-bound and soluble forms. Thus, proteolysis of the ligands may regulate whether they act on distant cells or whether they are able to act only on neighboring cells. Some of the proteolytic enzymes responsible for the release of membrane-bound precursors have been identified, such as the metalloproteinases releasing HB-EGF, and they represent yet another mechanism for regulating the activity of the EGF system.

In conclusion, the general picture emerging for the EGF ligands is that of a group of structurally related growth factors synthesized as precursor molecules and most likely able to act as membrane-bound peptides as well as after proteolytic release. The activity is mediated by binding to one or more of the receptors belonging to the EGF system.

## MODE OF ACTION

The EGF system promotes a number of actions at the cellular level. The earliest described and the most comprehensively studied function is the stimulation of proliferation. In particular, an increased activity of the EGF system is found associated with the increased proliferation observed in a number of cancer cells. Other functions include cellular migration, chemotaxis, cellular adhesion, and cellular differentiation. Activation of the EGF system may also prevent cells from entering the programmed cell death pathway. Though it is evident that many different signals may be the result of interaction between EGF ligands and their receptors, the exact pathways leading to growth as compared to lack of apoptosis or differentiation have yet to be described.

The activity of the EGF ligands requires the presence of membrane-bound receptors. Four receptors, HER1–4 (human EGF receptor), are known. Three of them (HER1, HER3, and HER4) are recognized by one or more of the EGF ligands, whereas no ligands able to bind to HER2 have been identified. Three of the receptors (HER1, HER2, and HER4) have tyrosine kinase activity and are able to phosphorylate tyrosine residues.

All of the receptors are membrane-spanning single-chain glycoproteins of approximately 200,000 kDa. The ligand-binding site is present in the extracellular part of the molecule, whereas the tyrosine kinase activity is mapped to the intracellular part of the molecule. A protein domain containing several tyrosine phosphorylation sites further characterizes the intracellular portion. Activation of these sites represents the first step in the signal transduction cascade. In addition, the intracellular part contains a number of other phosphorylation sites involved in the regulation of the EGF receptor by other proteins. The intracellular part of the receptor also contains binding sites for tyrosine phosphatases. These enzymes are responsible for down-regulating the number of tyrosine-phosphate residues formed and are thereby able to modify the signal induced by the interaction of the ligand and the receptor.

Binding of the EGF ligands to the EGF receptor induces dimerization either in the form of a homodimer of two receptors, e.g., two HER1 molecules, or in the form of a heterodimer, where, e.g., HER1 dimerizes with HER2. All the ligands described in this article bind to HER1 but some of them also bind to HER4 (HB-EGF, betacellulin, epiregulin) (Table 1). Increased signaling diversity is added to the system by heterodimerization with either HER2 or HER3. HER2 apparently has no activating ligand but potentiates signaling and HER3 increases the signaling diversity of the system although it has no active kinase domain.

The model for transmitting the activity of the EGF ligand present outside the cell into a signal within the cell involves several steps. First, the ligand binds to the preferred receptor. This induces both dimerization and activation of the tyrosine kinase activity of the receptor. The activation induces phosphorylation of tyrosine residues present on the intracellular part of the receptors though a reaction often referred to as autophosphorylation. But in fact, it is thought that one partner of the formed receptor dimer phosphorylates the other and vice versa.

The phosphorylated tyrosines on the intracellular part of the activated receptors generate docking sites for intracellular proteins containing SH2 (Src homology 2) or other phosphotyrosine-binding domains. These activated adapter molecules then activate a series of signaling molecules in the cytoplasm, which results in the activation of many different signal transduction pathways, such as the mitogen-activated protein kinases, phosphoinositol kinase, and protein kinase A. Ultimately, the activation of the signal transduction pathways results in regulation of the expression of specific genes in the nucleus.

One of the immediate results of signaling through the EGF pathway is an up-regulation of the EGF ligands synthesized by the cells, with HB-EGF and amphiregulin being induced earliest. This self-induction is considered to represent a local amplification of the signal.

Thus far, only activation of the EGF system has been discussed, but obviously it is as important to understand how this signal is turned off again. One mechanism is dephosphorylation of the phosphorylated tyrosine residues on the activated receptors by intracellular phosphatases. In fact, it has been discussed whether the phosphatases are the main determinants of the activity of the EGF receptors. In this scenario, a low phosphatase level will result in a high degree of activity, whereas a high phosphatase level will prevent the presence of phosphorylated tyrosines

on the receptors and thereby prevent the initiation of the signaling cascade.

Another mechanism for termination of signaling through the EGF system is internalization of the ligand-receptor complex by endocytosis. After activation of the receptor by a ligand, the ligand-receptor complex is taken up by the cell through the clathrin-coated regions of the cell membrane. In the endosome, the internalized ligand is degraded and the receptors may be either recycled or degraded. A key protein determining whether a receptor is recycled or degraded is Cbl. Cbl binds to the activated receptor after internalization, which results in ubiquitination and degradation of the receptor. In the absence of Cbl, the receptor is transported back to the cell membrane and is able to engage in the signal transduction process again. The stability of the internalized receptor dimers is an important determinant of the recycling pathway controlled by Cbl. Studies on the degradation of the different forms of receptor homodimers and heterodimers have shown that HER1 receptor homodimers have a high level of stability, whereas heterodimers have a lower level of stability. Consequently, the unstable heterodimer complex uncouples before it is bound by Cbl and thereby the receptors are recycled to the cell membrane. This mechanism contributes to the signaling superiority of the heterodimeric receptor complexes.

In summary, the EGF system with its many receptors and ligands represents a complicated alphabet able to transmit a wide variety of messages from outside the cell into the nucleus. The net result of the signaling can be observed, but how to interpret the many possible combinations between ligands and receptors has yet to be determined.

## LESSONS FROM GENE-MANIPULATED MICE AND PHARMACOLOGICAL STUDIES

At first it was believed that the majority of EGF was synthesized in the submandibular glands and that deficiency could be studied by removal of these glands. Then the picture emerged of a number of ligands belonging to the EGF system, which were able to replace one another's function and which were synthesized locally in all regions of the body. In this scenario, deficiency of the EGF system can be studied only employing knockout animals and severe effects can be expected only if the common key point, the receptor, is nonfunctioning.

The TGF- $\alpha$  knockout mice have been thoroughly studied. Small but distinct features characterize these animals. They develop changes in the growth of the fur, including curly whiskers, and their eyes do not develop normally. Also, an EGF knockout animal has been developed, but lack of EGF does not impose a characteristic feature on these animals.

Quite a different picture is observed for animals lacking the EGF receptor (HER1). Most of these animals die *in utero* and those that are born display severe defects in the lungs and in the gastrointestinal system. Also, animals lacking HER2 or HER4 die *in utero*. In these animals, the major malformations occur in the heart. In addition, correct development of the brain is dependent on the presence of EGF receptors and here HER2 and HER3 seem to be the key players.

Drugs able to inhibit the activity of HER1 or HER2 have been developed and used for the treatment of cancer. From a physiological point of view, it is interesting that drugs interfering with HER1 result in side effects related to the skin, such as acne, whereas drugs directed toward HER2 may induce cardiac symptoms. Thus, in the adult human organism, the skin seems to be the tissue most sensitive to a lack of HER1 signaling, whereas the heart seems to be the organ most sensitive to signaling through HER2.

An excess supply of EGF ligands has been studied in two ways. In transgenic animals, increased expression of TGF- $\alpha$  leads to premalignant or malignant transformations in, e.g., the skin and the mammary glands. Less dramatic is the result of an overexpression of amphiregulin. It decreases the final maturation of the keratinocytes and leads to a condition similar to psoriasis.

Another path has been to inject large doses of EGF into rats or mini-pigs. Such studies have shown that EGF promotes the growth of the urinary tract in particular. After 3 weeks of treatment, the ureter of the mini-pig more than doubles its size, but after cessation of therapy, the ureter slowly decreases in size toward pretreatment values. It is not known whether the potential role of the EGF ligands in tumor development precludes the use of these ligands in situations where growth of tissue is required. The potential of the EGF ligands in this regard is obvious as judged from the studies performed thus far.

Studies on gene-modified animals have confirmed the notion that the EGF ligands by and large may replace one another and have also stressed the role of the ligands in connection with the development of cancer. The few data available also indicate that even though the ligands may often replace one another, each of them is likely to play a unique role.

## EGF LIGANDS AND CANCER

It has already been noted that there is a tight connection between the EGF system and the development and growth of cancer. The EGF receptor is a proto-oncogene and one or more of the receptors are overexpressed in numerous types of cancer, including breast cancer and other cancers of epithelial origin. The activity of the EGF system is regulated both by the expression of the receptors and by the ligands and it is well known that coexpression of both a receptor and a ligand from the EGF system is related to a poor prognosis for the patient. It has been observed that the expression of a subset of the EGF ligands (epiregulin, TGF- $\alpha$ , and HB-EGF) is a stronger predictor for survival in patients with bladder cancer than the expression of the receptors. The precise mechanism behind the increased expression of the EGF ligands is unknown; gene amplification for any of the ligands has not been described. This is in contrast to the situation for HER2, where gene amplification is observed in approximately 20% of breast cancer patients.

Hundreds of papers describe the possible usefulness of measurement of the EGF ligands and their receptors as prognostic factors in patients with cancer, but a far more exciting development has taken place. The first drug able to inhibit signaling through the EGF system (Herceptin or trastuzumab) has now been registered for treatment of patients with breast cancer that overexpresses HER2. Herceptin is a humanized monoclonal antibody able to inhibit HER2. The first clinical phase III trial on breast cancer patients published in 2001 showed a remarkable effect. The time to disease progression increased from 4.6 to 7.4 months in patients receiving Herceptin plus chemotherapy as compared to patients receiving chemotherapy alone. The same drug is also being tested for the treatment of a variety of tumors overexpressing HER2, including pancreatic cancer, colorectal cancer, and bladder cancer. Similarly, monoclonal antibodies against HER1 are currently being tested for the treatment of patients with head and neck cancer.

Another strategy for decreasing signaling through the EGF system is to use drugs that are able to inhibit the tyrosine kinase activity of the receptors. A number of such drugs are being tested for the treatment of non-small-cell lung cancer and head and neck cancer. Other strategies involve the introduction into the tumor of a transcription factor, E1A, that is able to decrease the expression of HER2. Initial experiments on patients with ovarian or breast cancer have shown that this principle may also work.

The role of the EGF system in the development and growth of cancer is well established and data showing that inhibition of the system is a new treatment modality for cancer are rapidly accumulating. The treatment is well accepted and improves the condition of the patient. Thus far, the reported side effects have been limited. In view of the widespread occurrence of the EGF system and in view of its potential role for tissue regeneration, this has been a pleasant surprise.

## UPCOMING CHALLENGES

Inhibition of the EGF system seems to be a challenging new treatment modality for cancer. To fully explore this possibility, there is a need to know more about the normal physiology of EGF and the related ligands. The mode of synthesis and release needs to be explored and more must be learned about the differences between normal and diseased tissue. Only with such knowledge at hand will it be possible to ensure a rational way to treat the cancer cells without damaging other tissues of the body. At the same time, it is to be expected that a more profound knowledge of the EGF ligands will allow this system to be modulated in order to promote growth and differentiation in situations such as restoration after tissue damage.

## See Also the Following Articles

Fibroblast Growth Factor (FGF) • GI Hormones as Growth Factors • GI Hormones in Cancer • Growth Factor Receptors • Hepatocyte Growth Factor • Insulin-like Growth Factors • Peptide Hormones and Growth Factors: Cellular Signaling Mechanisms • Platelet-Derived Growth Factor (PDGF)

## Further Reading

- Cohen, S. (1986). Nobel lecture. Epidermal growth factor. *Biosci. Rep.* **12**, 1017–1028.
- Prenzel, N., Fischer, O. M., Streit, S., Hart, S., and Ullrich, A. (2001). The epidermal growth factor receptor family as a central element for cellular signal transduction and diversification. *Endocr. Relat. Cancer* **1**, 11–31.
- Slamon, D. J., Leyland-Jones, B., Shak, S., Fuchs, H., Paton, V., Bajamonde, A., Fleming, T., Eiermann, W., Wolter, J., Pegram, M., Baselga, J., and Norton, L. (2001). Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N. Engl. J. Med.* **344**, 783–792.
- Thogersen, V. B., Sorensen, B. S., Poulsen, S. S., Orntoft, T. F., Wolf, H., and Nexø, E. (2001). A subclass of HER1 ligands are prognostic markers for survival in bladder cancer patients. *Cancer Res.* **61**, 6227–6233.
- Vinter-Jensen, L. (1999). Pharmacological effects of epidermal growth factor (EGF) with focus on the urinary and gastrointestinal tracts. *Acta Patbolog. Microbiol. Immunol. Scand. Suppl.* **93**, 1–42.

required for differentiation of the fetal gonads into testes, which then produce the hormones necessary for sexual differentiation of other structures. The configuration will be female unless the fetal testis secretes Müllerian-inhibiting substance (MIS) to induce regression of the Müllerian duct and testosterone to both maintain the Wolffian duct and masculinize external genitalia.

This process can be illustrated with a pathological example, testicular feminization mutation (i.e., androgen insensitivity syndrome), in which genetic males have an X-linked mutation that renders their androgen receptors (ARs) insensitive to the actions of their endogenous androgenic hormones. Because they have a functional SRY gene, individuals with this mutation will develop testes that secrete testosterone and MIS; however, their tissues cannot be masculinized without functional ARs, so they will appear to be female at birth. Incomplete androgen insensitivity leads to intermediate sexual differentiation, with hypospadias (displacement of urethral opening from the tip of the phallus) and cryptorchidism (undescended testes) being common outcomes.

### Effects Seen in the Laboratory: Purposeful Disruption

Disruption of male sexual differentiation similar to that seen in individuals with incomplete androgen insensitivity syndrome can be mimicked in the laboratory by exposing animals to chemicals that antagonize the action of endogenous androgens at the AR (antiandrogens). Perinatal exposure to flutamide (a powerful pharmaceutical AR antagonist) or to pesticides with antiandrogenic activity (e.g., vinclozolin, linuran) causes female-like anogenital distance, hypospadias, cryptorchidism, decreased weights of sex accessory glands, and retained mammary tissue in male rodents and rabbits, even at low doses. Interestingly, some of the same developmental defects can be induced in males by perinatal exposure to estrogenic pharmaceuticals such as diethylstilbestrol (DES) and ethinyl estradiol. The mechanism of estrogen action is unclear but has been speculated to involve abnormal gene imprinting, interference with androgen production, androgen/estrogen ratio imbalance, and/or interference with testis development. Abnormal testis development potentially can lead to abnormal sexual differentiation (because normal differentiation depends on normally functioning testes) but also can have functional effects that are not apparent until puberty or later (e.g., abnormal sperm production,

testicular cancer). Cryptorchidism itself can lead to abnormal sperm production and increased risk of testicular cancer. In controlled laboratory settings, perinatal exposure to antiandrogens or environmental estrogens also produces functional effects, including changes in basal and luteinizing hormone-releasing hormone (LHRH)-induced gonadotropin release and circulating levels of sex steroids, reduced sperm count, and changes in average testis weight.

## THE ENDOCRINE DISRUPTION HYPOTHESIS

Consideration of morphological and functional abnormalities detected in wildlife and epidemiological studies in the context of results from the laboratory studies mentioned previously gave rise to the endocrine disruption hypothesis, that is, that previously unexplained reproductive system abnormalities in wildlife and human populations might be caused by exposure to environmental chemicals shown, under controlled conditions, to disrupt endocrine function.

## EVIDENCE

### Wildlife and Epidemiological Indicators

Regarding EDC effects on male sexual differentiation, most evidence from studies on wildlife comes from nonmammalian species such as demasculinized/feminized Florida alligators exposed to a pesticide spill and agricultural runoff and feminized fish exposed to estrogenic sewage effluent.

Most scientists agree that testicular cancer rates among humans have been rising. An analysis of sperm count data from 1934 to 1996 confirmed earlier studies indicating that sperm densities have been declining in Western countries, although some scientists have pointed out that methodological factors (e.g., geographical variability) might confound the analysis. Congenital abnormalities, such as hypospadias and cryptorchidism, appear to have increased in frequency in some populations, notably in England and Denmark, but also appear to have increased and then leveled off around 1985 in many populations.

For most studies, there are little or no data on EDC exposure or tissue loads; even for those studies that do have this information, the results are only correlative. Performing prospective double-blind controlled studies on the effects of EDCs on humans is not possible, so the best causal evidence comes from industrial accidents and an unfortunate medical

treatment mistake: the use of DES to prevent miscarriage. Males exposed to DES *in utero* have a higher prevalence of cryptorchid testes, epididymal cysts, hypoplastic testes, and abnormal sperm than do unexposed men, and these effects mimic those seen in rodents exposed to DES *in utero*. Studies on the effects of this potent pharmacological estrogen provide insight to possible mechanisms of EDC action, but typical EDC loads in the general public are unlikely to reach this potency and also might include EDCs with other modes of action (e.g., antiandrogenic). However, the general public is at risk for some exposure; one study found that one-third of routine amniocentesis samples (taken during the second trimester) had detectable levels of EDCs. Possible EDC exposure sources for children *in utero* and/or postnatally include occupational exposure, industrial spills, plasticizers in plastic products, agricultural runoff and overspray, diet (e.g., fish from contaminated waters, meat from animals treated with hormones), and some cosmetic products (e.g., hair care). Proponents of the endocrine disrupter hypothesis warn that laboratory studies on EDC effects, combined with epidemiological evidence of increasing rates of male reproductive disorders, provide cause for concern about EDC exposure, particularly *in utero* during the critical periods for organogenesis and sexual differentiation.

### Caveats

The difficulty in establishing EDC causation for abnormal health trends, and in some cases in verifying the trends themselves, has led to criticism of the endocrine disrupter hypothesis. Critics recognize that the hypothesis is biologically plausible but call for stronger evidence than is currently available.

### See Also the Following Articles

Androgen Biosynthesis Inhibitors and Androgen Receptor Antagonists • Androgen Insensitivity Syndrome • Androgens, Gender and Brain Differentiation • Germ Cell Differentiation Signaling Events, Male • Goitrogens, Environmental • Müllerian Inhibiting Substance: New Insights • Testes, Embryology of • Undescended Testes

### Further Reading

- Crews, D., Willingham, E., and Skipper, J. K. (2000). Endocrine disruptors: Present issues, future directions. *Q. Rev. Biol.* **75**, 243–260.
- Gray, L. E., Jr., Ostby, J., Furr, J., Wolf, C. J., Lambright, C., Parks, L., Veeramachaneni, D. N., Wilson, V., Price, M., Hotchkiss, A., Orlando, E., and Guillette, L. (2001). Effects of environmental antiandrogens on reproductive development in experimental animals. *Hum. Reprod. Update* **7**, 248–264.
- Gulledge, C. C., Burow, M. E., and McLachlan, J. A. (2001). Endocrine disruption in sexual differentiation and puberty: What do pseudohermaphroditic polar bears have to do with the practice of pediatrics? *Pediatr. Clinics North America* **48**, 1223–1240.
- International Programme on Chemical Safety (2002). “Global Assessment of the State-of-the-Science of Endocrine Disruptors” (T. Damstra *et al.*, eds.). World Health Organization. Available online at [www.who.int/pcs/emerga\\_site/edc/global\\_edc\\_TOC.htm](http://www.who.int/pcs/emerga_site/edc/global_edc_TOC.htm).
- Joffe, M. (2001). Are problems with male reproductive health caused by endocrine disruption? *Occup. Environ. Med.* **58**, 281–288.
- McLachlan, J. A. (2001). Environmental signaling: What embryos and evolution teach us about endocrine disrupting chemicals. *Endocrine Rev.* **22**, 319–341.
- Sharp, R. M. (2001). Hormones and testis development and the possible adverse effects of environmental chemicals. *Toxicol. Lett.* **120**, 221–232.
- Swan, S. H., and vom Saal, F. S. (2001). Alterations in male reproductive development: The role of endocrine disrupting chemicals. In “Endocrine Disruptors in the Environment” (M. Metzler, ed.), vol. 3, pp. 131–170. Springer-Verlag, Heidelberg, Germany.

for endometriosis. Many women will present with no physical symptoms, yet endometriosis will be discovered during the course of an infertility evaluation. Other women may present with a pelvic mass on routine annual gynecological examination.

Another common clinical manifestation is that of cyclic symptomatology. With classic pelvic endometriosis, pelvic pain, dysmenorrhea, and dyspareunia may worsen with menses. This is presumed to be due to localized bleeding that causes irritation at the sites of the endometriotic implants. Locally, the bowel may be affected, causing dyschezia or diarrhea with menses. In rare cases, endometriotic foci may be located in other parts of the body, such as a surgical scar (causing cyclical incision bleeding), in the lungs (causing cyclic hemoptysis or pneumothorax), or in the brain (causing catamenial seizures). When unusual, seemingly nongynecological symptoms occur only with menses, endometriosis often becomes part of the differential diagnosis.

## DIAGNOSIS

The diagnosis of endometriosis requires direct visualization by laparoscopy or laparotomy, preferably with biopsy for histologic confirmation. However, when the clinician suspects endometriosis, the first step is to obtain a history and perform a physical examination. Abdominal examination may reveal diffuse or focal tenderness in the lower quadrants or occasionally rebound tenderness or guarding. Pelvic examination including rectovaginal examination may reveal cervical motion tenderness, adnexal tenderness, or masses or nodularity of the uterosacral ligaments. Although the above findings are suggestive of endometriosis, they are insufficient to make a definitive diagnosis.

Other less invasive modalities have been studied including CA-125 levels and radiologic imaging studies. CA-125 is a tumor marker for certain types of ovarian tumors and has been studied for use in the diagnosis of endometriosis. Although levels are elevated in advanced disease, the rise is nonspecific and is not seen in minimal or mild disease. Therefore, the test is of limited value except in estimating the risk that a pelvic malignancy exists. Pelvic ultrasonography is useful for evaluating pelvic masses that are suspected to be endometriomas. An ovarian cyst with homogenous low-level internal echoes suggests the presence of an endometrioma. Magnetic resonance imaging may occasionally detect endometriosis, but has a very poor sensitivity.

**Table I** Locations of Endometriosis

Common sites	Less common sites
Ovary	Brain
Pelvic peritoneum	Lung
Uterine serosa	Bowel
Fallopian tube	Ureter
Posterior cul-de-sac	Kidney
Uterosacral ligaments	Appendix
Rectovaginal septum	Vulva
Anterior cul-de-sac	Incision

The diagnosis of endometriosis is made by visualization, most often at the time of laparoscopy. Most authors advocate biopsy to confirm the diagnosis, but this is not always carried out. A thorough examination of the pelvis is performed, paying particular attention to sites such as the ovaries, ovarian fossae, uterosacral ligaments, and pouch of Douglas, which are the most common sites for endometriotic lesions. These lesions may have a myriad of appearances including powder burn lesions, fleshy lesions, adhesions, and peritoneal pockets. Some of the sites where endometriosis may be found are shown in [Table I](#).

At this time in the diagnosis, the extent of the disease may be staged. The most commonly used staging system is the American Society for Reproductive Medicine (ASRM) revised classification system. The system involves a point system based on the size, depth, and location of both implants and adhesions. Points are tallied on a form and a stage is assigned based on the number of points. These include stage I (minimal, 1–5 points), stage II (mild, 6–15 points), stage III (moderate, 16–40 points), and stage IV (severe, 40+ points). Although the staging system may be useful for assessment of fertility potential, the ASRM stage does not correlate well with the degree of pelvic pain.

## TREATMENT

The treatment of endometriosis depends on the individual patient's presenting symptoms. A woman who is asymptomatic but is diagnosed with endometriosis at the time of an unrelated surgical procedure (e.g., a laparoscopic tubal sterilization) does not necessarily require treatment. In the absence of pain or infertility, the disease may be managed expectantly. When a patient presents with endometriosis-related pain, she may be treated using a myriad of various medical or surgical therapies. On the other hand, a patient who presents with endometriosis and infertility may be

**Table II Medical Treatments for Endometriosis**

<b>Oral contraceptives</b>
Continuous
Cyclic
<b>Progestins</b>
Medroxyprogesterone acetate, norethindrone, gestrinone
<b>Danazol</b>
<b>GnRH agonists</b>
Leuprolide, nafarelin, goserilin, buserelin, histrelin
<b>Antiprogestins</b>
Mifepristone
<b>Aromatase inhibitors</b>
Anastrozole, letrozole

treated using surgical means or other types of infertility treatments that are not directly related to endometriosis.

Medical treatment for endometriosis-related pain involves the use of hormonal therapies to suppress ovulation, as listed in [Table II](#). The rationale for this type of treatment is that the implants of endometriosis are similar to normal eutopic endometrium and will respond to estrogen and progesterone in the same way. Thus, the same hormonal therapies that can induce amenorrhea in women with dysfunctional bleeding would be expected to decrease pain in women with pelvic endometriosis.

Most commonly, a “pseudo-pregnancy-type” regimen is used with oral contraceptives for initial treatment. Due to the benign nature of this therapy, treatment is often initiated prior to obtaining a definitive diagnosis. A standard monophasic 30 or 35 µg birth control pill is taken continuously, with no days off and no placebo pills. This will mimic the hormonal milieu found during pregnancy in which high levels of estrogen and progesterone are observed. In general, pregnant women with endometriosis experience a significant abatement of symptoms during pregnancy. Alternatively, progestational agents may be used. Depot formulations of progestational agents, such as medroxyprogesterone acetate (Depo-Provera), or daily oral progestins, such as medroxyprogesterone acetate (Provera, Cycrin) and norethindrone (Aygestin), may be used.

An even more effective way to treat endometriosis pain involves what is called a “pseudo-menopause” regimen. Various medications can be used to suppress the hypothalamic–pituitary axis, resulting in a hypogonadotropic, hypoestrogenic state. One of the first pseudo-menopausal agents to be widely used was Danazol (Danocrine). Danazol is an orally active androgenic steroid believed to both suppress ovulation

and act directly to cause atrophy of the endometriotic implants. Given in doses of 200 to 400 mg twice daily, it is accompanied by side effects that may include hot flashes, vaginal dryness, muscle aches, deepening of the voice, and hirsutism. Gonadotropin-releasing hormone (GnRH) agonists, such as leuprolide (Lupron), goserilin (Zoladex), nafarelin (Synarel), and buserelin, have become the agents of choice to induce a pseudo-menopause. These medications are all administered parenterally. Some are designed to be used monthly or even at 3-month intervals for prolonged suppression of endogenous gonadotropins. This suppression will create a hypoestrogenic milieu and will cause the involution of the endometriotic implants, which are estrogen-dependent. The side effects of the GnRH agonists include headaches, hot flashes, vaginal dryness, and bone loss. In order to prevent the sequelae of osteoporosis later in life, Danazol and GnRH analogues are limited to 6 to 9 months of use.

The GnRH agonists may also be used on a longer term basis when accompanied by low doses of estrogen plus progestin replacement. This type of regimen is referred to as add-back therapy and its rationale is based on the estrogen threshold hypothesis. This hypothesis states that endometriosis and bone loss respond to different circulating estrogen levels. The goal thus is to attain an estrogen level that is high enough to conserve bone density yet low enough to prevent regrowth of the endometriosis lesions. The patient is started on a long-acting GnRH agonist, and concurrently or at a later date, low-dose estrogen/progestin replacement therapy is added in a dose that is similar to that given to menopausal women. Alternatively, a norethindrone-only regimen may be used. Add-back therapy has been demonstrated to be safe and effective when used for up to 1 year. Although there exists no general consensus on how long a patient may be kept on such a regimen, many clinicians will continue such therapy for a number of years, with intermittent monitoring of bone density by DEXA scan to ensure that there is no significant bone loss.

Treatment with anti-progestins and aromatase inhibitors is considered experimental and will not be reviewed in this article.

Surgical therapy may also be employed for the treatment of endometriosis. Frequently, surgical treatment is performed at the time of the initial laparoscopy when the diagnosis of endometriosis is made. After the diagnosis is established, the lesions discovered may be fulgurated or resected. Various modalities have been used to ablate the implants including various lasers (CO<sub>2</sub>, KTP, Nd-YAG), bipolar or monopolar cautery, endocoagulation, the harmonic

scalpel, and sharp resection. Although many physicians have advocated the use of laser for ablation, there are no data that show better outcomes associated with a specific energy source. At the time of surgery, any pelvic adhesions noted may be lysed in order to reduce the pain. Additionally, the uterosacral ligament, which carries sensory nerve fibers from the uterus, may be ablated at this time to assist in pain relief. If an endometrioma or chocolate cyst is encountered, when possible the cyst is opened, the chocolate fluid is aspirated, and the cyst wall is excised.

Other surgical procedures are available for the treatment of endometriosis. The procedures described above for laparoscopy may also be performed at the time of laparotomy. A presacral neurectomy may be performed to interrupt the sympathetic nerves at the level of the superior hypogastric plexus. In some instances, one or both ovaries may be removed and sometimes a hysterectomy may be required. Many patients with intractable endometriosis pain and multiple prior surgeries benefit from a "pelvic clean-out," which involves total abdominal hysterectomy with bilateral salpingoophorectomy. It should be noted that even this most definitive of therapies is only approximately 90% effective in the resolution of pain.

## ENDOMETRIOSIS AND INFERTILITY

The incidence of endometriosis in infertility patients undergoing diagnostic laparoscopy is relatively high (38%), yet the association between endometriosis and infertility remains unclear. Advanced cases of endometriosis (stage III and IV disease) are associated with decreased fertility, presumably on the basis of pelvic adhesions and mechanical factors. Less advanced cases (stages I and II) are considered by some authors to be causal of infertility and others as merely coincidental. The suggested mechanisms include interference of peritoneal fluid with sperm function and fertilization and luteal phase dysfunction.

A meta-analysis by Hughes on the treatment of infertility in early stage endometriosis found no benefit of medical treatment on pregnancy rates, but a possible benefit of surgical treatment. Medical treatment appears not to be beneficial and in fact may be detrimental by causing a delay in pregnancy due to suppression of ovulation. The use of laparoscopic surgery to treat early stage endometriosis was evaluated in two randomized controlled trials. A large Canadian multicenter study showed an increase in monthly fecundity in patients in whom endometriosis was ablated compared with those in whom it was left

*in situ*. A much smaller Italian trial failed to show any benefit. Thus, it appears likely though not certain that early stage endometriosis may benefit from surgical treatment.

Infertility in endometriosis patients can be treated by the same modalities used to treat those with unexplained infertility. Superovulation with gonadotropins accompanied by intrauterine insemination resulted in a significant increase in pregnancy rates in women with unexplained infertility in a large multicenter study. *In vitro* fertilization is also very effective in women with early and advanced stages of endometriosis.

## SUMMARY

Endometriosis is an enigmatic disease that is defined by the presence of endometrial tissue in an ectopic location. The etiology remains unknown, though there are certain changes in the immune system as well as an association with abnormal outflow tracts. The symptoms of endometriosis include various manifestations of pelvic pain and infertility and are often temporally related to menses. Diagnosis is made by surgical visualization and treatment may involve hormonal manipulation and/or surgical ablation or resection.

## See Also the Following Articles

Infertility, Overview • Menstrual Cycle: An Integrative View • Polycystic Ovary Syndrome: Implications for Cardiovascular, Endometrial, and Breast Disease • Pregnancy Endocrinology

## Further Reading

- Guzick, D. S., Sullivan, M. W., Adamson, G. D., Cedars, M. I., Falk, R. J., Peterson, E. P., and Steinkampf, M. P. (1998). Efficacy of treatment for unexplained infertility. *Fertil. Steril.* **70**, 207–213.
- Hornstein, M. D., Surrey, E. S., Weisberg, G. W., and Casino, L. A. (1998). Leuprolide acetate depot and hormonal add-back in endometriosis: A 12-month study. *Lupron Add-Back Study Group. Obstet. Gynecol.* **91**, 16–24.
- Kennedy, S., Bennett, S., and Weeks, D. E. (2001). Affected sib-pair analysis in endometriosis. *Hum. Reprod. Update* **7**, 411–418.
- Marcoux, S., Maheux, R., and Bérubé, S. (1997). The Canadian Collaborative Group on Endometriosis. Laparoscopic surgery in infertile women with minimal or mild endometriosis. *N. Engl. J. Med.* **337**, 217–222.
- Moghissi, K. M., and Winkel, C. A. (eds.) (1999). American College of Obstetricians and Gynecologists Practice Bulletin No. 11: Medical Management of Endometriosis. American College of Obstetrics and Gynecology, Washington, DC.
- Olive, D. L., and Pritts, E. (2001). Drug therapy: Treatment of endometriosis. *N. Engl. J. Med.* **345**, 266–275.

## ANATOMY AND PHYSIOLOGY OF PENILE ERECTION

Because of its anatomy, the penis can rapidly engorge with blood, increase dramatically in size, and develop rigidity. Within the penis are two paired corpora cavernosa and a corpus spongiosum. The corpus spongiosum, which surrounds the urethra, terminates in the glans penis. A thick fibrous sheath, the tunica albuginea, encases the sponge-like cavernosal tissue, which consists of interconnecting lacunar spaces lined by vascular endothelium. Trabeculae form the walls of the lacunae and are composed of thick bundles of smooth muscle fibroblasts, collagen, and elastin. This arrangement enables very high intracavernosal pressure generation and penile rigidity when venous outflow is reduced and the lacunae are engorged.

When flaccid, the penile smooth muscle cell, arteries, and corpora cavernosa are in a state of tone (contraction). Relaxation of the smooth muscle (arterial and cavernosal) increases blood flow into the corpora cavernosa. The arterial pressure expands the relaxed trabecular walls, and the venules are compressed against tunica albuginea. This mechanism of veno-occlusion restricts the outflow of blood through these channels. After ejaculation or cessation of the erotic stimuli, the smooth muscle in the arteries and lacunar spaces contracts. Inflow of blood is reduced, venous outflow from the corporeal spaces resumes, and the penis returns to a flaccid state.

### Neuromuscular Influence

Increased arterial flow causing the penile erection is controlled by the autonomic nervous system. Peripheral innervation consists of sympathetic nerves arising from the T11 to L2 and parasympathetic and somatic nerves arising from the S2 to S4. They reach the pelvic plexus through hypogastric and pudendal nerves. The glans penis is rich in nerve endings that convey touch, pain, temperature, and vibration. Although neural centers in the lumbosacral spinal cord can cause erection in response to penile stimulation, erections usually are stimulated by olfactory, auditory, visual, and other sensory inputs. These inputs are perceived centrally by the frontal, temporal, parietal, or occipital lobe. Furthermore, they may be stimulated by memory or fantasy. Signals are sent to the visceral brain, to the rhinencephalon, and subsequently to the spinal cord centers.

Several neurotransmitters are involved in penile erection. Histochemical studies indicate that the corpora cavernosa are rich in adrenergic fibers,

acetylcholinesterase-containing (probably cholinergic) nerves, and nerves that are immunoreactive to vasoactive intestinal polypeptide (VIP) and neuropeptide Y. Nitric oxide is the primary endothelium-derived relaxing factor. It causes the smooth muscle in the corpora cavernosa to relax. It is produced from L-arginine by nitric oxide synthase. Nitric oxide activates guanine cyclase, forming intracellular cyclic guanosine monophosphate (cGMP). cGMP decreases intracellular calcium, resulting in smooth muscle relaxation. cGMP is degraded by phosphodiesterase-5, and inhibition of this enzyme potentiates the erectile process.

### Endocrine Influence

In humans, testosterone and dihydrotestosterone (DHT) play critical roles during embryogenesis of the testes, epididymis, vas deferens, seminal vesicles, prostate, and external genitalia. Sex steroids are implicated in imprinting gender identification on the central nervous system. Development of accessory glands, testes, external genitalia, and secondary sex characteristics at puberty depends on normal circulating androgen levels and on target androgen tissue responsiveness. In addition, androgen mediates the development and maintenance of libido and erectile function.

Early studies in mixed populations indicate several age-related changes. Serum total testosterone levels fall, sex hormone-binding globulin (SHBG) levels increase, free testosterone levels fall, and estradiol levels tend to remain constant after the fifth decade of life. Health status and age can affect total and free testosterone levels. Androgen deficiency most commonly occurs in men over 60 years of age.

## ETIOLOGY OF ERECTILE DYSFUNCTION

### Vascular

ED frequently is due to vascular dysfunction. Altered blood flow to and from the penis is thought to be the most frequent organic cause of impotence. Hypogastric-cavernous arterial bed occlusion can occur, but more commonly, penile endothelial dysfunction at the level of the penis decreases blood flow to the lacunar spaces and prevents complete relaxation of the trabecular smooth muscle. This also contributes to a "venous leak," decreasing penile rigidity and increasing the latency to achieve maximal erection. The most common associations with arteriogenic

impotence include diabetes mellitus, hypertension, hyperlipidemia, cigarette smoking, perineal or pelvic trauma, and pelvic irradiation.

### **Diabetes Mellitus**

The prevalence of ED in men with diabetes mellitus is in the range of 50% at 50 years of age, and it increases with age. An individual with diabetes who smokes cigarettes is 1.4 times more likely to have ED than is a nonsmoker. In men with diabetes, ED progresses gradually, and other symptoms and signs of autonomic neuropathy are frequently present. ED is more common in men with severe vascular or neurological complications. Thus, ED in men with diabetes may be caused by vascular disease, neuropathy, or both.

### **Hypertension**

Men with hypertension frequently report difficulty in initiating or maintaining erections that are adequate for sexual intercourse. Clinicians widely believe this to be a side effect of antihypertensive medication. The reported incidence of erectile failure is 14% with propranolol, 9% with hydrochlorothiazide, and 1% to 41% with clonidine. However, untreated hypertension is independently associated with increased ED.

### **Neurogenic Impotence**

Disorders affecting the spinal cord or the peripheral efferent autonomic fibers to the penis cause partial or complete ED. This may be due to afferent limb interruption of reflexogenic erections as well as to interruption of tactile perception, with projections to supraspinal centers that may be important in maintaining psychogenic erections. Although the reflexogenic erectile mechanism is preserved in men with supraspinal lesions, the erections usually are poorly maintained without constant tactile stimulation.

Common neurological disorders associated with ED include spinal cord injury, multiple sclerosis, and peripheral neuropathy due to diabetes mellitus or alcoholism. In addition, surgical procedures (e.g., radical prostatectomy, cystoprostatectomy, proctocolectomy) may disrupt the autonomic nerve supply to the corporal bodies.

In patients with spinal cord injury, preservation of erectile function is largely dependent on the completeness and level of the spinal lesion. Although 75% of patients with spinal cord injury have some capacity to have penile erections, the erections are adequate for penetration in only 25%.

## **Endocrine Impotence**

### **Androgen Deficiency**

Androgen deficiency is a widely recognized cause of sexual dysfunction. Diminished libido, reduced ejaculate volume, and ED commonly accompany androgen deficiency that develops after puberty. ED symptoms usually become manifest when testosterone levels are less than 300 ng/dl. The estimated prevalence of androgen deficiency in unselected men with ED ranges from 2 to 23% but is higher in older men. Many patients with ED and testosterone deficiency have secondary or tertiary hypogonadism. Pituitary and hypothalamic imaging is recommended when the morning testosterone level is less than 150 to 200 ng/dl. On average, older patients are less likely to experience an improvement in erections due to comorbid diseases that cause ED.

### **Hyperprolactinemia**

ED is common in men with pituitary tumors. Fully 76% of patients with a tumor in the region of the sella turcica reported decreased or absent libido or potency. These patients often have subnormal serum testosterone levels. Prolactin levels are more than 50 ng/ml in most patients with prolactinomas. Normalizing minimal hyperprolactinemia does not ameliorate ED. Therefore, prolactin levels less than 50 ng/ml should alert the physician to evaluate medications and thyroid, liver, and renal function. Studies performing routine measurement of serum prolactin found a low yield of prolactinomas (<1% of patients). Measuring the morning prolactin level when the testosterone level is less than 300 ng/dl and luteinizing hormone (LH) is normal or low, when sexual desire is low, or when gynecomastia is present is recommended.

## **DIAGNOSTIC WORKUP OF ERECTILE DYSFUNCTION**

The evaluation of patients with ED begins with a thorough history and physical examination. Initial questions should focus on the onset, progression, and duration of the symptoms. The presence of nocturnal erections and the ability to attain erections with a different partner may help to differentiate psychogenic causes from organic causes. It also is important to assess libido, which may indicate androgen deficiency, depression, or excessive stress. The medical history is essential to assess comorbidities commonly associated with ED (especially diabetes, hypertension, peripheral vascular disease, neurological disorders, and coronary artery disease).

Because the use of certain drugs is associated with ED, a detailed and complete drug history should be taken. Recent changes in social status, alcohol consumption, and cigarette smoking should be noted.

The physical examination is essential. It should focus on the cardiovascular, endocrine, neurological, and genitourinary systems. Blood pressure, palpation of peripheral pulses, and auscultation for abdominal and femoral bruits should be assessed. Special attention to signs of androgen deficiency (e.g., decreased body and facial hair, gynecomastia, testicular atrophy or decreased testicular consistency, regression in prostate size) must be undertaken. The penis should be examined for fibrous plaques (Peyronie's disease), and the rectal exam should assess the sphincter tone and the prostate.

Appropriate laboratory testing should complement the history and physical examination. Patients should be screened for diabetes as well as for liver and renal disease. Routine measurements of serum testosterone are recommended, and thyroid-stimulating hormone (TSH) should be obtained if there are signs suggesting thyroid or pituitary disease.

In refractory patients, nocturnal penile tumescence testing (NPT) and duplex ultrasonography in conjunction with intracavernosal administration of vasoactive agents may be indicated. Patients with ED and normal NPT usually have psychogenic impotence; whereas abnormal NPT in conjunction with normal sleep usually indicates organic cause (vascular or neurological). These tests should be performed in centers where there are facilities and expertise to perform and interpret them.

## TREATMENT

There are many treatment considerations and options for men with ED. The goal of treatment is to restore libido and the capacity to acquire and sustain an erection. Treatment should address the cause of ED. Patients should be asked to avoid cigarette smoking, excessive alcohol consumption, and drug abuse. Changes in medication should be tried. Medication may contribute to ED in up to 25% of patients with ED. The major drugs that induce ED are thiazide, beta-blockers, and some antidepressants.

### Psychosexual Therapy

When depression and excessive anxiety are associated with ED, psychosexual counseling is indicated. When it is selected appropriately, the success rate for psychosexual therapy can reach 70%. Furthermore, some

patients with organic ED can benefit from sex therapy because it reduces anxiety and sexual inhibition. Yohimbine, an alpha-2-adrenergic receptor-blocker, may be effective in some men with psychogenic ED. In a double-blind, placebo-controlled study that evaluated men with psychogenic ED, it was found that 37% of men treated with yohimbine responded to treatment. However, other studies have not found yohimbine to be beneficial.

### Hormonal Therapy

Testosterone replacement therapy is the treatment of choice for men with hypogonadism. Because of the potential complications of testosterone replacement therapy in men over 50 years of age, we suggest that the hematocrit, serum prostate-specific androgen (PSA), and digital rectal examination of the prostate be done at baseline, 3 and 6 months, and then annually. Testosterone should not be provided to eugonadal men with ED or to men with a history of prostate or breast cancer.

Hyperprolactinemia may decrease libido by suppressing gonadotropin-releasing hormone (GnRH), resulting in reduced LH and testosterone levels. Treatment of hyperprolactinemia with a dopamine agonist or surgical removal of the tumor may restore testosterone levels and potency.

### Medical Treatment

After excluding conditions treatable with specific therapies, the clinician should consider treatment with a type 5-phosphodiesterase inhibitor. These drugs provide effective oral therapy for a wide range of disorders causing ED. Sildenafil, the first available type 5 inhibitor, acts by inhibiting cGMP phosphodiesterase type 5, resulting in increased intracellular cGMP. cGMP increases nitric oxide-induced smooth muscle relaxation of arteries to the penis and of vascular tissue in the corpora cavernosum. Sildenafil was evaluated in 532 men with ED of organic, psychogenic, and mixed causes, and 69% of all attempts of sexual intercourse were successful for the men receiving sildenafil compared with 22% for those men receiving placebo. For maximum effectiveness, sildenafil should be taken 1 h before planned intercourse and preferably at least 2 h after a meal. It should be started at a 50-mg dose. If tolerated but not fully efficacious, the dose can be increased to 100 mg. Sildenafil is absolutely contraindicated in patients taking nitrates. Some level of exercise tolerance should be documented; otherwise

a treadmill test should be considered for patients at risk for cardiac disease.

Penile self-injections were more popular before the introduction of sildenafil. This treatment involves injecting vasoactive drugs (e.g., alprostadil, papaverine, phentolamine) that induce relaxation of the smooth muscles. Penile self-injections are especially effective in men with neurogenic impairments such as spinal cord injury and diabetes mellitus. The major side effects are penile pain and priapism.

Alprostadil also is available for intraurethral application (Medicated Urethral System for Erection [MUSE]). In a placebo-controlled study, 65% of patients treated with MUSE successfully completed intercourse compared with 19% of men treated with placebo. However, this efficacy rate has not been replicated in unselected populations of men with ED.

## Medical Devices

Vacuum devices increase the volume of blood in the penis by creating a vacuum around the penis. After the penis is engorged, a constriction ring is placed at the base of the penis to prevent venous outflow. The induced erection usually is sufficient for vaginal penetration and sexual intercourse. Adverse events include pain, bruising, and retrograde ejaculation. Vacuum constriction devices represent an alternative for patients who cannot tolerate type 5 phosphodiesterase inhibitors or who do not wish to use intracavernosal injection.

## Surgery

Surgical therapy is offered to patients who are refractory to other less invasive treatments. Vascular reconstruction is sometimes indicated in young patients with a history of pelvic trauma. Skepticism exists regarding other vascular reconstruction surgeries. Penile prostheses are a viable option for patients who cannot be treated medically. There are two kinds of penile prostheses: the semi-rigid prosthesis and the inflatable prosthesis. The inflatable prosthesis is better accepted but is more expensive. It also is associated with more frequent mechanical failure than are the semi-rigid devices.

## See Also the Following Articles

Aging and the Male Reproductive System • Fertilization • Gynecomastia • Hypertension and Diabetes • Impotence and Aging • Sexual Function and Androgens

## Further Reading

- Burnett, A. L., Lowenstein, C. J., Bredt, D. S., Chang, T. S., and Snyder, S. H. (1992). Nitric oxide synthase: A physiologic mediator of penile erection. *Science* **257**, 401–403.
- Davidson, J. M., Camargo, C. A., and Smith, E. R. (1979). Effects of androgen on sexual behavior in hypogonadal men. *J. Clin. Endocrinol. Metab.* **45**, 955.
- Goldstein, I., Lue, T. F., Padma-Nathan, H., Rosen, R. C., Steers, W. D., and Wicker, P. A. (1998). Oral sildenafil in the treatment of erectile dysfunction. *N. Engl. J. Med.* **338**, 1397–1404.
- Katchadorrean, H. A. (1990). "The Biological Aspects of Human Sexuality," 4th ed. Holt, Rinehart & Winston, New York.
- Kwan, M., Greenleaf, W. J., Mann, J., Crapo, L., and Davidson, J. M. (1983). The nature of androgen action on male sexuality: A combined laboratory–self-report study on hypogonadal men. *J. Clin. Endocrinol. Metab.* **57**, 557–562.
- Laumann, E. O., Paik, A., and Rosen, R. C. (1999). Sexual dysfunction in the United States: Prevalence and predictors. *JAMA* **281**, 537–544.
- Lue, T. F. (2000). Erectile dysfunction. *N. Engl. J. Med.* **342**, 1802–1813.
- Masters, W. H., and Johnson, E. (1970). "Human Sexual Inadequacy." Little, Brown, Boston.
- O'Keefe, M., and Hunt, D. K. (1995). Assessment and treatment of impotence. *Med. Clin. North America* **79**, 415–434.
- Padma-Nathan, H., Hellstrom, W. J., Kaiser, F. E., Labasky, R. F., Lue, T. F., Nolten, W. E., Norwood, P. C., Peterson, C. A., Shabsigh, R., and Tam, P. Y. (1997). Medicated Urethral System for Erection (MUSE) Study Group: Treatment of men with ED with transurethral alprostadil. *N. Engl. J. Med.* **337**, 1–7.
- Riley, A. J., Goodman, R. E., Kellett, J. M., et al. (1989). Double blind trial of yohimbine hydrochloride in erection inadequacy. *J. Sex. Mar. Ther.* **4**, 17.
- Saenz de Tejada, I., Goldstein, I., Azadzoi, K., Krane, R. J., and Cohen, R. A. (1989). Impaired neurogenic and endothelium-mediated relaxation of penile smooth muscle from diabetic men with impotence. *N. Engl. J. Med.* **320**, 1025–1030.
- Slag, M. F., Morley, J. E., Elson, M. K., Trencle, D. L., Nelson, C. J., Nelson, A. E., Kinlaw, W. B., Beyer, H. S., Nuttall, F. Q., and Shafer, R. B. (1983). Impotence in medical clinic outpatients. *JAMA* **249**, 1736–1740.
- Virag, R., Bouilly, P., and Frydman, D. (1985). Is impotence an arterial disorder? A study of arterial risk factors in 440 impotent men. *Lancet* **1**, 181.
- Wyndaele, J. J., De Meyer, J. M., de Sy, W. A., and Claessens, H. (1986). Intracavernous injection of vasoactive drugs: An alternative for treating impotence in spinal cord injury patients. *Paraplegia* **24**, 271–275.

ER $\beta$  is more numerous than ER $\alpha$  in the testis. Within the prostate, ER $\alpha$  is located primarily in the stroma, whereas ER $\beta$  is present in the epithelial cells. ER $\alpha$  is expressed in both osteoblasts and osteoclasts, and ER $\beta$  is expressed mostly in osteoblasts.

Pituitary secretion of luteinizing hormone (LH) controls the secretion of E<sub>2</sub> and testosterone from the testis. E<sub>2</sub> and testosterone, in turn, provide negative feedback at the level of the hypothalamus and anterior pituitary to regulate gonadotropin secretion. Testicular E<sub>2</sub> secretion is also stimulated whenever plasma LH/human chorionic gonadotropin (hCG) levels are elevated. Elevated levels of estrogen, either endogenous or exogenous, can reduce testicular testosterone secretion by suppression of LH release. Furthermore, excess estrogen stimulates SHBG synthesis in the liver, resulting in increased bound testosterone with a reduction in plasma-free testosterone.

## ROLES OF ESTROGEN IN MALE PHYSIOLOGY

### Lessons from Males with Estrogen Deficiency States

Descriptions of a man with estrogen resistance and three men with aromatase deficiency provide insights into our understanding of estrogen action in the male. In 1994, Smith and colleagues described a 28-year-old

man with estrogen resistance due to an autosomal recessive missense mutation in exon 2 of ER $\alpha$ , resulting in a complete loss of protein function. He had eunuchoid skeletal proportions with a height of 204 cm, delayed bone age (15 years), open epiphyses, persistent linear growth, and accelerated bone turnover with marked osteopenia. Biochemical testing revealed normal serum concentration of testosterone and elevated circulating levels of follicle-stimulating hormone (FSH), LH, E<sub>2</sub>, and E<sub>1</sub>.

Subsequently, from 1995 to 2002, three unrelated men with severe aromatase deficiency secondary to distinct autosomal dominant missense mutations of the CYP19 gene were reported. Like the man with estrogen resistance, the aromatase-deficient men (24–31 years of age) all had a tall stature (height of 187–204 cm) with continued linear growth, delayed skeletal maturation (bone age of 14.0–16.5 years), open epiphyses, and evidence of severe osteopenia. Biochemical studies revealed normal to elevated levels of testosterone, LH, and FSH and showed very low levels of circulating E<sub>2</sub> and E<sub>1</sub>. The details of the phenotype and biochemical data of the estrogen-resistant and aromatase-deficient men are summarized in [Table I](#).

When the ER $\alpha$ -negative man was treated with high doses of estrogen, there were no clinical, hormonal, or metabolic changes. In contrast, the aromatase-deficient men responded dramatically to relatively low doses of estrogen, resulting in correction of the

**Table I** Summary of the Phenotype and Biochemical Data of a Man with ER $\alpha$  Defect and Three Men with Aromatase Deficiency

	ER $\alpha$ resistant	Aromatase deficient #1	Aromatase deficient #2	Aromatase deficient #3
Publications	Smith <i>et al.</i> (1994)	Morishima <i>et al.</i> (1995)	Carani <i>et al.</i> (1997)	Herrmann <i>et al.</i> (2002)
Inheritance	Autosomal recessive	Autosomal recessive	Autosomal recessive	Autosomal recessive
Mutation	ER $\alpha$ gene (missense)	CYP19 gene (missense)	CYP19 gene (missense)	CYP19 gene (missense)
Stature	Tall (204 cm)	Tall (204 cm)	Tall (187 cm)	(Tall 197 cm)
Bone age	Delayed	Delayed	Delayed	Delayed
Bone density	Severe osteopenia	Severe osteopenia	Severe osteopenia	Severe osteopenia
Testicular size	Normal (20–25 ml)	Large (> 25 ml)	Small (8 ml)	Normal (13–14 ml)
Sperm analysis	18 % viable normal count	Not done	0 % immobile low count	0 % immobile low count
FSH/LH	Very high	Very high	High/Normal	High/Normal
Testosterone	Normal	Very high	Normal	High
Estradiol	Very high	Very low	Very low	Low
Estrone	Very high	Very low	Very low	Low
Total cholesterol	Normal	High	High	Normal
LDL	Normal	High	High	Normal
HDL	Low	Low	Low	Low
Triglycerides	Normal	High	High	High
Insulin resistance	Yes	Yes	No	Yes

described hormonal abnormalities without side effects such as gynecomastia.

Observations in estrogen-resistant and aromatase-deficient men have redefined our view of estrogen action and the male skeleton. Other findings from these human models, including the role of estrogen in lipid/carbohydrate metabolism and spermatogenesis, are discussed in subsequent sections of this article.

### Estrogen and the Male Skeleton

Despite normal or elevated levels of testosterone, men with impaired estrogen synthesis or action all had delayed epiphyseal closure, severe osteopenia, and evidence of increased bone turnover. Although these men had normal pubertal onset and normal male secondary sexual characteristics, they did not experience the classic growth spurt associated with puberty. Estrogen stimulates pubertal growth spurt in both sexes by enhancing growth hormone secretion during puberty. Contrary to prior belief, epiphyseal fusion in males is not mediated by testosterone; rather, it is mediated by estrogen. With estrogen treatment, despite declining levels of testosterone, all of the men had complete epiphyseal fusion within 6 to 9 months and had a progressive increase in bone mass. In one man, after 3 years of continuous estrogen therapy, bone mass increased by 20.7% in the lumbar spine, 15.7% in the femoral neck, and 12.9% in the distal radius. Thus, androgen alone is clearly not sufficient to promote normal skeletal growth and maturation. Both androgen and estrogen are required for accrual of optimal peak bone mass in men. Estrogen appears to increase bone mass by exerting anabolic as well as antiresorptive effects.

Evidence from mouse knockout models ( $\alpha$ -ERKO,  $\beta$ -ERKO, and ArKO mice) has also contributed to our understanding of estrogen and the male skeleton during recent years. Deletion of the ER $\alpha$  or aromatase gene in the male mouse results in osteopenia with increased markers of bone remodeling. In contrast, the skeleton of the ER $\beta$  knockout male mouse is phenotypically identical to that of the wild type. It appears that estrogen mediates its effects on the bone primarily through ER $\alpha$ , although both ER $\alpha$  and ER $\beta$  are expressed in the bone.

Estrogen is clearly critical for bone mass acquisition in the growing skeleton. However, the inherited states of impaired estrogen synthesis or action do not clearly reflect the role of estrogen in maintaining bone density in the mature adult skeleton. Additional data from nearly all of the cross-sectional observational

studies looking at the relationships between sex steroid levels and bone mineral density (BMD) show that E<sub>2</sub>, especially the non-SHBG-bound E<sub>2</sub>, has a better correlation with BMD than do either total or free testosterone levels.

In a prospective study, it was found that the BMD in young men (20–40 years of age), when compared with that in elderly men (60–90 years of age), was most closely correlated with E<sub>2</sub> levels. In particular, the decreased bioavailable E<sub>2</sub> levels in elderly men appear to be a good predictor of age-related bone loss. This relationship was further investigated by giving elderly men physiological hormone replacement after pretreatment with a gonadotropin-releasing hormone (GnRH) agonist and an aromatase inhibitor. Estrogen alone was far more effective than testosterone alone in preventing an increase in bone resorption markers. Estrogen and testosterone combined have positive synergistic effects on bone remodeling. Based on the data, it was estimated that estrogen is responsible for about 70% of the total effect of sex steroids on bone resorption, whereas testosterone (in the absence of conversion to estradiol) accounts for no more than 30% of the effect. Another study reported that when elderly men were treated with an aromatase inhibitor for 9 weeks, they had a significant increase in bone resorption markers and a decrease in bone formation markers when compared with baseline. Therefore, a subset of elderly men with low bioavailable E<sub>2</sub> are the most susceptible to age-related bone loss similar to the increased risk of bone loss seen in postmenopausal women. More studies are needed in this area.

### Effects of Estrogen on Lipids and Carbohydrates

Despite differences of total cholesterol, low-density lipoprotein (LDL) cholesterol, and triglyceride levels in estrogen-resistant and aromatase-deficient men, serum concentrations of high-density lipoprotein (HDL) cholesterol are uniformly low (Table I). The low HDL cholesterol is likely the result of unopposed action of androgen as well as the decreased estrogen. When estrogen therapy was initiated, levels of HDL cholesterol and other lipids improved due to an increase in estrogen with a concomitant decrease in androgen.

There appears to be an association between estrogen deficiency and the propensity of developing insulin resistance (Table I). The estrogen-resistant man had axillary acanthosis nigricans, a cutaneous marker of insulin resistance, in addition to elevated

glycosylated hemoglobin (HbA1C 9.5%). Two of the men with aromatase deficiency also had inappropriately elevated serum insulin levels. When estrogen was given, serum insulin levels were lowered in all except the estrogen-resistant man.

### Estrogen and the Prostate

Benign prostatic hypertrophy (BPH) is usually found in men over 50 years of age. Estrogen concentration in the stroma increases with age with contributions from peripheral sources and from aromatization of androgen within the stroma. Estrogen can induce squamous epithelial metaplasia and can stimulate growth of the fibromuscular stroma, resulting in BPH. Estrogen also induces synthesis of insulin-like growth factor-1 (IGF-1), which acts to prime the response of epithelium to androgen. Therefore, estrogen can increase epithelial proliferation indirectly by enhancing androgen action.

By increasing the levels of circulating SHBG, estrogen can exert an inhibitory effect on the prostate by reducing the amount of free testosterone available for the androgenic stimulation of prostate. The decrease in the plasma concentration of testosterone is further mediated by the negative feedback action of estrogen on LH secretion. Therefore, in contrast to its growth-promoting effect in BPH, estrogen in pharmacological amounts can reduce the size and function of the epithelial cells of the prostate and has been a form of therapy in prostate cancer. Alternatively, estrogen may exert antiandrogen effects within the tumor cells, or it may be directly inhibitory to the tumor cells.

Epidemiological studies demonstrate geographical differences in incidence rates of prostate cancer, with a 30-fold greater incidence in the United States than in Japan and a 120-fold greater incidence in the United States than in Shanghai, China. The geographical differences are thought to be due to dietary factors such as soy (phytoestrogens), which may be protective against the development of prostate cancer, in addition to genetic factors.

### ESTROGEN-EXCESS STATE IN THE MALE

The major clinical signs of estrogen excess in males are gynecomastia, testicular atrophy, erectile dysfunction, and infertility. Estrogen-secreting tumors and other endogenous or exogenous sources of estrogen can induce inhibition of testicular function by inhibiting gonadotropin secretion (Table II).

### Gynecomastia

The crucial factor in the development of gynecomastia from any cause is not the absolute level of estrogen but rather the ratio of estrogen to testosterone; the higher the ratio, the greater the likelihood of developing gynecomastia. Table II summarizes the various causes of gynecomastia.

Physiological gynecomastia occurs during the newborn period, during adolescence, and with advanced age. Breast enlargement occurs in approximately 60 to 90% of newborns due to stimulation by maternal placental estrogen, and it usually resolves within a few weeks. Pubertal gynecomastia usually has an onset of between 10 and 12 years of age and peaks at approximately 13 to 14 years of age. It affects 30 to 40% of adolescent boys, and in the majority of cases it resolves by 17 years of age. The high plasma ratio of E<sub>2</sub> to testosterone in boys with pubertal gynecomastia is likely the result of normal aromatase activity in the testis and the extraglandular tissues before maximum production of testosterone is achieved. The prevalence of gynecomastia in older men ranges from 40 to 70%. With advanced age, the increased ratio of estrogen to testosterone favors feminization. However, in elderly men, it is a diagnosis of exclusion given that many elderly patients take medication or have medical problems that may contribute to breast enlargement.

Pathological gynecomastia can be due to a variety of causes, resulting in an alteration of the estrogen/testosterone ratio (Table II). Gynecomastia can be seen in men with androgen deficiency or resistance. When testosterone production or action is compromised, elevated plasma gonadotropin levels further alter the plasma estrogen/testosterone ratio by stimulating the testis to produce more estrogen. Likewise, men with testicular tumors or bronchogenic carcinoma have increased testicular estrogen production largely from the stimulating affect of hCG produced by the tumor. Liver disease, starvation, thyrotoxicosis, and adrenocortical tumors are conditions where increased amounts of aromatizable androgen (i.e., androstenedione) are produced and gynecomastia can occur due to increased availability of this substrate for extraglandular aromatization.

There are several reports of familial gynecomastia due to excessive extraglandular aromatase activity. The initial description was a boy, age 10 years 7 months, who had accelerated linear growth (height of 157 cm), advanced bone maturation (bone age of 15 years), and severe gynecomastia (Tanner stage IV). The clinical signs of hyperestrogenism occurred at

Table II Causes of Gynecomastia

---

<b>Physiological gynecomastia</b>
Newborn
Adolescence
Elderly
<b>Pathological gynecomastia</b>
<i>Testosterone deficiency</i>
Hypergonadotropic hypogonadism
Klinefelter's syndrome
Enzyme defect of testosterone synthesis (17 $\beta$ -hydroxysteroid dehydrogenase deficiency, 3 $\beta$ -hydroxysteroid dehydrogenase deficiency)
Secondary testicular failure (viral orchitis, trauma, infiltrative disease, renal failure)
Hypogonadotropic hypogonadism (pituitary or hypothalamic causes)
Androgen resistance
Hyperprolactinemia
<i>Estrogen overproduction</i>
Testicular tumors
hCG-secreting tumors (bronchogenic carcinoma)
Adrenocortical adenoma or carcinoma
Severe liver disease
Hyperthyroidism
Recovery from malnourishment
Excess aromatase activity (hereditary)
<i>Drug induced</i>
Estrogens or estrogen agonists (digitalis, phytoestrogens, estrogen creams, diethylstilbestrol)
Clomiphene, gonadotropins
Aromatizable androgens
Antiandrogens or inhibitors of androgen synthesis (spironolactone, ketoconazole, flutamide, metronidazole, cimetidine, etomidate, cyproterone)
Drugs that cause elevated prolactin (phenothiazines, haloperidol, methyl dopa, reserpine, metoclopramide, tricyclic antidepressants)
Others (ranitidine, omeprazole, alkylating agents, cisplatin, busulfan, isoniazid, penicillamine, captopril, enalapril, amiodarone, nifedipine, verapamil, diazepam, phenytoin, amphetamines, heroin, marijuana, alcohol)
<b>Idiopathic gynecomastia</b>

---

8 years of age, around the time of adrenarche when androstenedione levels began to rise. The conversion of androgens to estrogens in this boy was about 50 times greater than that observed in normal prepubertal boys. Another kindred of five males spanning two generations had a similar condition, resulting in marked prepubertal gynecomastia and accelerated bone maturation (bone age of 14.5 years in a 10-year-old). An autosomal dominant pattern of inheritance was established in another kindred of four affected members. A 9-year-old boy from this kindred had prepubertal gynecomastia, and his 712-year-old sister had precocious puberty. His father and his paternal grandmother had prepubertal gynecomastia and macromastia, respectively. The genetic defect of the aromatase excess syndrome has not been elucidated.

Drugs that alter the ratio of estrogen to androgen at the level of the breast can cause gynecomastia

(Table II). Exposure to estrogen in men, either in the form of a pill or transdermally via ointments or cream, can result in gynecomastia. There also have been reports of gynecomastia in children who ingested dairy or meat products from estrogen-injected cows. Digitalis administration may result in gynecomastia, presumably by the binding of digitalis to the ER. As mentioned previously, gonadotropin-like substance such as hCG may cause gynecomastia by increasing testicular estrogen synthesis. In addition, drugs that impair testosterone synthesis or action can also cause gynecomastia. Approximately 50% of the men who receive high-dose spironolactone (150 mg/day) develop gynecomastia due to inhibition of both testosterone synthesis and testosterone action. Spironolactone at low doses (50 mg/day) may cause gynecomastia, although less commonly, by blocking androgen binding to the receptor, with less effect on testosterone synthesis.

## Estrogen and the Male Reproductive System

FSH and androgens are known to be important in the regulation of spermatogenesis. However, emerging evidence suggests that estrogen also plays an important role in spermatogenesis and male fertility. Estrogen appears to play a role in the development and maintenance of the efferent ductules and epididymides. On the cellular level, estrogen may play a role in controlling the proliferation, development, and function of the Leydig, Sertoli, and germ cells.

Men with estrogen deficiency due to a defect in estrogen action or synthesis have abnormal spermatogenesis (Table I). Conversely, men with either exogenous or endogenous estrogen excess can experience azospermia. The ratio of testosterone to E<sub>2</sub> is perhaps more important in spermatogenesis and fertility than is the absolute level of estrogen.

Increased environmental exposure to estrogen in males during fetal development may also have detrimental effects on development and function of the male reproductive system. Pregnant women who took diethylstilbestrol (DES) during pregnancy from 1945 to 1971 had male offspring with a significantly higher incidence of disorders of the reproductive tract, decreased sperm counts, and decreased sperm activity.

## Male Breast Cancer

Breast cancer is responsible for less than 1% of all cancer deaths in men. The ratio of female breast cancer to male breast cancer is 100:1 in Caucasians and 70:1 in African Americans. A higher incidence is reported in Jewish men and in men living in a stretch of Bantu-speaking countries in central Africa. Other risk factors include a family history of breast cancer, history of exogenous estrogen exposure, increased circulating endogenous estrogen from chronic liver disease, testicular pathology, and prior chest wall irradiation. The greatest risk factor for developing male breast cancer is Klinefelter's syndrome. The risk in affected men is about 50-fold higher than in men with a normal genotype. The risk of male breast cancer appears to be associated with conditions with relative estrogen excess and/or lack of androgen, resulting in an altered ratio of estrogen to testosterone.

## CONCLUSION

During recent years, some of the critical roles played by estrogen in the male have been recognized. We have just begun to understand estrogen action in men. More investigations are needed to clarify the mechanisms by which estrogen mediates its effects in male physiology.

## See Also the Following Articles

Benign Prostatic Hyperplasia (BPH) • Bone Remodeling, Dynamics of • Brain, Effects of Steroid Hormones • Gynecomastia • Hormone Replacement Therapy, Male • Klinefelter's Syndrome • Postnatal Normal Growth and Its Endocrine Regulation • Prostate Cancer • Puberty: Physical Activity and Growth • Skeletal Development

## Further Reading

- Bilezikian, J. P., Morishima, A., Bell, J., *et al.* (1998). Increased bone mass as a result of estrogen therapy in a man with aromatase deficiency. *N. Engl. J. Med.* **339**, 599–603.
- Carani, C., Qin, K., Simoni, M., *et al.* (1997). Effect of testosterone and estradiol in a man with aromatase deficiency. *N. Engl. J. Med.* **337**, 91–95.
- Farnsworth, W. E. (1996). Roles of estrogen and SHBG in prostate physiology. *Prostate* **28**, 17–23.
- Frantz, A. G., and Wilson, J. D. (1998). In "Williams Textbook of Endocrinology," 9th ed. W. B. Saunders, Philadelphia.
- Grumbach, M. M., and Auchus, R. J. (1999). Estrogen: Consequences and implications of human mutations in synthesis and action. *J. Clin. Endocrinol. Metab.* **84**, 4677–4694.
- Herrmann, B. L., Saller, B., Janssen, O. E., *et al.* (2002). Impact of estrogen replacement therapy in a male with congenital aromatase deficiency caused by a novel mutation in the CYP19 gene. *J. Clin. Endocrinol. Metab.* **87**, 5476–5484.
- Khosla, S., Melton, L. J., and Riggs, B. L. (2002). Estrogen and the male skeleton. *J. Clin. Endocrinol. Metab.* **87**, 1443–1450.
- Morishima, A., Grumbach, M. M., Simpson, E. R., *et al.* (1995). Aromatase deficiency in male and female siblings caused by a novel mutation and the physiological role of estrogens. *J. Clin. Endocrinol. Metab.* **80**, 3689–3698.
- O'Donnell, L., Robertson, K. M., Jones, M. E., *et al.* (2001). Estrogen and spermatogenesis. *Endocrine Rev.* **22**, 289–318.
- Sharpe, R. M. (1998). The roles of oestrogen in the male. *Trends Endocrinol. Metab.* **9**, 371–377.
- Smith, E. P., Boyd, J., Frank, G. R., *et al.* (1994). Estrogen resistance caused by a mutation in the estrogen-receptor gene in a man. *N. Engl. J. Med.* **331**, 1056–1061.
- Taxel, P., Kennedy, D., Fall, P. M., *et al.* (2001). The effect of aromatase inhibition on sex steroid, gonadotropins, and markers of bone turnover in older men. *J. Clin. Endocrinol. Metab.* **86**, 2869–2874.

recurrence when the medication is stopped. Vaginal dryness occurs at menopause and is improved by ERT.

Many women experience symptoms that develop long after menopause starts, such as vaginal atrophy, dyspareunia, and urinary tract symptoms, including stress incontinence, urinary tract infections, and increased urinary frequency. ERT, along with Kegel exercises to strengthen the vaginal muscles, improves these symptoms.

## Osteoporosis Prevention and Treatment

Estrogen depletion, which occurs at menopause, increases bone resorption. There is a loss of bone mineral density of approximately 3% per year in the first 5 years of menopause. In later years, the rate of bone loss decreases. ERT prevented this bone loss and actually improved bone density in several randomized controlled double-blind studies. However, as soon as ERT is stopped, bone loss resumes at the initial high rate. ERT may be started at any age after menopause to prevent bone loss.

Osteoporotic fractures are reduced in patients on ERT compared to nonusers. The addition of progesterone does not alter these beneficial effects. It is estimated that HRT causes 172 fewer fractures per 100,000 users per year.

Selective estrogen receptor modulators (SERMs), raloxifene and tamoxifen, also have beneficial effects on bone. Raloxifene causes a greater increase in bone mineral density than tamoxifen and has been shown to decrease fracture risk.

## POTENTIAL BENEFITS

### Cardiovascular

The relative risk of heart disease increases at menopause, coinciding with estrogen depletion. Estrogen has many beneficial effects on the cardiovascular system, including vasodilation, diminished arterial plaque formation, decreased platelet aggregation, improved carbohydrate metabolism, increased fibrinolysis, antioxidant properties, and improved lipid profile. Progesterone may diminish some but not all of these beneficial effects. The lipid profile was examined in a prospective randomized 3-year trial (Postmenopausal Estrogen/Progestin Interventions), which showed a decrease in total cholesterol, an increase in high-density lipoprotein (HDL), and a decrease in low-density lipoprotein in estrogen users. The addition of medroxyprogesterone acetate diminished

the beneficial increase in HDL, whereas micronized progesterone did not alter the beneficial effect.

Oral estrogen (not transdermal estrogen) use also increases triglyceride levels, which are associated with increased cardiovascular risks. Severe familial hypertriglyceridemia is a contraindication to ERT.

Most epidemiologic studies indicate an almost 50% reduction in heart disease with current ERT and HRT. These studies may have a selection bias that overestimates the beneficial effects.

Randomized clinical trials for secondary prevention of heart disease have not shown similar results. Dr. D. M. Herrington performed a 3-year randomized placebo-controlled trial on women with heart disease, which failed to show a beneficial effect of ERT or HRT by quantitative coronary angiography. HERS (Heart and Estrogen/Progestin Replacement Study), a randomized, double-blind trial treating women with preexisting coronary artery disease with HRT found a 50% increase in cardiac events in the first year of the study in the treatment group. By the fourth year, there was a decreased risk in the treatment group. Overall, there was no difference in cardiac events over 4 years. Possible reasons for the increased risk in the first year may be an increase in thromboembolic events and systemic inflammation in HRT users. ERT should not be started in the first year after a cardiovascular event due to the increased risks. If a patient was already on ERT, there were no increased risks in continuing the medication. Randomized clinical trials on the effects of HRT in women without cardiovascular disease are being conducted.

### Cognition

Estrogen has been shown to have beneficial effects in the brain, such as increasing synapses, augmenting neuronal growth, antioxidant protection of neurons, and decreasing amyloid P concentrations (found in Alzheimer's neurofibrillary tangles). Estrogen stimulates pathways that have important functions in memory and learning. However, ERT does not improve cognitive function in normal healthy subjects.

Meta-analysis indicates approximately one-third less dementia in estrogen users than in nonusers. Small short-term randomized studies have not shown a benefit of ERT in Alzheimer's patients but it may help to enhance the positive treatment effect of tacrine.

### Other Potential Benefits

Meta-analysis indicates that estrogen use may decrease the risk of colon cancer. Studies indicate a

potential benefit for ERT to prevent falls, improve depression, reduce osteoarthritis, preserve teeth, accelerate wound healing, decrease insulin resistance, prevent age-related hearing loss, decrease skin wrinkles, reduce dryness of the skin, improve visual acuity, reduce dryness of the eyes, lower intraocular pressure, and decrease cataract formation. Multiple epidemiologic studies indicate decreased mortality rates but selection bias is a concern. More evidence is needed to assess these claims.

## RISKS

### Endometrial Hyperplasia and Cancer

Stimulation of the uterine lining with ERT is well documented. This can lead to endometrial hyperplasia and progress to endometrial cancer in women who have not had a hysterectomy. Endometrial hyperplasia can be demonstrated in 20% of women using estrogen alone within 1 year. Women who use unopposed estrogen have a 4 to 15 times increased incidence of uterine cancer. Endometrial hyperplasia and cancer rates increase with higher doses and longer duration of estrogen use. Adequate progesterone supplementation reduces this risk to baseline levels. Women with late-stage estrogen receptor-positive endometrial cancer should not be given HRT until they are in remission for at least 5 years.

### Breast Cancer

Women with prolonged high estrogen states, such as early menarche, late menopause, and nulliparity, are known to be at higher risk for breast cancer. Therefore, there is concern about whether ERT increases the risk of breast cancer. There have been over 50 epidemiologic studies, at least five meta-analyses, and a reanalysis of the risk of breast cancer with postmenopausal ERT. The results are inconclusive but it is likely that there is a slight increase in risk. It is estimated that estrogen use may increase breast cancer by approximately 200 cases per 100,000 individuals in the first 5 years of estrogen use. After 15 years of ERT, an increase of 1200 cases is estimated. Progesterone is known to have stimulatory effects on the breast but progesterone supplementation has not been shown to increase the risk of breast cancer.

The breast cancer that develops in women on HRT tends to be diagnosed earlier (perhaps due to increased surveillance with mammography), to be less virulent, and to be less fatal. SERMs may have a better risk profile for women with a personal or family

history of breast cancer. Breast cancer risk needs to be discussed in perspective with the benefits of HRT.

### Other Risks

HERS showed a threefold increase in the risk of thromboembolism during the first year of HRT. The risk returns to baseline levels after the first year. Only 1% of the time is venous thrombosis fatal. Screening for thrombotic disorders is appropriate when warranted by patient or family history. Evidence suggests that there is approximately twice the risk of gallbladder disease, a 30% increased risk of ovarian cancer, and twice the risk of systemic lupus erythematosus. There may be exacerbations of asthma, migraines, and seizure disorders.

## PREPARATIONS AND USAGE

### Estrogen

The estrogen with the highest potency in the body is estradiol. Micronized estradiol can be used for ERT. However, a preparation of daily conjugated equine estrogens (0.625 mg) is most commonly prescribed in the United States. It is mainly estrone but also includes many forms of estrogen derived from pregnant mares. Equivalent dosages for bone of oral estrogens are shown in Table 1. It is unclear whether these products have equivalent actions in all organ systems. Estrogens are also given in transdermal and transvaginal forms.

Phytoestrogens are natural weak estrogens found in soy products, which, in high doses, reduce hot flashes. No research has been performed on other benefits or risks.

### Supplemental Progesterone

Progesterone should be used as adjunct treatment with estrogen whenever a woman has a uterus to

Table 1 Equivalent (in Bone) Estrogen Products

Category	Brand	Dose (mg)
Conjugated equine estrogen	Premarin	0.625
Conjugated plant estrogen	Cenestin	0.625
Esterified estrogens	Menest	0.625
Piperazine estrone sulfate (estropipate)	Ogen, Ortho-est, generic	0.625
Micronized estradiol	Estrace, generic	1
Ethinyl estradiol	Estinyl	0.05

protect against endometrial pathology. It is not necessary when a woman has had a hysterectomy since progesterone does not prevent the risk breast cancer and may adversely affect some of the cardiac benefits of estrogen alone. In the United States, medroxyprogesterone acetate is most commonly used.

Progesterone can be given cyclically for 12 to 14 days of the month (medroxyprogesterone 5–10 mg) or continuously (2.5–5 mg) to protect against endometrial hyperplasia. Using progesterone for less than 12 days or every 3 months is less protective against endometrial hyperplasia.

Medroxyprogesterone acetate can have unpleasant side effects, such as bloating, breast tenderness, irritability, and depression. Other progestins may have a better side effect profile. Micronized progesterone at doses of 200–400 mg administered cyclically or 100–200 mg administered continuously have been shown to minimize endometrial hyperplasia. It has an added benefit of less reduction in the beneficial HDL rise provided by estrogen alone. Norethindrone (0.7 mg cyclic or 0.35 mg continuous administration), norgestimate (0.09 mg), and norethindrone acetate (5 mg daily) are also acceptable alternatives. Progestin can also be given as a vaginal suppository, transdermal patch, intrauterine device, or intramuscular injection.

### Treatment Regimens

As suggested above, progesterone supplementation can be given cyclically or continuously to prevent endometrial pathology. Estrogen should be prescribed alone to women who have had a hysterectomy.

Cyclic use of progesterone usually causes cyclic vaginal bleeding in 80–90% of women. Some may have enjoyed years without this inconvenience. However, half of the women who use continuous HRT experience unexpected vaginal bleeding in the first 6 months. After 1 year, 10–50% continue to have this side effect. Bleeding should be investigated if it occurs at unexpected times in the cyclic regimen or after 6 months with the continuous regimen by ultrasound measurement of the endometrial lining and/or by endometrial biopsy.

Some women desire only ERT despite the risks of endometrial cancer due to intolerable side effects from progesterone. These patients must be monitored closely with yearly endometrial biopsy to ensure that no pathology develops.

### SERMs

Tamoxifen is prescribed to women who have had estrogen receptor-positive breast cancer to reduce recurrence and to high-risk women to prevent breast cancer. It improves the lipid profile and reduces bone loss. Negative effects include worsening of hot flashes, increasing thromboembolism, and increasing the risk of endometrial polyps, hyperplasia, and cancer. Intrauterine progesterone has been given to prevent endometrial pathology.

Raloxifene (60 mg) improves bone mass, decreases fracture rates, improves the lipid profile, and protects the endometrium. A 3-year trial showed a decrease in breast cancer rates but longer term studies are needed. Unfortunately, hot flashes and thromboembolic disorders are exacerbated. Leg cramps are an unexpected side effect.

### See Also the Following Articles

Breast Disease: Impact of Sex Steroid Replacement • Cardiovascular Disease: Impact of Sex Steroid Replacement • Estrogen Replacement, Vaginal • Hormone Replacement, Transdermal • Hot Flash: Impact of Sex Steroid Replacement • Osteoporosis in Older Women • Osteoporosis in Reproductive Age Women: Impact of Sex Steroid Replacement • SERMs (Selective Estrogen Receptor Modulators)

### Further Reading

- Christiansen, C., Christensen, M. S., and Transbol, I. (1981). Bone mass in postmenopausal women with withdrawal of estrogen/gestagen replacement therapy. *Lancet* **8218**, 459–461.
- Gambrell, R. D. (1982). The menopause: Benefits and risks of estrogen-progestogen replacement therapy. *Fertil. Steril.* **37**, 457–474.
- Herrington, D. M., Reboussin, D. M., Brosnihan, K. B., *et al.* (2000). Effects of estrogen replacement on the progression of coronary artery atherosclerosis. *N. Engl. J. Med.* **343**, 522–529.
- Hulley, S., Grady, D., Bush, T., Furberg, C., Herrington, D. B., and Vittinghoff, E., for the Heart and Estrogen/Progestin Replacement Study (HERS) Research Group (1998). Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. *J. Am. Med. Assoc.* **280**, 605–613.
- LeBlanc, E. D., Janowsky, J., Chan, B. K., and Nelson, H. D. (2001). Hormone replacement therapy and cognition: Systemic review and meta-analysis. *J. Am. Med. Assoc.* **285**, 1489–1499.
- Manson, J. E., and Martin, K. A. (2001). Clinical practice: Postmenopausal hormone-replacement therapy. *N. Engl. J. Med.* **345**, 34–40.
- The Writing Group for the PEPi Trial. (1995). Effects of estrogen or estrogen/progestin regimens on heart disease risk factors in postmenopausal women: The Postmenopausal Estrogen/Progestin Interventions (PEPi) Trial. *J. Am. Med. Assoc.* **273**, 199–208.

Table I Vaginal Compounds Available in the United States

	Brand name	Pharmaceutical company	Estrogen	Available dosage	Common starting dosage	Maintenance dosage
Estrace	Estrace	Warner Chilcott	Estradiol	0.1 mg E <sub>2</sub> /g	2–4 g/day	1 g thrice weekly
Estring vaginal ring	Estring	Pharmacia Upjohn	Estradiol	7.5 µg/24 <sup>h</sup> in 2 mg reservoir		Stable for 90 days
Dinestrol cream	Ortho Dinestrol cream	Ortho-McNeil	Dinestrol	0.01% cream	1–2 applicators/day	1–3 times weekly
Premarin vaginal cream	Premarin	Wyeth-Ayerst	Conjugated equine estrogens	0.625 mg/g	0.5–2 g/day	1 g/day
Vagifem	Vagifem	Pharmacia Upjohn	Estradiol	25 µg/tablet	1 tablet/day	1 tablet twice weekly

Changes in endometrial thickness are even larger than those caused by similar oral doses of estradiol. Vaginal estrogen therapy should therefore be no different than other methods of estrogen replacement therapy: unopposed estrogen stimulation should be discouraged in all cases and close follow-up should be maintained on a regular basis.

## VAGINAL THERAPY CONSIDERATIONS

Most estrogens are readily absorbed across mucous membranes. Consequently, vaginal absorption can be achieved via creams, saline suspensions, tablets, or hormonally impregnated pessaries and rings. All of these modes of transmission, save the use of saline suspension, have been officially approved for clinical usage (see Table I). Each of these methodologies has its own unique list of advantages and disadvantages. Micronized estradiol suspended in a saline buffer has the swiftest absorption profile, with peak systemic levels occurring at 2 h, compared with 4 h for the same dosage compounded in a vaginal cream. Conversely, vaginal rings impregnated with estradiol produce systemic levels that persist for as long as 3 months. Generally, estrogenic compounds in vaginal creams will achieve approximately one-fourth of the systemic levels gained by an equivalent oral dosage of the same estrogen.

## VAGINAL CREAMS

Vaginal creams remain the mainstay of transvaginal estrogen therapy. A variety of estrogenic compounds have been formulated for admixture in a water-soluble base (see Table I). Most commonly initiated for urogenital atrophy and atrophic vaginitis, treatment is generally instituted with higher doses on a daily basis. Therapy should be continued for 1 to 2 weeks at this dosage. A small amount of the cream should be

lightly applied to the vulva simultaneously to assist in restoration of tissue integrity. At that point, dosage should be decreased in a gradual manner to one-half the starting dose for a similar period of time. A maintenance dosage of 1 g (refer to Table I) of either estradiol or conjugated equine estrogens can be given one to three times weekly once restoration of the vaginal mucosa has been achieved. The most common adverse sequelae of therapy remain the same as with other methods of estrogen replacement: changes in the nature of vaginal bleeding patterns, breast tenderness, allergic reactions, and headaches. Close follow-up and addition of an oral or vaginal progestin should be considered in all cases of maintenance therapy.

## VAGINAL RINGS/PESSARIES

Vaginal rings consist of crystalline estradiol mixed with a liquid polymer, polydimethylsiloxane. This mixture is then molded into rings similar in shape and size to the contraceptive diaphragm. After an initial burst effect, sustained and reasonably constant levels of estradiol, approximately 60 pg/ml, are seen for 3 months. Such devices have been shown in open, parallel-group, comparative trials to be effective in the treatment of urogenital atrophy in greater than 75% of patients. Due to their similarity to the diaphragm, they can be placed by either a medical professional or the patient herself and are tolerated far better than the relatively stiffer structure of estrogen-containing pessaries. Side effects can include fever, pain, pruritis, urticaria, and a modest foreshortening of the vagina. Ring or pessary systems may be unsuitable for women with narrow, short, or stenosed vaginal canals. Narrowing of the vagina, vaginal stenosis, prolapse of pelvic structures, and active vaginitis make the vagina more susceptible to adverse side effects of these systems. Patients with these limiting factors should be considered for alternative modalities.

## See Also the Following Articles

Breast Disease: Impact of Sex Steroid Replacement • Cardiovascular Disease: Impact of Sex Steroid Replacement • Estrogen Replacement, Oral • Hormone Replacement, Transdermal • Hot Flash: Impact of Sex Steroid Replacement • Osteoporosis in Older Women • Osteoporosis in Reproductive Age Women: Impact of Sex Steroid Replacement • SERMS (Selective Estrogen Receptor Modulators)

## Further Reading

Fanchin, R., Righini, C., *et al.* (2001). Vaginal versus oral E<sub>2</sub> administration: Effects on endometrial thickness, uterine perfusion and contractility. *Fertil. Steril.* **76**, 994–998.

Goebelsmann, U., Maschak, C. A., and Mishell, M. A. (1985). Comparison of hepatic impact of oral and vaginal

administration of ethinyl estradiol. *Am. J. Obstet. Gynecol.* **151**, 868–877.

Handa, V. L., Bachus, K. E., *et al.* (1994). Vaginal administration of low-dose conjugated estrogens: Systemic absorption and effects on the endometrium. *Obstet. Gynecol.* **84**, 215–218.

Henriksson, L., Sternquist, M., Boquist, L., Ålander, U., and Selinus, I. (1994). A comparative multicenter study of the effects of continuous low-dose estradiol released from a new vaginal ring versus Estriol vaginal pessaries in postmenopausal women with symptoms and signs of urogenital atrophy. *Am. J. Obstet. Gynecol.* **171**, 624–632.

Johnston, A. (1996). Estrogens—Pharmacokinetics and pharmacodynamics with special reference to vaginal administration and the new estradiol formulation-Estring. *Acta Obstet. Gynecol. Scand.* **63**, 16–25.

Jones, K. P. (1992). Estrogens and progestins: What to use and how to use it. *Clin. Obstet. Gynecol.* **35**, 871–883.

Stumpf, P. G. (1986). Selecting constant serum estradiol levels achieved by vaginal rings. *Obstet. Gynecol.* **67**, 91–94.

does not alter the progression of diabetic retinopathy. All diabetic retinopathy eventually reaches an involuntional stage in which visual loss may relate to macular detachment, ischemia, long-standing macular edema, or optic nerve disease.

The treatment of diabetic retinopathy is focused on the prevention of visual loss. Clinical trials clearly indicate that good control of blood glucose reduces the risk of the onset and progression of retinopathy, with the possibility of transient worsening of retinopathy at the beginning of tight control. Focal or grid laser treatment should be considered when there is clinically significant macular edema regardless of the stage of diabetic retinopathy. Clinically significant macular edema is present when funduscopic findings meet any of the following three criteria: retinal thickening within 500  $\mu\text{m}$  of the center of the macula; hard exudates within 500  $\mu\text{m}$  of the center of the macula associated with thickening of the adjacent retina; or retinal thickening more than one disc area in size, part of which is within one disc diameter of the center of the macula. Panretinal laser photocoagulation, a procedure in which laser burns are applied to the retina outside the macular and optic disc, is indicated for PDR with high-risk characteristics, defined as one of the following three findings: neovascularization of the disc that is more than one-fourth of the disc area in size, any degree of neovascularization of the disc associated with preretinal or vitreous hemorrhage, or neovascularization elsewhere in the retina that is more than one-half of the disc area in size associated with a preretinal or vitreous hemorrhage. Some ophthalmologists also include neovascularization of the iris as a high-risk characteristic. Pars plana vitrectomy for the removal of vitreous and replacement with saline, combined with endolaser therapy with or without cataract extraction and lens implant, may be considered when the ocular media are too hazy for an adequate fundus view for laser therapy. Other conditions suitable for vitrectomy and endolaser include dense, nonclearing vitreous/premacular hemorrhage, tractional retinal detachment involving the macula, severe fibrovascular proliferation or macular epiretinal membrane, retinal neovascularization refractory to laser photocoagulation, and red blood cell-induced glaucoma. Peripheral retinal cryotherapy is used for advanced proliferative diabetic retinopathy with hazy ocular media by some ophthalmologists. Patients with type 1 diabetes have more aggressive proliferative diabetic retinopathy than those with type 2 diabetes and benefit from earlier laser photocoagulation and vitrectomy. Noncompliant patients should be treated early as well.

It is recommended that patients with type 1 diabetes be referred for an eye examination within 5 years after diagnosis and reexamined at least annually. Those with type 2 diabetes should have an eye exam at the time of diagnosis and be reexamined at least annually. Since retinopathy can become more aggressive during pregnancy, these women should have an eye exam before conception or early in the first trimester and at least every 3 months thereafter until parturition. Patients with mild NPDR should have an eye exam every 9 months, those with moderate NPDR every 6 months, those with severe NPDR every 4 months, and those with PDR every 2 or 3 months.

## GLAUCOMA AND HYPHEMA

Small isolated tufts of neovascularization at the pupillary border are relatively common in patients with diabetes and can be carefully followed without initial treatment. Patients with contiguous neovascularization of the iris or the anterior chamber angle may develop hyphema (bleeding in the anterior chamber) and glaucoma. They require prompt panretinal photocoagulation regardless of the presence of high-risk characteristics. Patients with diabetes are also at an increased risk for primary open-angle glaucoma as well as pupillary block glaucoma after cataract surgery. Note that topical beta-blockers for glaucoma in diabetic patients may mask the warning symptoms of hypoglycemia, such as sweating, shaking, and restlessness.

## CATARACT

Typical nuclear sclerosis, cortical opacities, and especially posterior subcapsular cataract occur earlier and more frequently in patients with diabetes. Rarely, the lens may become completely opaque in several days to weeks in patients with diabetes (true diabetic cataract). Although diabetic retinopathy may progress after cataract surgery, patients enrolled in clinical trials who underwent cataract surgery usually had improved visual acuity postoperatively.

## DIABETIC NEUROPATHY

Diabetes affects multiple cranial nerves, including II (optic nerve)–VI, and is thus called diabetic polyneuropathy. Diabetic papillopathy is benign disc edema associated with mild visual loss. It has no correlation with the severity of diabetic retinopathy. Spontaneous

resolution usually occurs after 3 months. The transient nature, frequent bilaterality, benign outcome, and the absence of subsequent disc atrophy are some of the features that help distinguish diabetic papillopathy from ischemic optic neuropathy. Diabetes is a risk factor for ischemic optic neuropathy, which may result in profound visual loss. Characteristic telangiectatic vessels at the optic disc can be observed in some patients with type 1 diabetes who experience sudden visual loss due to ischemic optic neuropathy. No effective treatment has been established for nonarteritic ischemic optic neuropathy.

An isolated third, fourth, or sixth cranial nerve palsy may develop from diabetic microvascular disease, resulting in diplopia, and it may be accompanied by intense periorbital pain. Rarely, two of these cranial nerves are involved simultaneously. Third cranial nerve involvement does not cause dilation of the pupil. A diabetic cranial nerve paralysis usually resolves within 3 months without intervention. Diabetic neuropathy of cranial nerve V may result in neurotrophic keratopathy and nonhealing corneal epithelial defects.

## REFRACTIVE CHANGES

Diabetes can affect the clarity of the lens, its refractive index, and its accommodative amplitude. The state of lenticular hydration can be affected by blood glucose level, causing transient changes of the refractive power of the lens, most commonly myopic but occasionally hyperopic. Patients with diabetes have a decreased amplitude of accommodation compared to age-matched controls, and presbyopia may occur at a younger age.

## MUCORMYCOSIS

In patients with diabetes, particularly those with ketoacidosis, a rare but life-threatening orbital infection may be caused by the fungus *Mucor*. Decreased vision, pain, proptosis, red-eye, and fever usually occur. Intravenous amphotericin B, surgical debridement of the necrotic tissue, and diabetic control are required to prevent death.

## MISCELLANEOUS

Diabetes has nonspecific corneal manifestations, such as punctate epithelial erosions, basement membrane changes leading to delayed epithelial healing, Descemet's folds, and decreased corneal sensation. Tear production is decreased in patients with diabetes.

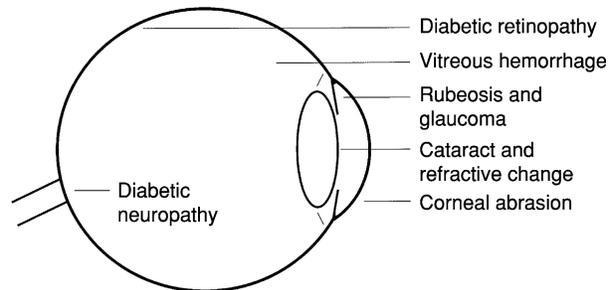


Figure 1 Eye disease in diabetes.

Glycogen infiltration of the pigment epithelium, sphincter, and dilator muscles of the iris may cause decreased pupillary responses. Autonomic neuropathy of diabetes can also be contributory. Patients with diabetes have a higher incidence of asteroid hyalosis, a condition with minute white opacities composed of calcium-containing phospholipids in the vitreous. Repeated vitreous hemorrhage in diabetic patients may result in cholesterosis, also called sychysis scintillans, with refractile cholesterol crystals in the inferior vitreous. Partial posterior vitreous detachment, which may cause floaters, flashes of light, vitreous hemorrhage, and retinal breaks, is seen far more commonly in patients with PDR than in individuals without diabetes or with NPDR. Children born to mothers who have diabetes during pregnancy may have superior segmental optic nerve hypoplasia (topless disc), with a matching inferior semialtitudinal visual field defect. Diabetes is also a risk factor for cystoid macular edema after cataract surgery.

## CONCLUSION

The eye diseases associated with diabetes are summarized in Fig. 1. Most can be prevented with tight control of blood glucose levels. Modalities have been developed to treat each of these eye diseases in diabetes. Since diabetic retinopathy remains the leading cause of visual loss despite the currently available treatments, clinical trials are under way to test the use of intravitreal steroids, oral and local antivascular endothelial growth factor agents, in preventing macular edema and retinal neovascularization, without the retinal destruction by laser photocoagulation.

## See Also the Following Articles

Cardiovascular Disease in Diabetes • Diabetic Nerve Disease, Neuropathy • Foot Disease in Diabetes • Graves' Ophthalmopathy • Kidney Disease in Diabetes • Neurological Disease and Diabetes, Autonomic

## Further Reading

- Aiello, L. M., Cavallerano, J., and Aiello, L. P. (2000). Diagnosis, management, and treatment of nonproliferative diabetic retinopathy and macular edema. *In* "Principles and Practice of Ophthalmology" (D. M. Albert and F. A. Jakobiec, eds.), 2nd ed., pp. 1900–1914. Saunders, Philadelphia.
- Flynn, H. W., Jr., and Smiddy, W. E. (2000). "Diabetes and Ocular Disease: Past, Present and Future Therapies. Ophthalmology Monograph 14." American Academy of Ophthalmology, San Francisco.
- Hayreh, S. S. (2002). Diabetic papillopathy and nonarteritic anterior ischemic optic neuropathy. *Surv. Ophthalmol.* 47(2), 183–188.
- Miller, J. W., and D'Amico, D. J. (2000). Proliferative diabetic retinopathy. *In* "Principles and Practice of Ophthalmology" (D. M. Albert and F. A. Jakobiec, eds.), 2nd ed., pp. 1915–1936. Saunders, Philadelphia.

assembled in liver from lipids synthesized *de novo* or from dietary lipids stored in liver between meals. Levels of lipoproteins are important predictors of atherosclerotic cardiovascular disease and are also critical in diagnosing disorders of lipoprotein transport, which afflict approximately 20% of the populations of Western societies. The vast majority of those affected have one of several hyperlipoproteinemias, but a number of syndromes afflicting a relatively small number of people exist, characterized by low levels of total cholesterol, either due to low levels of apoA1-containing lipoproteins [e.g., high-density lipoprotein (HDL)] or due to low levels of apoB-containing lipoproteins (e.g., LDL). This article is devoted to low-apoB syndromes, with special emphasis on familial hypobetalipoproteinemia.

## GENETICS

Abetalipoproteinemia is due to a set of mutations in the gene for microsomal triglyceride transfer protein (MTP). The MTP protein is crucial to the initial assembly of VLDL and chylomicron particles in liver and intestine, respectively. Early during assembly, apoB is synthesized on polyribosomes and cotranslationally transferred into the lumen of the endoplasmic reticulum (ER). Triglycerides synthesized nearby associate with MTP and are transferred onto apoB to form primary VLDL particles. In a second step, the primary particles "mature" as they acquire more lipids from secondary VLDL particles in the *trans*-ER and Golgi and the mature particles are secreted from hepatocytes (or in the case of chylomicrons, from enterocytes). MTP dysfunction renders assembly ineffective and lipids destined for cellular export accumulate in hepatocytes and enterocytes. Chylomicron retention disease (Anderson's disease) is similar symptomatically, but its molecular defect is in the Sar 1 GTPase of COPII vesicles that transport lipoproteins from ER to Golgi.

Similarly, the genetic cause(s) is not known for the vast majority of cases of FHBL. Less than 5% of cases are due to mutations of the apoB gene (APOB). Most mutations are insertions or deletions of a few base pairs and produce frameshifts of translation resulting in the introduction of premature stop codons and truncated proteins. The truncations are designated according to a centile nomenclature: the normal protein of 4536 amino acid residues, secreted from liver as part of VLDL particles, is designated apoB100. The normal intestinal variant associated with chylomicrons is apoB48. Over 40 different truncations, some as short as apoB2 and some as long as apoB89,

have been described. Null alleles probably result in untranslated mRNA. A fascinating group of seven families has been identified in which the FHBL trait seemed to segregate in a Mendelian dominant fashion, but in which no truncated forms of apoB were detectable in plasma. A genome scan was performed followed by linkage analysis. The FHBL trait was not linked to any locus on chromosome 2 (the site of APOB). Rather, linkage was established to a susceptibility locus on chromosome 3p21, between markers D3S2407 and D3S1767. In a third group of five families, genome scanning and linkage studies indicated linkage to neither chromosome 2 nor chromosome 3. Thus, FHBL appears to be genetically heterogeneous.

## METABOLISM OF ApoB-CONTAINING LIPOPROTEINS IN FHBL

Plasmas of homozygotes bearing mutations specifying apoB truncations longer than apoB48 contain three subfamilies of apoB-containing lipoproteins. For example, the plasmas of subjects heterozygous for the apoB54.8 mutation contain apoB100, apoB48, and apoB54.8 particles. Plasmas of homozygous individuals contain only the truncated form. Lipoproteins bearing truncated forms of apoB vary in size directly with the lengths of the truncations; i.e., apoB89-bearing particles have sizes and densities similar to those of normal apoB100-bearing VLDL and LDL, whereas apoB38.9-bearing particles have sizes intermediate between LDL and HDL and densities similar to those of large HDLs. Furthermore, short truncation-containing particles contain smaller molar amounts of triglycerides per particle than apoB100-containing particles. The presence of various forms of apoB on distinct particles permits the performance of concurrent *in vivo* metabolic studies on the kinetic parameters of apoB100 and apoB truncation-bearing particles, with apoB100 particles serving as internal controls for the truncations. Truncated forms are usually present at <20% of the levels of apoB100 particles, due to reduced production rates and enhanced clearance relative to apoB100 particles. This is due to a combination of low production rates and rapid clearance rates. The relative importance of production and clearance rates in setting the plasma levels of truncation-bearing particles depends on the truncation in question. For example, the production rate of apoB89-containing particles is only 15% lower than normal, but their clearance rate is more than twice normal due to the enhanced affinity of interaction of apoB89 with the LDL receptor. For apoB75,

production is more impaired than for apoB89, but clearance is still rapid due to enhanced interaction with the LDL receptor. The shorter truncations, such as apoB38.9, are cleared very rapidly, mostly by the kidney. Clearance is probably mediated by the megalin/gp330 receptor located in proximal tubule cells. Based on the presence of one normal allele specifying the production of apoB100, the average concentrations of apoB100 in plasmas of heterozygotes would be expected to be  $\approx 50\%$  of levels in unaffected controls. In fact, levels are closer to  $\sim 30\%$  of normal. This is due to production rates that are  $\sim 30\%$  of normal.

## HEPATOSTEATOSIS

As a result of low rates of production of the normal lipid transporter protein apoB100 and the impaired capacities particularly of short truncations of apoB to transport triglycerides, the export system for hepatic lipids via VLDL particles is impaired. This would be expected to result in the accumulation of triglycerides in liver. Indeed, several groups have reported cases of hepatic steatosis detected by ultrasound or in rare instances by liver biopsy. Schonfeld *et al.* examined 22 individuals with various truncations of apoB ranging from apoB4 to apoB89 and 13 controls, using magnetic resonance spectroscopy. Results were calculated from energy spectra and represent a precise noninvasive method for quantifying liver fat. The mean value for liver fat in FHBL was five times that of controls. The differences between FHBL and controls could not be accounted for by dietary intake, indexes of total body and abdominal fat, or indexes of glucose tolerance and insulin sensitivity, suggesting that the APOB mutations per se were responsible.

## LESSONS FOR FHBL FROM RECOMBINANT MICE AND HEPG2 CELLS

### Mice and Cells

Several recombinant mice meant to mimic human FHBL have been produced. The author's group created two types of apoB truncation-harboring mice by targeted homologous recombination in embryonic stem cells, using the Cre-loxP system to excise any extraneous genomic sequences that may have been introduced during recombination. The resulting apoB38.9 and apoB27.6 mice closely resemble their human counterparts with respect to the genomic sites of the mutations. Maeda's group has produced an

apoB81 mouse and Young's group has produced apoB83 and apoB39 mice. The author's group has also produced an apoB82-expressing HepG2 cell line.

## Plasma Lipoproteins

In contrast with humans, mouse livers produce not only apoB100 and the apoB truncations, but also apoB48. Thus, apoB48-containing lipoproteins arise from both liver and intestine. The mice exhibit hypobetalipoproteinemia, with low levels of cholesterol and apoB. Similar to human heterozygotes, three sub-families of apoB-containing lipoproteins circulate in plasmas of apoB<sup>+38.9</sup> and apoB<sup>+27.6</sup> heterozygous mice: those containing apoB100, apoB48, and the apoB truncation. Plasma levels of apoB100 and apoB48 are approximately equal; plasma levels of the truncations are much lower. In the plasmas of apoB<sup>38.9/38.9</sup> and apoB<sup>27.6/apoB27.6</sup> homozygous mice, only apoB truncation-containing lipoproteins circulate and their plasma lipid levels are very low.

## Mouse ApoB Lipoprotein Metabolism

Studies performed *in vivo* and in primary cultures of hepatocytes demonstrate that livers of apoB<sup>+38.9</sup> mice produce apoB100 only in small amounts, that apoB48 is the dominant normal variant, and that apoB38.9 is produced in equimolar amounts with apoB48. In apoB<sup>+27.6</sup> mice, the relative amounts of apoB100 and apoB48 produced are the same as for apoB38.9 mice, but the relative amount of apoB27.6 produced exceeds that of apoB48 by approximately fivefold. Since plasma levels of both the apoB truncations are less than the levels of apoB48, the truncation-containing lipoproteins must be cleared more rapidly than apoB48 lipoproteins. The more rapid clearance of the mutant apoBs is compatible with previous *in vivo* studies in humans and rabbits. Similarly, although more apoB48 is produced than apoB100, plasma levels of apoB48 lipoproteins are equal or less than levels of apoB100 lipoproteins, confirming older studies that apoB48 is cleared more rapidly than apoB100. The secretion of apoB100 in humans at rates lower than those expected based on the number of normally functioning alleles is due to enhanced intracellular degradation prior to secretion of apoB100 in the presence of the defective apoB81 or apoB82 alleles.

## Hepatosteatosis in Mice

The low rate of synthesis and secretion of normal apoBs, the rapid clearance of truncation-containing

lipoproteins, and the limited ability of the apoB38.9 and apoB27.6 truncations to ferry triglycerides led the author's group to predict that just as in humans, the mice would also have fatty livers. Indeed, liver triglycerides were increased 1.5- and 3-fold in apoB<sup>+38.9</sup> and apoB<sup>38.9/38.9</sup> mice, respectively, over age- and sex-matched apoB<sup>+/+</sup> wild-type mice. Similarly, liver triglycerides of apoB<sup>+27.6</sup> and apoB<sup>27.6/27.6</sup> were increased 3- and 5-fold. The greater accumulation of triglycerides in the animals bearing the shorter truncation is compatible with the more severe defect in the transport capacity of the shorter truncation.

### Feedback Inhibition of Hepatic Triglyceride Synthesis

It is interesting that the APOB defect-induced disturbance of triglyceride transport produces feedback inhibition of fatty acid synthesis in liver in the recombinant mice, in a gene-dose-dependent fashion. This is accompanied by decreases in hepatic mRNA levels for the transcription factor SREBP-1c that regulates enzymes in the fatty acid synthetic pathway and in the mRNA levels for the enzymes fatty acid synthase and steroyl coenzyme A desaturase-1. This feedback would tend to limit the amount of fat accumulated in the face of the APOB mutation-induced defect in the triglyceride export pathway. Thus, the amount of fat accumulated is under the control of several genes; the expression of the genes relative to one another probably sets hepatic triglyceride levels.

### The Region between ApoB38.9 and ApoB27.6 Supports Embryogenesis

ApoB-containing lipoproteins appear to be critical in supporting embryogenesis, at least in mice. Heterozygous crosses are expected to produce homozygous wild-type, heterozygous, and homozygous affected offspring in proportions of 25, 50, and 25%. However, crosses between mice heterozygous for null mutations of apoB (apoB<sup>0/0</sup>) yield more than 25% wild-type (apoB<sup>+/+</sup>) pups, more than the expected 50% heterozygotes, and ~2–3% apoB<sup>0/0</sup> homozygotes, which is ~10% of the expected number. Thus, the induction of a null mutation in mouse apoB results in embryonic lethality for homozygotes. By contrast, apoB<sup>+38.9</sup> × apoB<sup>+38.9</sup> crosses yield ~13% apoB<sup>38.9/38.9</sup> homozygotes, i.e., ~50% of the expected number. By contrast, heterozygous crosses of apoB<sup>+27.6</sup> yield only 3–4% apoB<sup>27.6/27.6</sup> homozygous offspring, similar to the yield of null

homozygotes, suggesting that the first 27.6% of the N-terminal region contains very little, if any, apoB sequence (or structure) able to support embryogenesis. In contrast, the next 11.3% (the stretch of sequence between apoB27.6 and apoB38.9) does contain such structures. This hypothesis was verified by making apoB<sup>+38.9</sup> × apoB<sup>+27.6</sup> crosses. The yields of compound heterozygotes (apoB<sup>38.9/27.6</sup>) were nearly identical to the yields of apoB 38.9 homozygotes; i.e., apoB38.9 was able to “rescue” apoB27.6 fetuses.

### ApoB38.9-Containing Lipoproteins Can Support Atherogenesis

ApoB-containing lipoproteins are the major atherogenic particles in plasma. After LDL particles enter the subendothelial space, some of the particles interact with proteoglycans and become oxidatively modified. This is an early event in atherogenesis. The interaction with proteoglycans is mediated by binding sites on apoB located near the N-terminal part of the molecule that were identified in studies carried out *in vitro*. The apoB38.9 mouse permitted the translation of these findings to the situation *in vivo*. The apoE<sup>-/-</sup> knockout mouse is a frequently used model of atherosclerosis. ApoE<sup>-/-</sup> mice develop florid aortic atherosclerosis even while eating normal low-cholesterol- and low-fat-containing mouse chow. ApoB-containing cholesterol-rich particles are responsible for the atherosclerosis in these animals. To assess whether apoB38.9-containing particles could support the development of aortic lesions, apoB<sup>38.9/38.9</sup> mice were crossed with apoE<sup>-/-</sup> mice. The resultant apoB<sup>38.9/38.9</sup>/apoE<sup>-/-</sup> mice developed just as much atherosclerosis as the apoB<sup>+/+</sup>/apoE<sup>-/-</sup> mice. Thus, the first 38.9% of the N-terminal end of apoB contains sufficient structure for lesion formation.

### Therapy

In abetalipoproteinemia, which produces severe symptoms related to malabsorption of fat and fat-soluble vitamins, low-fat diets and replacement therapy are mandatory. Since most subjects with FHBL are asymptomatic, the diagnosis is usually made on routine cholesterol screening. After the serious secondary causes for low cholesterol and apoB levels are ruled out, usually no therapy is needed. However, given the frequency of fatty liver in these subjects, periodic evaluations of liver function may be warranted.

## See Also the Following Articles

Abetalipoproteinemia • Anderson's Disease (Chylomicron Retention Disease) • Dysbetalipoproteinemia and Type III Hyperlipidemia • Hypercholesterolemias, Familial Defective ApoB (FDB) and LDL Receptor Defects • Lipid Disorders in the Elderly • Lipoprotein(a) • Low HDL/High HDL Syndromes • Mixed Lipemias

## Further Reading

- Berriot-Varoqueaux, N., *et al.* (2000). The role of the microsomal triglyceride transfer protein in abetalipoproteinemia. *Annu. Rev. Nutr.* **20**, 663–697.
- Chen, Z., *et al.* (1999). Truncated apo B-70.5-containing lipoproteins bind to megalin but not the LDL receptor. *J. Clin. Invest.* **103**, 1419–1430.
- Chen, Z., *et al.* (2000). A targeted apolipoprotein B-38.9-producing mutation causes fatty livers in mice due to the reduced ability of apolipoprotein B-38.9 to transport triglycerides. *J. Biol. Chem.* **275**, 32807–32815.
- Dannoura, A. H., *et al.* (1999). Anderson's disease: Exclusion of apolipoprotein and intracellular lipid transport genes. *Arterioscler. Thromb. Vasc. Biol.* **19**, 2494–2508.
- Elias, N., Patterson, B. W., and Schonfeld, G. (2000). *In vivo* metabolism of ApoB, ApoA-I, and VLDL triglycerides in a form of hypobetalipoproteinemia not linked to the ApoB gene. *Arterioscler. Thromb. Vasc. Biol.* **20**, 1309–1315.
- Jones, B., *et al.* (2003). Mutations of a Sar 1 GTPase of COPII vesicles are associated with lipid absorption disorders. *Nat. Genet.* **34**, 29–31.
- Kane, J. P., and Havel, R. J. (1995). Disorders of the biogenesis and secretion of lipoproteins containing the B apolipoproteins. In "The Metabolic and Molecular Bases of Inherited Disease," pp. 1853–1885. McGraw-Hill, New York.
- Lin, X., Schonfeld, G., and Chen, Z. (2001). Molecular adaptation of the fatty acid synthetic pathway to the presence of fatty livers in gene-targeted mice producing apolipoprotein B-39.8. *Arterioscler. Thromb. Vasc. Biol.* **21**, 661.
- Neuman, R. J., *et al.* (2002). Replication of linkage of familial hypobetalipoproteinemia to chromosome 3p in six kindreds. *J. Lipid Res.* **43**, 407–415.
- Schonfeld, G., *et al.* (2003). Fatty liver in familial hypobetalipoproteinemia: Triglyceride assembly into VLDL particles is affected by the extent of hepatic steatosis. *J. Lipid Res.* **44**, 470–478.
- Tarugi, P., *et al.* (2001). Phenotypic expression of familial hypobetalipoproteinemia in three kindreds with mutations of apolipoprotein B gene. *J. Lipid Res.* **42**, 1552–1561.
- Welty, F. K., Guida, K. A., and Andersen, J. J. (2001). Donor splice-site mutation (210 + 1G\_C) in the ApoB gene causes a very low level of ApoB-100 and LDL cholesterol. *Arterioscler. Thromb. Vasc. Biol.* **21**, 1864–1865.
- Yuan, B., *et al.* (2000). Linkage of a gene for familial hypobetalipoproteinemia to chromosome 3p21.1–22. *Am. J. Hum. Genet.* **66**, 1699–1704.
- Yue, P., *et al.* (2002). Novel mutations of APOB cause ApoB truncations undetectable in plasma and familial hypobetalipoproteinemia. *Hum. Mutat.* **20**, 110–116.

normal development) and position of the urethral meatus (for normal deposition of semen). The prostate exam enables one to assess prostate size, firmness, tenderness, and the presence of cysts (congenital and ejaculatory duct), which may be associated with ejaculatory duct obstruction.

The scrotal exam should be performed in a warm examining room to promote relaxation of the scrotum. It is important to record the position (scrotal and inguinal), volume (normal, ~15–25 ml), and consistency (normally firm) of the testis. Normally, more than 70% of the testis volume is from germ cells alone. Therefore, a soft and/or small testis is indicative of abnormal spermatogenesis. The diagnosis of a varicocele is generally made on physical examination. The detection of a varicocele is greatly facilitated by examining the patient standing. Varicoceles are graded as follows: grade I, palpable only with Valsalva; grade II, palpable without Valsalva; and grade III, visible. The presence or absence of the vas deferens can be confirmed by physical examination.

### Semen Analysis

The semen analysis is the cornerstone of male infertility evaluation. Two or three samples collected by masturbation after a minimum of 3 days of abstinence should be examined. Assessment of semen volume is important because it is indicative of seminal vesicle function (~70% of the semen volume is of seminal vesicle origin) and of the patency of the ducts distal to this gland (ejaculatory ducts). In azoospermic men, a low semen volume may be due to a distal obstruction (e.g., ejaculatory duct), seminal vesicle agenesis (associated with congenital absence of the vas deferens), hormonal deficiency, or, rarely, retrograde ejaculation.

Sperm concentration is indicative of quantitative spermatogenesis and the normal cutoff is 20 million sperm per milliliter. However, spontaneous pregnancies are possible for couples in which the man's sperm concentration is less than 20 million sperm/ml. Reports have suggested that a sperm concentration cutoff of 14 million sperm/ml may be more appropriate. Percentage sperm motility and sperm morphology are indicators of qualitative spermatogenesis. Typically, the percentage of sperm motility should be >50% and the percentage of sperm with normal morphology should be >30%. It is generally accepted that sperm morphology is a better predictor of fertility than sperm concentration or sperm motility. Although each of the conventional semen parameters (sperm concentration, motility, and morphology) may

help to differentiate infertile from fertile men, none of these parameters is a powerful discriminator because there is significant overlap between semen parameters of infertile and fertile men. Indeed, some infertile men have normal semen parameters and are said to have unexplained infertility. Moreover, conventional semen parameters exhibit a high degree of variability from sample to sample; as such, there is a need for multiple analyses.

Specialized sperm function tests have been developed in the hope of better discriminating fertile from infertile men. It is also hoped that these tests may help predict fertility potential (*in vivo* and *in vitro*) better than conventional semen parameters. Finally, evaluating sperm function may increase our understanding of spermatogenesis and gamete biology. However, no single specialized sperm function test has been universally accepted.

### Hormonal Evaluation

The initial hormonal evaluation should include measurement of serum follicle-stimulating hormone (FSH) and testosterone (T) levels. Abnormal FSH or T levels should prompt further evaluation with serum luteinizing hormone (LH) and prolactin determination.

Typically, most men with spermatogenic failure have an elevated serum FSH, with or without an elevated serum LH and low serum T. Although there is no absolute FSH cutoff above which active spermatogenesis is ruled out, most clinicians agree that men with FSH levels greater than twice the upper limit of normal (generally associated with small, soft testes) probably have nonobstructive azoospermia (with the predominant histology being maturation arrest or Sertoli cell-only pattern), and a confirmatory testis biopsy is not indicated. As such, men with high FSH and small testes are not candidates for surgical reconstruction. However, men with elevated serum FSH may harbor small foci of active spermatogenesis amid a histologic background that is predominantly characterized by maturation arrest or Sertoli cell-only pattern. In these men, a diagnostic biopsy may be useful to predict the presence of sperm for a subsequent therapeutic testis biopsy [used in conjunction with *in vitro* fertilization (IVF) and intracytoplasmic sperm injection (ICSI)].

Low serum FSH, T, and LH is suggestive of hypogonadotropic hypogonadism with secondary spermatogenic failure. In these men, measurement of prolactin level is indicated together with pituitary

imaging to rule out a pituitary lesion. These men often respond to hormonal replacement therapy (in the form of intramuscular human chorionic gonadotropin (hCG) and FSH or pulsatile subcutaneous gonadotropin-releasing hormone). A normal hormonal evaluation is suggestive of normal, active spermatogenesis but does not rule out spermatogenic failure.

### Imaging Studies

Further assessment of reproductive tract anatomy is possible with the aid of ultrasonography. A scrotal/testicular ultrasound is used for assessment of testicular volume, epididymal anatomy, and the presence or absence of varicoceles. A transrectal ultrasound (TRUS) is primarily used to assess prostate anatomy. In infertile men, a TRUS is also used to evaluate the seminal vesicles and ejaculatory ducts in order to determine whether there is a genital tract obstruction at this level. Vasography is used (at the time of surgical reconstruction) to confirm and localize a ductal obstruction.

### Genetic Evaluation

Many cases of male infertility may be associated with a genetic defect. One concern is that the genetic defect responsible for the man's infertility may be transmitted to his child through assisted reproduction. In many cases, this causes the same problem in the next generation (infertility in a male child), but some genetic defects may cause other diseases, miscarriage, or even early death. As such, appropriate genetic testing and counseling are an essential part of the infertility evaluation. The available clinical genetic tests for male-factor infertility include chromosome analysis, Y microdeletion analysis, and cystic fibrosis mutation analysis.

#### **Chromosomal Analysis or Karyotype**

The prevalence of chromosomal abnormalities in the general population is less than 1%. The prevalence of chromosomal abnormalities in men with azoospermia is up to 14%. Klinefelter's syndrome (47,XXY) is the most common abnormality, observed in approximately 10% of azoospermic men. Other chromosome abnormalities described in infertile men include 47,XYY; 46,XX; X-autosome, Y-autosome, and auto-some-autosome translocations; and pericentric inversions. Sex chromosome abnormalities are seen in approximately 12% and autosome abnormalities in

1% of azoospermic men. In men with oligospermia, the prevalence of chromosomal abnormalities is approximately 5%: Sex chromosome abnormalities are seen in 1 or 2% and autosome abnormalities constitute the other 3%. In clinical practice, a karyotype is recommended in infertile men with severe oligospermia (<10 million sperm/ml) and nonobstructive azoospermia. Clearly, the karyotype analysis is important given the serious reproductive consequences of transferring an abnormal chromosomal complement to the offspring (e.g., miscarriage and birth malformations).

#### **Y Microdeletion**

In approximately 10% of men with azoospermia or severe oligospermia (<5 million sperm/ml), a portion of the long arm of the Y chromosome is missing. It is speculated that one or more genes important in spermatogenesis reside within this deleted region of the Y chromosome, which explains the observed phenotype associated with Y microdeletions. It does not appear that men with this condition have any other physical problems apart from infertility. There is evidence that men with a Y microdeletion pass on the genetic abnormality to their male offspring through IVF/ICSI. Genetic counseling and Y microdeletion testing are recommended for infertile men with severe oligospermia and nonobstructive azoospermia.

#### **Cystic Fibrosis Gene and Male Infertility**

The cystic fibrosis transmembrane conductance regulator (CFTR) gene is located on chromosome 7. More than 800 different mutations of the CF gene have been reported. In the Caucasian population, the carrier frequency of these autosomal recessive mutations is approximately 4%. Mutations of this gene have been associated with CF and male infertility. CF mutations have been associated with congenital bilateral absence of the vas deferens (CBAVD) (80%), epididymal obstruction (30%), and oligospermia (~20%). They have not been associated with testicular failure or nonobstructive azoospermia. CF mutations result in a defective chloride channel and this is associated with production of thick respiratory and gastrointestinal secretions, leading to the CF phenotype. In male infertility, CF generally results in obstruction of the excurrent duct system, resulting in obstructive azoospermia. Genetic counseling and CF mutation testing are recommended for infertile men with CBAVD, idiopathic epididymal obstruction, and oligospermia.

## Testicular Biopsy

A testicular biopsy is indicated for men with azoospermia and normal (or slightly elevated) serum FSH to confirm active spermatogenesis on testicular histology and, hence, diagnose obstructive azoospermia. Men with prior vasectomy and men with CBAVD generally have normal spermatogenesis, and a testis biopsy is not usually indicated. Men with elevated serum FSH (at least twice the upper limit of normal) likely have nonobstructive azoospermia, and a testis biopsy is not indicated because it is unlikely to indicate active spermatogenesis. However, with the advent of advanced assisted reproductive technologies, it is possible to harvest sperm from the testis of men with nonobstructive azoospermia because they may harbor rare foci of spermatogenesis and to use the sperm for IVF with ICSI. As such, the indications for testicular biopsy have broadened.

## MANAGEMENT

### General Recommendations

Prior to initiating therapy (medical or surgical) for the treatment of male infertility, efforts should be made to eliminate any potential gonadotoxins (agents toxic to the gonads or testicles). A number of medications are known to have adverse effects on sperm production, including hormonal agents (e.g., testosterone or its derivatives), antineoplastic agents (chemotherapeutic drugs), and other miscellaneous drugs. Cigarette smoking is associated with decreased sperm counts and sperm function. Similarly, excessive alcohol consumption (binge drinking) has a detrimental effect on sperm production in the testis. Proper nutrition is also important to maintain adequate sperm production. Finally, chronic exposure to excessive heat (hyperthermia) is detrimental to normal sperm production in the testis.

### Medical Therapy for Spermatogenic Failure

Specific medical therapy (targeted to specific pathology, e.g., hormonal replacement for hypogonadotropic hypogonadism) is generally effective in improving spermatogenesis and semen quality. On the other hand, nonspecific or empiric therapy (targeted to unknown pathology) is largely ineffective and should remain experimental until studies prove the contrary. As such, men with idiopathic spermatogenic failure are not candidates for empiric medical therapy.

For most men with idiopathic male infertility associated with spermatogenic failure, assisted reproduction (AR) is the treatment of choice.

A small subset (<5%) of infertile men with spermatogenic failure suffer from endocrine infertility. Specific causes of secondary testicular failure include hypogonadotropic hypogonadism, hyperprolactinemia, hypothyroidism, and congenital adrenal hyperplasia.

Hypogonadotropic hypogonadism can be congenital (Kallman's syndrome) or acquired (secondary to a pituitary adenoma, pituitary infarction, radiotherapy, or idiopathic). The standard treatment involves replacement of deficient hormones. Typically, patients are treated with intramuscular human chorionic gonadotropin (hCG, 1000–2000 IU two to three times per week) and human menopausal gonadotropin (100 IU two or three times per week) for at least 1 year. Alternatively, pulsatile GnRH (2–40 µg every 2 h by pump) can be administered. The newer recombinant gonadotropins have also been used. Generally, this form of treatment results in increased testicular volume and improved semen quality, with many men obtaining normal fertility.

Hyperprolactinemia can be secondary to a pituitary tumor (generally benign), hypothyroidism, liver disease, medication (antidepressants and cimetidine), or idiopathic. Typically, patients are treated with bromocriptine (2.5–10 mg/day), which suppresses growth of prolactin-secreting tumors, or the longer acting form cabergoline (1 mg/week). For large tumors, transphenoidal resection of the pituitary adenoma is recommended.

Hypothyroidism or hyperthyroidism may be associated with depressed sperm production. Treatment consists of thyroid hormone replacement or ablation of the hyperactive thyroid. Generally, this leads to improved sperm production and return of fertility.

Congenital adrenal hyperplasia is a condition caused by a congenital enzyme deficiency leading to depressed steroid levels. These patients are treated with steroid replacement and generally this leads to improved sperm production and return of fertility.

### Surgical Therapy for Spermatogenic Failure

#### Varicocelectomy

Varicocelectomy is the only male infertility surgery specifically designed to improve spermatogenesis. Varicocelectomy is indicated in men with a clinical varicocele, abnormal semen parameters and couple infertility. This is based on the demonstration that varicocele is associated with a progressive decline in

testicular function and that repair of varicocele can improve spermatogenesis. Varicocelectomy is also indicated in men with a clinical varicocele and testicular pain and in the child or adolescent with a clinical varicocele and decreased ipsilateral testicular volume (>2 ml difference between the right and left testis).

Many studies have evaluated the outcome of varicocelectomy on fertility parameters, and most have demonstrated an improvement in semen quality and pregnancy rates. However, because the bulk of the outcome data on varicocelectomy are derived from uncontrolled or poorly designed controlled studies, the value of these results is limited. Overall, varicocelectomy results in significant improvement in semen analysis in 60–80% of men, and pregnancy rates after varicocelectomy vary from 20 to 60%.

Studies indicate that there may be some benefit in repairing varicoceles in infertile men with azoospermia and clinical varicocele. Although significant improvement in semen quality (the appearance of sperm in the semen) is reported in approximately 50% of these men, a clinically significant outcome (with spontaneous pregnancy) is reported in less than 20% of these cases. Preoperative testicular biopsy is predictive of outcome in these cases. Only men with mature spermatids or spermatozoa on testicular biopsy had a good outcome (the appearance of sperm in the semen). Men with maturation arrest or Sertoli cell-only pattern on testicular biopsy remained azoospermic postoperatively.

### **Surgical Sperm Retrieval**

The advent of advanced AR has made it possible for men with spermatogenic failure to father children. In the subset of men with spermatogenic failure and azoospermia, only testicular surgical sperm retrieval coupled with AR will allow these men to father children. Although a standard open testicular biopsy can be used to extract testicular sperm in men with spermatogenic failure, microsurgical testicular sperm extraction (micro-TESE) has been advocated as the gold standard approach for sperm retrieval in men with nonobstructive azoospermia. With the testicle opened, the operating microscope enables the surgeon to identify and extract the most dilated seminiferous tubules (indicative of active spermatogenesis) while thinner tubules (indicative of germ cell aplasia) are spared. As such, micro-TESE allows for maximal sperm recovery and minimal tissue extraction.

Spermatozoa recovered by testicular biopsy can only be used in conjunction with IVF/ICSI (in which a single sperm is injected into the oocyte). Pregnancy rates are in the range of 30–40%.

## **Assisted Reproduction**

Assisted reproduction is indicated for couples in which the man suffers from idiopathic spermatogenic failure. AR is also indicated for couples in which the man has a specific cause of spermatogenic failure but has not responded to conventional therapy (e.g., varicocelectomy for varicocele) or has chosen AR instead of conventional therapy.

In men with a mild degree of spermatogenic failure and borderline semen parameters, three to six cycles of sperm washing with intrauterine insemination (IUI) should be attempted first. IUI is a minimally invasive form of AR; however, pregnancy rates are in the range of 5–15% per cycle and are highly dependent on the concentration of motile sperm (the minimum threshold for effective IUI is 3–5 million motile sperm).

For couples in which the man has severe oligospermia or nonobstructive azoospermia, only the most advanced AR (IVF/ICSI) will allow them to achieve a pregnancy. In men with nonobstructive azoospermia, a surgical testicular sperm retrieval procedure is necessary to obtain spermatozoa. IVF/ICSI is also indicated for couples that have failed to achieve a pregnancy through IUI. However, the pregnancy rates with advanced AR (e.g., IVF/ICSI) remain modest, with most centers reporting IVF/ICSI pregnancy rates of 30–40% when female factors (e.g., advanced female age) are excluded.

## **See Also the Following Articles**

Klinefelter's Syndrome • Spermatogenesis, Endocrine Control of

## **Further Reading**

- Barratt, C. L., and St. John, J. C. (1998). Diagnostic tools in male infertility. *Hum. Reprod.* **13**(Suppl. 1), 51–61.
- Bonduelle, M., Camus, M., DeVos, A., Staessen, C., Tournaye, H., Van Assche, E., Verheyen, G., Devroey, P., Liebaers, I., and Van Steirteghem, A. (1999). Seven years of intracytoplasmic sperm injection and follow-up of 1987 subsequent children. *Hum. Reprod.* **14**(Suppl.), 243–264.
- de Kretser, D. M., Loveland, K. L., Meinhardt, A., Simorangkir, D., and Wreford, N. (1998). Spermatogenesis. *Hum. Reprod.* **13**(Suppl. 1), 1–8.
- Fraser, L. R. (1999). New insights into possible causes of male infertility. *Hum. Reprod.* **14**(Suppl. 1), 38–46.
- Greenberg, S. H., Lipschultz, L. I., and Wein, A. J. (1978). Experience with 425 subfertile male patients. *J. Urol.* **119**, 507–510.
- Guzick, D. S., Overstreet, J. W., Factor-Litvak, P., Brazil, C. K., Nakajima, S. T., Coutifaris, C., Carson, S. A., Cisneros, P., Steinkampf, M. P., Hill, J. A., Xu, D., and Vogel, D. L. (2001).

- Sperm morphology, motility, and concentration in infertile and fertile men. *N. Engl. J. Med.* **345**, 1388–1393.
- Kamischke, A., and Nieschlag, E. (1999). Analysis of medical treatment of male infertility. *Hum. Reprod.* **14**(Suppl. 1), 1–23.
- Schlegel, P. N. (1999). Testicular sperm extraction: Microdissection improves sperm yield with minimal tissue excision. *Hum. Reprod.* **14**, 131–135.
- Schlesinger, M. H., Wilets, I. F., and Nagler, H. M. (1994). Treatment outcome after varicocelectomy. A critical analysis. *Urol. Clin. North Am.* **21**, 517–529.
- World Health Organization (1999). “WHO Laboratory Manual for the Examination of Human Semen and Sperm–Cervical Mucus Interaction,” 34th ed. Cambridge Univ. Press, Cambridge, UK.

the fluidity of the phospholipid bilayer and facilitates the acrosome reaction. As sperm progress through the reproductive tract, cholesterol is removed from the plasma membrane by albumin and high-density lipoproteins found in the female genital tract. Membrane fluidity may also be increased by removal of phospholipids in a second mechanism. In some species, extracellular glycoproteins are stripped from sperm traversing the genital tract. When glycoproteins are removed from the membrane, associated phospholipids are also removed. These modifications of the plasma membrane may expose membrane-bound enzymes and ligands utilized in sperm penetration of the cumulus oophorus and anchoring to the zona pellucida.

Other hallmarks of capacitation are an increase in adenylate cyclase, a subsequent increase in cyclic AMP (cAMP) and cAMP-dependent protein kinases, and an increase in the intracellular calcium concentration. Capacitation is absolutely dependent on an influx of extracellular calcium, which may be at least partially mediated by T-type (voltage-sensitive) calcium channels present in the sperm membrane. The elevation of cAMP affects numerous cellular events and is necessary for the acquisition of hyperactivated motility.

Hyperactivation of the sperm occurs concomitantly with capacitation. Capacitation and hyperactivation have been delineated as separate processes by studies showing that sperm can be hyperactivated but not capacitated by incubation in culture medium containing calcium but not containing serum or albumin. Both events are triggered in part by extracellular calcium, but hyperactivation manifests as changes in motility, whereas capacitation is principally seen as membrane remodeling. Hyperactivated sperm have a distinct figure-eight tail whip and move in a less progressive fashion than nonhyperactivated sperm. The changes in motility associated with hyperactivation are likely requisite for sperm penetration of the oocyte vestments.

Physical and functional barriers to fertilization exist and may serve as selection mechanisms. Prior to capacitation, sperm must have sufficient motility to move to the site of fertilization. Only a minority of ejaculated sperm undergo capacitation. In the human, the cumulus facilitates the selection of both morphologically normal and capacitated sperm. Only capacitated sperm can undergo zona binding and the acrosome reaction, but a premature acrosome reaction will preclude zona binding. Studies have shown that severely infertile patients may have a deficient capacitation ability and increased incidence

of chromosome aneuploidies. Therefore, proper capacitation appears to be one of the key selection processes of fertilization. The significance of the selection process is not well understood and has considerable ramifications in assisted reproductive technologies.

## SPERM BINDING TO THE ZONA PELLUCIDA

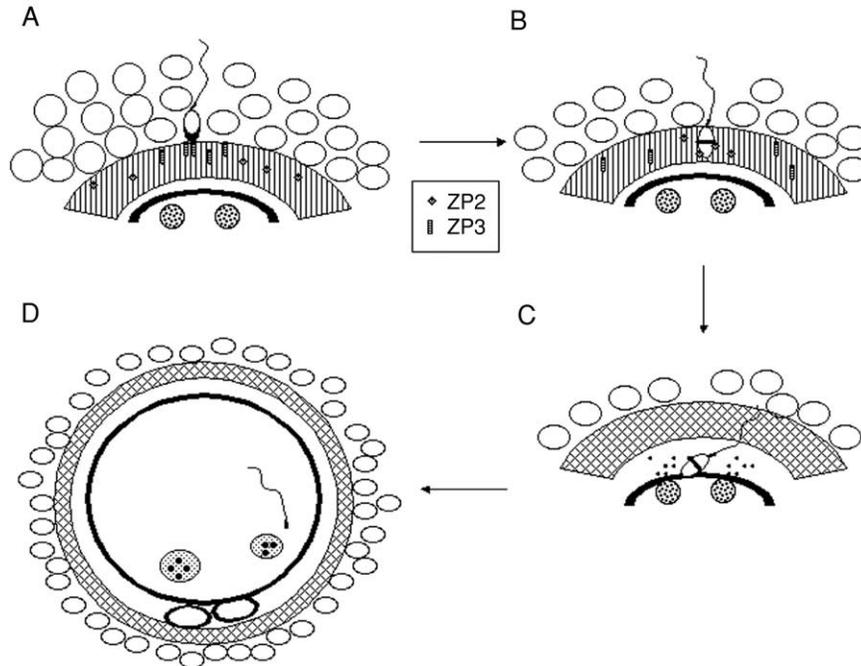
Following penetration of the cumulus, the sperm cell must bind and penetrate the zona pellucida, which is both a barrier and a facilitator of fertilization. The zona is an integral part of the block to polyspermy, provides species specificity, and is a selective barrier for capacitated sperm. It facilitates fertilization by stimulating the acrosome reaction, which is necessary for binding to the oolemma. In mammals, the zona pellucida is composed of three secreted glycoproteins, ZP1, ZP2, and ZP3. ZP2 and ZP3 are heavily cross-linked by ZP1, the "structural" zona protein, creating a porous matrix around the oocyte.

Glycosylation of ZP3 near the carboxy terminus is slightly varied in different species. Variation of the O-linked oligosaccharides, found at the carboxy terminus, provides species specificity by allowing only complementary sperm to bind and penetrate the zona. Sperm egg-binding proteins (EBPs) bind to the zona pellucida through the O-linked oligosaccharides on ZP3 (Fig. 1A). Several candidate EBPs, such as sp56, p95, and GalTase, likely become available for binding on the surface of the sperm through membrane remodeling as a result of capacitation.

## ACROSOME REACTION AND PENETRATION OF THE ZONA PELLUCIDA

The penetration of the zona pellucida is driven by two physiological events, sperm motility in the hyperactive state and the acrosome reaction. Although the acrosome reaction is stimulated by the binding of sperm to O-linked oligosaccharides of ZP3, ZP3 will not induce uncapacitated cells to undergo the acrosome reaction. Binding to ZP3 stimulates two pathways: G protein-linked signal cascades and voltage-sensitive T-type calcium channels. Activated second messengers and T-type channels raise the intracellular calcium concentration and activate phospholipase C, inducing the acrosome reaction.

The acrosome reaction involves the fusion of the outer acrosomal membrane and the plasma



**Figure 1** An overview of fertilization. (A) Capacitated sperm penetrate the cumulus cells and bind to ZP3 on the zona pellucida. (B) Following induction of the acrosome reaction by ZP3, ZP2 ligands facilitate zona penetration. (C) The equatorial region of the sperm plasma membrane binds to the oolemma and fuses. (D) Decondensation of the sperm head and formation of the male and female pronuclei.

membrane. The fusion and vesiculation of the membranes cause the release of the enzymes stored within the acrosome. These enzymes include acid glycosylhydrolases, acid phosphatases, aryl sulfatases, esterases, and acrosin. When released, the proteases may disrupt the zona pellucida matrix and facilitate zona penetration. The penetration of the zona is dependent on ZP2, which also has specific sperm surface receptors (Fig. 1B).

The acrosome reaction results in a modification of the phospholipid and protein content of the resulting plasma membrane. Additionally, several transmembrane and membrane-associated proteins are localized to the equatorial region of the sperm head. These proteins create equatorial binding motifs on the sperm recognized by the oolemma.

### OOLEMA–SPERM BINDING AND FUSION AND THE BLOCK TO POLYSPERMY

Sperm–oocyte binding is mediated by proteins from the ADAM (a disintegrin and metalloproteinase domain) family. They include fertilin and crytistin. The receptor on the oocyte is thought to be from the

integrin protein family. Once bound to the oocyte, the sperm plasma membrane begins to integrate with the oolemma (Fig. 1C). The fusion of the gametes is dependent on the sperm protein CD9, an integral membrane protein that associates with the integrin proteins of the oocyte.

There are two postulates proposed for the mechanism of gamete fusion, viral-type or SNARE (soluble *N*-ethylmaleimide-sensitive fusion attachment protein receptor)-type fusion. Viral-type fusion relies on the presence of a fusion peptide region found in fertilin- $\alpha$ , formerly PH30. The fusion peptide causes a destabilization of the lipid bilayer and fusion pores are created in a manner similar to the fusion mechanism of viruses. SNARE-type fusion has been proposed but has not been described in gamete fusion. In this mechanism, transport vesicles and subsequent fusion are directed by t- and v-SNAREs located on the target membrane and the vesicle, respectively.

Oocyte activation, in response to sperm–oocyte fusion, is the first step to incorporation of the sperm nucleus and the block to polyspermy. Two pathways are involved in oocyte activation. First, as the sperm fuses with the oocyte, soluble proteins or peptides capable of activating calcium channels in the

membrane are released into the oocyte. Second, G protein activation mobilizes intracellular calcium stores through inositol 1,4,5-triphosphate/diacylglycerol activity. The calcium transients or waves originate from the site of sperm entry and then propagate through the oocyte. Calcium waves signal the onset of oocyte activation and prepare the oocyte cytoplasm to receive the sperm nucleus. The calcium waves continue through pronuclear formation.

The first calcium waves are followed by the release of cortical granules, which remodel the oolemma and modify the zona. In addition, there is an increase in the perivitelline space. These changes preclude further sperm penetration of the zona pellucida or binding of the oolemma, thus blocking polyspermy.

### PRONUCLEAR FORMATION AND SYNGAMY

The sperm nucleus is largely inactive at the time it enters the oocyte. During spermiogenesis, nuclear histones are replaced by protamines to increase the chromatin packaging and the sperm becomes transcriptionally repressed. After oocyte penetration, the sperm nucleus and chromatin must be made ready for fusion with its maternal counterpart. This requires decondensation of the sperm nucleus, which is initiated by degradation of the nuclear envelope by oocytic lipases. During decondensation, protamines bound to the DNA are replaced with maternally derived histones. Decondensation results in a swelling of the sperm head and a release of the sperm tail. The sperm pronucleus is completed by addition of a nuclear membrane consisting of oocytic lipid vesicles and nuclear pore complexes (Fig. 1D).

The newly formed pronucleus migrates centrally toward the simultaneously forming female pronucleus (Fig. 2). The migration is regulated by microtubule formation, which is directed by the microtubule organizing center (MTOC). The main component of the MTOC is the centrosome, which is paternally derived in the human. After the pronuclei have aligned centrally in the oocyte, the paternal pronuclear envelope is degraded and the maternal pronuclear envelope encompasses both fractions. Syngamy, the final step of the fertilization process, is completed with the alignment of homologous chromosomes, creating a unique genome for the embryo. During fertilization, the oocyte is in metaphase II; one copy of the maternal chromosomes is extruded in the second polar body and fertilization is complete.

### CLINICAL IMPLICATIONS

Any abnormalities in the fertilization process may lead to infertility. The most common abnormality of fertilization is diminished capacitation, which results in a lower probability of a competent sperm binding and penetrating the zona. Although there is limited availability of diagnostic techniques to evaluate the other steps of fertilization, abnormalities of ZP3, the acrosome reaction, oocyte activation pronuclear formation, syngamy, and the block to polyspermy have all been reported.

Polyspermy, fertilization by more than one sperm, is a common fertilization error. The mechanisms to block polyspermy are less efficient in older women undergoing *in vitro* fertilization (IVF) and may be absent in some women. Most dispermic embryos undergo early demise and account for approximately 10% of detected spontaneous abortions. However, most polyspermic embryos will not implant or be detected when occurring *in vivo*. Triploid concepti may also be of maternal origin, as a result of abnormal meiosis in the oocyte. Some cases of triploidy result in hydatidiform moles, concepti with hyperplasia of the trophoblast. Moles are more likely to form if the triploidy is of paternal origin. Complete moles are diploid, but completely androgenic and lacking any fetal tissue.



**Figure 2** The male and female pronuclei migrate toward the center of the oocyte through microtubule-assisted movement. The nucleoli are visible and align vertically.

Two techniques have greatly facilitated therapy of some types of fertilization failure. *In vitro* sperm capacitation techniques increase the percentage of capacitated sperm available for artificial insemination or IVF. Sperm capacitation techniques vary from simple washing, which is the removal of sperm from semen and resuspension in culture medium, to more complex preparations, including density gradient isolation of motile sperm and/or incubation in culture medium containing calcium ionophores, phosphodiesterase inhibitors, or a high concentration of albumin. The second therapy, intracytoplasmic sperm injection (ICSI), involves the direct injection of sperm into the cytoplasm of the oocyte during IVF. ICSI will effectively treat defects of capacitation, zona binding, the acrosome reaction, zona penetration, and oolemma-sperm fusion. Future therapies to treat fertilization failure may include such diverse procedures as cloning, oocyte cytoplasmic transfer, and artificial oocyte activation.

### See Also the Following Articles

Assisted Reproductive Technology (ART) • Fertility in Men with Spermatogenesis Abnormalities • Implantation • In

Vitro Fertilization (IVF) • Infertility, Overview • Ovarian Failure Treatment Strategies: Egg Donation • Pregnancy Endocrinology • Premature Ovarian Failure • Spermatogenesis, Endocrine Control of • Superovulation and Intrauterine Insemination

### Further Reading

- Bailey, J. L., Tardif, S., and Dubé, C. (2001). "Andrology in the 21st Century" (B. Robaire, H. Chemes, and C. R. Morales, eds.). Medimond, Englewood, NJ.
- Küpker, W., Diedrich, K., and Edwards, R. G. (1998). Principles of mammalian fertilization. *Hum. Reprod.* **13**, 20–32.
- Plachot, M. (2000). Fertilization. *Hum. Reprod.* **15**, 19–30.
- Schatten, H., Hueser, C. H., and Chakrabarti, A. (2000). From fertilization to cancer: The role of centrosomes in the union and separation of genomic material. *Microsc. Res. Tech.* **49**, 420–427.
- Toshimori, K. (2001). "Andrology in the 21st Century" (B. Robaire, H. Chemes, and C. R. Morales, eds.). Medimond, Englewood, NJ.
- Wassarman, P. M. (1999). Fertilization in animals. *Dev. Genet.* **25**, 83–86.
- Wassarman, P. M., Jovine, L., and Litscher, E. S. (2001). A profile of fertilization in mammals. *Nat. Cell Biol.* **3**, E59–E64.
- Yanagimachi, R. (2001). Gamete manipulation for development: New methods for conception. *Reprod. Fertil. Dev.* **13**, 3–14.

third zone, between the definitive and fetal zones of the adrenal, known as the transitional zone. After mid-gestation, transitional zone cells have the capacity to synthesize cortisol, and by 30 weeks gestation in the human, the definitive and transitional zones take on the structural and functional characteristics of the zona glomerulosa and fasciculata, respectively, of the adult adrenal cortex.

## FUNCTIONAL DEVELOPMENT OF THE FETAL ADRENAL

During mid-gestation (16–22 weeks), the fetal adrenal lacks the enzyme 3 $\beta$ -hydroxysteroid dehydrogenase/isomerase (3 $\beta$ HSD), a critical enzyme for the *de novo* synthesis of glucocorticoid and mineralocorticoid hormones. By 28 weeks gestation, 3 $\beta$ HSD is expressed within the definitive and transitional zones, and it appears that the fetal adrenal can synthesize and secrete cortisol from at least 30 weeks gestation. Interestingly, it appears that the transitional zone is the major source of cortisol production given that 3 $\beta$ HSD and CYP17 both are localized within this zone, whereas CYP17 is not expressed in the definitive zone that, therefore, produces aldosterone during late gestation. 3 $\beta$ HSD is not expressed in the fetal zone of the adrenal cortex throughout gestation. However, there is high level of expression of the enzyme CYP17 in the fetal zone; therefore, this zone produces large quantities of the androgenic C19 steroid dehydroepiandrosterone sulfate (DHEA-S), which is the major steroid produced by the primate adrenal during gestation. There is evidence from a range of experimental studies that adrenocorticotrophic hormone (ACTH) derived from the fetal pituitary plays a key role in the regulation of adrenal growth and in the zone-specific changes in steroid output during mid- and late gestation. For example, treatment with an agent, metyrapone (which increases fetal ACTH concentrations), results in induction of expression of 3 $\beta$ HSD in the transitional zone of the fetal rhesus monkey adrenal gland. Furthermore stimulation of the hypothalamic–pituitary–adrenal axis results in a premature stimulation of hypertrophic growth and steroidogenesis in both the definitive and transitional zones of the fetal adrenal during late gestation. These data suggest that the ontogenetic changes in fetal pituitary ACTH secretion and in adrenal responsiveness to ACTH play an important role in the induction of 3 $\beta$ HSD expression in the transitional zone.

ACTH also plays an important role in the marked growth of the fetal zone, which comprises 80 to 90%

of the gland during most of gestation and in the stimulation of DHEA-S secretion from this zone. Disruption of the fetal hypothalamo–pituitary axis in the human fetus results in a failure of the fetal zone to grow after 15 weeks gestation and dexamethasone administration to pregnant rhesus monkeys during late gestation, suppressing fetal hypothalamo–pituitary function and resulting in atrophy of the fetal zone. There is also substantial evidence that the actions of ACTH on fetal adrenal growth may be mediated by intra-adrenal growth factors, including insulin-like growth factor-2 (IGF-2).

After birth, there is a rapid remodeling of the fetal zone of the adrenal cortex; therefore, it has been suggested that the placenta may play a role in the differentiation, maintenance, or growth of this zone. The specific role of placental factors in the regulation of fetal adrenal steroid output is discussed in the next section.

## THE FETAL ADRENAL, DHEA, AND THE PLACENTA

The primate placenta does not produce estrogens *de novo* from cholesterol and cannot convert pregnenolone or progesterone into C19 steroids because it lacks CYP17. Therefore, the human placenta produces estrogens by aromatization of the C19 precursors DHEA-S and the 16-hydroxyl form of DHEA-S, which are synthesized predominantly in the fetal adrenal and liver, respectively. In the placenta, sulfatase acts to produce DHEA and 16-hydroxy DHEA, which are then aromatized to estradiol, estrone, and estriol. Therefore, maternal estrogen levels rise in parallel with the increase in the output of DHEA-S from the fetal adrenal during late gestation. It has been demonstrated in the rhesus monkey that infusion of androstenedione at 0.8 of gestation results in an increase in maternal plasma estrogen concentrations and in premature delivery. The coadministration of the aromatase inhibitor 4-hydroxyandrostenedione prevented the androstenedione-induced changes in maternal estrogen concentrations and delivery. Because estrogen infusion alone does not induce premature birth, this suggests that it is the autocrine/paracrine functions of estrogen at its site of production that play critical and central roles in delivery in the primate.

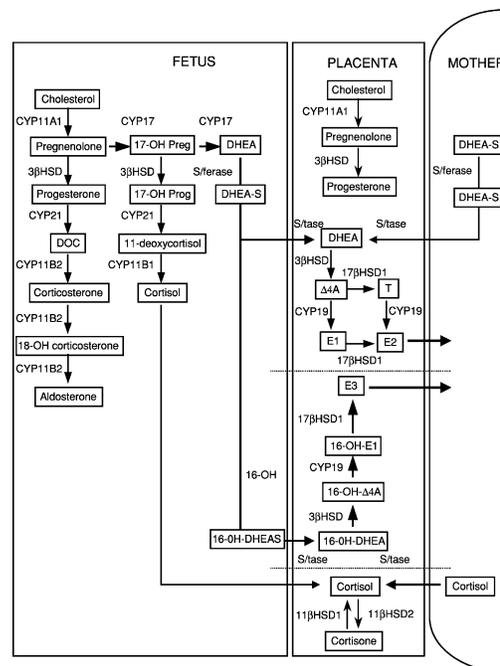
## THE FETAL ADRENAL, CORTISOL, AND THE PLACENTA

In most mammalian species, including the primate, there is an increase in fetal plasma glucocorticoid

concentrations during late gestation due to activation of the fetal hypothalamo–pituitary axis. Glucocorticoids play a key role in inducing maturational changes in the fetal lungs, kidneys, liver, and gut—therefore, in a successful transition from intrauterine life to extra-uterine life. Although cortisol is secreted by the fetal adrenal during late gestation, it is also produced locally within the placenta and a range of fetal tissues. The interconversion of biologically active cortisol and inactive cortisone is catalyzed by two isoforms of the enzyme 11 $\beta$ -hydroxysteroid dehydrogenase (11 $\beta$ HSD-1 and 11 $\beta$ HSD-2). It has been proposed that the placenta can regulate fetal adrenal function through the regulation of the amount of maternal cortisol that crosses the placenta to the fetus via regulation of the activity of placental 11 $\beta$ HSD-2. Studies in the baboon have shown that during mid-gestation there is an increase in the reduction of cortisone to cortisol in the placenta and have proposed that this would result in a decrease in fetal ACTH secretion (via increased negative feedback action of the maternally derived glucocorticoids on the fetal hypothalamo–pituitary axis) and that factors other than ACTH then stimulate the growth and function of the fetal zone. During later pregnancy, there is an increase in the oxidation of cortisol to cortisone in the baboon placenta, and this would allow ACTH concentrations to increase with an associated increase in adrenal growth and DHEA-S production. Interestingly, placental estrogens can stimulate increased conversion of cortisol to cortisone, and this would result in a positive feedback loop leading to an increased secretion of ACTH from the fetal pituitary, DHEA-S from the adrenal, and enhanced placental estrogen production.

During late gestation, there is also an increase in prostaglandin (PG) synthesis by the intrauterine tissues, and administration of PG synthase inhibitors suppress uterine activity and prolong gestation, whereas exogenous PGs stimulate uterine activity. It has been proposed that glucocorticoids may be one of several regulatory factors that stimulate the synthesis and secretion of intrauterine PGs. Although cortisol may be derived from the fetal adrenal, placental 11 $\beta$ HSD-2 activity is decreased by PGs. Thus, an increase in PG output during late gestation would result in a decrease in the conversion of cortisol to cortisone in the placenta, resulting in a further increase in cortisol stimulation of PG production. [Figure 1](#) provides an illustration of the steroidogenic pathway.

Corticotropin-releasing hormone (CRH) is present in placental tissue and the decidua in increasing amounts during human pregnancy, and CRH concentrations also increase in umbilical cord blood



**Figure 1** Schematic of the steroidogenic pathway describing the sites of synthesis of the key adrenal and placental steroids in the fetal–placental unit. DOC, deoxycorticosterone; 17-OH-Preg, 17-hydroxy-pregnenolone; 17-OH-Prog, 17-hydroxy-progesterone; DHEA, dehydroepiandrosterone; DHEA-S, dehydroepiandrosterone sulfate; D4A, androstenedione; T, testosterone; E1, estrone; E2, estradiol; E3, estriol; CYP11A1, cytochrome P450 cholesterol side-chain cleavage; CYP17, cytochrome P450 17 $\alpha$ -hydroxylase, CYP21, cytochrome P450 21-hydroxylase; CYP11B1, cytochrome P450 11 $\beta$ -hydroxylase; CYP11B2, aldol synthase; CYP19, cytochrome P450 aromatase; 3 $\beta$ HSD, 3 $\beta$ -hydroxysteroid dehydrogenase; 17 $\beta$ HSD1, 17 $\beta$ -hydroxysteroid dehydrogenase type 1; 16-OH, 16-hydroxylase; S/ferase, DHEA sulfotransferase; S/tase, sulfatase.

samples. In addition, it has been shown that maternal plasma CRH concentrations are predictive of the subsequent length of gestation. One proposal is that placental CRH influences fetal adrenal growth and steroidogenesis and that, in turn, fetal cortisol can stimulate placental CRH production, creating a positive feedback loop between the placenta and the fetal hypothalamo–pituitary–adrenal axis. Importantly, it has been shown that CRH can act directly on the fetal zone of the fetal adrenal gland to increase the output of DHEA.

## INTRAUTERINE STRESS AND THE FETO-PLACENTAL UNIT

There has been considerable recent interest in the impact of poor maternal nutrition and placental insufficiency on the functional interactions between the

fetal hypothalamic–pituitary–adrenal axis and the placenta. There is extensive epidemiological evidence that a suboptimal intrauterine environment or increased exposure to glucocorticoids *in utero* is associated with poor adult health outcomes. It has been proposed that prolonged maternal or fetal stress may be associated with enhanced exposure of the fetus to glucocorticoids of maternal or fetal origin that results in changes in an array of fetal tissues and the placenta. Experimental animal studies have demonstrated that maternal undernutrition results in increased maternal and fetal cortisol concentrations, but it is not clear whether such increases are sustained in poorly nourished pregnant women and their fetuses. Interestingly, restriction of placental growth and function results in an increase in fetal cortisol concentrations, and it is apparent that the increase in cortisol is not a consequence of activation of the fetal hypothalamo–pituitary axis. Plasma ACTH concentrations are not elevated and CYP17 expression is not up-regulated in the adrenals of growth-restricted fetuses. It has been proposed that placental factors such as PG E<sub>2</sub> may stimulate adrenal glucocorticoid output. This would represent an important mechanism for cross-talk between a failing placenta and the fetal adrenal and could set in motion a series of positive feed-forward

endocrine loops within the feto–placental unit designed to mature key fetal organs in preparation for a preterm delivery.

### See Also the Following Articles

Adrenal Cortex, Development • Adrenal Cortex Development, Regulation of • Corticotropin-Releasing Hormone, Placenta • Intrauterine Growth Retardation • Pregnancy Endocrinology

### Further Reading

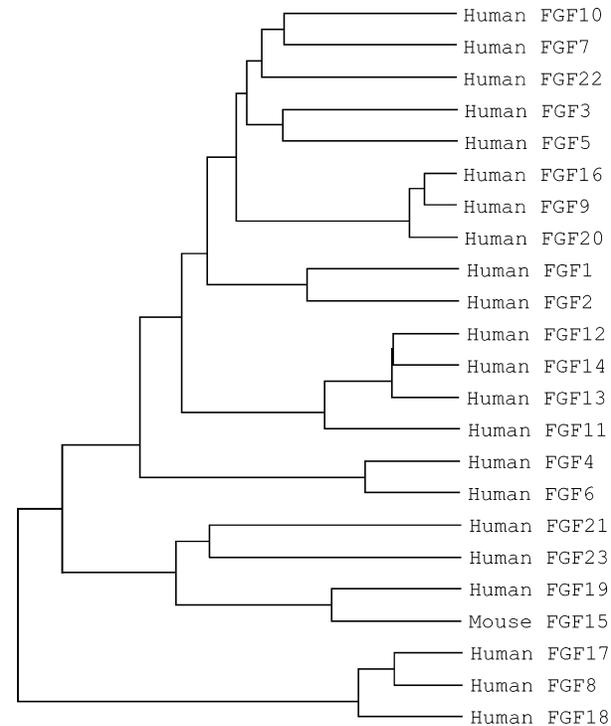
- Challis, J. R. G., Matthews, S. G., Gibb, W., and Lye, S. J. (2000). Endocrine and paracrine regulation of birth at term and preterm. *Endocrine Rev.* **21**, 514–550.
- Larsen, P. R., Kronenberg, H. M., Melmed, S., and Polonsky, K. S. (2003). “Williams Textbook of Endocrinology,” 10th ed. W. B. Saunders, Philadelphia.
- McMillen, I. C., Adams, M. B., Ross, J. T., Coulter, C. L., Simonetta, G., Owens, J. A., Robinson, J. S., and Edwards, L. J. (2001). Fetal growth restriction: Adaptations and consequences. *Reproduction* **122**, 195–204.
- Mesanio, S., and Jaffe, R. B. (1997). Developmental and functional biology of the primate fetal adrenal cortex. *Endocrine Rev.* **18**, 378–403.
- Pepe, G. J., and Albrecht, E. D. (1995). Actions of placental and fetal adrenal steroid hormones in primate pregnancy. *Endocrine Rev.* **16**, 608–648.

**Table I** Chromosomal Localizations of Human FGF Genes

Gene	Localization
FGF1	5q31
FGF2	4q26-q26
FGF3	11q13
FGF4	11q13.3
FGF5	4q21
FGF6	12p13
FGF7	15q15-q21.1
FGF8	10q24
FGF9	13q11-q12
FGF10	5p12-p13
FGF11	17p31.1
FGF12	3q28
FGF13	X26
FGF14	13q34
(FGF15)	
FGF16	—
FGF17	8p21
FGF18	5q34
FGF19	11q13.1
FGF20	8p21.3-p22
FGF21	19q13.1-qter
FGF22	19p13.3
FGF23	12p13.3

closely linked to the FGF3 and FGF4 genes in orthologous chromosomal regions. These findings indicate that human FGF19 is the human orthologue of mouse Fgf15.

Human FGFs range in molecular mass from 17 to 34 kDa and have a conserved approximately 120-amino acid residue core with approximately 30 to 60% amino acid identity. Most FGFs (3–8, 10, 15, 17–19, and 21–23) have amino-terminal signal peptides and are readily secreted from cells. FGF9, FGF16, and FGF20 lack obvious amino-terminal signal peptides but are secreted nonetheless. These FGFs have a noncleaved amino-terminal hydrophobic sequence that is required for secretion. FGF1 and FGF2 also lack signal peptides but, unlike FGF9, FGF16, and FGF20, are not secreted. FGF1 and FGF2 may be released from damaged cells or by an exocytotic mechanism that is independent of the endoplasmic reticulum–Golgi pathway. FGF11 through FGF14 lack signal peptides and remain intracellular. It is unknown whether FGF11 through FGF14 interact with known FGFRs or function in a receptor-independent manner within cells.



**Figure 1** Apparent evolutionary relationships of human FGF genes.

Most human FGF genes are scattered throughout the genome. The chromosomal locations of all except FGF16 are known (Table I). Several FGF genes are clustered within the genome. FGF3, FGF4, and FGF19 are located at 11q13 and are separated by only 40 and 10 kbp, respectively. FGF6 and FGF23 are located within 55 kbp at 12p13, whereas FGF17 and FGF20 map to 8p21–p22. These gene locations indicate that the FGF gene family was generated both by gene and by chromosomal duplication and translocation during evolution.

The apparently evolutionary relationships of the 22 known human FGFs are shown in Fig. 1. FGFs can be classified into several subgroups or subfamilies. Members of a subgroup share increased sequence similarity and biochemical and developmental properties. For example, members of the FGF8 subfamily (FGF8, FGF17, and FGF18) have 70 to 80% amino acid sequence identity, similar receptor-binding properties, and some overlapping sites of expression. Members of FGF subfamilies are not closely linked in the genome, indicating that the subfamilies were generated by gene translocation or genome duplication events but not by location duplication events.

Table II The Major Phenotypes of Fgf Knockout Mice

Gene	Survival of null mutant	Phenotype
Fgf1	Viable	None identified
Fgf2	Viable	Mild cardiovascular, skeletal, neuronal
Fgf3	Viable	Mild inner ear, skeletal (tail)
Fgf4	Lethal, E4–5	Inner cell mass proliferation
Fgf5	Viable	Long hair, angora mutation
Fgf6	Viable	Subtle, muscle regeneration
Fgf7	Viable	Hair follicle growth, ureteric bud growth
Fgf8	Lethal, E7	Gastrulation defect, central nervous system development, limb development
Fgf9	Lethal, P0	Lung branching, XY sex reversal
Fgf10	Lethal, P0	Multiorgan development, including limb and lung
Fgf12	Viable	Neuromuscular phenotype
Fgf14	Viable	Neurological phenotype
Fgf15	Lethal, E9.5	Not clear
Fgf17	Viable	Cerebellar development
Fgf18	Lethal, P0	Skeletal development

## THE FGFR FAMILY

Most FGFs mediate their biological responses by binding to and activating cell surface receptors. The human FGFR gene family consists of four members, FGFR1 through FGFR4. FGFRs, transmembrane receptors, contain three extracellular immunoglobulin-like domains and two intracellular tyrosine kinase domains and share approximately 55 to 70% amino acid sequence identity. Alternative splicing of FGFR1 through FGFR3 generates two isoforms: the “b” and “c” isoforms. The FGFR protein family consists of seven members: FGFR1b, FGFR1c, FGFR2b, FGFR2c, FGFR3b, FGFR3c, and FGFR4. This alternative splicing event is regulated in a tissue-specific manner and dramatically affects ligand–receptor binding specificity. The “b” and “c” isoforms tend to be expressed in epithelial and mesenchymal lineages, respectively.

## INTERACTION OF FGFS WITH HEPARIN OR HEPARAN SULFATE

FGFs interact with heparin or heparan sulfate proteoglycan. The interaction stabilizes FGFs, making them more resistant to thermal denaturation and proteolysis, and may severely limit their diffusion and release into interstitial spaces. The interaction also results in the formation of dimers and higher order oligomers. Although the biologically active form of FGF is poorly defined, it has been established that heparin is required for FGF to effectively activate FGFR. Additional studies have shown that heparin and/or

heparan sulfate act to increase the affinity and half-life of the FGF–FGFR complex.

## THE FGF SIGNALING

FGFRs transmit extracellular FGF signals to cytoplasmic signal transduction pathways through tyrosine phosphorylation. By binding ligand, FGFRs dimerize and phosphorylate specific tyrosine residues on their own and others' cytoplasmic tails.

The phosphorylation of FGFRs triggers the activation of cytoplasmic signal transduction pathways, including the MAPK signaling pathway, the PLC $\gamma$  signaling pathway, the Src signaling pathway, the Crk-mediated signaling pathway, and the SNT-1/FRS signaling pathway.

## Fgf KNOCKOUT MICE

Many Fgf genes have been disrupted by homologous recombination in mice. The phenotypes range from very early embryonic lethality to subtle phenotypes in adult mice. The major phenotypes observed in Fgf knockout mice are shown in [Table II](#).

## HUMAN DISEASES CAUSED BY FGF SIGNALING DISORDERS

### Chondrodysplasia Syndromes

Achondroplasia (ACH) is characterized by deduced growth of long bones with proximal elements affected

more severely than distal elements. Hypochondroplasia (HCH) is characterized by mild short stature and shares some clinical features with ACH. Thanatophoric dysplasia (TD) is characterized by a very severe skeletal dysplasia and is clinically similar to homozygous cases of ACH. Most cases of these syndromes originated from amino acid substitutions of FGFR3. These mutations constitutively activate FGFR3.

### Craniosynostosis Syndromes

Craniosynostosis is characterized by premature fusion of cranial sutures. Associated phenotypes in some craniosynostosis syndromes include malformations in the appendicular skeleton and nonskeletal phenotypes such as mental retardation. Mutations in FGFR1 through FGFR3 have been associated with craniosynostosis syndromes. However, most craniosynostosis syndromes are associated with mutations in FGFR2. Most of the mutations are missense or small in-frame insertions or deletions. These mutant FGFRs are dominant with gain-of-function mutations.

### Tumor-Induced Osteomalacia and Autosomal Dominant Hypophosphatemic Rickets

Tumor-induced osteomalacia (TIO) is a renal phosphate wasting disorder resulting in low serum phosphorus concentration and osteomalacia. Removal of the tumors responsible for TIO normalizes phosphate metabolism. These tumors secrete a heat-sensitive molecule of approximately 25 kDa designated "phosphatonin" that specifically inhibits sodium-dependent phosphate transport in cultured renal proximal epithelial cells. Phosphatonin was found to be FGF23. Autosomal dominant hypophosphatemic rickets (ADHR) is also a renal phosphate wasting disorder resulting in low serum phosphorus concentration, rickets, and osteomalacia. The ADHR gene was also found to be the FGF23 gene with missense mutations.

### Borjeson–Forssman–Lehmann Syndrome

The chromosomal location (Xq26) and tissue-specific expression pattern suggest that the FGF13 gene may be a causal gene for Borjeson–Forssman–Lehmann syndrome, an X-linked mental retardation syndrome.

### Cancers

The expression of FGFs in human cancers has been documented on numerous occasions, and in many instances overexpression of mRNA and/or protein has been found in conjunction with amplification of the respective FGF genetic locus.

The oncogenicity of FGF3 and FGF4 genes following transfection into NIH 3T3 cells has been documented. These genes are not expressed in adult tissues. Amplification of these genes has been demonstrated in a variety of human cancers, including Kaposi's sarcoma and stomach tumors.

In prostate, bladder, and renal cancers, FGFs regulate the induction of metalloproteinases that degrade extracellular matrix proteins, thereby facilitating tumor metastasis. Probably because of their potent angiogenic properties, aFGF and bFGF have received the most attention. However, there is increasing evidence that other FGFs also play crucial roles in the tumors. The recent identification in transitional cell carcinoma of a high frequency of mutations in FGFR3 predicted to activate the kinase activity of the receptor indicates a potential role as an oncogene in the urothelium.

It has become well accepted that solid tumors must create a vascular system for nutrient delivery and waste removal to grow appreciably. Angiogenesis is critical to the progression of gliomas. A role for vascular endothelial growth factor in glioma angiogenesis has been demonstrated convincingly. However, FGFs are also expected to influence glioma angiogenesis by directly stimulating endothelial cell proliferation and/or by mediating the expression of key proteases on endothelial cells necessary for angiogenesis.

### FGFS AND FGFRS AS THERAPEUTIC TARGETS IN HUMAN CANCERS

Because FGFs and/or FGFRs have been identified in tumor vasculature and supporting stromal cells as well as in tumor cells, FGF and FGFR targeting reagents may be useful for targeting tumor vasculature/stroma and tumor cells. A variety of FGF-targeting reagents have been expected to inhibit tumor growth, including antisense FGF oligonucleotides, soluble FGFRs, neutralizing antibodies, peptides corresponding to FGF functional domains, and small molecules such as pentosan polysulfate, suramin, and Suradista. FGFR-targeting reagents, including antisense FGFR oligonucleotides, dominant negative FGFRs, toxin-conjugated anti-FGFR antibodies, and small molecule FGFR protein kinase inhibitors, have also been expected to inhibit tumor growth.

**See Also the Following Articles**

Angiogenesis • Bone Remodeling, Dynamics of • Chondrodysplasias • Collagen Metabolism • EGF and Related Growth Factors • Hepatocyte Growth Factor • Insulin-like Growth Factors

**Further Reading**

Hogan, B. L. M. (1999). Morphogenesis. *Cell* **96**, 225–233.  
 Jeffers, M., LaRochele, W. J., and Lichenstein, H. S. (2002). Fibroblast growth factors in cancer: Therapeutic possibilities. *Expert Opin. Therapeutic Targets* **6**, 469–482.

Ornitz, D. M., and Itoh, N. (2001). Fibroblast growth factors. *Genome Biol.* **2**, 3005.1–3005.12.  
 Ornitz, D. M., and Marie, P. J. (2002). FGF signaling pathways in endochondral and intramembrous bone development and human genetic disease. *Gene Dev.* **16**, 1446–1465.  
 Powers, C. J., McLeskey, S. W., and Wellstein, A. (2000). Fibroblast growth factors, their receptors, and signaling. *Endocrine-Related Cancer* **7**, 165–197.  
 Szebenyi, G., and Fallon, J. F. (1999). Fibroblast growth factor as multifunctional signaling factors. *Intl. Rev. Cytol.* **185**, 45–106.  
 Wilkie, A. O. M., Patey, S. J., Kan, S., van den Ouweland, A. M. W., and Hamel, B. C. J. (2002). FGFs, their receptors, and human limb malformations: Clinical and molecular correlations. *Am. J. Med. Genet.* **112**, 266–278.



**Fibromyalgia**

see Chronic Fatigue Syndrome and Fibromyalgia



**Fine Needle Aspiration Biopsy**

see Thyroid Fine Needle Aspiration Biopsy



**Follicle-Stimulating Hormone**

see FSH

to 46 times greater risk for lower extremity amputation than persons without diabetes. Foot wounds are the proximate cause of 90% of these nontraumatic amputations. In addition to the physical and emotional problems they cause, amputations have been associated with an increased mortality in the subsequent 3 years. By 2025, there will be more than 300 million persons with diabetes worldwide, emphasizing the urgency of this problem.

### Financial Cost

As with the clinical morbidity, the financial cost of treating diabetic foot ulcers is substantial. One database of approximately 7 million patients followed over a 2-year period revealed that the total direct cost for treated diabetic foot ulcers was \$16 million or an average of \$4595 per ulcer episode. Another study identified attributable costs for a 40- to 65-year-old male with a new foot ulcer as \$27,987 for the 2 years after its occurrence. Studies of diabetes-related lower extremity amputation in Europe and the United States over the past decade have reported direct related costs ranging from \$20,000 to \$27,000. Indirect costs, such as loss of productivity and long-term care at skilled facilities or by family members, may double the total cost to society.

## PATHWAYS TO ULCERATION AND AMPUTATION

### Risk Factors for Foot Problems

Many factors, especially poor foot care and inappropriate footwear, increase the risk of foot ulceration. Most studies have shown that foot lesions generally arise from a combination of three factors. Peripheral neuropathy (the insensate foot) is the most important and can be identified by the patient's inability to feel 10 g of pressure applied with a Semmes-Weinstein monofilament. The other two factors are foot deformity, largely related to motor neuropathy, and repetitive moderate pressure or stress. In a case-control model assessing 225 diabetic patients, Lavery and co-workers noted that the presence of neuropathy alone increased the risk of ulceration 1.7-fold, neuropathy and deformity increased the risk 12.1-fold, and neuropathy, deformity, and a history of previous ulcer or amputation increased the risk of developing another ulcer approximately 36-fold. The major precipitating cause of foot amputation is an ulceration that fails to heal. An additional contributing factor is the presence of arterial vasculopathy, most frequently involving the

infrageniculate blood vessels. Accelerated atherosclerosis associated with diabetes can lead to critical foot ischemia that impairs wound healing and infection resolution.

### The Role of Infection

All open wounds are colonized by microorganisms, but when the lesion is complicated by clinical signs and symptoms of infection (which occurs in perhaps half the cases), the likelihood of a poor outcome greatly increases. Diabetic foot ulcers with concomitant infection and ischemia are up to 90 times more likely to result in a high-level amputation than wounds without those two additional risk factors. Infection is usually caused by aerobic gram-positive cocci, particularly staphylococci; chronic wounds, especially those previously treated with antibiotics, may become infected with aerobic gram-negative rods and anaerobes. Deep infections, which often involve the underlying bone, can be both limb- and life-threatening. Diagnosing bone infection can be difficult and may require imaging procedures (preferably magnetic resonance studies) or bone biopsy.

## TREATING FOOT WOUNDS

Once an ulcer develops, it is critical to get it healed as quickly as possible. This requires attention to local, systemic, and social or psychological issues. The wound must be carefully assessed for neurologic, vascular, and infectious complications. Some patients will need to be hospitalized, whereas others can be safely treated at home. Almost all wounds will require some debridement of necrotic material and callus; some may require further surgical intervention (e.g., incision and drainage, or bone resection). Infected wounds should be cultured (preferably by sending a sample of tissue, not a swab) and empiric antibiotic therapy should carefully selected. If the limb is ischemic, a vascular evaluation may be needed. Appropriate dressings must be selected, based on the type of wound, and the patient or caregiver must be instructed on how to properly change them. Limb edema and dry skin should be treated. Any systemic metabolic problems (e.g., poor glycemic control, malnutrition) should be addressed. The patient should never walk out of the office in the same shoe that caused an ulceration; proper footwear and methods or devices to offload the involved site are crucial to healing the wound. Foot lesions present the provider with a teachable moment to instruct the patient on

the causes of foot problems and how they may be prevented.

## HOW ULCERS AND AMPUTATIONS CAN BE PREVENTED

Intervention at any point in the critical pathway to ulceration may prevent the serious foot complications of diabetes. For instance, reducing foot pressure points with appropriate footwear or orthoses and modulating activity should help avoid ulceration. Treating infected wounds with appropriate (and not excessive) antimicrobial therapy should prevent deeper infections. Improving blood flow (e.g., with a distal bypass procedure) to a critically ischemic wound should lessen the likelihood of amputation. Other key preventative factors include the following: regular foot inspection by the patient and health care providers; for patients with risk factors for foot ulcerations, regular foot examination and care by primary care and podiatric providers; wearing appropriate shoe gear; and, in some instances, undergoing judicious surgical interventions to reduce deformity may be helpful. Lower extremity ulcerations and amputations in persons with diabetes take a terrible toll on the individual and on the world's health care systems, but the good news is that they are almost entirely preventable with appropriate care and education.

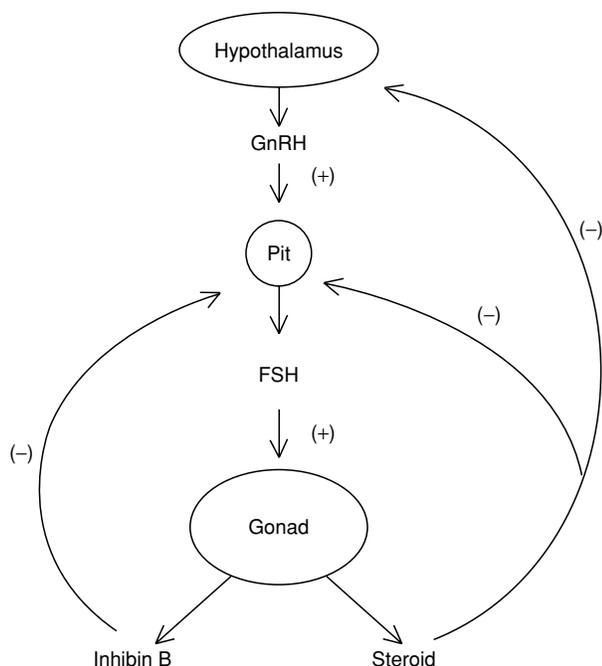
### See Also the Following Articles

Cardiovascular Disease in Diabetes • Diabetic Nerve Disease, Neuropathy • Eye Disease in Diabetes • Kidney

Disease in Diabetes • Neurological Disease and Diabetes, Autonomic

### Further Reading

- Akbari, C. M., Pomposelli, F. B., Gibbons, G. W., *et al.* (2000). Lower extremity revascularization in diabetes: Late observations. *Arch. Surg.* **135**, 452–456.
- Armstrong, D. G., Lavery, L. A., and Harkless, L. B. (1998). Validation of a diabetic wound classification system: The contribution of depth, infection, and ischemia to risk of amputation. *Diabetes Care* **21**, 855–859.
- Armstrong, D. G. (2001). Is diabetic foot care efficacious or cost-effective? *Ostomy Wound Manage.* **47**, 28–32.
- Holzer, S. E. S., Camerota, A., Martens, L., Crystal-Peters, J., and Zagari, M. (1998). Costs and duration of care for lower extremity ulcers in patients with diabetes. *Clin. Ther.* **20**, 169–181.
- Lavery, L. A., Armstrong, D. G., Vela, S. A., Quebedeaux, T. L., and Fleischli, J. G. (1998). Practical criteria for screening patients at high risk for diabetic foot ulceration. *Arch. Intern. Med.* **158**, 158–162.
- Lipsky, B. A. (2001). Infections of the foot in patients with diabetes. In "The Diabetic Foot" (M. A. Pfeifer and J. H. Bowker, eds.), 6th ed., pp. 467–480. Mosby, St. Louis, MO.
- Mayfield, J. A., Reiber, G. E., Maynard, C., Czerniecki, J. M., Caps, M. T., and Sangeorzan, B. J. (2000). Trends in lower limb amputation in the Veterans Health Administration, 1989–1998. *J. Rehab. Res. Dev.* **37**, 23–30.
- Ramsey, S. D., Newton, K., Blough, D., *et al.* (1999). Incidence, outcomes, and cost of foot ulcers in patients with diabetes. *Diabetes Care* **22**, 382–387.
- Reiber, G. E., Vileikyte, L., Boyko, E. J., *et al.* (1999). Causal pathways for incident lower-extremity ulcers in patients with diabetes from two settings. *Diabetes Care* **22**, 157–162.
- Williams, D. R. R., and Airey, C. M. (2000). The size of the problem: Epidemiological and economic aspects of foot problems in diabetes. In "The Foot in Diabetes" (A. J. M. Boulton, H. Connor, and P. R. Cavanagh, eds.), 3rd ed., pp. 15–24. Wiley, New York.



**Figure 1** This schematic is a general introduction to the concept of the regulation of FSH secretion and action. The hypothalamus secretes GnRH, which acts on the pituitary gland (Pit) to stimulate the secretion of FSH. FSH stimulates (+) the gonad to secrete steroids that act on the hypothalamus, pituitary gland, or both, to inhibit (-) the secretion of FSH. FSH also stimulates (+) the gonad to synthesize a protein, inhibin B, that inhibits (-) the secretion of FSH by the pituitary gland.

by the hypothalamic polypeptide, GnRH. Only one hypothalamic secretagogue stimulating the synthesis and secretion of both gonadotropins has been identified, although two releasing hormones, one for each gonadotropin, were initially postulated and the designation of GnRH rather than luteinizing hormone-releasing hormone is used throughout this article. The secretion of this hypothalamic secretagogue is episodic and regular. GnRH is released by neurons that terminate on the blood vessels running along the pituitary stalk. These vessels constitute a specialized form of the circulatory system, called a portal system. Portal systems are characterized as special veins that arise in one organ system and terminate in another before returning to the heart. They carry the pulses of GnRH to the gonadotropes in the pituitary gland and stimulate the synthesis and release of FSH. Accordingly, FSH is also released from the pituitary gland in pulses. Changing the pattern of the GnRH pulses results in coincidental changes in the pattern of pulsatile secretion of FSH. GnRH binds to a G protein-coupled membrane receptor on the surface of the gonadotrope, activating the  $G_{q/11}$  proteins that, in

turn, stimulate phospholipase C. This enzyme generates inositol-1,4,5-triphosphate and diacylglycerol. The increased levels of these intracellular messengers activate protein kinase C (PKC) and increase intracellular calcium also. PKC and calcium regulate the expression of the  $\beta$ -subunit gene, whereas secretion of FSH is regulated by the increase in intracellular calcium.

## RECEPTOR AND RECEPTOR DEFECTS

The receptor for FSH is found only in the gonads and has been described in several mammalian species. The FSH receptor (75–80 kDa) is a member of the family of G protein-coupled receptors. Like all members of this family of receptors, it is composed of three domains, a transmembrane domain, an extracellular domain, and an intracellular domain. The transmembrane domain is composed of seven hydrophobic stretches of 20–25 amino acids that form  $\alpha$ -helices traversing the cellular membrane alternating between intracellular and extracellular loops. The extracellular domain of the receptor binds FSH, which, in turn, causes the intracellular domain to activate the  $G_s$  protein. This is the first step in an intracellular cascade that continues with activation of the enzyme adenylyl cyclase and production of cyclic AMP (cAMP). In turn, cAMP-dependent protein kinase (PKA) is activated and phosphorylation of the transcription factor CREB (cAMP regulatory element-binding protein) leads to the expression of genes coding for several factors regulating the cellular functions of proliferation, differentiation, and survival. This pathway has been expanded to include various isoforms of adenylyl cyclase, phosphodiesterase, and anchoring proteins of PKA. Other transcription factors, e.g., stimulatory protein 1, upstream stimulatory factor, estrogen receptor- $\alpha$ , and estrogen receptor- $\beta$ , have been shown to be activated by high intracellular concentrations of cAMP. In addition, CREB can be phosphorylated by kinases other than PKA.

The gene of the FSH receptor has 10 exons, the last encoding both the transmembrane and intracellular domains. Only a few mutations have been identified. A Finnish family has a mutation in the FSH receptor gene that prevents its transport to the cellular membrane. The women of this family have amenorrhea and atrophic ovaries with follicles that do not develop beyond primary follicles. The men homozygous for this mutation are normally virilized and have sufficient testosterone to initiate and maintain spermatogenesis. A second mutation leads to a partially

defective receptor. These women also present with primary amenorrhea and the ovaries are of normal size, but folliculogenesis proceeds only to antral follicles. The larger follicles degenerate. In addition to these mutations in the receptor, mutations in the gene for the FSH  $\beta$ -subunit in two men have been reported. These men had small testes and were azoospermic. The difference in the phenotype of men with inactivating mutations of FSH or its receptor remains an unresolved paradox.

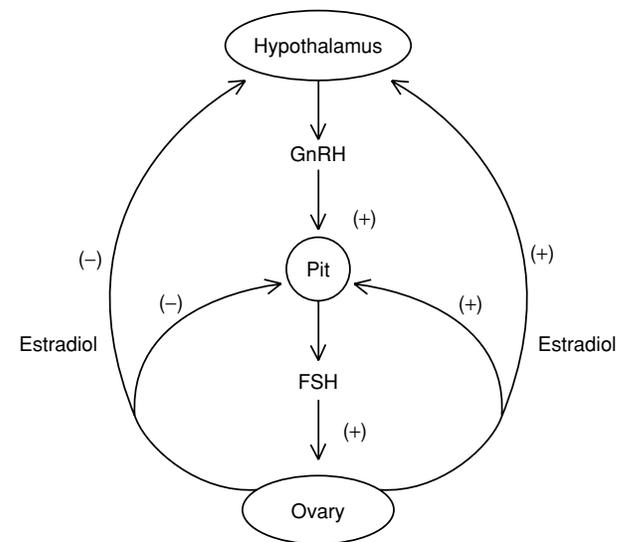
## ACTION IN THE OVARY

FSH, as the full name suggests, acts on the ovary by stimulating folliculogenesis. Before the role of FSH in the ovary is discussed, a brief description of the ovarian cycle is necessary. During fetal development of the female mammal, the primitive germ cells in the ovary enter meiosis and arrest in the prophase of meiosis I. Each oocyte is surrounded by a single layer of cells, and soon after formation, these primordial follicles enter a resting phase. During fetal life in primates and ruminants and within the first 2 weeks after birth in rodents, some of the primordial follicles are activated and become primary follicles. The mechanisms that stimulate the primordial follicles to initiate folliculogenesis are unknown, but are clearly independent of gonadotropin action. Two layers of somatic cells develop around the oocyte. The outermost layer is composed of theca cells that are separated from the innermost layer, comprising granulosa cells, by a basement membrane. Both types of cells proliferate, increasing the size of the follicle until a fluid-filled cavity, or antrum, forms in the layers of the granulosa cells. The mature, or antral, follicle ruptures, marking the end of folliculogenesis.

The rupture of the follicle releasing the oocyte and some of the granulosa cells surrounding it is called ovulation. During the periovulatory period, anti-inflammatory factors are produced and appear to aid in rapid healing and vascularization of the ruptured follicle. The theca and granulosa cells remaining in the ruptured follicle differentiate into luteal cells, forming the corpus luteum that characterizes the luteal phase of the ovarian cycle. If no pregnancy occurs, after a finite time specific to each species, the corpus luteum dies, marking the end of the luteal phase. The rapid involution of the corpus luteum is called luteolysis and is regulated by specific factors. Luteolysis is necessary before a subsequent ovarian cycle can begin. The follicular phase, ovulation, luteal phase, and luteolysis constitute the ovarian cycle and

FSH plays an important role only during the first phase of the ovarian cycle, namely, folliculogenesis.

The theca cells of the developing ovary produce androgens in response to LH. FSH stimulates the granulosa cells to produce the enzyme aromatase, which converts the androgens produced by the theca cells into estradiol (Fig. 2). As the number of granulosa cells increases by mitosis, a greater amount of estradiol is produced by the developing follicle. FSH also increases the number of LH receptors on theca cells. Estradiol inhibits the secretion of FSH and LH. With the increase in LH receptors on the theca cells of one or more follicles in the presence of decreased LH, only those follicles with sufficient gonadotropin stimulation will continue folliculogenesis. As the follicle enlarges, a cavity forms around the oocyte so that it is surrounded by a few layers of granulosa cells and fluid, but a peduncle of granulosa cells remains



**Figure 2** The feedback mechanism regulating FSH in the female mammal. The hypothalamus secretes GnRH, which acts at the pituitary gland (Pit) to stimulate (+) the secretion of FSH. FSH stimulates the granulosa cells of the ovarian follicle to proliferate and synthesize the steroidogenic enzyme aromatase. This enzyme converts androstenedione, produced by the theca cells of the follicle under LH stimulation, into estradiol. During early folliculogenesis, estradiol acts at the hypothalamus, pituitary gland, or both, to inhibit (-) the secretion of FSH (left-hand side of figure). At the end of folliculogenesis, the high concentrations of estradiol in the blood produced by the large graafian follicle cause an abrupt release of LH (not shown) and FSH (right-hand side of figure). The abrupt release of FSH may result from a direct action of the steroid at the pituitary or indirectly at the hypothalamus to increase the secretion of GnRH that, in turn, causes a marked increase in pituitary secretion of the gonadotropins. The former action of the estradiol is an example of positive feedback.

attached to the rest of the follicle. Once the level of circulating estradiol produced by the graafian follicles is elevated, the steroids cause an abrupt release of LH and FSH, which results in ovulation, that is, the extrusion of the oocyte and the granulosa cells immediately surrounding it, which is subsequently referred to as the oophorus cumulus. Once ovulation has occurred, the role of FSH is greatly diminished and the formation of the corpus luteum and maintenance of the endometrium of the uterus are under the influence of LH.

## ACTION IN THE TESTIS

The role of FSH in regulating spermatogenesis in adult mammals has been controversial for several decades. FSH appears to have little direct role in regulating spermatogenesis in the testes of adult male rodents. It, however, is the primary determinant of the spermatogenic capacity of the rodent testis. This capacity is defined as the number of sperm potentially produced by the male gonad and intuitively this must directly relate to the number of spermatogenic stem cells and the capacity of the structure of the gonad to nourish and support the progeny of these stem cells. Before further discussion, a brief description of Sertoli cell proliferation and establishment of the seminiferous epithelium is presented.

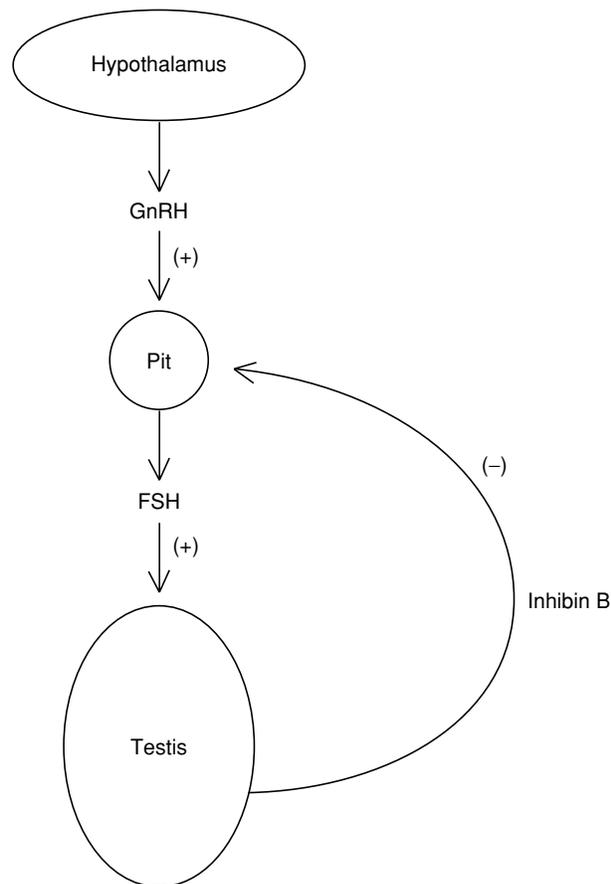
The Sertoli cells of the mammalian testis are the only somatic cells in the seminiferous epithelium and are known to proliferate only during fetal and peripubertal development in rodents. A third period of Sertoli cell proliferation occurs during infancy in higher primates. After these somatic cells have attained a population size associated with adulthood, proliferation ceases and the cells terminally differentiate. The mature Sertoli cells serve to organize and nourish the germ cells that make up the remainder of the cell types in this epithelium. Moreover, they are the mediators of the action of FSH because they are the only cells in the male that have receptors for FSH. FSH is the hormone that stimulates Sertoli cell proliferation in the rodent testis, but, in primates, both FSH and LH are equally effective in stimulating the proliferation of Sertoli cells. Because the Sertoli cells provide the nutrients and structural support to spermatogenesis, it was immediately recognized that FSH regulates the spermatogenic capacity of the testis. In contrast to this situation in rodents, FSH plays a more direct role in regulating spermatogenesis in higher primates. Before this role is discussed, a brief review of spermatogenesis is presented.

Spermatogenesis is a series of processes comprising three phases: stem cell renewal, germ cell proliferation, and spermiogenesis. Stem cell renewal is the mechanism that guarantees that a large and undiminishing number of undifferentiated germ cells are continually available for the subsequent waves of spermatogenesis. The spermatogenic stem cells are called type A spermatogonia and, depending on the species, several type A spermatogonia have been identified. In addition to renewing their population, the type A spermatogonia also produce more differentiated spermatogonia, often referred to as type B spermatogonia. These latter cells, depending again on species, may produce two or more generations of differentiated spermatogonia. The differentiated spermatogonia in the final generation divide and produce the primary spermatocytes. These germ cells enter prophase of meiosis I, lasting from 1 to 2 weeks, and the primary spermatocytes synthesize fourfold more DNA than is found in a somatic cell. During the remainder of the long prophase, the homologous chromosomes form bivalents and crossing over of the genetic material occurs. The primary spermatocytes rapidly complete meiosis I, producing two haploid secondary spermatocytes each, but with twice the haploid content of DNA. The secondary spermatocytes enter meiosis II and form the haploid spermatids containing the haploid amount of DNA, often referred to as C. The spermatids enter spermiogenesis, during which the nuclear contents condense, the acrosome is formed, a tail is produced, and the amount of cytoplasm is reduced. This process requires approximately 2 to 3 weeks to complete.

The central role of the pituitary to stimulate spermatogenesis in adult rodents and primates was established nearly 60 years ago by Phillip E. Smith. Surgical removal of the pituitary gland, hypophysectomy, of adult rats led to a decrease in testicular weight and the seminiferous epithelium comprised Sertoli cells and germ cells as mature as spermatids. Hypophysectomy of adult rhesus monkeys resulted in a precipitous decline in testicular size associated with the complete regression of the seminiferous epithelium to the extent that the tissue comprised only Sertoli cells and type A spermatogonia, that is, only stem cells. Additional experiments on adult rodents and monkeys have confirmed Dr. Smith's earlier observation. Moreover, his observation was extended by demonstrating that replacement of testosterone in hypogonadotropic hypogonadal rhesus monkeys resulted in a stimulation of testicular growth to approximately 60% of the pretreatment size. The gonadal growth was due primarily to the stimulation of

spermatogenesis by androgen, but morphometric analysis of the seminiferous epithelium of the monkeys revealed that the smaller testicular size was accounted for by a deficit in the numbers of all type B spermatogonia. Replacement of FSH in testosterone-treated hypophysectomized adults resulted in a greater number of all four generations of type B spermatogonia. These results led to the conclusion that testosterone alone stimulates spermatogenesis, but FSH is necessary to restore spermatogenesis completely. This action of FSH is posited to be the rescue of type B spermatogonia from programmed cell death.

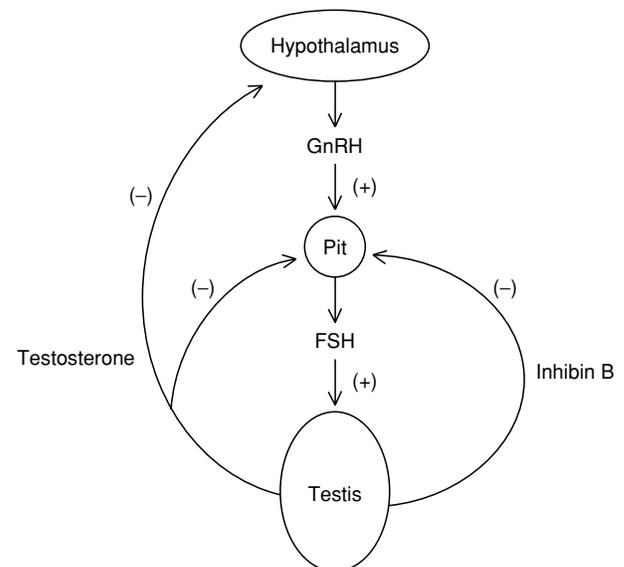
Unilateral orchidectomy in adult macaques results in a compensatory growth of the remaining testis. The number of Sertoli cells per testis was identical in the gonad removed at the time of unilateral orchidectomy



**Figure 3** The regulation of FSH secretion in the male primate. GnRH, secreted by the hypothalamus, stimulates (+) the pituitary gland (Pit) to synthesize and release FSH. FSH is transported by the circulatory system to the testis, where the hormone binds to receptors on the Sertoli cells and stimulates the transcription and translation of several gene products, including the protein hormone, inhibin B. In male primates, inhibin B acts at the pituitary gland to inhibit (-) FSH secretion.

and the gonad that remained in the animal for 45 days. The number of all germ cells more mature than type Ap spermatogonia was greater in the remaining testis than in the removed testis. Moreover, the removal of one testis was occasioned by a transient decline in testosterone that in turn led to a transient increase in LH. By 4 days after surgery, the testosterone and LH concentrations were restored to those measured prior to surgery. The removal of one gonad in these primates was marked also by a decline in inhibin B, which is secreted by Sertoli cells into the circulatory system. The concentrations of FSH in the circulation increased, confirming the role of inhibin B in the negative feedback on FSH secretion by the pituitary gland of the primate (Fig. 3). The role of inhibin B is less clear in rodents and the steroid testosterone may play a greater role in the regulation of FSH secretion in those species (Fig. 4).

In contrast to the results reported above, other investigators have presented results indicating that FSH increases the number of germ cells in the primate testis indirectly, by acting on type Ap spermatogonia, the renewing stem cells. The administration of FSH to adult macaques resulted in an increase in the type Ap spermatogonia in the seminiferous epithelium, but an analysis using the kinetics of spermatogenesis does not



**Figure 4** The negative feedback (-) of the androgen testosterone and the protein inhibin B on FSH secretion in male mammals. As explained in the legend to Fig. 3 and in the text, inhibin B from the Sertoli cells of the testis inhibits the secretion of FSH by the pituitary (Pit) in higher primates (right-hand side of figure). In other male mammals, this action of inhibin B may be less important and the steroid testosterone may play the major role in regulating FSH secretion (left-hand side of figure).

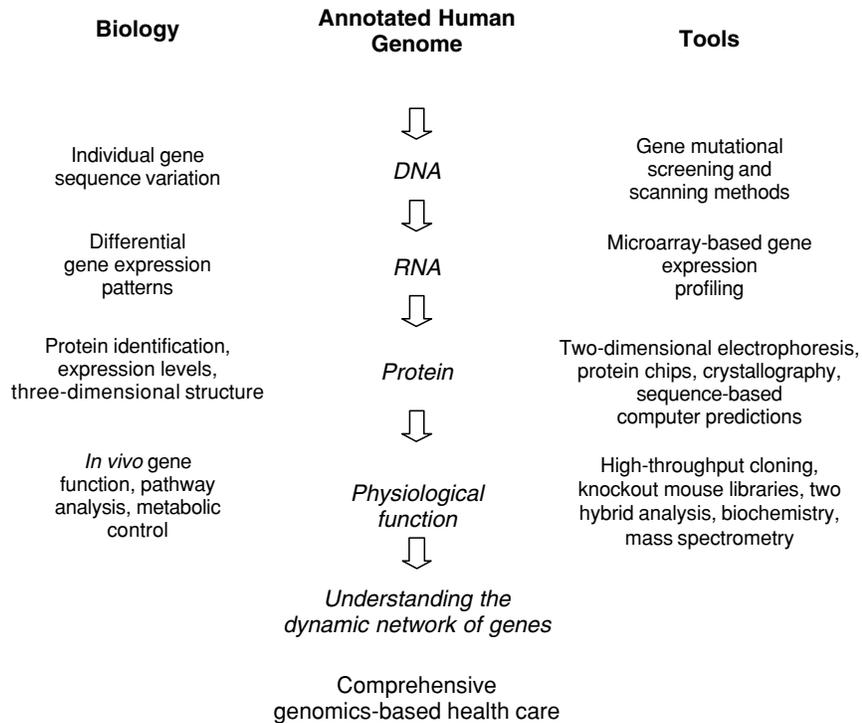
support the notion that the increase in germ cell production is due to an action of FSH to increase the number of Ap spermatogonia.

### See Also the Following Articles

Fertilization • Gonadotropin-Induced Ovulation • Gonadotropin-Releasing Hormone, Family of • Implantation • LH (Luteinizing Hormone) • Ovarian-Follicular Apparatus • Pituitary Gland Anatomy and Embryology • Pregnancy Endocrinology • Puberty: Physical Activity and Growth • Sexual Maturation, Female • Sexual Maturation, Male • Spermatogenesis, Endocrine Control of

### Further Reading

- Achermann, J. C., Weiss, J., Lee, E.-J., and Jameson, J. L. (2001). Inherited disorders of the gonadotropin hormones. *Mol. Cell. Endocrinol.* **179**, 89–96.
- Ando, H., Hew, C. L., and Urano, A. (2001). Signal transduction pathways and transcription factors involved in the gonadotropin-releasing hormone-stimulated gonadotropin subunit gene expression. *Comp. Biochem. Physiol. Part B Biochem. Mol. Biol.* **129**, 525–532.
- De Roux, N., and Milgrom, E. (2001). Inherited disorders of GnRH and gonadotropin receptors. *Mol. Cell. Endocrinol.* **179**, 83–87.
- Fimia, G. M., Morlon, A., Macho, B., de Cesare, D., and Sassone-Corsi, P. (2001). Transcriptional cascades during spermatogenesis: Pivotal role of CREM and ACT. *Mol. Cell. Endocrinol.* **179**, 17–23.
- Fortune, J. E., Cushman, R. A., Wahl, C. M., and Kito, S. (2000). The primordial to primary follicle transition. *Mol. Cell. Endocrinol.* **163**, 53–60.
- Habert, R., Lejune, H., and Saez, J. M. (2001). Origin, differentiation and regulation of fetal and adult Leydig cells. *Mol. Cell. Endocrinol.* **179**, 47–74.
- Hillier, S. G. (2001). Gonadotropic control of ovarian follicular growth and development. *Mol. Cell. Endocrinol.* **179**, 39–46.
- Mendis-Handagama, S. M. I. C., and Ariyaratne, H. B. S. (2001). Differentiation of the adult Leydig cell population in the post-natal testis. *Biol. Reprod.* **65**, 660–671.
- Plant, T. M., and Marshall, G. R. (2001). The functional significance of FSH in spermatogenesis and the control of its secretion in male primates. *Endocr. Rev.* **22**, 764–786.
- Richards, J. S. (2001). New signaling pathways for hormones and cyclic adenosine 3',5'-monophosphate action in endocrine cells. *Mol. Endocrinol.* **15**, 209–218.
- Simoni, M., Gromoll, J., and Nieschlag, E. (1997). The follicle-stimulating hormone receptor: Biochemistry, molecular biology, physiology, and pathophysiology. *Endocr. Rev.* **18**, 739–773.
- Stocco, D. M. (2001). StAR protein and the regulation of steroid hormone synthesis. *Annu. Rev. Physiol.* **63**, 193–213.
- Tilbrook, A. J., and Clarke, I. J. (2001). Negative feedback regulation of the secretion and actions of gonadotropin-releasing hormone in males. *Biol. Reprod.* **64**, 735–742.



**Figure 1** Schematic representation of the various levels of functional genomics as discussed in the text.

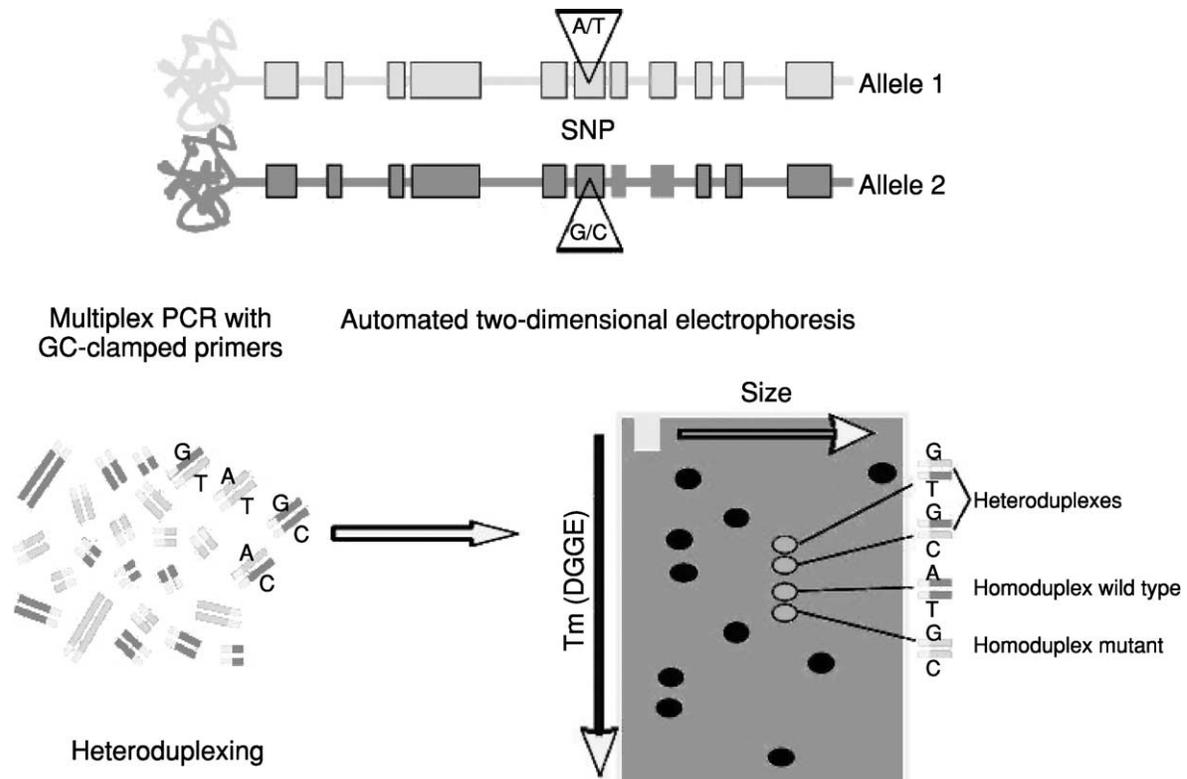
## GENETIC VARIATION

First-level functional genomics activities on an annotated genome mainly involve individual genetic variation and its association with particular phenotypes. The most common type of such variation is called a single-nucleotide polymorphism (SNP) and accounts for heritable interindividual differences such as susceptibility to disease and response to medication. Because most of the common human diseases are now age related, the identification of SNPs in genes and their association with aging-related phenotypes is important. Assessing the genetic component of aging will contribute to our understanding of functional diversity in aging human populations. In addition to environmental factors, many distinct aging-related phenotypes involving functional decline or age-related disease risk are likely to have a genetic component. Similarly, healthy aging could also have a genetic component. Examples of the latter are unusual longevity, extended preservation of function such as cognitive or vascular function, and resistance to age-related disease. Genetic control of normal aging in humans is likely to be determined by SNPs in many genes involved in multiple functional pathways. The challenge is to identify these gene variants and to apply the tools of modern bioinformatics to associate combinations of

alleles with particular phenotypes. This is illustrated by some recent work in our laboratories.

## SNP HAPLOTYPE ANALYSIS OF AN AGING-RELATED DISEASE

Methods that are capable of analyzing entire gene-coding regions for unknown mutations and polymorphisms are still rare. Indeed, semiautomated nucleotide sequencing, according to the principle originally described by Sanger and colleagues, is still the gold standard. Despite some obvious advantages of DNA sequencing, including its high level of automation and ability to provide complete information about the location and nature of sequence variants, its high cost makes it less suitable as a primary DNA variation detection method in large genetic epidemiological studies. Two-dimensional gene scanning (TDGS), developed in our laboratory, is a prescreening method based on denaturing gradient gel electrophoresis (DGGE), allowing for the separation of DNA fragments differing by as little as 1 bp on the basis of differences in their melting temperature in a two-dimensional format (Fig. 2). TDGS has been applied to detect mutational variants in a number of large human genes as well as the mitochondrial genome. With the recent introduction



**Figure 2** Schematic representation of the various steps involved in two-dimensional gene scanning. The target regions, such as gene coding regions, are amplified by polymerase chain reaction (PCR) in two steps. First, a long-distance PCR generates overrepresentation of the target regions as part of large PCR amplicons (i.e., about 10 kb). Then, using the large amplicons as template, target regions of 250 bp on average are amplified in an extensive multiplex PCR, with the primers chosen in a way as to create optimal conditions for detecting all possible sequence variants using the denaturing gradient principle. For this purpose, it is also essential that all primers be attached to a GC-rich fragment approximately 40 bases long. At the end of the PCR process, all amplicons are heteroduplexed by a round of complete denaturation and renaturation and then are separated based on size and base pair sequence by a denaturing gradient gel electrophoresis (DGGE). The presence of sequence variation is revealed by four different double-stranded fragments: two homoduplex molecules and two heteroduplex molecules.

of high-speed, two-dimensional electrophoresis with multicolor fluorescent detection, TDGS has become at least an order of magnitude more cost-effective than nucleotide sequencing, at equal accuracy, for large-scale SNP discovery and genetic testing.

Using TDGS, we have compared SNP haplotypes of Korean diabetes patients with normal controls in a case-control study in collaboration with Kyong Soo Park of Seoul National University in Korea. Mitochondrial DNA (mtDNA) variation has emerged as a candidate susceptibility factor in aging-related degeneration and disease, including premature aging, cancer, diabetes mellitus, Parkinson's disease, Alzheimer's disease, epilepsy, sensory losses (hearing and vision), and a variety of syndromes involving the muscles and the central nervous system. MtDNA is a maternally inherited, 16,568-bp haploid genome that encodes 13 subunits of the respiratory chain complexes. Its mutation rate

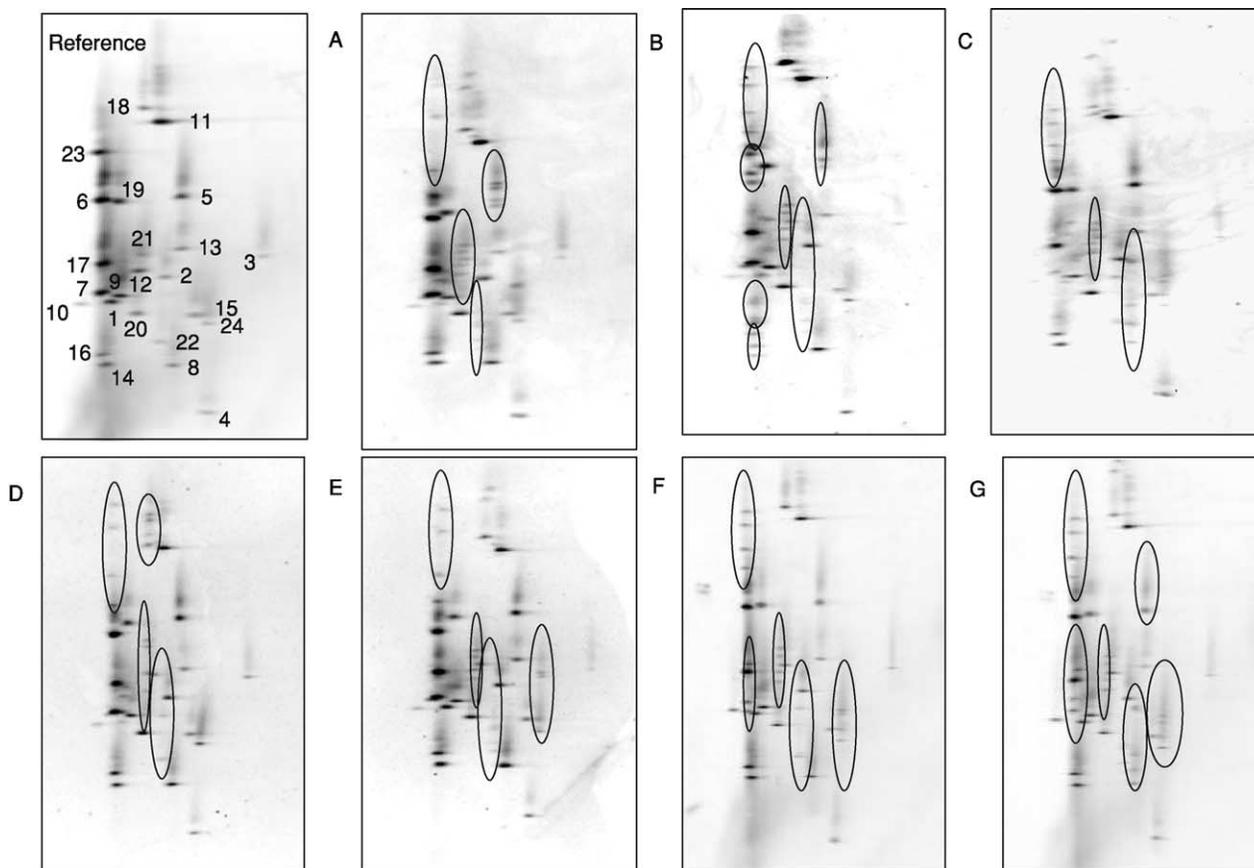
is 10 times faster than that of the nuclear genome, and this high mutation rate is attributable partly to the lack of an effective DNA repair mechanism and partly to the constant exposure of mtDNA to oxygen-free radicals from oxidative phosphorylation (OXPHOS). Polymorphisms in mtDNA may cause subtle differences in the encoded proteins and, therefore, subtle changes in OXPHOS activity and free radical production. Therefore, variant mtDNA sequences may predispose to an earlier onset of degenerative cellular processes or could be beneficial. Such SNP variation may exert its effect, for example, by influencing the energy production or by interacting with the nuclear genome.

We used a TDGS test for mtDNA that covers all of the regions implicated in human diseases. The test involves a two-step multiplex polymerase chain reaction (PCR) amplification: a long-distance PCR to

amplify the entire mitochondrial genome, except for the D-loop region, from total genomic DNA, which then serves as a template for the amplification of 25 short PCR fragments in two multiplex groups. Because mtDNA is a maternally inherited haploid genome, TDGS cannot be used to detect heterozygotes (heteroduplexes) unless there is heteroplasmic variation. To detect all possible SNPs in mtDNA using TDGS, we chose to use a reference DNA to create heteroduplexes with the individual DNA samples to be screened to detect variants that are different from the reference DNA. Our reference DNA sample was chosen on the basis of its lack of heteroplasmy (Fig. 3) and was fully sequenced. After PCR, the amplicons of the reference DNA were combined with the amplicons of an individual DNA sample and then were heteroduplexed by a round of complete denaturation and renaturation. TDGS analysis identified heteroduplexed mtDNA fragments

(amplicons) resulting from sequence differences between the reference DNA and the individual DNA (ellipses in Figs. 3A–G).

After analyzing 195 Korean individuals (65 normal controls and 130 patients with type 2 diabetes), a total number of 215 unique variations were identified, approximately 50% of which were previously unknown novel variants, indicating the highly polymorphic nature of mtDNA. In addition, we found 10310A→G, a synonymous SNP in the ND3 gene-coding region, significantly less frequently in diabetes patients (44.6%) than in normal controls (96.9%,  $P = 1.1 \times 10^{-14}$ ). Based on 9 common SNPs with a frequency of more than 10% in the two groups (Table I), mtDNA haplotypes were determined. A total of 15 common haplotypes were detected in the two groups, and a total of 18 rare haplotypes were group specific (“others” in Fig. 4) (data not shown). The frequencies of 3 haplotypes (haplotypes 6, 8, and 11) differed



**Figure 3** Detection of mitochondrial DNA variation by two-dimensional gene scanning (TDGS). To detect all possible SNPs in mtDNA by TDGS, a reference DNA, chosen on the basis of its lack of heteroplasmy and fully sequenced, was used to create heteroduplexes with the individual DNA samples to be screened to detect variants that are different from the reference DNA (A–G). The ellipses contain heteroduplexed mtDNA fragments (PCR amplicons) resulting from sequence differences between the reference DNA and individual DNAs.

Table I Haplotyping of Mitochondrial DNA Variation

Haplotype	12S rRNA	ND2		NC	ATPase 8	ND3			Cytb
	G709A	Syn C4883T	Leu-Met C5178A	Δ8271-81	Leu-Phe C8414T	Syn A10310G	Thr-Ala A10398G	Syn C10400T	Syn G15301A
1		•	•			•	•	•	•
2		•	•		•	•	•	•	•
3	•			•		•	•		
4				•		•			
5						•	•	•	•
6									
7			•			•	•	•	•
8						•			
9	•					•	•		
10	•					•	•	•	•
11		•	•		•				•
12									•
13				•					
14	•								•
15				•		•	•		

The mtDNA haplotypes were determined based on nine common SNPs with more than 10% minor allele frequency in the two groups. The nine common SNPs used to generate the 15 haplotypes are indicated as coding region, amino acid change, and position and nature of the nucleotide change. Cyt b, cytochrome b; ND2, NADH dehydrogenase subunit 2; NC, noncoding intergenic region; ND3, NADH dehydrogenase subunit 3.

significantly between the two groups. Haplotype 8 ( $P=0.00000025$ ) is significantly more frequent in normal controls than in diabetes patients, whereas haplotypes 6 ( $P=0.0000042$ ) and 11 ( $P=0.0172$ ) are significantly less frequent in normal controls than in diabetes patients (Fig. 4).

This example illustrates how a functional genomics approach can be built on the fruits of the human genome project by associating gene variants with medically relevant phenotypes of aging, in this case type 2 diabetes. A logical next step is to obtain information about the mechanistic basis of the functional

significance of such gene variants. When, as is the case in our example, the SNP variants occur in a gene-coding region and are nonsynonymous, it would be relevant to study the possible effect of the deduced protein alteration. This can be done by predictive modeling and biocomputational means as well as by using binding assays such as protein chips to study possible alterations in protein folding and protein-protein or protein-DNA interactions. Finally, SNP variants in gene regulatory regions such as promoter regions could affect the level of expression. Variation in gene expression levels involves the second and third

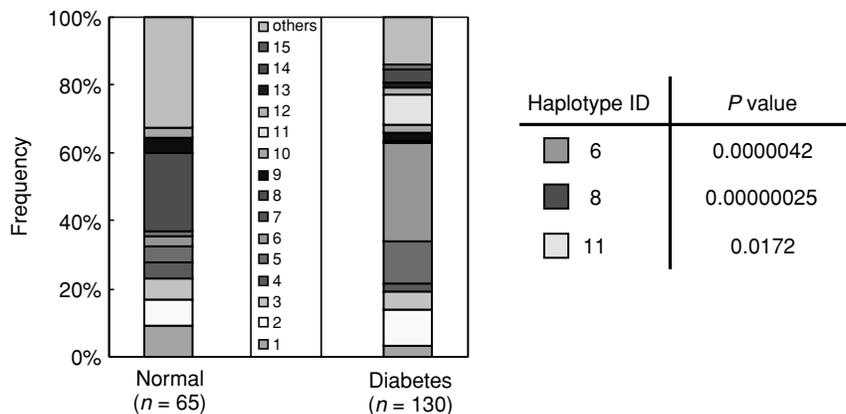


Figure 4 The frequency of common mitochondrial DNA haplotypes in normal controls and diabetes patients.

Table II Applications of Functional Genomics in Health Care

Application	Purpose	Benefit
Gene-based diagnostics	Disease susceptibility/drug responses	Screening, specific preventive measures
Increase predictability of preclinical and clinical testing	Accelerate drug development	More and better drugs
Disease subtype-specific drugs	Refine therapy	High-quality personalized medicine

levels of functional genomics analyses and has its own technological requirements.

## GENE EXPRESSION

One level down from the DNA of the genome is the analysis of the entire message of each cell type. Whereas SNP maps can help to identify sequence variation in groups of genes, including their regulatory sequences, gene expression patterns are a quick way in which to bring the different gene functions together in patterns characteristic for a specific phenotype, including disease. The most popular vehicle for gene expression profiling is the microarray chip, in which hundreds to thousands of cDNAs or oligonucleotides specific for parts of individual mRNAs are attached to a glass slide. Hybridization of such slides with labeled RNA probes obtained from the tissues of interest permits the analysis of changes in expression of a large number of genes simultaneously.

As time goes on, it is recognized more and more frequently that analyzing genes and their mRNAs will not be sufficient to understand cell function and dysfunction. Many alterations in proteins are not reflected in corresponding changes in mRNAs, for example, due to posttranslational modifications. This has resulted in a return to protein analysis, most notably to the two-dimensional protein separation system first reported by Patrick O'Farrell in 1975. This method remains the only one suitable for protein discovery. Equipped with more powerful ancillary tools (e.g., mass spectrometry), it now forms the backbone of many proteomics platforms. However, a wealth of additional proteomics tools are currently under development. Ultimately, it will be necessary to recognize the need to link gene networks, as well as variations therein, to physiological end points. This brings us back to old-fashioned biochemistry but also to advanced gene knockout and knock-in techniques of mouse modeling.

Functional genomics approaches at the analysis levels described previously and applied in an integrated manner should greatly facilitate the unraveling of the intricate relationship between the degenerative processes of aging and the increased incidence of

disease among the elderly. This should ultimately lead to a comprehensive, genomics-based health care in which the complexity of the aging process and its many associated disease sequelae, as well as the bewildering individual variability of geriatric patients, is taken into account (Table II). Such personalized geriatric medicine can be successful only when integrated with new computational and informatics approaches. Indeed, the new discipline of "bioinformatics" will play a critical role in setting the pace for this endeavor. As indicated in Table III, bioinformatics is the true sine qua non for success in functional genomics. Bioinformatics uses information technology to organize, visualize, interpret, and distribute biological information to answer complex biological questions. It allows workers in functional genomics to cope with an enormous flood of data and to answer a variety of biological questions in a fraction of the time it would take using traditional analysis techniques. It is bioinformatics that enables functional genomics to bring order out of chaos.

## CONCLUSION AND FUTURE PROSPECTS

In view of its extreme complexity and strong association with a variety of complex disease phenotypes, aging is the ideal target for the new discipline of functional genomics. Functional genomics can best be defined as a comprehensive approach to studying the dynamic network of genes that ultimately determines the physiology of an individual organism. As such, it is capable of overcoming the formidable barrier of the current lack of tools to define and understand complex functional organizations, consisting of multiple interacting subsystems, such as aging, the

Table III Bioinformatics: From Tools to Dominance

• Gene identification (e.g., ab initio)
• Gene ontology
• Interpretation of expression patterns (e.g., clustering analysis)
• Interpretation of SNP maps
• Three-dimensional modeling of proteins
• Prediction of protein function and protein-protein interaction
• Studying digital organisms

most complex phenotype that we know. Functional genomics of aging is likely to greatly increase our knowledge of human aging, including its relationship with disease, and revolutionize geriatric medicine by changing the ways in which we diagnose, monitor, and treat elderly patients.

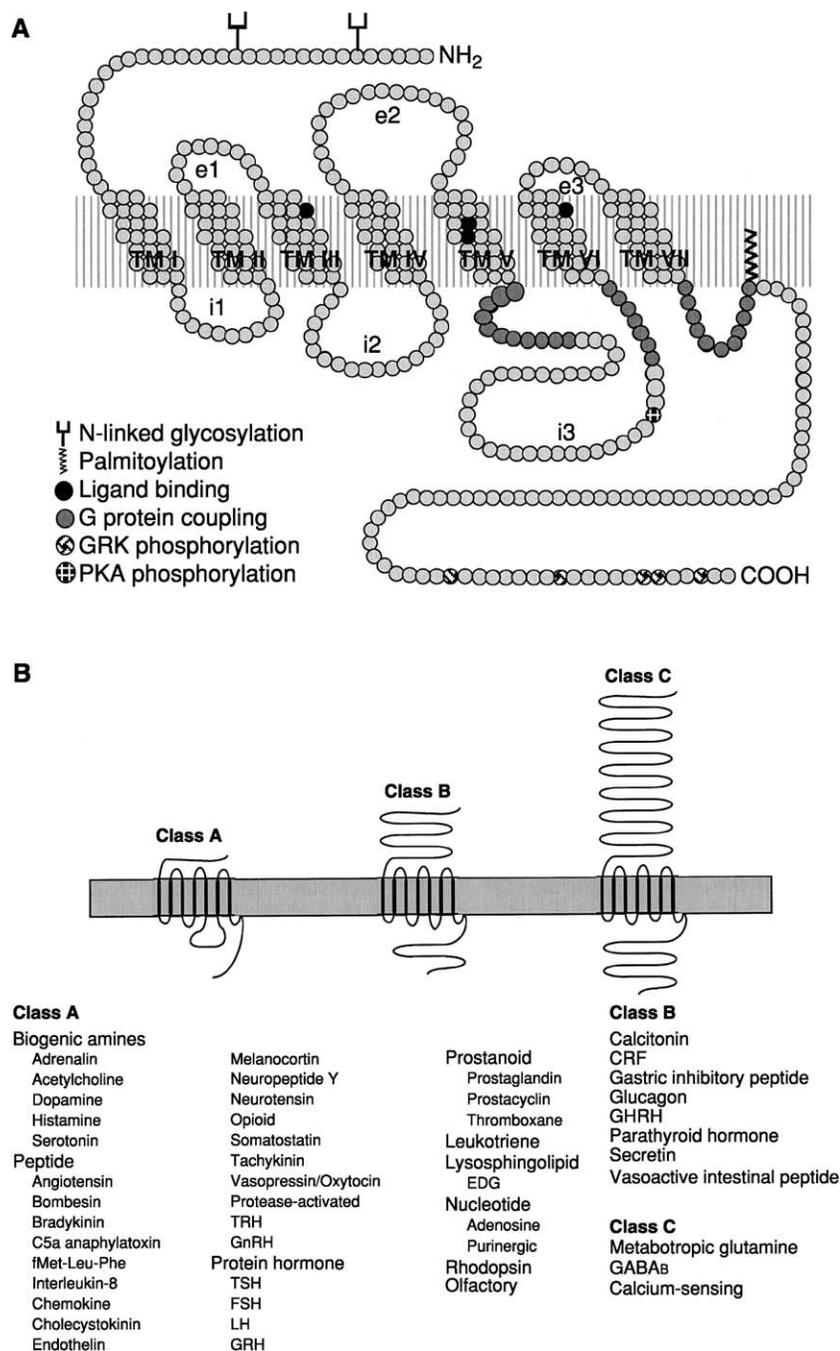
### See Also the Following Articles

Aging and Longevity of Human Populations • Aging and the Male Reproductive System • Aging, Animal Models for • Aging, Immunology and • Alzheimer's Disease and Hormones • Autonomic Nervous System, Aging and • Neuroendocrine System and Aging • Oxidative Stress and Aging • Stress, Aging, and Central Nervous System Interactions

### Further Reading

Berdanier, C. D., and Everts, H. B. (2001). Mitochondrial DNA in aging and degenerative disease. *Mut. Res.* **475**, 169–184.

- Butte, A. (2002). The use and analysis of microarray data. *Nat. Rev.* **1**, 951–960.
- Collins, F. S., Green, E. D., Guttmacher, A. E., and Guyer, M. S., for U.S. National Human Genome Research Institute. (2003). A vision for the future of genomics research. *Nature* **422**, 835–847.
- McGrath, S. B., Bounpheng, M., Torres, L., Calavetta, M., Scott, C. B., Suh, Y., Rines, D., van Orsouw, N., and Vijg, J. (2001). High-speed, multicolor fluorescent two-dimensional gene scanning. *Genomics* **78**, 83–90.
- Nair, P. N., Golden, T., and Melov, S. (2003). Microarray workshop on aging. *Mech. Ageing Dev.* **124**, 133–138.
- Slonim, D. K. (2002). From patterns to pathways: Gene expression data analysis comes of age. *Nat. Genet.* **32**(Suppl.), 502–508.
- Stein, L. D. (2003). Integrating biological databases. *Nat. Rev. Genet.* **4**, 337–345.
- Tyers, M., and Mann, M. (2003). From genomics to proteomics. *Nature* **422**, 193–197.
- Vijg, J., and Suh, Y. (2003). Functional genomics of ageing. *Mech. Ageing Dev.* **124**, 3–8.
- Vijg, J., and van Orsouw, N. (2002). Searching for genetic determinants of human aging and longevity: Opportunities and challenges. *Mech. Ageing Dev.* **123**, 195–205.



**Figure 1** Architecture of G protein-coupled receptors. (A) Schematic diagram of the predicted heptahelical structure of the  $\beta_2$ -adrenergic receptor, a Class A GPCR. Seven membrane-spanning domains (TMI–TMVII) are connected by three extracellular (e1–e3) and three intracellular (i1–i3) loops. The approximate positions of posttranslational modification, including N-glycosylation of the extracellular amino terminus and palmitoylation of the intracellular carboxy terminus, are indicated. Residues involved in epinephrine binding and predicted sites of PKA and GRK phosphorylation are shown. (B) Schematic comparison of the structure of the three classes of mammalian GPCR along with examples of the receptors belonging to each class.

transmembrane  $\alpha$  helices (TMI–TMVII), linked by three alternating intracellular and extracellular loops (e1–3 and i1–3). The transmembrane domains share the highest degree of sequence conservation, whereas

the intracellular and extracellular domains exhibit great variability in size and complexity. The extracellular and transmembrane regions of the receptor are involved in ligand binding, whereas the intracellular

domains are important for signal transduction via heterotrimeric G proteins and for feedback modulation of receptor function.

The gene structure of GPCRs is highly variable. The genes encoding some GPCRs, such as the  $\beta$ 2-adrenergic receptor, are unusual among eukaryotic genes in that they are intronless. Others show typical intron–exon structure. Usually, the intron–exon borders correspond to functional domains, allowing for the insertion of large extracellular or intracellular domains into the basic heptahelical receptor architecture. Variation in the complexity of the extracellular and intracellular domains that connect to the more conserved transmembrane helices accounts for extensive variability in the size of GPCRs (Fig. 1B). The pattern of GPCR expression also varies widely, with some receptors (e.g., the  $\beta$ -adrenergic receptors) being nearly ubiquitous in distribution, whereas others (e.g., rhodopsin) are confined to a single cell type.

GPCRs undergo posttranslational modifications that can affect receptor function. One or more sites for N-glycosylation are present within the amino terminus or, less often, the extracellular loops. Most GPCRs have in common two Cys residues that form a disulfide bridge between e1 and e2 that is critical for normal protein folding and another Cys residue in the carboxy-terminal domain that serves as a site for palmitoylation. This lipid modification leads to the formation of a putative fourth intracellular loop.

### Classification of Heptahelical Receptors

Structurally, GPCRs have been divided into three main classes (A, B, and C) that contain all of the known human receptors. Two additional classes (D and E) are made up of fungal pheromone and *Dictyostelium* GPCRs. Each class has been further divided into subclasses based on a combination of sequence homology, functional domains, and ligand-binding properties (Fig. 1B).

GPCRs of Class A, of which rhodopsin and the  $\beta$ 2-adrenergic receptor are members, make up the largest family. Class A GPCRs share approximately 20 conserved amino acid residues, including the three Cys residues in e1 and e2 and the carboxy terminus, an Asp–Arg–Tyr (DRY) motif in i2, and an Asn–Pro–x–x–Tyr (NPxxY) motif in TMVII. Within class A, the rhodopsin subclass contains a light reactive 11-*cis*-retinal moiety covalently linked to TMVII that serves as a detector for incoming photons of light. In a second subclass that contains receptors for biogenic

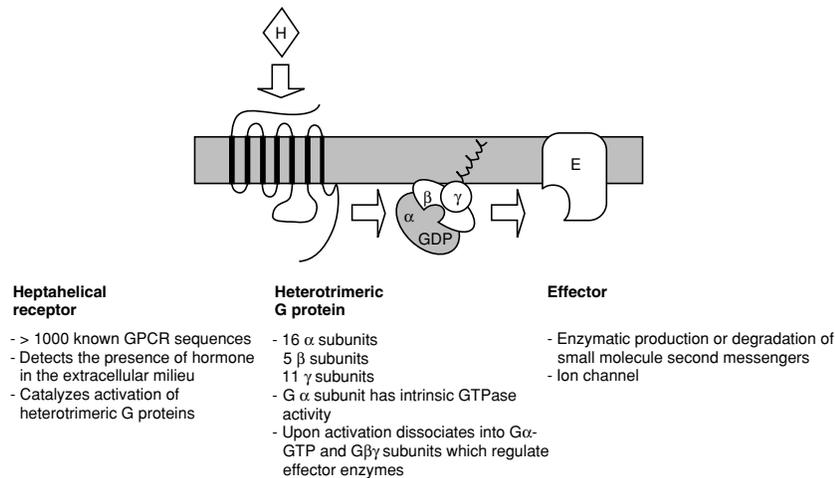
amines and other small nonpeptide ligands, the bound ligand resides within a ligand-binding “pocket” formed by the seven transmembrane helices. Another subclass includes receptors that respond to chemokines, opioids, and other small peptides. These ligands bind to the amino terminus, e1–3, and regions of the TMs that are close to e1–3. This group also contains protease-activated receptors such as the PAR1 thrombin receptor. The amino terminus of these receptors includes a peptide sequence that serves as a tethered ligand once it is exposed by regulated proteolysis. Class A receptors also include several GPCRs in which ligand binding occurs within a relatively large amino-terminal domain. This subclass includes receptors for glycoprotein hormones such as luteinizing hormone (LH), follicle-stimulating hormone (FSH), and thyroid-stimulating hormone (TSH).

Class B GPCRs, the second largest class, contain receptors that bind to hormones of higher molecular weight such as glucagon, calcitonin, and parathyroid hormone. Class B GPCRs are characterized by long amino-terminal domains involved in ligand binding, six well-conserved Cys residues in the amino terminus, two highly conserved Cys residues in e1 and e2, and approximately 15 other residues that are identical in all members of this class. Class C, the smallest class of mammalian GPCRs, contains the metabotropic glutamate receptors, the GABA<sub>B</sub> receptor, and the calcium-sensing receptor. These receptors possess extremely long extracellular amino-terminal domains that are involved in ligand binding and contain several conserved Cys residues in the transmembrane and extracellular regions that are essential for receptor function.

## G PROTEIN-COUPLED RECEPTOR SIGNALING VIA HETEROTRIMERIC G PROTEINS

### The Receptor–G Protein–Effector Model of Signal Transduction

The binding of a “first messenger” hormone to the extracellular or transmembrane domains of a GPCR triggers conformational changes that are transmitted to the intracellular receptor domains and lead to coupling between the receptor and heterotrimeric G proteins. Ligand-bound GPCRs promote the activation of heterotrimeric G proteins by catalyzing the exchange of GTP for GDP on the G $\alpha$ -subunit and



**Figure 2** The receptor–G protein–effector model of G protein-coupled receptor signaling. The three principal components of GPCR signaling are the heptahelical receptor, heterotrimeric G protein, and effector enzyme. The receptor detects the presence of a hormone (H) or “first messenger” in the extracellular milieu. The heterotrimeric G protein dissociates into a GTP-bound G $\alpha$  subunit and G $\beta\gamma$  heterodimer on interaction with a ligand-bound receptor. The effector, which is typically an enzyme or ion channel, is activated by free G $\alpha$ -GTP or G $\beta\gamma$  subunits and produces small molecule “second messengers” that transmit signals intracellularly.

dissociation of the GTP-bound G $\alpha$ -subunit from the G $\beta\gamma$ -subunit heterodimer. Once dissociated, free G $\alpha$ -GTP and G $\beta\gamma$ -subunits regulate the activity of enzymatic effectors, such as adenylate cyclases, phospholipase C isoforms, and ion channels, to generate small molecule “second messengers.” These second messengers, in turn, control the activity of protein kinases that regulate key enzymes involved in intermediary metabolism. Signaling continues until the intrinsic GTPase activity of the G $\alpha$ -subunit returns the G protein to the inactive heterotrimeric state. This three-component system, originally discovered as the mechanism for hormonal control of adenylate cyclase, forms the basis of the classical GPCR signaling paradigm (Fig. 2).

### Heterotrimeric G Proteins

The heterotrimeric G proteins are a group of GTPases that share a common multi-subunit structure. They function as intermediates in the transduction of external stimuli by linking GPCRs, which lack intrinsic enzymatic activity, to enzymatic effectors that convey information about the presence of an external stimulus to the cell interior. Heterotrimeric G proteins are composed of a 39- to 52-kD GTP-binding G $\alpha$ -subunit that possesses intrinsic GTPase activity and a G $\beta$ - and G $\gamma$ -subunit that form a tightly linked heterodimer that is noncovalently bound to the G $\alpha$ -subunit in the inactive state.

In contrast to the immense diversity found in GPCRs, there are comparatively few genes encoding each of the heterotrimeric G protein subunits. The 16 known mammalian G $\alpha$ -subunit genes are grouped by sequence homology into four families. With splice variants, approximately 20 distinct G $\alpha$ -subunit proteins are expressed. The G $\alpha_s$  family contains the adenylate cyclase-stimulatory  $\alpha$ -subunit, G $\alpha_s$ , and the olfactory  $\alpha$ -subunit, G $\alpha_{olf}$ . The G $\alpha_i$  family includes the adenylate cyclase-inhibitory  $\alpha$ -subunits, G $\alpha_{i1}$ , G $\alpha_{i2}$ , G $\alpha_{i3}$ , and G $\alpha_o$ ; two isoforms of the retinal  $\alpha$ -subunit, transducin or G $\alpha_t$ ; the taste  $\alpha$ -subunit, gustducin or G $\alpha_{gust}$ ; and G $\alpha_z$ , an  $\alpha$ -subunit whose function is not well understood. The G $\alpha_q$  family includes two  $\alpha$ -subunits that regulate phospholipase C activity, G $\alpha_q$  and G $\alpha_{11}$ , as well as G $\alpha_{14}$  and G $\alpha_{15}$ . Finally, the G $\alpha_{12}$  family contains G $\alpha_{12}$  and G $\alpha_{13}$ , two  $\alpha$ -subunits whose function is poorly understood.

All G $\alpha$ -subunits share considerable sequence homology (~40%) that corresponds to the regions of the protein that form the guanine nucleotide-binding pocket. More divergence is found in the amino terminus, which is required for G $\beta\gamma$ -subunit binding, and in the carboxy terminus, which is involved in the association with both receptors and effectors. Although none of the G $\alpha$ -subunits possess membrane-spanning domains, all associate with the plasma membrane. G $\alpha_i$  and G $\alpha_o$  are posttranslationally modified by the addition of the fatty acid myristate to an amino-terminal glycine residue, and this

lipid modification is required for membrane association. All  $G\alpha$ -subunits, with the exception of  $G\alpha_t$ , also undergo palmitoylation of one or more amino-terminal cysteine residues, and this is also important for membrane localization. Unlike myristoylation, however, this is a reversible modification. On binding of GTP and  $G\beta\gamma$ -subunit dissociation, palmitate is cleaved. Reassociation of the  $G\alpha$ - and  $G\beta\gamma$ -subunits promotes repalmitoylation.

In addition to the  $G\alpha$ -subunits, there are 5 known  $G\beta$ -subunits and 12  $G\gamma$ -subunits. All  $G\beta$ -subunits are approximately 36 kD in size and share a common three-dimensional structure with an amino-terminal, coiled-coil structure that associates tightly with the  $G\gamma$ -subunit. The remainder of the protein resembles a seven-bladed propeller, with each blade composed of interlocking  $\beta$ -sheets. All  $G\gamma$ -subunits are 7 to 8 kD and possess a carboxy-terminal Cys-Ala-Ala-x motif, which serves a site for prenylation, a lipid modification that is essential for membrane localization. Most, but not all,  $G\beta$ - and  $G\gamma$ -subunit pairs can form stable heterodimers *in vitro*, suggesting that they may combine *in vivo* to generate a diverse array of  $G\beta\gamma$ -subunit heterodimers. Some combinations make up the predominant form of the heterodimer in specific tissues; for example,  $G\beta_1\gamma_1$  is the predominant form associated with  $G\alpha_t$  in the retina. Some evidence suggests that specific  $G\beta\gamma$ -subunit heterodimers differentially regulate certain effectors; however, the overall contribution of  $G\beta\gamma$  heterodimer diversity to GPCR signal transduction is poorly understood.

Although GPCRs generally exhibit preferential coupling to certain G proteins, the degree of coupling specificity varies widely. In some cases, such as the chemoattractant C5A receptor and the M4 muscarinic acetylcholine receptor, coupling is limited to specific members of a single G protein subfamily. In other cases, coupling is much more promiscuous, involving members of different G protein families. For example, the thyroid-stimulating hormone receptor couples to both  $G_s$  and  $G_{q/11}$  family G proteins and stimulates both adenylyl cyclase and phospholipase activity, whereas the thrombin receptor couples to both  $G_{q/11}$  and  $G_i$  family G proteins.

## G Protein Effectors

The third component of the GPCR signaling system is the G protein-regulated effector, which is typically an enzyme or ion channel. Although it was once thought that only  $G\alpha$ -subunits interacted with

effectors, it is now clear that both GTP-bound  $G\alpha$  and  $G\beta\gamma$ -subunits regulate effector activity.

The most widely studied G protein-regulated enzymes are the adenylyl cyclases, which catalyze the conversion of ATP to the intracellular second messenger cyclic-adenosine 3',5'-monophosphate or cyclic AMP (cAMP). cAMP regulates the activity of the cAMP-dependent protein kinase (PKA). The 10 cloned adenylyl cyclases all share a common 12-membrane-spanning domain architecture but vary significantly in tissue distribution and in regulation by G protein subunits. All of the adenylyl cyclases are stimulated by  $G\alpha_s$ . Some are also stimulated by  $G\beta\gamma$ -subunits or by calcium-calmodulin, but only in the presence of active  $G\alpha_s$ , creating the prospect of "conditional stimulation" in which activation of more than one type of GPCR might be required for full activation of the cyclase.  $G\alpha_i$  mediates inhibition of some, but not all, adenylyl cyclase isoforms, as does  $G\alpha_o$  and intracellular calcium. The activity of some adenylyl cyclases is also modulated through phosphorylation by the second messenger-dependent protein kinases PKA and protein kinase C (PKC).

The four phospholipase C (PLC) $\beta$  isoforms are also regulated by G proteins. Phospholipase C catalyzes the hydrolysis of membrane phosphatidylinositol to yield two intracellular second messengers: inositol 1,4,5-trisphosphate, which controls calcium efflux from the endoplasmic reticulum, and diacylglycerol, which along with calcium controls the activity of several isoforms of PKC. PLC $\beta_1$ –3 are independently activated by both  $G\alpha$ -subunits of the  $G_q$  family and  $G\beta\gamma$ -subunits. The PLC $\beta_2$  and PLC $\beta_3$  isoforms are more sensitive to  $G\beta\gamma$ -subunit regulation than is PLC $\beta_1$  and usually account for the activation of PLC by GPCRs coupled to  $G_i$  family proteins.  $G\beta\gamma$ -subunits may also directly regulate phospholipase A2 activity in some settings.

In the rod cells of the retina, activated transducin stimulates a cGMP phosphodiesterase, leading to closure of plasma membrane sodium channels, hyperpolarization of the plasma membrane, and signaling of second-order retinal neurons. Other ion channels may function as direct G protein-regulated effectors. Direct interaction between channel components and G protein subunits may account for the inhibition of high-voltage "N-type" calcium channels by GPCRs acting via subtypes of  $G\alpha_o$  and the stimulation of "L-type" calcium channels by GPCRs acting via  $G\alpha_s$ .  $G\beta\gamma$ -subunits appear to be the physiological activator of the inward-rectifying muscarinic-gated potassium channel IKACH, which mediates the cardiac response to vagal stimulation.

Finally, some kinases, notably the G protein-coupled receptor kinases (GRK)2 and GRK3, are directly regulated by G $\beta\gamma$ -subunits. These GRKs, which play a central role in receptor desensitization and may also phosphorylate nonreceptor substrates, possess carboxy-terminal regulatory domains that bind free G $\beta\gamma$ -subunits. Binding to the G $\beta\gamma$ -subunits of activated G proteins allows the kinase to translocate from the cytosol to the plasma membrane, where it gains access to substrate.

GPCRs are classically thought to act catalytically in that one GPCR may activate multiple G proteins. Free G protein subunits then act on enzymatic effectors or ion channels, producing a pool of second messengers that control protein kinase activity. The presence of multiple enzymatic steps in the GPCR signaling cascade creates the potential for tremendous signal amplification. As a result of this amplification process, a full biological response can often be obtained with as little as 5% receptor occupancy, a phenomenon that early pharmacologists referred to as “spare receptors.”

## MODULATION OF G PROTEIN-COUPLED RECEPTOR SIGNALING

### Positive Regulation of G Protein-Coupled Receptor Signaling

Over the past several years, it has become evident that the physical interaction of GPCRs with one another and with certain accessory proteins serves to expand the repertoire of GPCR signaling. Two examples of such phenomena are receptor dimerization and the interaction of GPCRs with receptor activity-modifying proteins (RAMPs).

#### **GPCR Dimerization**

A range of biochemical and biophysical data indicate that many, if not most, GPCRs can exist in the form of homo- or hetero-oligomers. In some cases, dimerization appears to be regulated by agonist binding. Limitations in currently available techniques have to this point precluded an accurate determination of the extent of GPCR oligomerization in intact cells. Nonetheless, some data support the concept that the formation of GPCR dimers or higher order oligomers may be required for receptor function or may contribute to diversity of ligand binding. The former concept is supported by the finding that the internal tethered ligand of one protease-activated receptor can

“transactivate” other family members, a signaling event that must occur over very short distances. The latter concept has been invoked to account for observed diversity in opioid receptor pharmacology that appears to extend beyond that attributable to any single receptor subtype.

The most compelling data for an obligatory functional role for receptor dimerization, however, comes from study of the class C GABA<sub>B</sub> receptor. The GABA<sub>B</sub>R1 receptor, which contains all of the structural determinants necessary for ligand binding and G protein activation, is retained in the endoplasmic reticulum as an immature glycoprotein unless it is coexpressed with a second GABA<sub>B</sub> receptor transcript, the GABA<sub>B</sub>R2. The GABA<sub>B</sub>R2 receptor alone can reach the cell surface but cannot bind ligand. Dimerization of the two receptors, which is mediated by their carboxy-terminal tails, leads to occlusion of an endoplasmic reticulum retention sequence located in the tail of the GABA<sub>B</sub>R1 and allows for correct processing and membrane transport of GABA<sub>B</sub>R1. These data indicate that the GABA<sub>B</sub>R2 functions as a chaperon for GABA<sub>B</sub>R1 through the formation of an obligatory GPCR dimer. Similarly, expression of truncated class A D2 dopamine and V2 vasopressin receptors inhibits the trafficking and function of the intact receptor, suggesting that GPCR dimerization prior to membrane delivery may be a general process.

#### **Receptor Activity-Modifying Proteins**

The pharmacology of at least two class B GPCRs is determined not strictly by the intrinsic structure of receptor but also by their interaction with members of a family of novel transmembrane proteins, called receptor activity-modifying proteins (RAMPs). The three known RAMP proteins are 148- to 174-amino acid single transmembrane domain glycoproteins with large extracellular domains and short intracytoplasmic domains. RAMPs form complexes with the calcitonin receptor-like receptor (CRLR) and calcitonin receptor and control receptor trafficking and function. RAMP binding to the CRLR is required for transport of nascent CRLR to the plasma membrane. Furthermore, the specific CRLR–RAMP complex determines the ligand specificity of the receptor. The CRLR–RAMP1 complex acts as a receptor for the calcitonin gene-related peptides, a pleiotropic family of neuropeptides with homology to calcitonin, amylin, and adrenomedullin. When CRLR is complexed with RAMP2 or RAMP3 it functions as an adrenomedullin receptor. Similarly, complexes between a naturally occurring splice variant of the calcitonin receptor and RAMP1 or RAMP3 yield a functional amylin

receptor. RAMP expression is modified under various forms of physiological stress and in response to glucocorticoids, suggesting that cellular responsiveness to certain hormones may be regulated through control of accessory proteins.

### Negative Regulation of G Protein-Coupled Receptor Signaling

A common principle of receptor biology is that of tachyphylaxis, the waning of signal intensity over time in the continued presence of agonist. Presumably, tachyphylaxis acts as a protective mechanism by allowing the cell to adjust its sensitivity to external stimuli and by preventing signal overload. Mechanisms to dampen GPCR signals exist at every level, from receptor to G protein to effector. Second messengers are enzymatically degraded, for example, by cAMP phosphodiesterases, phosphatidylinositol phosphatases, and diacylglycerol kinases. Effector activity is modulated both by the availability of free G protein subunits and by feedback inhibition, as in the regulation of certain isoforms of adenylate cyclase by PKA or PKC phosphorylation. G $\alpha$ -subunit activity is modulated by protein-protein interactions that accelerate the normally slow intrinsic GTPase activity of the protein. In some cases, this GTPase-activating protein (GAP) activity is an inherent property of the effector, such that interaction of the G $\alpha$ -GTP complex with the effector leads to an acceleration of GTP hydrolysis and a return to the inactive state. In other cases, GAP activity is conferred by members of a large family of proteins called regulators of G protein signaling (or RGS proteins) that interact with the G $\alpha$ -subunit and stabilize the transition state for GTP hydrolysis. Approximately 19 mammalian genes containing RGS core domains are known. Most act as GAPs toward Gi proteins, and some also act as GAPs for Gq proteins. G $\beta\gamma$ -subunit function is also under control. In the retina, the G $\beta\gamma$ -subunit-binding protein phosducin can sequester free G $\beta\gamma$ -subunits, thereby modulating the G $\beta\gamma$ -G $\alpha$ t interaction.

The GPCR itself is the target for extensive negative regulation. The processes that regulate GPCR responsiveness at the receptor level are typically divided, based on mechanism, into heterologous desensitization, homologous desensitization, receptor sequestration, and receptor down-regulation. Together, they lead sequentially to the uncoupling of receptor from G protein, removal of receptors from the plasma membrane, receptor degradation, and reduced synthesis of new receptors (Fig. 3).

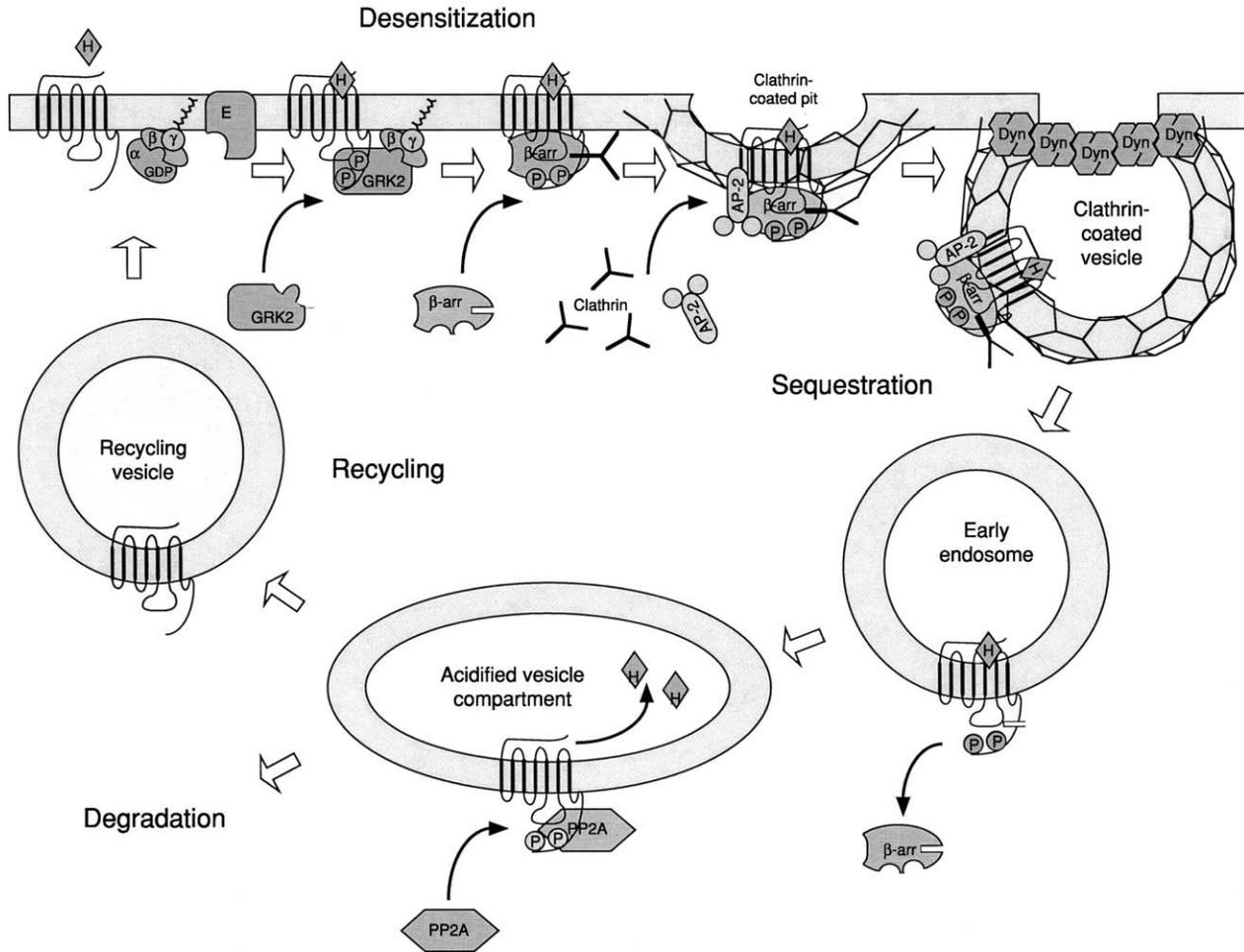
### Heterologous Desensitization

Desensitization begins within seconds of agonist exposure and is initiated by phosphorylation of the receptor. Second messenger-dependent protein kinases, including PKA and PKC, phosphorylate serine and threonine residues within the cytoplasmic loops and C-terminal tail domains of many GPCRs. Phosphorylation of these sites directly impairs receptor-G protein coupling. For example, PKA phosphorylation of purified  $\beta$ 2-adrenergic receptors *in vitro* is sufficient to impair receptor-stimulated Gs activation in the absence of other proteins. Agonist occupancy of the target GPCR is not required for this process. Thus, receptors that have not bound agonist, including receptors for other ligands, can be desensitized by the activation of second messenger-dependent protein kinases. This lack of requirement for receptor occupancy has led to the use of the term "heterologous desensitization" to describe the process.

For some GPCRs, phosphorylation may also alter the specificity of G protein coupling, leading to feedback inhibition or activation of secondary G protein-mediated events. For example, phosphorylation of the  $\beta$ 2-adrenergic receptor by PKA shifts receptor coupling from the adenylate cyclase-stimulatory G protein, Gs, to the adenylate cyclase-inhibitory G protein, Gi. This leads to reduced adenylate cyclase activation and couples the receptor to pathways leading to activation of mitogen-activated protein (MAP) kinases.

### Homologous Desensitization and the Arrestins

Homologous desensitization is mediated by phosphorylation of the receptor by specialized kinases called G protein-coupled receptor kinases (or GRKs), followed by the binding of arrestins. There are seven known GRKs. GRK1 (rhodopsin kinase) and GRK7 are retinal kinases involved in the regulation of rhodopsin photoreceptors, whereas GRK2 to GRK6 are more widely expressed. Membrane targeting of all the GRKs apparently is critical to their function and is conferred by a carboxy-terminal tail domain. GRK1 and GRK7 each possesses a carboxy-terminal Cys-Ala-Ala-x motif. Light-induced translocation of GRK1 from the cytosol to the plasma membrane is facilitated by the posttranslational farnesylation of this site. The  $\beta$ -adrenergic receptor kinases (GRK2 and GRK3) have carboxy-terminal G $\beta\gamma$ -subunit-binding and pleckstrin-homology domains, and they translocate to the membrane as a result of interactions between these domains and free G $\beta\gamma$ -subunits and inositol phospholipids. Palmitoylation of GRK4 and GRK6 on carboxy-terminal cysteine



**Figure 3** Desensitization, sequestration, and recycling of G protein-coupled receptors. Within seconds of hormone (H) binding, GPCRs are phosphorylated by both second messenger-dependent protein kinases and specialized G protein-coupled receptor kinases (GRKs). Heterologous desensitization is due to phosphorylation of the receptor by second messenger-dependent protein kinases that leads to impaired receptor–G protein coupling. Homologous desensitization of GPCRs results from the binding of  $\beta$ -arrestins ( $\beta$ -arr) to agonist-occupied receptors following phosphorylation of the receptor by GRKs.  $\beta$ -arrestin binding blocks coupling between the receptor and G protein, leading to termination of signaling by G protein effectors (E). Receptor-bound  $\beta$ -arrestins also bind to components of the clathrin endocytic machinery, including clathrin,  $\beta$ 2-adaptin (AP-2), and NSF. Receptor sequestration reflects the dynamin (Dyn)-dependent endocytosis of GPCRs via clathrin-coated pits. Once internalized, receptors may traffick to an acidified endosomal compartment, where the ligand is dissociated and the receptors are dephosphorylated by a GPCR-specific protein phosphatase PP2A isoform, after which they are recycled to the plasma membrane. Alternatively, receptors are targeted for degradation as part of the process of receptor down-regulation.

residues leads to constitutive membrane localization. Targeting of GRK5 to the membrane is thought to involve the electrostatic interaction of a highly basic 46-residue carboxy-terminal domain with membrane phospholipids.

Like PKA and PKC, GRKs phosphorylate GPCRs on serine and threonine residues in the i3 loop and carboxy-terminal tail. Unlike the second messenger-dependent kinases, GRKs preferentially phosphorylate receptors that are in the agonist-occupied conformation. Furthermore, GRK phosphorylation alone has little direct effect on receptor–G protein

coupling. Rather, the role of GRK phosphorylation is to increase the affinity of the receptor for arrestins. It is the binding of an arrestin to the receptor domains involved in G protein coupling, rather than GRK phosphorylation, that leads to uncoupling of receptor and G protein.

To this point, four functional members of the arrestin gene family have been cloned. Two arrestins, visual arrestin (arrestin 1) and cone arrestin, are expressed almost exclusively in the retina, where they regulate photoreceptor function. The  $\beta$ -arrestins,  $\beta$ -arrestin 1 and  $\beta$ -arrestin 2 (also called arrestin 2 and

arrestin 3) are ubiquitously expressed proteins whose highest levels of expression are in brain and spleen. All arrestins are 48- to 50-kD proteins that bind specifically to light-activated or agonist-occupied GPCRs that have been phosphorylated by GRKs. The crystal structure of visual arrestin indicates that the arrestins contain two major domains, an amino-terminal domain and a carboxy-terminal domain, each composed of a seven-stranded  $\beta$  sandwich and joined by a phosphate sensor region that maintains the protein in an inactive conformation until it interacts with a GRK-phosphorylated GPCR. Additional regulatory motifs reside at the amino- and carboxy-termini of the protein.

### **Receptor Sequestration, Resensitization, and Down-Regulation**

Internalization of GPCRs, also termed receptor sequestration or endocytosis, occurs more slowly than does desensitization, occurring over a period of several minutes after agonist exposure. It is likely that several mechanisms exist for the removal of GPCRs from the membrane, but the most universal and best-understood mechanism involves  $\beta$ -arrestins acting as adapter proteins to target GPCRs to clathrin-coated pits for endocytosis.

The carboxy-terminal tail of the  $\beta$ -arrestins, but not the visual arrestins, contains structural motifs that allow them to link GPCRs to components of the clathrin-dependent endocytic machinery.  $\beta$ -arrestins bind with high affinity to both clathrin and the  $\beta$ 2-adaptin subunit of the heterotetrameric AP-2 adapter complex. In addition,  $\beta$ -arrestin 1 binds to the N-ethylmaleimide-sensitive fusion protein (NSF), an ATPase involved in intracellular transport. Through these interactions, agonist-occupied GPCRs are directed to clathrin-coated pits and subsequently removed from the cell surface.  $\beta$ -arrestin 1 function is also modulated by phosphorylation of a carboxy-terminal serine, whereas  $\beta$ -arrestin 2 has been shown to undergo rapid ubiquitination, catalyzed by an E3 ubiquitin ligase, Mdm2, which binds directly to the  $\beta$ -arrestin. Ubiquitination of  $\beta$ -arrestin 2 apparently is required for receptor internalization, although its exact role is not currently understood.

Once internalized, GPCRs either are degraded or are dissociated from ligand, dephosphorylated, and returned to the cell surface through a process termed "resensitization" or "recycling." For many GPCRs, most notably the  $\beta$ 2-adrenergic receptor, resensitization requires  $\beta$ -arrestin binding and internalization because receptor dephosphorylation must occur within acidified endosomal vesicles. The stability of the receptor- $\beta$ -arrestin complex appears to play an

important role in determining rate of receptor recycling. Recent data suggest that GPCRs can be grouped based on their pattern of interaction with  $\beta$ -arrestins. One group of receptors, represented by the  $\beta$ 2- and  $\alpha$ 1B-adrenergic,  $\mu$  opioid, endothelin A, and dopamine D1A receptors, bind to  $\beta$ -arrestin 2 *in vitro* with higher affinity than to  $\beta$ -arrestin 1 and do not bind to visual arrestin. In addition, their interaction with  $\beta$ -arrestin is transient.  $\beta$ -arrestin is recruited to the receptor at the plasma membrane and translocates with it to clathrin-coated pits, but the receptor- $\beta$ -arrestin complex dissociates on internalization of the receptor, such that the  $\beta$ -arrestin recycles to the plasma membrane as the receptor proceeds into an endosomal pool. A second group of receptors, which includes the angiotensin AT1a, neurotensin 1, vasopressin 2, thyrotropin-releasing hormone, and neurokinin NK-1 receptors, bind  $\beta$ -arrestin 1 and  $\beta$ -arrestin 2 with equal affinity *in vitro* and also interact with visual arrestin. These receptors form stable complexes with  $\beta$ -arrestin, such that the receptor- $\beta$ -arrestin complex internalizes as a unit that is targeted to endosomes. The presence of specific clusters of serine and threonine residues in the carboxy-terminal tail of the receptor apparently dictates the stability of the receptor- $\beta$ -arrestin complex. Data obtained using chimeric receptors, in which the carboxy-terminal tail of receptors of various groups have been exchanged, suggest that the formation of a transient receptor- $\beta$ -arrestin complex favors rapid dephosphorylation and return to the plasma membrane, whereas formation of a stable receptor- $\beta$ -arrestin complex retards resensitization and may favor targeting of the receptor for degradation.

Down-regulation of GPCRs, the persistent loss of cell surface receptors that occurs over a period of hours to days, is the least understood of the processes controlling GPCR responsiveness. Control of cell surface receptor density occurs at least partially at the transcriptional level, but the removal of agonist-occupied receptors from the cell surface and their sorting for either degradation or recycling to the membrane is also important, at least during the early stages of down-regulation. Consistent with this is the finding that down-regulation of  $\beta$ 2-adrenergic receptors does not occur in murine fibroblasts lacking both  $\beta$ -arrestin 1 and  $\beta$ -arrestin 2.

Other factors contribute in as yet poorly understood ways to regulate receptor recycling and down-regulation. Ubiquitination of the  $\beta$ 2-adrenergic receptor, which appears to be a  $\beta$ -arrestin-dependent event but is mediated by an as yet unidentified ubiquitin ligase, is not required for receptor internalization

but does appear to determine the rate of proteosomal degradation of receptors. In addition, an interaction between the distal carboxy-terminal tail of some GPCRs and the PDZ domain containing protein NHERF1/EBP-50 may regulate endosomal sorting. The NHERF1/EBP-50 protein links  $\beta$ 2-adrenergic receptors to the actin cytoskeleton, and disruption of this interaction leads to missorting of the endocytosed receptor and accelerated receptor degradation.

## NOVEL MECHANISMS OF G PROTEIN-COUPLED RECEPTOR SIGNALING

Certain physiological responses to GPCR stimulation have been difficult to explain in the context of the classical receptor–G protein–effector paradigm of heptahelical receptor signaling. Notable among these is the ability of GPCRs to stimulate cell proliferation, which in many systems is demonstrably independent of the activation of second messenger-dependent protein kinases. Study of the mechanisms underlying these responses has led to the discovery that GPCRs can employ novel signal transduction mechanisms, some of which appear not to require G protein activation.

### Heterologous Receptor Cross-Talk

In many cell types, GPCR-mediated activation of the extracellular signal-regulated kinase (ERK) MAP kinase cascade, a central pathway in the control of cell proliferation, shares many early steps with signals arising from classical receptor tyrosine kinase growth factor receptors. The basis for this convergence of signaling is that at least two receptor tyrosine kinases, those for platelet-derived growth factor and for epidermal growth factor (EGF), can be “transactivated” by GPCRs.

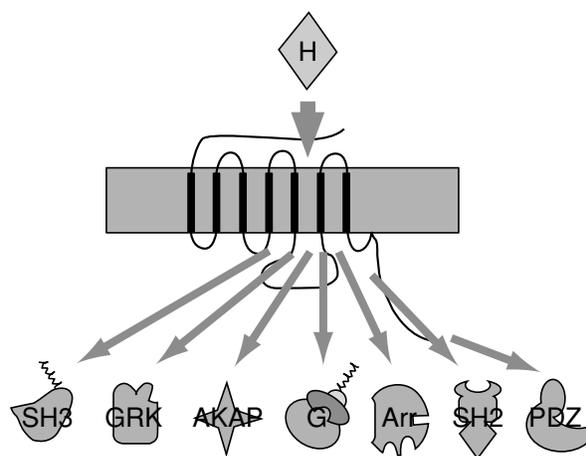
The best-understood mechanism of cross-talk between GPCRs and receptor tyrosine kinases is the transactivation of EGF receptors resulting from the autocrine/paracrine release of EGF receptor ligands from the cell surface via a process termed “ectodomain shedding.” Each of the known ligands for the EGF receptor—EGF, transforming growth factor  $\alpha$ , heparin-binding (HB)–EGF, amphiregulin, betacellulin, and epiregulin—is synthesized as a transmembrane precursor that is proteolyzed to produce a soluble growth factor. In fibroblasts, both  $G_i/o$ - and  $G_q/11$ -coupled receptors have been shown to stimulate HB–EGF release. For  $G_i/o$ -coupled GPCRs,

HB–EGF shedding and EGF receptor transactivation are mediated by  $G\beta\gamma$ -subunits, whereas HB–EGF shedding in response to stimulation of  $G_q/11$ -coupled receptors is mediated by  $G_q/11\alpha$ -subunits. Proteolysis of the HB–EGF precursor is mediated by GPCR-regulated matrix metalloprotease activity, possibly by a member of the membrane-associated ADAM family of matrix metalloproteases. The immediate G protein effectors that regulate ectodomain shedding are incompletely defined, although phosphatidylinositol 3'-kinases and Src family nonreceptor tyrosine kinases have been implicated in the process.

### GPCRs as Scaffolds

A recent development in the field of GPCR signal transduction has been the discovery that GPCRs can signal by interacting with proteins other than heterotrimeric G proteins. This model of signaling, in which the agonist-occupied receptor recruits proteins to form a signaling complex, is fundamentally different from the classical model G protein-mediated signaling, which is thought of as a catalytic event (Fig. 4).

In some cases, the GPCR itself has been proposed to act as a scaffold for the binding of signaling molecules. Assembly of these complexes involves noncovalent interaction between the intracellular domains of a GPCR and putative effector enzymes or adapters.



**Figure 4** Alternative mechanisms of G protein-coupled receptor signaling. Recent research has indicated that GPCRs use many intermediates in addition to G proteins (G) to signal intracellularly. Potential novel mediators of GPCR signals include *src* homology 2 (SH2) and *src* homology 3 (SH3) domain-containing proteins, G protein-coupled receptor kinases (GRKs), protein kinase A anchoring proteins (AKAP), arrestins (Arr), and PDZ domain-containing proteins.

Some GPCRs are substrates for tyrosine phosphorylation, which can lead to their association with proteins containing *src* homology 2 (SH2) domains, motifs that confer specific recognition of phosphotyrosine residues. The  $\beta$ 2-adrenergic receptor can be phosphorylated on tyrosine residues by the insulin receptor tyrosine kinase. This has been reported to lead to direct association of the receptor with the adapter proteins Grb2 and Shc, central elements in the control of the low-molecular-weight G protein, Ras. Stimulation of the JAK-STAT pathway of transcriptional regulation by angiotensin AT1a receptors apparently involves angiotensin-dependent tyrosine phosphorylation of Tyr319 in the AT1a receptor tail by an Src family kinase, followed by association of JAK2 with the receptor. In this case, the binding of JAK2, which does not have an SH2 domain, appears to be indirect but may be mediated by members of the SHP family of SH2 domain-containing tyrosine phosphatases.

Some GPCRs contain relatively large proline-rich inserts in their intracellular loops that may mediate binding to proteins containing *src* homology 3 (SH3) domains. For example, the  $\beta$ 3-adrenergic receptor, which does not recruit  $\beta$ -arrestin, binds to the c-Src SH3 domain in an agonist-dependent manner via proline-containing motifs in its i3 loop and carboxy-terminus. This interaction may contribute to the ability of  $\beta$ 3-adrenergic receptors to activate the ERK pathway.

Direct GPCR-adapter protein interactions may also regulate the activity of serine/threonine kinases. The PKA anchoring protein, AKAP79/150, coprecipitates with the  $\beta$ 2-adrenergic receptor from brain lysates and interacts with peptides derived from the i3 and carboxy-terminal tail of the receptor *in vitro*. This interaction, which does not appear to be agonist regulated, promotes PKA-mediated phosphorylation of the receptor following G protein-dependent activation of adenylate cyclase and facilitates downstream activation of MAP kinases.

### PDZ Domain Interactions

The intracellular tails of a small number of heptahelical receptors terminate in variations of the Thr/Ser-x-Val motif required for binding to PDZ domains. More than 50 PDZ domain-containing proteins have been cloned that selectively recognize variants of this motif, and a few GPCR-PDZ domain-binding interactions have been described. The  $\beta$ 2-adrenergic receptor has been shown to bind the  $\text{Na}^+/\text{H}^+$  exchanger regulatory factor (NHERF/EBP-50) in an agonist-dependent manner. This association, which

is apparently G protein independent, accounts for  $\beta$ 2-adrenergic receptor-mediated stimulation of renal  $\text{Na}^+/\text{H}^+$  exchange. Another PDZ domain-containing protein, PSD-95, binds to the carboxy-terminus of the  $\beta$ 1-adrenergic receptor. This interaction is regulated by receptor phosphorylation in that GRK5 phosphorylation of the receptor dramatically enhances binding. PSD-95 facilitates the physical association of  $\beta$ 1-adrenergic receptors with other synaptic proteins such as N-methyl-D-aspartate receptors and, thus, may provide the physical basis for another form of heterologous receptor cross-talk.

Rhodopsin is another GPCR that has been found to associate with a PDZ domain-containing protein in a functionally relevant manner. Rhodopsin binds to InaD, a multiple PDZ domain-containing scaffolding protein that associates with a number of signaling intermediates involved in rhodopsin-initiated pathways such as phospholipase C, protein kinase C, and the TRP ion channel. Analogous to the proposed function of PSD-95 in the synapse, the rhodopsin-InaD interaction may allow the formation of signaling complexes necessary for signal transduction in the visual system.

### GRKs and $\beta$ -Arrestins as Scaffolds

GRKs have been shown to phosphorylate protein substrates other than GPCRs. For example, GRK2, a  $\text{G}\beta\gamma$ -subunit-regulated kinase, associates with and phosphorylates tubulin. In this respect, GRK2 functions as G protein-regulated kinase. In addition, GRKs may act as noncatalytic adapters that recruit signaling intermediates into complex with GPCRs. GRKs have been shown to associate with actin and with a novel ARF6 GTPase-activating protein called GIT1. ARFs are small GTPases involved in vesicle trafficking and sorting, and both GIT1 and ARF6 are involved in GPCR internalization.

As discussed previously,  $\beta$ -arrestins bind tightly to agonist-occupied, GRK-phosphorylated GPCRs and act as adapters between the receptor and components of the cellular endocytic machinery.  $\beta$ -arrestins also serve to recruit regulatory enzymes to the GPCR complex. Endogenous  $\beta$ -arrestin is found in complex with ARNO, an ARF nucleotide exchange factor that, along with GIT-1, regulates ARF function. However, the role of  $\beta$ -arrestins as adapter proteins appears to extend well beyond their role in the regulation of receptor trafficking. Data obtained by a variety of techniques indicate that  $\beta$ -arrestins are able to bind to a number of proteins involved in signal transduction. By simultaneously binding to Src family tyrosine

kinases and to agonist-occupied  $\beta$ 2-adrenergic, neurokinin NK-1, chemokine CXCR-1, or endothelin ET1a receptors,  $\beta$ -arrestins have been shown to confer tyrosine kinase activity on the GPCR- $\beta$ -arrestin complex. The  $\beta$ -arrestin-Src interaction involves the SH3 domain and the amino-terminal portion of the catalytic domain of Src, which interact with epitopes in the amino-terminal half of  $\beta$ -arrestin molecule. The formation of  $\beta$ -arrestin-Src complexes has been implicated in several aspects of GPCR signaling, including  $\beta$ 2-adrenergic and NK-1 receptor-mediated activation of the ERK cascade;  $\beta$ 2-adrenergic receptor internalization, a process that is modulated by Src-dependent tyrosine phosphorylation of dynamin; CXCR-1 receptor-mediated neutrophil degranulation; and ET1a receptor-mediated translocation of the GLUT4 glucose transporter.

The MAP kinases are regulated via a series of parallel kinase cascades. Each MAP kinase pathway is composed of three kinases, each of which phosphorylates and activates the next kinase in succession. In many cases, MAP kinase activation is controlled by binding of the component kinases to a scaffolding protein. These scaffolds serve at least three functions in cells: to increase the efficiency of signaling between successive kinases in the phosphorylation cascade, to ensure signaling fidelity by dampening cross-talk between parallel MAP kinase cascades, and to target MAP kinases to specific subcellular locations. Recent data suggest that  $\beta$ -arrestins function as scaffolds for some MAP kinase modules. Overexpression of  $\beta$ -arrestin 1 or  $\beta$ -arrestin 2 increases the efficiency of angiotensin II AT1a receptor-mediated ERK activation by binding to each of the three kinases of the ERK cascade: Raf-1, MEK, and ERK. For both the AT1a receptor and the proteinase-activated PAR2 receptor, receptor activation triggers assembly of a multiprotein complex containing the internalized receptor,  $\beta$ -arrestin, Raf-1, and activated ERK1/2. Perhaps significantly, ERK activated in the context of a  $\beta$ -arrestin scaffold is targeted, along with the receptor, into endosomal vesicles rather than translocating to the nucleus, as is the case when it is activated via  $\beta$ -arrestin-independent mechanisms. Thus,  $\beta$ -arrestins may control both MAP kinase activity and MAP kinase function by targeting it to a specific subcellular location.

$\beta$ -arrestins may also be involved in scaffolding other MAP kinase pathways.  $\beta$ -arrestin 2 forms complexes with the components of the neuronal c-Jun N-terminal kinase (JNK) cascade: Ask1, MKK4, and JNK3. As with the ERK cascade, Ask1-dependent

JNK3 activation is dramatically enhanced by  $\beta$ -arrestin 2 overexpression, and in the presence of  $\beta$ -arrestin 2, activated JNK3 is retained in the cytosol following stimulation of AT1a receptors. Because JNK pathways are involved in apoptotic signaling, these data suggest that  $\beta$ -arrestin scaffolds may provide a mechanism for the GPCR regulation of programmed cell death.

## CONCLUSION

As the most diverse type of cell surface receptor, the importance of GPCR signaling to clinical medicine cannot be overestimated. Visual, olfactory, and gustatory sensation; intermediary metabolism; and cell growth and differentiation all are influenced by GPCR signals. The basic receptor-G protein-effector mechanism of GPCR signaling is tuned by a complex interplay of positive and negative modulatory events that amplify the effect of a hormone binding the receptor or that adjust the gain on cellular responsiveness. The association of heptahelical receptors with a variety of intracellular partners other than G proteins has led to the discovery of potential mechanisms of GPCR signaling that extend beyond the classical paradigms. Although the physiological relevance of many of these novel mechanisms of GPCR signaling remains to be established, their existence suggests that the mechanisms of GPCR signaling are even more diverse than was imagined previously.

## See Also the Following Articles

G Proteins and Effectors • Peptide Hormones and Growth Factors: Cellular Signaling Mechanisms

## Further Reading

- Berman, D. M., and Gilman, A. G. (1998). Mammalian RGS proteins: Barbarians at the gate. *J. Biol. Chem.* **273**, 1269-1272.
- Birnbaumer, L. (1990). Transduction of receptor signal into modulation of effector activity by G proteins: The first 20 years or so ... *EASEB J.* **4**, 3068-3078.
- Bouvier, M. (2001). Oligomerization of G-protein-coupled transmitter receptors. *Nature Rev. Neurosci.* **2**, 274-286.
- Brzostowski, J. A., and Kimmel, A. R. (2001). Signaling at zero G: G-protein-independent functions for 7-TM receptors. *Trends Biochem. Sci.* **26**, 291-297.
- Clapham, D. E., and Neer, E. J. (1993). New roles for G-protein  $\beta\gamma$ -dimers in transmembrane signaling. *Nature* **365**, 403-406.
- Downes, G. B., and Gautam, N. (1999). The G protein subunit gene families. *Genomics* **62**, 544-552.

- Ferguson, S. S. (2001). Evolving concepts in G protein-coupled receptor endocytosis: The role in receptor desensitization and signaling. *Pharmacol. Rev.* **53**, 1–24.
- Freedman, N. J., and Lefkowitz, R. J. (1996). Desensitization of G protein-coupled receptors. *Recent Progr. Horm. Res.* **51**, 319–351.
- Hall, R. A., Premont, R. T., and Lefkowitz, R. J. (1999). Heptahelical receptor signaling: Beyond the G protein paradigm. *J. Cell Biol.* **145**, 927–932.
- Horn, F., Weare, J., Beukers, M. W., Hörsch, S., Bairoch, A., Chen, W., Edvardsen, Ø., Campagne, F., and Vriend, G. (1998). GPCRDB: An information system for G protein-coupled receptors. *Nucleic Acid Res.* **26**, 277–281.
- Lefkowitz, R. J. (2000). The superfamily of heptahelical receptors. *Nature Cell Biol.* **2**, E133–E136.
- Luttrell, L. M., and Lefkowitz, R. J. (2002). The role of  $\beta$ -arrestins in the termination and transduction of G-protein-coupled receptor signals. *J. Cell Sci.* **115**, 455–465.
- Miller, W. E., and Lefkowitz, R. J. (2001). Expanding roles for beta-arrestins as scaffolds and adapters in GPCR signaling and trafficking. *Curr. Opin. Cell Biol.* **13**, 139–145.
- Sexton, P. M., Albiston, A., Morfis, M., and Tilakaratne, N. (2001). Receptor activity modifying proteins. *Cell Signalling* **13**, 73–83.
- Sprang, S. R. (1997). G protein mechanisms: Insights from structural analysis. *Annu. Rev. Biochem.* **66**, 639–678.
- Sunahara, R., Dessauer, C., and Gilman, A. G. (1996). Complexity and diversity of mammalian adenylyl cyclases. *Annu. Rev. Pharmacol. Toxicol.* **36**, 461–480.

it takes several minutes to be inactivated. The activation of the  $G_\alpha$  subunit is, however, usually reversed more quickly under physiological conditions, because its GTPase activity is greatly enhanced by the binding of a second protein, which can be either its target protein (e.g., phospholipase  $C\beta$ ) or a specific modulator known as a regulator of G protein signaling (RGS). RGS proteins act as  $\alpha$ -subunit-specific GTPase-activating proteins (GAP) and they have been shown to play a crucial role in shutting off G protein-mediated cell responses in eukaryotes.

Heterotrimeric G proteins are usually grouped into four subfamilies, e.g.,  $G_s$ ,  $G_i$ ,  $G_q$ , and  $G_{12}$ , on the basis of their amino acid sequences and functions (Table I).  $G_s$  (stimulatory G protein) was first defined functionally by its ability to activate adenylyl cyclases. GTP-bound  $G_{s\alpha}$  dissociates from the  $\beta\gamma$ -subunits and binds instead to adenylyl cyclase, which is then activated to produce cAMP.  $G_{s\alpha}$  is sensitive to cholera toxin, an enzyme that catalyzes the transfer of ADP ribose from intracellular  $NAD^+$  to  $\alpha_s$ . The ADP ribosylation alters the  $\alpha_s$  so that it can no longer hydrolyze its bound GTP, resulting in an indefinite activation. The prolonged elevation in cAMP levels within intestinal epithelial cells causes a large efflux of  $Na^+$  and water into the gut, which leads to severe diarrhea.

Another G protein, called inhibitory G protein ( $G_i$ ), is known to inhibit the adenylyl cyclases.  $G_i$  signaling not only decreases the enzyme activity of adenylyl cyclases, it also decreases the direct regulation of ion channels, such as cardiac  $K^+$  and neuronal  $Ca^{2+}$  channels. Characteristically,  $G_{\beta\gamma}$  subunits, not

$G_{i\alpha}$  subunits, play major roles in this regulation. Pertussis toxin, which is made by the bacterium that causes pertussis (whooping cough), catalyzes ADP ribosylation of the  $\alpha$ -subunit of  $G_i$ , preventing the subunit from interacting with receptors, which results in the continuous inactivation of the  $G_{i\alpha}$  subunit.

Some G protein-coupled receptors exert their effects through the G proteins,  $G_q$ , that activate the plasma membrane-bound enzyme, phospholipase  $C\beta$ . The phospholipase acts on a phosphoinositide, phosphatidylinositol 4,5-bisphosphate [ $PIP_2$ ]. The activated phospholipase  $C\beta$  cleaves  $PIP_2$  to generate two products, inositol 1,4,5-trisphosphate ( $IP_3$ ) and diacylglycerol.  $IP_3$  is a small, water-soluble molecule that diffuses rapidly in the cytosol. When  $IP_3$  reaches the endoplasmic reticulum (ER), it binds to the  $IP_3$  receptor on the ER membrane. The  $IP_3$  receptor is a  $Ca^{2+}$  channel, through which the  $Ca^{2+}$  stored in the ER is rapidly released on the binding of  $IP_3$ . On the other hand, diacylglycerol plays a role in the activation of a family of serine/threonine protein kinases, the protein kinase C (PKC) family.

The  $G_{12}$  subfamily has two members,  $\alpha_{12}$  and  $\alpha_{13}$ . These are thought to participate in cell transformation and embryonic development, but their signaling pathways have remained unknown for a long time. p115 RhoGEF, which is a GEF for the small GTPase Rho, has been shown to function as a RGS for  $G_{12}$  and  $G_{13}$ . It was also shown that the GEF activity of p115 is enhanced by  $G_{12}$  and  $G_{13}$ . These results may provide evidence for the novel concept that  $G_{12}$  and  $G_{13}$  signaling directly couples with

Table I The Family of G Protein  $\alpha$ -Subunits and Their Effectors

Subfamily		Toxin sensitivity	Tissue distribution	Target effectors
$G_s$	$\alpha_s$	Cholera toxin	Ubiquitous	Adenylyl cyclase $\uparrow$
	$\alpha_{olf}$	Cholera toxin	Olfactory cell	Adenylyl cyclase $\uparrow$
$G_i$	$\alpha_{i-1}$	Pertussis toxin	Ubiquitous	Adenylyl cyclase $\downarrow$
	$\alpha_{i-2}$	Pertussis toxin	Ubiquitous	
	$\alpha_{i-3}$	Pertussis toxin	Ubiquitous	
	$\alpha_o$	Pertussis toxin	Neuron	$K^+$ channel $\uparrow$ (N, P/Q type) $Ca^{2+}$ channel $\downarrow$ (via $\beta\gamma$ -subunits)
	$\alpha_{gust}$	Pertussis toxin	Taste cell	
	$\alpha_{t-1}$	Pertussis toxin/cholera toxin	Rod photoreceptor cell	cGMP phosphodiesterase $\uparrow$
	$\alpha_{t-2}$		cone photoreceptor cell	
$\alpha_z$	None	Neuron, adrenal gland, platelet	Adenylyl cyclase $\downarrow$	
$G_q$	$\alpha_q$	None	Ubiquitous	Phospholipase $C\beta$ $\uparrow$
	$\alpha_{11}$	None	Ubiquitous	Phospholipase $C\beta$ $\uparrow$
	$\alpha_{14}$	None	Lung, kidney, liver	Phospholipase $C\beta$ $\uparrow$
	$\alpha_{15}$	None	Lymphocyte	Phospholipase $C\beta$ $\uparrow$
$G_{12}$	$\alpha_{12}$	None	Ubiquitous	p115 RhoGEF $\uparrow$
	$\alpha_{13}$	None	Ubiquitous	p115 RhoGEF $\uparrow$

small (monomeric) G protein signaling via p115 RhoGEF.

## ADENYLYL CYCLASE

At least 10 isoforms of adenylyl cyclases, which are responsible for the synthesis of cAMP from ATP, have been identified in mammals. Each isoform consists of two hydrophobic domains, each of which comprises six transmembrane-spanning domains and two cytoplasmic catalytic domains, resulting in a pseudo-symmetrical structure. The regulatory properties and tissue distributions of the different adenylyl cyclases are shown in Table II. Almost all of the adenylyl cyclases can be activated directly by the  $G_{s\alpha}$  subunit and by forskolin, which is a toxin isolated from the root of a herb, whereas they are differentially modulated by proteins such as  $G_{i\alpha}$ ,  $G_{\beta\gamma}$ , protein kinases [cAMP-dependent protein kinase (PKA) or PKC], and  $Ca^{2+}$ /calmodulin (CaM).  $G_{s\alpha}$ -,  $G_{i\alpha}$ -,  $Ca^{2+}$ /CaM-, and forskolin-insensitive adenylyl cyclase (type X) is found only in testis.

Direct interaction of  $G_{i\alpha}$  with adenylyl cyclase is the most obvious mechanism for inhibition of cAMP synthesis.  $G_{s\alpha}$ - and forskolin-stimulated adenylyl cyclase activity of type V or VI is substantially (60–80%) inhibited by the  $G_{i\alpha_1}$ ,  $G_{i\alpha_2}$ ,  $G_{i\alpha_3}$ , and  $G_{z\alpha}$  subunits.

A G protein  $\beta\gamma$ -subunit complex is a very potent inhibitor of type I adenylyl cyclase. The  $\beta\gamma$ -subunits appear to associate directly with this enzyme. On the other hand, the  $\beta\gamma$  complex exhibits remarkable stimulatory effects on type II and IV adenylyl cyclases, although concurrent activation with  $G_{s\alpha}$  is a prerequisite for these stimulatory effects.

Adenylyl cyclase activity of type V and VI is directly inhibited by low concentrations ( $\sim 0.5 \mu M$ ) of calcium ions. According to these results, it has been proposed that the wave of intracellular calcium could lead to a rhythmic mobilization of the cAMP concentration. On the other hand, certain isoforms of adenylyl cyclase (types I, III, and VIII) are directly activated by  $Ca^{2+}$ /CaM. A 10 nM concentration of CaM is sufficient for activation of type I and type VIII adenylyl cyclase, whereas the effect of  $Ca^{2+}$ /CaM on type III is less obvious and micromolar concentrations of  $Ca^{2+}$ /CaM and concurrent stimulation of  $G_{s\alpha}$  are required. Type IX adenylyl cyclase is inhibited by calcium-dependent protein phosphatase (calcineurin).

Type V and type VI adenylyl cyclases are phosphorylated by PKA, which results in the reduction of the enzyme activity. Thus, this modulation is thought to be a negative feedback mechanism of adenylyl cyclase signaling. Stimulation of  $G_{q-}$  and phospholipase

C-coupled receptor leads to the mobilization of intracellular  $Ca^{2+}$ , the synthesis of diacylglycerol, and the activation of PKC. PKC phosphorylates and activates some isoforms (types II, IV, V, and VI) of adenylyl cyclases. In particular, type V enzyme is shown to be a direct substrate for phosphorylation *in vitro* by PKC $\alpha$  and PKC $\zeta$ .

As shown above, the regulatory mechanisms of the adenylyl cyclases are diverse. In addition to their ability to respond to  $G_{s\alpha}$ , the different isoforms can receive signals from a variety of sources, such as  $G_{i\alpha}$ ,  $G_{\beta\gamma}$ , protein kinases (PKA and PKC), protein phosphatase (calcineurin), and  $Ca^{2+}$ /CaM. Such diversity in the control of adenylyl cyclase activity may contribute to the significant variation in the production of cAMP in different cell types on stimulation of multiple receptor-mediated cell signal systems.

## PHOSPHOLIPASE C

Eleven distinct isoforms of phospholipase C (PLC), which can be grouped into four subfamilies (PLC $\beta$ , PLC $\gamma$ , PLC $\delta$ , and PLC $\epsilon$ ), have been identified in mammals. These enzymes catalyze the hydrolysis of PIP $_2$  to IP $_3$  and diacylglycerol in response to the stimulation of various kinds of membrane receptors. The  $\alpha$ -subunits ( $\alpha_q$ ,  $\alpha_{11}$ ,  $\alpha_{14}$ , and  $\alpha_{16}$ ) of all four members of the  $G_q$  subfamily activate PLC $\beta$  isozymes but not PLC $\gamma$ , PLC $\delta$ , or PLC $\epsilon$ . The four mammalian PLC $\beta$  isozymes differ in their tissue distribution as well as in their sensitivity to G proteins. Expression of PLC $\beta_2$  is restricted to hematopoietic cells and that of PLC $\beta_4$  is limited to the retina and neuronal cells, whereas PLC $\beta_1$  and PLC $\beta_3$  are expressed ubiquitously. G protein-coupled receptors that activate the  $G_q$ -PLC $\beta$  pathway include those for  $\alpha_1$ -adrenergic agonists, acetylcholine (muscarinic  $m_1$  or  $m_3$ ), endothelin-1, angiotensin II, histamine, and C-X-C chemokines.

With the exception of PLC $\beta_4$ , PLC $\beta$  isozymes are also activated by the  $G_{\beta\gamma}$  dimer. The relative sensitivity of PLC $\beta$  isozymes to  $G_{\beta\gamma}$  subunits differs from the sensitivity to the  $G_{q\alpha}$  subunit. PLC $\beta_2$  is the most sensitive to  $G_{\beta\gamma}$  and PLC $\beta_1$  is the least sensitive.

A characteristic feature of PLC $\beta_1$  is its GTPase-activating property for  $G_{q\alpha}$ . In the reconstituted system of  $m_1$  muscarinic receptor, G protein, and PLC $\beta_1$  in lipid vesicles, PLC $\beta_1$  increases the rate of GTP hydrolysis by  $G_{q\alpha}$  more than 10-fold. A fragment of the C-terminal region of PLC $\beta_1$  (amino acids 903–1042) has also been shown to exhibit GAP activity. In addition, the GTPase activity of  $G_{q\alpha}$  has been shown to be facilitated by a family of RGS proteins.

Table II Regulatory Properties and Tissue Distribution of Mammalian Adenylyl Cyclases

Isoforms	Modulators of enzyme activity				Protein kinases	Tissue distribution	Specific function	Chromosomal location
	G <sub>αs</sub>	G <sub>αi</sub>	G <sub>βγ</sub>	Calcium				
Type I	↑	↓	↓	↑ (CaM)		Brain (neuron), adrenal gland	Synaptic plasticity, learning, memory	7p12–p13
Type II	↑	→	↑ (When stimulated with G <sub>αq</sub> )		↑(PKC)	Brain, lung, skeletal muscle	Synaptic plasticity	5q15.3
Type III	↑	↓		↑ (CaM)		Brain, olfactory cell	Odorant stimulation	2q22–q24
Type IV	↑	↓	↑		↑ (PKC)	Brain		14q11.2
Type V	↑	↓	↓	↓ (Free Ca <sup>2+</sup> )	↑ (PKC) ↓ (PKA)	Heart, brain		3q13.2–q21
Type VI	↑	↓	↓	↓ (Free Ca <sup>2+</sup> )	↓ (PKA, PKC)	Ubiquitous	Cell proliferation	12q12–q13
Type VII	↑	↓	↑		↑(PKC)	Brain, platelet		16q12–q13
Type VIII	↑	↓		↑ (CaM)		Brain, lung	Synaptic plasticity	8q24.2
Type IX	↑	↓		↓ (Calcineurin)		Brain, skeletal muscle		16q13.3
Type X	→	→				Testis		1q24

## CGMP PHOSPHODIESTERASE

Vision begins when the chromophore 11-*cis*-retinal, covalently attached to the visual pigment, is isomerized to its all-*trans* form by a captured photon. Within a millisecond, rhodopsin (R) undergoes a conformational change into metarhodopsin (R\*), which is capable of activating the photoreceptor-specific G protein, transducin (G<sub>t</sub>) (G<sub>t</sub> belongs to the G<sub>i</sub> subfamily). R\* triggers the exchange of GDP to GTP on G<sub>αt</sub>, which leads to the dissociation of active G<sub>αt</sub>-GTP and G<sub>βγ</sub>. G<sub>αt</sub>-GTP stimulates the activity of its effector enzyme, cGMP phosphodiesterase (PDE). PDE is a heterotetramer consisting of two catalytic subunits (αβ in rods and αα in cones) and two identical regulatory γ-subunits (γγ), which serve as inhibitors of PDE activity in the resting state. Activation of cGMP PDE causes a reduction of the intracellular cGMP concentration, which leads to the closure of cGMP-gated cation channels located in the plasma membranes of photoreceptor cells. These channels are nonselective cation channels, through which Na<sup>+</sup> and Ca<sup>2+</sup> flow into the cell in the resting (dark) state and thus the closure of these channels causes a membrane hyperpolarization and decreased release of the neurotransmitter (glutamate) at the synaptic termini.

GTP hydrolysis on the G<sub>αt</sub> subunit is accelerated by a RGS protein, RGS9, which is predominantly expressed in photoreceptor cells. It has been shown that RGS9 constitutes a complex with the long splice variant G protein β-subunit type 5 via the G protein γ-subunit-like domain, located next to the RGS domain, of RGS9. RGS9 facilitates GTPase activity on the G<sub>αt</sub> subunit and thus accelerates the time course of recovery from the flash response approximately 45-fold.

## ION CHANNELS

Modulation of ion channels by G proteins can be direct (via physical coupling between G protein and channels) or indirect (via protein kinases). A direct regulation has been firmly established for the G protein-gated inwardly rectifying potassium channel and for the neuronal (N- and P/Q-type) voltage-dependent calcium channel. G<sub>βγ</sub> subunits, but not G<sub>α</sub> subunits, have been shown to play major roles in these examples of direct regulation.

G protein-gated inwardly rectifying K<sup>+</sup> (K<sub>G</sub>) channels are directly activated by the βγ-subunits released from pertussis toxin-sensitive G<sub>i</sub> proteins, hyperpolarize the membrane potential, and contribute to neurotransmitter (acetylcholine)-induced deceleration of the heart beat, formation of slow inhibitory

postsynaptic potentials in neurons, and inhibition of hormone release in endocrine cells.

K<sub>G</sub> channels are tetramers of inwardly rectifying K<sup>+</sup> (Kir) channel subunits, Kir3.x. The combination of Kir3.x subunits constituting each K<sub>G</sub> channel varies among tissues and cell types. The cardiac K<sub>G</sub> channel is composed of Kir3.1 and Kir3.4, whereas neuronal K<sub>G</sub> channels mainly exist as heterotetramers of Kir3.1 and Kir3.2. Each four-subunit channel possesses one G<sub>βγ</sub>-binding site within the N- and C-terminal regions. The binding of G<sub>βγ</sub> does not alter the fast open-close gating transition of the K<sub>G</sub> channel, but increases the number of functional K<sub>G</sub> channels via a mechanism that can be described by the Monod-Wyman and Monod-Changeux allosteric models.

Electrophysiological and pharmacological characterizations define multiple classes of voltage-dependent calcium channels, i.e., T, N, L, P/Q, and R types. Among these, the N and P/Q types, which are expressed predominantly in neurons, are inhibited by direct interaction with G<sub>βγ</sub> subunits under the effects of a variety of neurotransmitters, including opioids, γ-aminobutyric acid, and somatostatin. The N-type calcium channel is inhibited to a greater extent than the P/Q-type channel. Because both N- and P/Q-type calcium channels mediate neurotransmitter release from presynaptic membranes, their differential modulation by G proteins may suggest a unique contribution of these channels to neurotransmission. Voltage-dependent calcium channels, including N and P/Q types, consist of a pore-forming α<sub>1</sub>-subunit, which has four domains, each containing six putative transmembrane segments, and accessory β, γ, and α<sub>2</sub>-δ subunits. G<sub>βγ</sub> subunits have been shown to interact with cytoplasmic linker regions between domains I and II (I-II linker) of the α<sub>1</sub>-subunit.

## See Also the Following Articles

Adenylyl Cyclase • G Protein-Coupled Receptors • Ion Channels • Receptor-Regulated Phospholipases

## Further Reading

- Arshavsky, V. Y., Lamb, T. D., and Pugh, E. N., Jr. (2002). G proteins and phototransduction. *Annu. Rev. Physiol.* **64**, 153–187.
- Gilman, A. G. (1987). G proteins: Transducers of receptor-generated signals. *Annu. Rev. Biochem.* **56**, 615–649.
- Rhee, S. G. (2001). Regulation of phosphoinositide-specific phospholipase C. *Annu. Rev. Biochem.* **70**, 281–312.
- Sunahara, R. K., Dessauer, C. W., and Gilman, A. G. (1996). Complexity and diversity of mammalian adenylyl cyclases. *Annu. Rev. Pharmacol. Toxicol.* **36**, 461–480.
- Yamada, M., Inanobe, A., and Kurachi, Y. (1998). G protein regulation of potassium ion channels. *Pharmacol. Rev.* **50**, 723–760.

investigation strategies to understand disease states and to devise therapeutic interventions to enhance dysfunctional physiological inhibition.

## PHYSIOLOGY OF THE GABA SYSTEM

The inhibitory neurotransmitter GABA affects neuronal activity mainly by gating selective  $\text{Cl}^-$  channels associated with  $\text{GABA}_A$  receptors. GABA function is also mediated by metabotropic receptors ( $\text{GABA}_B$ ), which exert their action by indirectly activating  $\text{K}^+$  efflux from neurons through specific channels. The most significant advances in GABA neurophysiology derived from the increased understanding of the role of GABA in the regulation of synaptic functions of various interneuron subtypes in many areas of the neo- and limbic cortices. Although principal inhibitory neurons such as Purkinje cells exist, advances in the understanding of neuronal networks have derived from the study of the properties of GABAergic interneurons.

GABAergic interneurons in the hippocampus and neocortex are typically nonpyramidal in shape and include a wide variety of cellular morphologies (basket, bitufted, horizontal, Martinotti and chandelier cells, etc.). They are distinguished by their voltage- and ligand-gated channels, afferent input, and a complement of neuropeptides colocalized in GABAergic interneurons such as parvalbumin and other calcium-binding proteins. Also, interneurons selectively innervate certain target cells at well-defined cellular regions. For instance, the axo-axonic connections of GABAergic chandelier cells of the neocortex and hippocampus innervate the initial axon segments of pyramidal neurons. In many cortical regions, GABAergic basket cells form dense networks of terminals on the somata of principal pyramidal neurons. The specificity of the chandelier and basket cell targets on principal neurons suggests that GABAergic inhibition in the first case controls somatic integration and in the second case regulates the initiation of action potentials; both actions reflect a major modulatory mechanism of the target cell function.

At the other extreme of the functional interactions between GABAergic interneurons and principal neurons (usually glutamatergic) are some interneurons that terminate on the distal dendrites of principal cells, often at a distance from the somata of principal neurons. Thus, in many cases, the synaptic activation of distal dendrites may not influence the action potential threshold. The main role of this distal inhibitory innervation by GABAergic neurons may be the

modification of local dendritic integration. In the neocortex and cerebellum, interneurons often establish synaptic contact with other interneurons. When the axons of interneurons have such a broad target, they may affect the functional integration among populations of principal cells.

Networks formed by interneurons play a major role in generating rhythmic activity in a neuronal circuit. However, sometimes the details of the functional role of interneurons in a circuit cannot be readily defined because it adapts to changing variables of the circuit afferent and/or efferent functions. The difficulties of developing criteria for classification of interneurons are demonstrated by a study in 1998 by Parra *et al.* on the functional variability of interneurons expressed in the hippocampal CA1 region. This variability is not unique to the hippocampus and supports the view that rigidly definable classes of interneurons subserving specific functions may not be expressed in cortex and hippocampus. However, the possibility must not be ignored that interneuron interactions may be flexible and may change and adapt to contribute to a physiologically important neuronal plasticity.

## ELECTROPHYSIOLOGICAL PROPERTIES OF INTERNEURONS

The response properties that distinguish different interneuron subtypes are determined in large part by the expression of different ion channels in their membranes. A distinctive signature of many interneurons is a “fast spiking” firing pattern. The action potential duration of an interneuron often lasts one-half that of the target principal neuron timed at half-maximal amplitude. However, since this definition cannot be extended to every interneuron, a converse argument cannot be made. A burst firing pattern mode is frequent in interneurons and it may be indicative of the participation of a low-threshold  $\text{Ca}^{2+}$  spike in the initiation of the burst. Efforts to identify the ion channel complement operative in the initiation of burst firing of the interneurons include investigations using single cell reverse-transcriptase polymerase chain reaction techniques to link the structure of a specific ion channel to electrophysiological findings. Such studies have indicated that the fast spiking properties of basket cells in the dentate gyrus of the hippocampus may be due to the translation of mRNA encoding  $\text{Kv}3,1/\text{Kv}3,2$  vs  $\text{Kv}4,2/\text{Kv}4,3$  channel subunits.

The brevity of the interneuron excitatory postsynaptic potential (EPSP) duration often depends on the presence of AMPA receptors and the absence of

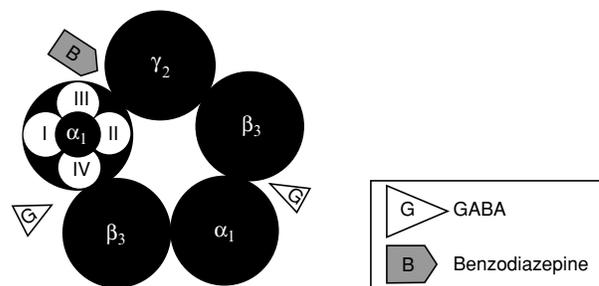
NMDA receptors in the synapses expressed on the interneuron membranes. When the interneuron postsynaptic densities express kainate receptors, the EPSP may last more than 100 ms. Temporal summation of long-lasting kainate-mediated EPSP may also be possible. Thus, an interneuron may act as a coincidence detector or as an integrator when glutamate secreted by the afferent terminals binds to AMPA kainate receptors.

In the early stages of CNS development, GABA often acts as a main excitatory transmitter. In adult neurons activation of GABA<sub>A</sub> receptors can stimulate biphasic responses; that is, it can depolarize (stimulate) or hyperpolarize (inhibit) neuronal membranes excitability. During the first week of postnatal life, spontaneous bicuculline-inhibited giant depolarizing potentials trigger an action potential in the neurons of the hippocampal CA3 region. This excitatory effect of GABA in immature neurons is due to a typical outward Cl<sup>-</sup> gradient; this action potential becomes inward after neuronal maturation. The switch from a depolarizing to a hyperpolarizing GABA<sub>A</sub> response during neuronal maturation is correlated with a change in ion transport at postsynaptic membranes from K<sup>+</sup> to Cl<sup>-</sup> transport.

At many central synapses during development, only one active zone for transmitter release is expressed in each nerve terminal and only one quantum of neurotransmitter is released by each impulse. In this case, the inhibitory postsynaptic current (IPSC) is short lasting. Multiquantal release may be established with neuronal maturation, leading to an increase in the duration of IPSC. In some hippocampal CA1 neurons, spontaneous quantal releases occur at low rates (~1 Hz); in others, rates of transmitter releases may be higher. A steady-state "tonic" inhibition may be caused by a summation due to a temporal overlap between two spontaneous events. When the summated background conductance is blocked by GABA<sub>A</sub> antagonists, increased cell excitability is usually observed. The effect of tonic inhibition may present as a subtle change in firing rates of the principal neuron. For instance, in the cerebellar Purkinje cell, tonic inhibition may elicit irregularities in Purkinje cell firing patterns.

## MOLECULAR ARCHITECTURE OF GABA<sub>A</sub> AND GABA<sub>B</sub> RECEPTORS

It is accepted that GABA is the major inhibitory transmitter in mammalian CNS. There are two important families of GABA receptors: GABA<sub>A</sub> and GABA<sub>B</sub>. The GABA<sub>A</sub> receptor (Fig. 1) is a ligand-gated Cl<sup>-</sup> ion



**Figure 1** A model of GABA<sub>A</sub> receptors showing localization of binding sites for GABA and benzodiazepine allosteric modulators. From Costa *et al.*, *Neuropharmacology* **43**, 925–937 (2002).

channel that functions as a polymorphic structure throughout the brain. This polymorphism is due to the expression of specific subunit subtypes (α<sub>6</sub>, β<sub>3</sub>, and γ<sub>3</sub>). These subunit subtypes have been cloned and form heteromeric (mainly pentameric) structures. Each GABA<sub>A</sub> receptor subunit includes four presumptive transmembrane domains. The GABA<sub>B</sub> metabotropic receptors consist of a single subunit with seven presumptive transmembrane domains (Fig. 1). The neurotransmitter binding pocket is buried within the bilayer. GABA<sub>B</sub> receptors are coupled to a G protein that regulates an enzymatic effector system (adenylyl cyclase, phospholipase C, or a K<sup>+</sup> channel).

## MODULATION OF GABA RECEPTORS BY NEUROSTEROIDS

In 1984, Harrison and Simmonds demonstrated that the steroidal anesthetic alpraxalone potently and selectively enhances the action of GABA at GABA<sub>A</sub> receptors. The rapidity of this modulatory action in single cell studies and the alpraxalone affinity in radioligand binding studies performed on membrane homogenates allow the exclusion of a traditional steroid action on genomic mechanisms. In fact, GABA<sub>A</sub> receptors harbor multiple binding sites for steroids that modulate GABA<sub>A</sub> receptor function nongenetically, very likely allosterically. An α-hydroxyl group at C3 and a ketone moiety at C20 of the steroid ring probably serve as points of attachment of neurosteroids within the primary binding pocket at the GABA<sub>A</sub> receptors, which function by donating and accepting hydrogen bonds, respectively. Additional important stabilizing influences include hydrophobic interactions between the steroid ring system and the receptor protein. Substitutions in β-orientation at C2 and C21 are well tolerated, whereas the effects of chemical modifications at C11 are dependent on a precise substitution. Steroid metabolism can be retarded by

substitution at the 3 $\beta$  position, which, in the case of phenyl ethylanyl derivatives, may also contribute to the potency of the steroidal modulation of GABA<sub>A</sub> receptors by contacting an auxiliary binding pocket.

### ANXIOLYTIC ACTION AND ALLOSTERIC MODULATION OF GABA<sub>A</sub> RECEPTORS BY BENZODIAZEPINES

GABA<sub>A</sub> receptors are molecular substrates for the regulation of vigilance, anxiety, muscle tension, epileptogenic activity, and the onset of anterograde amnesia, which is evident from the spectrum of actions elicited by clinically effective drugs acting at the benzodiazepine (BZ) binding sites expressed by GABA<sub>A</sub> receptors. These sites are located at the interfaces of two subunits ( $\alpha_1$ ,  $\alpha_2$ ,  $\alpha_3$ , or  $\alpha_5$  and  $\gamma_2$ ,  $\gamma_3$ ) (Fig. 1). The GABA<sub>A</sub> receptor opening frequency is enhanced by BZ receptor agonists; this enhancement is the basis for the BZ efficacy in the treatment of anxiety disorders. Among the BZs, diazepam interacts indiscriminately with every BZ-sensitive GABA<sub>A</sub> receptor subtype. Usually, the presence of  $\alpha_6$  or  $\gamma_1$  subunits in the GABA<sub>A</sub> receptors tends to obliterate the BZ-induced positive allosteric modulation of GABA at GABA<sub>A</sub> receptors. A conserved histidine residue is critical for the function of ligand binding at BZ receptors expressed in GABA<sub>A</sub> receptors. In fact, substitution of the histidine residue with arginine abates the BZ binding site function. Note that other compounds that do not have a BZ chemical structure can also act at BZ sites of GABA<sub>A</sub> receptors (e.g., zolpidem and various  $\beta$  carbolines) modulating these receptor positively or negatively (Table I). No endogenous positive or negative modulators of GABA<sub>A</sub> receptors have been identified. BZs are also prescribed for absence epilepsy (clonazepam) and status epilepticus (diazepam and lorazepam).

### ANTICONVULSANTS ACTING ON GABA SYSTEM

GABA<sub>A</sub> receptors are also an important target for anti-convulsants. There are a number of modulatory sites on GABA<sub>A</sub> receptors through which various drugs can act to potentiate or inhibit allosterically the actions of GABA. A number of these sites are being exploited to generate useful drugs in the treatment of convulsions, anxiety, sleep disorders, and psychoses.

Barbiturates have been used as anticonvulsants since the early 20th century. Phenobarbital is prescribed for both generalized and partial motor seizures. The primary site of action of barbiturates, such as phenobarbital, is the GABA<sub>A</sub> receptor. The barbiturates increase GABA<sub>A</sub> receptor open time. At low concentrations, barbiturates positively modulate the GABA<sub>A</sub> receptor response to GABA via an allosteric mechanism, increasing the hyperpolarization of the postsynaptic membranes. At higher concentrations, barbiturates are GABA-mimetic. In fact, they are able to activate GABA receptors in the absence of GABA. This ability reflects in part the poorer safety profile of these drugs compared with that of benzodiazepines, which are never agonists per se but allosterically modulate GABA<sub>A</sub> receptor responsiveness to GABA at a wide range of doses.

### GABAERGIC SYSTEM CHANGES IN DISEASE

It is widely accepted that pathological anxiety depends on a neurobiological and genetic underpinning. A crucial role has been delineated for the amygdala and its array of connections to higher cortical and subcortical areas, particularly the brainstem and hippocampus (both structures are operative in the acquisition and retention of conditioned fear in animals). These connections facilitate acquisition of sensory and interpretative information necessary to select fear responses according to context and allow the coordinated expression of the cognitive, affective, motor, and autonomic components of anxiety.

GABA<sub>A</sub> receptor deficits have been identified in patients with anxiety disorders and panic attacks; these deficits were localized in the hippocampus, parahippocampus, and orbitofrontal area. A GABA<sub>A</sub> receptor deficit has also been identified in generalized anxiety disorders. Whether GABA<sub>A</sub> receptor dysfunction is sufficient to induce anxiety still is debated. Since  $\gamma_2$  GABA<sub>A</sub> receptor subunits are required for GABA<sub>A</sub> receptor clustering in postsynaptic densities, mice heterozygous for this  $\gamma_2$  subunit show decreased binding of benzodiazepines in the hippocampal and parahippocampal area. These heterozygous mice were shown to have decreased binding for BZs. However, there is insufficient evidence that anxiety depends on a GABA<sub>A</sub> receptor abnormality.

Another disease in which the function of GABAergic neurons is related to GABA is schizophrenia. Documented in this disease is a GABAergic

neuron downregulation of  $GAD_{67}$  and reelin mRNA expression. Reelin is an important protein that during embryonic corticogenesis guides the layering in the cortex of neurons migrating via radial glia toward the cortical marginal zone from a ventricular proliferative center. It was formerly believed that reelin disappeared from the brain when the embryonic corticogenesis was terminated (i.e., approximately the sixth month of gestation in humans). However, it has been established that reelin persists throughout the life span and may play a crucial role in activating the translation of extrasomatic mRNA in the dendritic spines. Thus, it has been suggested that reelin plays a fundamental role in activating event-related protein synthesis leading to spine morphological changes that establish memory traces related to these events. It has also been found that in schizophrenia, there is an increase in the mRNA encoding for a DNA methylating enzyme in cortical GABAergic neurons. From this seminal finding, it can be speculated that GABAergic neurons may become an important target for new treatments of psychosis.

### See Also the Following Articles

Brain, Effects of Steroid Hormones • Catecholamines • Normetanephrine and Metanephrine • Somatostatin Analogs

### Further Reading

- Alger, B. E., and Nicoll, R. A. (1979). GABA-mediated biphasic inhibitory responses in hippocampus. *Nature* **281**, 315–317.
- Barnard, E. A. (2001). The molecular architecture of GABA<sub>A</sub> receptors. In “Pharmacology of GABA and Glycine Neurotransmission: Handbook of Experimental Pharmacology” (H. Möhler, ed.), Vol. 150, pp. 79–96. Springer, Berlin.
- Costa, E., Chen, Y., Davis, J., Dong, E., Noh, J. S., Tremolizzo, L., Veldic, M., Grayson, D. R., and Guidotti, A. (2002). Reelin and schizophrenia: A disease at the interface of the genome and the epigenome. *Mol. Interventions* **2**, 47–57.
- Dong, E., Caruncho, H., Liu, W. S., Smalheiser, N. R., Grayson, D. R., Costa, E., and Guidotti, A. (2003). The reelin–integrin interaction regulates Arc mRNA translation in synaptoneuroosomes. *Proc. Natl. Acad. Sci. USA* **18**, 235–240.
- Guidotti, A., Auta, J., Davis, J. M., Dwivedi, Y., Gerevini, V. D., Impagnatiello, F., Pandey, G., Pesold, C., Sharma, R., Uzunov, D. P., and Costa, E. (2000). Decrease in reelin and glutamic acid decarboxylase<sub>67</sub> ( $GAD_{67}$ ) expression in schizophrenia and bipolar disorder: A postmortem brain study. *Arch. Gen. Psychiatr.* **57**, 1061–1069.
- Harrison, N. L., and Simmonds, M. A. (1984). Modulation of the GABA<sub>A</sub> receptor complex by a steroid anesthetic. *Brain Res.* **323**, 287–292.
- Macdonald, R. L., and Olsen, R. W. (1994). GABA<sub>A</sub> receptor channels. *Annu. Rev. Neurosci.* **17**, 509–602.
- Möhler, H. (2001). Pharmacology of GABA and glycine neurotransmission. In “Pharmacology of GABA and Glycine Neurotransmission: Handbook of Experimental Pharmacology” (H. Möhler, ed.), Vol. 150, pp. 101–112. Springer, Berlin.
- Parra, P., Gulyas, A. I., and Miles, R. (1998). How many subtypes of inhibitory cells in the hippocampus? *Neuron* **20**, 983–993.

Key hormonal regulators of these processes are cholecystokinin (CCK) and secretin, both representing gastrointestinal peptide hormones. Both are produced in endocrine cells scattered among the mucosal cells lining the upper intestine. They are secreted in response to various components of chyme in the lumen of the intestine and they go on to act on mucosal epithelial cells and smooth muscle within the gallbladder and bile duct.

## REGULATION OF GALLBLADDER SECRETION AND CONTRACTILITY

The major physiologic stimulant of gallbladder emptying is CCK. This peptide hormone is secreted from I cells scattered throughout the duodenum and proximal two-thirds of the jejunum in response to the products of digestion of protein and fat components of a meal. The secreted hormone then binds to high-affinity type A CCK receptors present on gallbladder smooth muscle and neurons in the gallbladder and pancreas. These receptors are members of the Class I guanine nucleotide-binding protein (G protein)-coupled receptor family. Key physiologic actions of this hormone are the stimulation of gallbladder contraction and pancreatic exocrine secretion, both representing events critical to the intraluminal digestion of fats and protein. As such, this represents a logical and appropriate regulatory loop.

The gallbladder is predominantly an absorptive organ, in which water and electrolytes are absorbed in the process of concentrating the bile, considered to represent a constitutively active process. Little is known about the hormonal regulation of absorption and secretion at the level of the gallbladder mucosa.

## REGULATION OF BILIARY SECRETION

Secretin is a key physiologic stimulant of the secretion of bicarbonate-rich fluid from biliary epithelial cells. This regulatory loop is also beautifully designed. Secretin is secreted from S cells scattered within the mucosal layer of the duodenum. This hormone is secreted in response to acid in the lumen. It binds to receptors that are prototypical of the Class II family of G protein-coupled receptors, all of which bind moderately large peptide hormones having diffuse and complex pharmacophoric domains. By stimulating alkaline secretion from the pancreaticobiliary ductular tree, the resulting exocrine secretion works to

neutralize the luminal contents. This is critically important to provide a nondamaging milieu for bile acids and for digestive enzymes.

## IMPLICATIONS OF DYSFUNCTION OF THE GALLBLADDER AND BILIARY TREE

An intact excretory/secretory pathway that provides a conduit for bile and links the liver with the upper gastrointestinal tract is critical. Obstruction of this ductular structure leads to cholestasis, with damage to the liver and with retention of toxic compounds in the circulation. Even partial obstruction of the biliary tree has substantial clinical implications. Gallstones are the major clinical problem in these structures, often being formed in the gallbladder, where the bile is concentrated and where stasis can occur. Gallstones can be expelled from the gallbladder and can obstruct the bile or pancreatic duct, leading to a cholestatic pattern or pancreatitis or both. Tumors of the bile ducts and gallbladder can also occur.

Surgical removal of the gallbladder eliminates the reservoir and bile concentration function, but normally has relatively little clinical effect. Bile continues to enter the duodenum, without the advantage of titration and regulation provided by hormonal stimulation of gallbladder contraction. Although the bile acid concentration achieved in the intestinal lumen is lower than normal, it is adequate for most critical digestive functions.

There are no common recognized abnormalities of CCK or secretin secretion. Maldigestion leads to inadequate stimulation of CCK secretion that only exacerbates the digestive problem. In this setting, it is key to correct the underlying digestive problem.

## See Also the Following Articles

CCK (Cholecystokinin) • GI Tract, General Anatomy (Cells) • G Protein-Coupled Receptors

## Further Reading

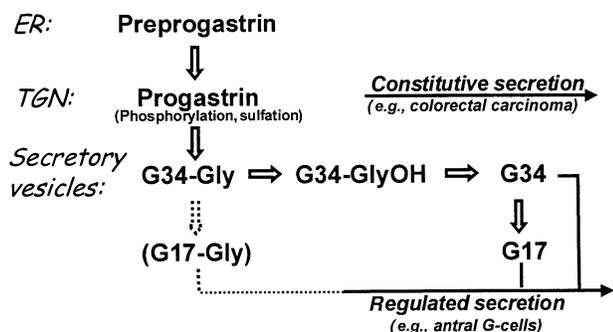
- Baiocchi, L., LeSage, G., Glaser, S., and Alpini, G. (1999). Regulation of cholangiocyte bile secretion. *J. Hepatol.* **31**, 179–191.
- Ding, X. Q., and Miller, L. J. (2001). Characterization of the type A cholecystokinin receptor hormone-binding domain: Use of contrasting and complementary methodologies. *Peptides* **22**, 1223–1228.
- Miller, L. J. (2003). Gastrointestinal hormones and receptors. In "Textbook of Gastroenterology" (T. Yamada, D. H. Alpers,

- L. Laine, C. Owyang, and D. W. Powell, eds.), 4th ed. pp. 48–77. Lippincott-Raven, Philadelphia, PA.
- Schjoldager, B., Molero, X., and Miller, L. J. (1989). Functional and biochemical characterization of the human gallbladder muscularis cholecystokinin receptor. *Gastroenterology* **96**, 1119–1125.
- Shaffer, E. A. (2000). Control of gall-bladder motor function. *Aliment. Pharmacol. Therap.* **14**(Suppl. 2), 2–8.
- Ulrich, C. D., Holtmann, M., and Miller, L. J. (1998). Secretin and vasoactive intestinal peptide receptors: Members of a unique family of G protein-coupled receptors. *Gastroenterology* **114**, 382–397.



## Gastric Inhibitory Polypeptide

see GIP

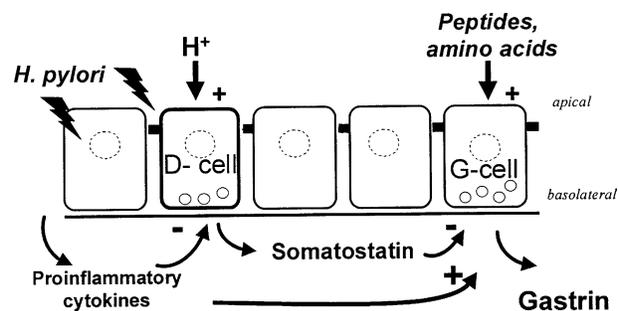


**Figure 1** Schematic representation of the biosynthesis of gastrin. Preprogastrin is translated in the endoplasmic reticulum (ER) and is converted there to progastrin, which translocates through the Golgi complex to secretory vesicles. In the trans-Golgi network (TGN), progastrin may be phosphorylated on Ser-96, sulfated on Tyr-86, or both. In secretory vesicles, it is cleaved to yield a COOH-terminal Gly peptide of 34 residues (G34-Gly), which may then be converted by COOH-terminal  $\alpha$ -amidation to G34 or cleaved to G17-Gly. In pyloric antral G cells, G17 and (to a lesser extent) G34 are the main products of regulated secretion. In some tumor cells, progastrin may be secreted by the constitutive pathway direct from the TGN to the cell surface.

in secretory vesicles, where it may be cleaved by enzymes of the prohormone convertase family. Biosynthetic intermediates with a COOH-terminal glycine residue (Gly-gastrins) provide substrates for the enzyme peptidyl  $\alpha$ -amino mono-oxygenase, which yields COOH-terminally amidated gastrins (Fig. 1). The amidated gastrins are typically peptides of 17 and 34 amino acid residues (G17 and G34), both of which occur as tyrosine-sulfated and unsulfated peptides. G34 corresponds to an NH<sub>2</sub>-terminal extended form of G17. Human gastrin (G) cells normally secrete mainly amidated gastrins, but small amounts of both progastrin and Gly-gastrins may also be released. The COOH-terminal pentapeptide amide of G17 (-Gly-Trp-Met-Asp-Phe-NH<sub>2</sub>) is identical to that in the related hormone cholecystokinin (CCK).

### Circulating Forms and Metabolism

The concentrations of amidated gastrins in the plasma of normal fasting humans are generally less than 30 pM. A mixed meal increases gastrin concentrations up to about threefold. The major circulating form of gastrin in humans is G34; however, this form is typically less than 10% of gastrin stored in G cells. In humans, it is cleared with a half-life of approximately 35 min, compared with approximately 7 min for G17. The difference in clearance accounts in part for the tendency of G34 to accumulate in plasma.



**Figure 2** Schematic representation of cellular relationships in the pyloric antral epithelium. The release of gastrin from G cells is stimulated by peptides and amino acids in the gastric lumen and by neurotransmitters (not shown). Acid releases somatostatin from D cells, and this inhibits gastrin release. In the presence of the gastric bacterium *Helicobacter pylori*, proinflammatory cytokines may inhibit D cells and stimulate G cells and so enhance gastrin release indirectly and directly.

### THE G CELL AND ITS CONTROL

G cells are found in the pyloric antral mucosa of all mammalian species. In humans, there are also some G cells in the duodenum. After a meal, gastrin is released in response to gastric luminal stimuli (mainly protein, peptides, and amino acids) and in response to nervous stimuli (Fig. 2). One candidate neurotransmitter regulating the G cell is gastrin-releasing peptide (GRP), which is a mammalian relative of the amphibian skin peptide bombesin. Gastric acid inhibits gastrin release, producing a negative feedback loop. Therefore, decreases in gastric acid secretion, either through disease (e.g., gastritis) or by drug treatment (e.g., proton pump inhibitors), tend to increase gastrin release. The effects of acid are mediated by release of somatostatin from D cells in the antrum, which then suppresses the G cell by a paracrine mechanism (Fig. 2). In patients with *Helicobacter pylori*, there is a tendency for increased gastrin release. This is attributable to proinflammatory cytokines that are thought to act partly by inhibiting D-cell function and partly by directly stimulating G cells (Fig. 2).

### GASTRIN RECEPTORS

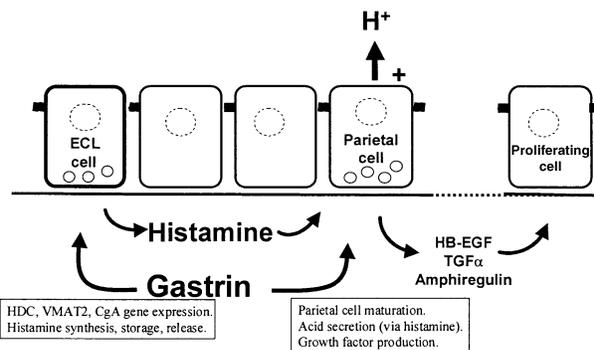
The effects of gastrin are exerted at gastrin-CCK<sub>B</sub> receptors (also known as CCK-2 receptors). These are found on gastric parietal and ECL cells, some smooth muscle cells, and some central nervous system (CNS) neurons. This receptor has high affinity for both gastrin and CCK; gastrin is the main ligand for receptors on parietal and ECL cells because its plasma concentrations are roughly 10 times higher than those of CCK. However, CCK is the main ligand for CNS

receptors because there is abundant CCK, but little or no gastrin, in the brain. The gastrin–CCK<sub>B</sub> receptor belongs to the family of G protein-coupled receptors, characterized by seven transmembrane domains. It acts via the G $\alpha_{q/11}$  group of G proteins to increase intracellular calcium concentrations and protein kinase C activity. Several antagonists have been generated, including YF476 and L-740,093.

## THE PHYSIOLOGY OF GASTRIN

### Acid Secretion

The primary action of gastrin is the stimulation of acid secretion. The question of whether gastrin acts primarily on the acid-producing parietal cells or indirectly via histamine release from ECL cells has been the subject of intense debate. It is now generally thought that gastrin acts mainly via histamine release from ECL cells (Fig. 3). Immunoneutralization of circulating gastrin, or administration of gastrin–CCK<sub>B</sub> receptor antagonists, inhibits gastric acid responses to food. In mice in which the gastrin or gastrin–CCK<sub>B</sub> receptor genes have been deleted by homologous recombination, there is a loss of acid secretion that is refractory to acute stimulation by gastrin, histamine, or cholinergic agonists. However,



**Figure 3** Schematic representation of some effects of gastrin in the gastric corpus. Gastrin stimulates histamine release from enterochromaffin-like (ECL) cells, and this in turn stimulates acid secretion from parietal cells. In ECL cells, gastrin also increases the production of histidine decarboxylase (HDC), which makes histamine from histidine; vesicular monoamine transporter type 2 (VMAT2), which sequesters histamine in secretory vesicles; and chromogranin A (CgA), which is co-stored with histamine. ECL cell numbers are regulated by gastrin. Parietal cells may also respond directly to gastrin. In addition to acute effects on acid secretion, gastrin controls the maturation of parietal cells and stimulates the production of growth factors, particularly members of the epidermal growth factor (EGF) family, including heparin-binding EGF, transforming growth factor (TGF)- $\alpha$ , and amphiregulin, which increase proliferation of other cells in the epithelium.

acid secretory capacity can be restored by infusion of gastrin, indicating a role for this hormone in maturation of the parietal cell as well as in acute secretory responses after a meal.

### ECL Cells

Gastrin increases the expression of several genes in ECL cells that are important for histamine synthesis and storage; these include histidine decarboxylase (HDC), which makes histamine from histidine; vesicular monoamine transporter type (VMAT2), which transports histamine into secretory vesicles; and chromogranin A, which is co-stored with histamine in secretory vesicles (Fig. 3). In addition, gastrin increases ECL cell number, probably by a direct effect on proliferation. In patients with prolonged increases in plasma gastrin, there is a tendency for ECL cell hyperplasia; in extreme cases, these patients may exhibit ECL cell carcinoid tumors that may resolve after removal of gastrin by antrectomy.

### Proliferation

Gastrin increases the proliferation of epithelial cells in the stomach. Apart from ECL cells, it is not clear whether proliferating cells express the gastrin–CCK<sub>B</sub> receptor. Therefore, proliferative responses are thought to be due to release of growth factors, including those of the epidermal growth factor (EGF) family (Fig. 3).

### Motility

Gastrin contracts smooth muscle and may regulate gastrointestinal motility, but the physiological significance of these effects is still uncertain.

### Central Nervous System

There are abundant gastrin–CCK<sub>B</sub> receptors on CNS neurons, but it is thought that the related peptide CCK is the main ligand for these receptors. There is little evidence to suggest that circulating gastrin released from antral G cells in mammals influences CNS function.

## GASTRIN IN GASTROINTESTINAL DISEASE

Elevated plasma gastrin concentrations occur in patients with gastrin-secreting tumors of the pancreas (Zollinger–Ellison syndrome, gastrinoma) and are

associated with increased acid secretion and intractable peptic ulcer. There are also high plasma gastrin concentrations in patients with pernicious anemia, which is attributable to autoimmune destruction of parietal cells and loss of acid inhibition of the G cell. In both groups of patients, there is a tendency for ECL cell carcinoid tumors. Moderately elevated plasma gastrin concentrations also occur in some patients infected with *H. pylori* (Fig. 2). When the infection is limited to the antrum, increased gastrin release may lead to increased acid secretion and duodenal ulcer.

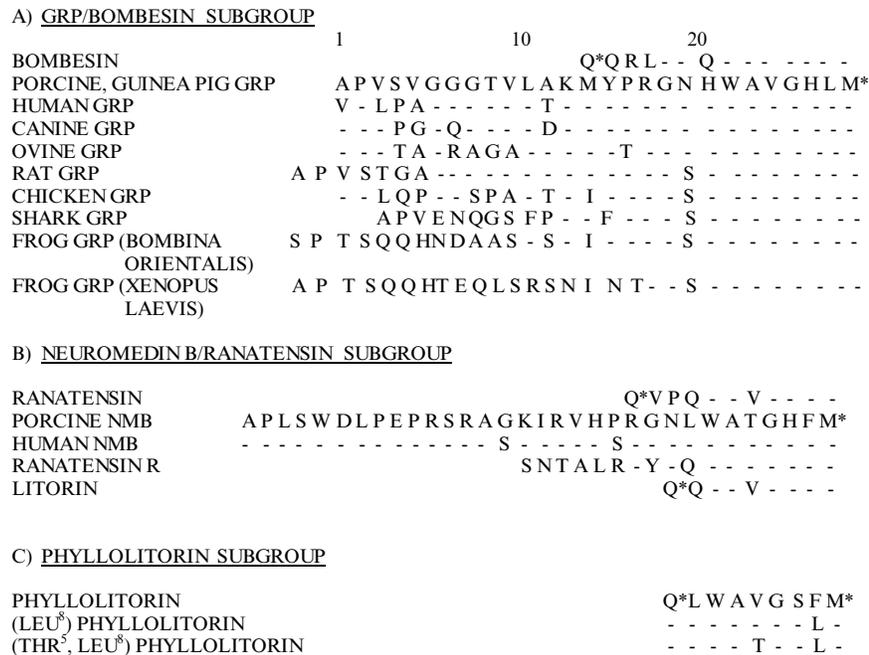
Recent work suggests that the gastrin gene is expressed in colorectal carcinoma, but the main products of expression are progastrin and Gly-gastrin (Fig. 1). These are thought to be growth factors for colon tumor cells. They do not act at gastrin-CCK<sub>B</sub> receptors, and the relevant receptor is still uncertain.

## See Also the Following Articles

CCK (Cholecystokinin) • ECL Cells • Gastrinomas • Gastrin-Releasing Peptide • GI Hormone Development (Families and Phylogeny) • GIP (Gastric Inhibitory Polypeptide)

## Further Reading

- Dockray, G. J. (1999). Topical review: Gastrin and gastric epithelial physiology. *J. Physiol. (Lond.)* **518**, 315–324.
- Dockray, G. J., Varro, A., Dimaline, R., and Wang, T. C. (2001). The gastrins: Their production and biological activities. *Ann. Rev. Physiol.* **63**, 119–139.
- Hinkle, K. L., and Samuelson, L. C. (1999). Lessons from genetically engineered animal models: III. Lessons learned from gastrin gene deletion in mice. *Am. J. Physiol.* **277**, G500–G505.
- Rozengurt, E., and Walsh, J. H. (2001). Gastrin, CCK, signaling, and cancer. *Annu. Rev. Physiol.* **63**, 49–76.
- Walsh, J. H. (1994). Gastrin. In "Gut Peptides" (J. H. Walsh and G. J. Dockray, eds.), pp. 75–121. Raven, New York.



**Figure 1** Bombesin-related peptide families. The amino acid sequences of the bombesin-related peptide family members are portrayed using the one-letter identification nomenclature (*Eur. J. Biochem.* 5, 151 [1968]). All members are aligned at their C-terminal residues. Amino acid residues identical to those in the sequence of porcine and guinea pig GRP are identified by a space (-). M\*, methionyl amide; Q\*, pyroglutamyl; NMB, neuromedin B.

multiple forms occur both in normal neural and endocrine tissue and in neoplastic cells, but the physiological significance of their presence is not fully understood.

## GRP/BOMBESIN FAMILY RECEPTOR SUBTYPES

Four distinct receptor subtypes interacting with members of the GRP/bombesin peptide family have been characterized. All are members of the seven-transmembrane G protein-linked receptor family, and each has approximately 50% structural homology with the others, acts through second messenger pathways and promotion of cellular calcium fluxes to produce the recognized bioactivities, and has distinct ligand-binding affinities and tissue distributions. The first subtype characterized was the GRP-preferring receptor (designated BB-2), which has a higher affinity for GRP than for bombesin. The second subtype (designated BB-1) has a greater affinity for neuromedin B than for GRP. The third receptor subtype (designated BB-3) was identified in the pregnant (but not the nonpregnant) guinea pig uterus and subsequently in the human pregnant uterus and other mammalian

tissues. The BB-3 receptor subtype is GRP rather than neuromedin B preferring but has a lower binding affinity for GRP than would be expected, suggesting that a natural ligand for this receptor has not yet been identified. Interestingly, synthetic peptide analogues of members of this family have been shown to have high affinity for BB-3. The fourth receptor subtype (designated BB-4), recently identified in frog brain, is a Phe-13 bombesin-recognizing receptor.

## LOCATIONS AND ACTIONS

The anatomical locations and recognized biological actions of GRP have been reviewed. In the mammal, GRP is principally confined to neural elements of the central, peripheral, and enteric nervous systems but is also found in neuroendocrine cells of peribronchial epithelial tissue, in certain neuroendocrine neoplastic cells (e.g., small cell lung cancer, medullary carcinoma of the thyroid, carcinoid tumors), and in certain epithelial-derived cancers.

Intracerebral administration of GRP to mammals results in potent and diverse biological effects, including the production of hypothermia and hypoglycemia, alteration of the sensation of satiety, and increased sympathetic tract outflow. Peripheral intravenous

administration of GRP results in a multitude of activities; it potently affects the release of a number of biologically active peptides (hormones and neuropeptides) and alters gastrointestinal smooth muscle activity and exocrine secretions of the pancreas and stomach. However, there is no single activity that has been demonstrated to be unequivocally physiological. There is evidence for a role for GRP in vagal mechanisms resulting in the stimulation of gastric acid secretion and gastrin release, although a physiological role for the latter activity has been questioned.

## OTHER DIRECTIONS OF STUDY

Having been first identified in neural tissues of the CNS and gastrointestinal tracts, most early studies were directed toward target tissues in these areas. A new direction of studies occurred with the finding that GRP was present in small cell carcinoma of the lung and that it might act as a trophic agent in these cells. These findings provided the impetus for other studies that demonstrated the presence of GRP and/or its receptor in a number of epithelial cell-derived and neural tissue-derived cancers and that GRP might act as a trophic agent in a number of these tissues. One study demonstrated the presence of GRP and its receptor in colon cancers. Both GRP and its receptor were often present in adenomas and well-differentiated colon cancers but not in normal colon epithelial cells or in poorly differentiated cancer cells, findings that are consistent with earlier observations made in small cell carcinoma of the lung. Of interest, although a number of studies have shown inhibition of tumor growth by blocking GRP's interaction with its receptor, certain studies have provided evidence that the presence of GRP and its receptors is associated with a survival advantage.

In a very different line of studies, GRP and GRP-like peptides have been demonstrated to be present and produced in the ovine pregnant uterus and ovine pregnancy fluids as well as in human maternal and fetal placental membranes, with rapid disappearance of the peptide postpartum. Given that a number of studies have demonstrated GRP's trophic effects on fetal tissues, GRP may have important but as yet unelucidated roles in developmental biology. Further clues to its potential importance in this regard are the early observations of GRP's presence in human fetal and pathological neonatal lung together with studies that demonstrate its trophic effects on lung. Early (and little noted) studies demonstrated that GRP is present not only in neoplastic adult lung tissues

but also in nonneoplastic pathological lung lesions, suggesting that a pathological process, but not necessarily a neoplastic process, will result in the reappearance of GRP. The findings that GRP has been seen in well differentiated but not in poorly differentiated neoplastic lesions, and that it has been seen in certain nonneoplastic but pathological tissues, lead to the suggestion that GRP has an undefined role during pathological processes that may be related to its roles in ontogeny. GRP may play an important role not solely as a mitogen but also as a differentiation agent in normal development, and such a regulatory role may be replayed in pathology. Hence, it has been suggested that GRP may play a role as a "morphogen" as well as a mitogen.

Because GRP appears to be present in certain neoplasms, a few studies have suggested that the presence of GRP in body fluids may act as an aid to the diagnosis of disease. In addition, given the presence of receptors for GRP on a number of neoplastic cells, it has been suggested that radiolabeled GRP and GRP analogues may be useful in the imaging of neoplasms.

Further studies into GRP's role as a developmental agent and its potential use as a diagnostic tool may prove to be fruitful.

## See Also the Following Articles

ECL Cells • Gastrin • Gastrinomas • GI Hormone Development (Families and Phylogeny) • GIP (Gastric Inhibitory Polypeptide) • G Protein-Coupled Receptors

## Further Reading

- Carroll, R. E., Matkowskyj, K. A., Cantave, I., Chakrabarti, S., McDonald, T. J., and Benya, R. V. (1999). Aberrant expression of gastrin-releasing peptide and its receptor by well-differentiated colon cancers in humans. *Am. J. Physiol.* **276**, G655–G665.
- Cullen, A., Emanuel, R. L., Torday, J. S., Asokanathan, N., Sikorski, K. A., and Sunday, M. E. (2000). Bombesin-like peptide and receptors in lung injury models: Diverse gene expression, similar function. *Peptides* **21**, 1627–1638.
- Cuttita, F., Carney, D. N., and Mulshine, J. L. (1985). Bombesin-like peptides can function as autocrine factors in human small cell lung cancer. *Nature* **316**, 823–826.
- Mantey, S. A., Coy, D. H., Pradhan, T. K., Igarashi, H., Rizo, I. M., Shen, L., Hou, W., Hocart, S. J., and Jensen, R. T. (2001). Rational design of a peptide agonist that interacts selectively with the orphan receptor, bombesin receptor subtype 3. *J. Biol. Chem.* **276**, 9219–9229.
- McDonald, T. J. (1988). The gastrin releasing polypeptide. In "Advanced Metabolic Disorders" (V. Mutt, ed.), Vol. 11, pp. 199–250. Academic Press, San Diego.

- McDonald, T. J., Jornvall, H., Nilsson, G., Vagne, M., Ghatei, M., Bloom, S. R., and Mutt, V. (1979). Characterization of a gastrin-releasing peptide from porcine non-antral gastric tissue. *Biochem. Biophys. Res. Commun.* **90**, 227-233.
- Spindel, E. R., Giladi, E., Segerson, T. P., and Nagalla, S. (1993). Bombesin-like peptides of ligands and receptors. *Rec. Prog. Horm. Res.* **48**, 365-391.
- Sunday, M. E. (1996). Pulmonary neuroendocrine cells and lung development. *Endocrinol. Patbol.* **7**, 173-201.
- Xiao, Q., Han, X., Challis, J. R. G., Hill, D. J., Spindel, E. R., Prasad, C. J., Akagi, K., and McDonald, T. J. (1996). Gastrin releasing peptide-like immunoreactivity is present in human maternal and fetal placental membranes. *J. Clin. Endocrinol. Metab.* **81**, 3766-3773.

Table I Clinical Features of Patients with ZES

Feature	NIH data (n = 261)	Literature data (range)
Gender (percentage male)	56	44–70
Mean age onset (years)	41	41–53
Initial symptom (percentage)		
Abdominal pain	75	26–98
Diarrhea	73	17–73
Heartburn	44	0–56
Nausea	30	8–37
Vomiting	25	26–51
Bleeding	24	8–75
Pain and bleeding	19	19–44
Pain and diarrhea	55	28–56
Mean duration of symptoms (years)	5.2	3.2–8.7
MEN1 present (percentage)	22	10–48
History-confirmed peptic ulcer (percentage)	71	71–93
Esophageal stricture (percentage)	4	4–6
Abdominal perforation (percentage)	5	5–18

Note. NIH data are from 261 patients with ZES. Literature data are from 10 series with 12 to 369 patients.

Gastrinomas were originally reported to be non-beta cell tumors of the pancreas, but currently approximately one-half are found in the duodenum, one-fourth in the pancreas, and the remainder in other sites (Table II). One controversial area is whether primary gastrinomas occur in abdominal lymph nodes. A long-term National Institutes of Health (NIH) study suggests that 10 to 11% of all primary gastrinomas may occur in lymph nodes. Other unusual intra-abdominal sites include the gastric antrum, biliary tract or liver, mesentery, ovary, and spleen. Extra-abdominal gastrinomas also occur rarely (<0.5%) in the cardiac intraventricular septum and due to a small cell lung cancer secreting gastrin. In the duodenum, 57% of the gastrinomas occur in the first portion and 31% occur in the second portion. Most patients have localized disease at presentation (83%) with no liver metastases, although 17% have or develop liver metastases. Approximately one-third of patients have a primary tumor only found at surgery, one-third have a primary tumor with lymph node metastases, one-fourth have a primary tumor with liver metastases, and one-fifth have either only liver metastases (3%) or only lymph node metastases (19%) (Table II).

Table II Characteristics of Gastrinomas (percentages)

Feature	NIH data	Literature data (range)
Location		
Pancreas	24	21–70
Duodenum	49	6–32
Lymph node	11	<1
Other	9	0–18
Unknown	16	7–48
Duodenal location		
D-1	57	ND
D-2	31	ND
D-3	6	ND
D-4	3	ND
Extent of disease		
No tumor found	13	7–48
Localized disease	70	23–51
Metastatic disease to liver	17	13–53
Extent metastases		
Primary only	36	23–51
Primary and lymph nodes	29	8–61
Primary and liver metastases	23	15–38
Liver metastases only	3	4–14
Lymph node metastases only	16	4–23

## MOLECULAR PATHOGENESIS OF GASTRINOMAS

Similar to other gastrointestinal neuroendocrine tumors (NETs) (e.g., carcinoids, pancreatic endocrine tumors [PETS]), abnormalities in common oncogenes (e.g., ras, c-Jun, 1-fos) are uncommon in gastrinomas, as are alterations in common tumor suppressor genes (e.g., retinoblastoma, p53). In contrast, studies suggest that alterations in the HER2/neu gene (a member of the Erb  $\beta$ -like oncogene family), the MEN1 gene on 11q13, and the p16<sup>INK4a</sup> gene on chromosome 9p21, as well as overexpression of epidermal growth factor and hepatocyte growth factor receptors (EGFRs and HGFRs, respectively), may be important in the molecular pathogenesis of gastrinomas and other PETS. Of these alterations, only overexpression of EGFRs has been shown to correlate with invasiveness.

## DIAGNOSIS/DIFFERENTIAL DIAGNOSIS

The diagnosis remains elusive, with a 4- to 7-year delay in diagnosis from onset of symptoms (Table I). This delay occurs because ZES is uncommon (1–3

new cases/1 million population/year) and its presentation closely mimics common disorders: idiopathic PUD (2300 cases/1 million population/year) and idiopathic gastroesophageal reflux disease (GERD), which occurs in 3 to 4% of the population. A recent study provides evidence that the widespread use of proton pump inhibitors (PPIs) is complicating and delaying the diagnosis of ZES further. This is occurring because PPIs are powerful acid suppressants that can control symptoms in patients with ZES, thereby masking the diagnosis. Furthermore, PPIs can cause hypergastrinemia in patients without ZES, and this can complicate the diagnosis of ZES in these patients. The diagnosis of ZES should be particularly suspected in patients with PUD with diarrhea, endocrinopathies, family history of PUD or endocrinopathies, and/or recalcitrant PUD disease as well as in patients with PUD without *Helicobacter pylori* or with PUD complications. Diagnosis is made by measuring fasting gastrin levels after stopping all antisecretory drugs (e.g., PPIs  $\times$  1 week, histamine H<sub>2</sub> antagonists  $\times$  1 day), which are elevated in 90 to 98% of patients. If the fasting gastrin level is elevated, gastric fluid pH should be assessed and the fasting gastrin should be repeated. If the gastric pH is more than 2.0 and the fasting serum gastrin is elevated, it is unlikely that ZES is present. If the fasting gastrin is elevated more than 10-fold with a gastric pH of less than 2.1, ZES is confirmed when there is no previous history of gastric resection. Unfortunately, two-thirds of patients have fasting gastrin levels increased less than 10-fold (101–999 pg/ml), and a number of other conditions can mimic ZES in this range (e.g., *H. pylori* infection, gastric outlet obstruction, antral G-cell hyperplasia/hyperfunction, chronic renal failure, short bowel syndrome). These can be excluded by both measuring a basal gastric secretory rate for 1 h and performing a secretin test. The secretin test yields no false positives except in patients with achlorhydria, and it is positive in 87% of patients with ZES in this fasting gastrin range.

## TUMOR LOCALIZATION

Localizing the primary tumor, as well as determining the extent of the tumor, is essential for all management steps. It is necessary to determine whether surgical exploration should be undertaken, the extent of possible resection, and the location of the primary tumor (to enhance the possibility of cure), as well as to assess changes in tumor size with patients with advanced metastatic disease, either in response to antitumor treatment or in determining the need

for it. Somatostatin receptor scintigraphy (SRS) with [<sup>111</sup>In-DTPA-DPhe<sup>1</sup>]octreotide (Figs. 1 and 2) and single photon emission computed tomography (SPECT) is now the initial imaging study of choice due to its greater sensitivity and ability to image all body areas at one time. SRS is more sensitive than all conventional imaging modalities combined (e.g., computed tomography [CT], magnetic resonance imaging

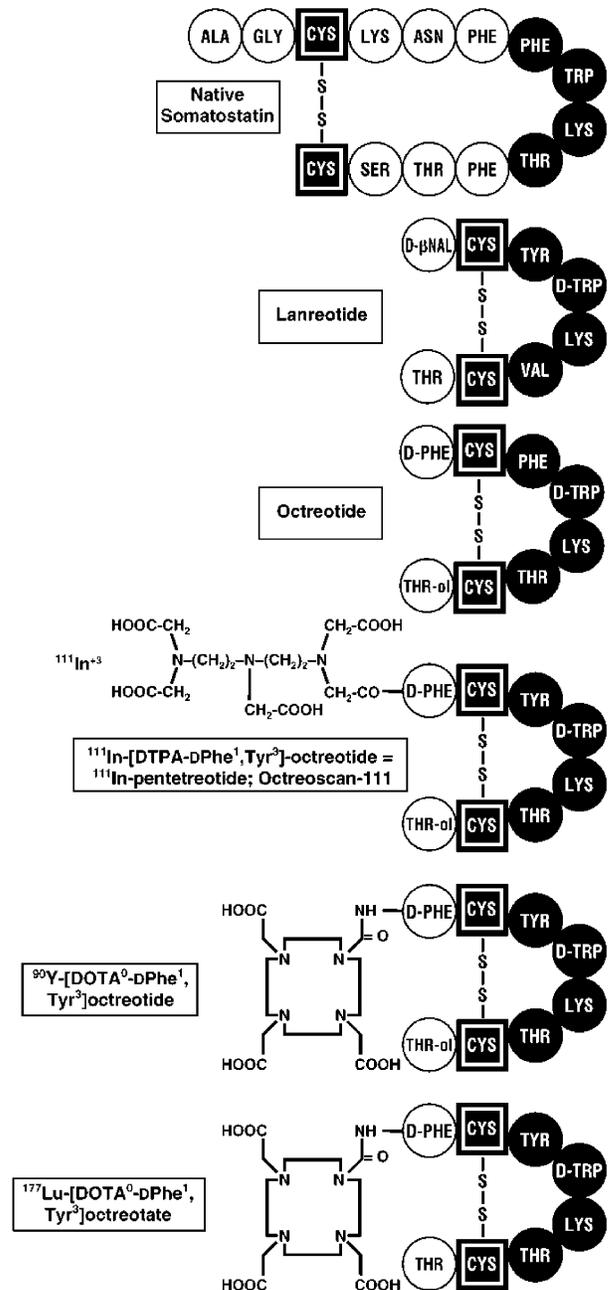
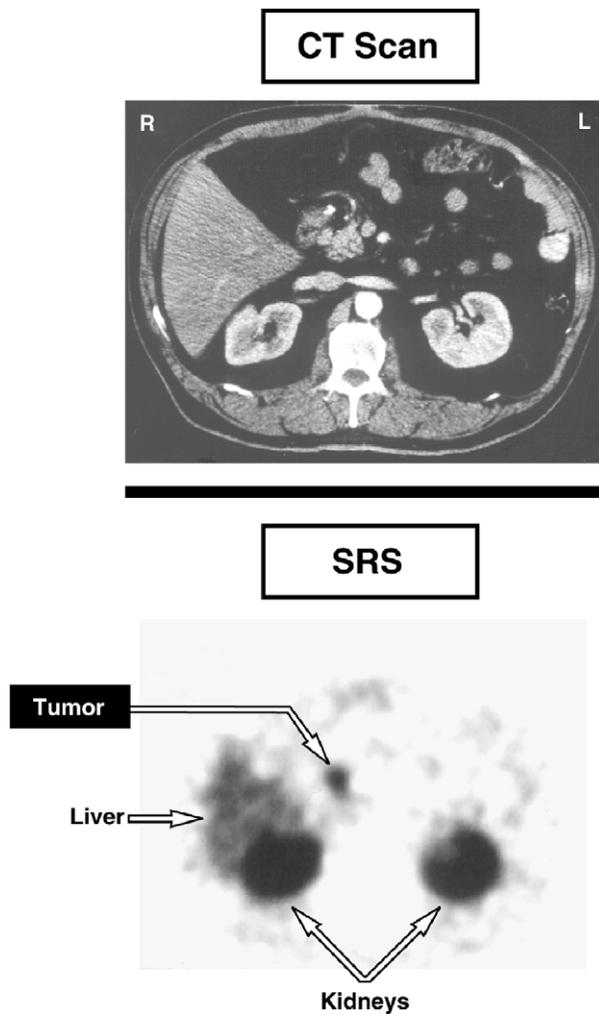


Figure 1 Structure of somatostatin and synthetic analogues used for tumor localization or radiotherapy.



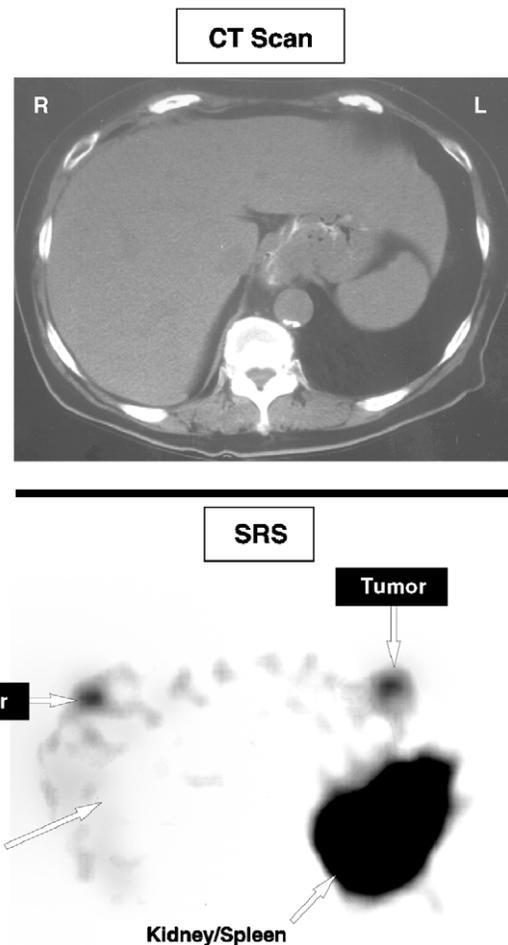
**Figure 2** Comparison of the ability of CT scan and SRS to localize a primary gastrinoma in a patient with ZES. Only the SRS is positive, showing its greater sensitivity.

[MRI], ultrasound), but it does not give reliable information on tumor size (Figs. 2 and 3). Therefore, a CT scan with contrast is also recommended. Endoscopic ultrasound is recommended by some but rarely localizes duodenal tumors; therefore, it is helpful in a minority of the patients. For patients with negative SRS, functional/localization studies measuring gastrin gradients with angiography are recommended.

## MANAGEMENT

### General

Treatment needs to be directed at controlling the gastric acid hypersecretion initially and long term as well as directed at the gastrinoma per se. It is also



**Figure 3** Comparison of the ability of CT scan and SRS to localize metastatic disease in the liver in a patient with ZES. Only the SRS localizes a metastasis in the left lobe, showing its greater sensitivity.

important to determine whether the patient has ZES due to the presence of MEN1 because this requires additional treatment strategies. MEN1 is generally recognized by a careful family and personal history for endocrinopathies (especially hyperparathyroidism, pituitary disease, and other PETs) and assessment for these endocrinopathies by appropriate hormonal assays. The hyperparathyroidism is usually present prior to ZES onset and is best diagnosed by measuring serum ionized calcium levels and plasma-intact PTH (Table III).

### Control of Acid Hypersecretion

Many patients with ZES have basal acid secretory rates elevated more than fivefold and can rapidly develop complications due to the gastric acid hypersecretion if

**Table III Management of ZES**

Treatment of gastric acid hypersecretion
Drugs of choice: PPIs (e.g., omeprazole, lansoprazole, pantoprazole, rebeprazole, esomeprazole)
PPIs control acid secretion >99%
Histamine H <sub>2</sub> receptor antagonist less effective (frequent high doses needed)
Total gastrectomy (needed only in rare patient who cannot take oral drugs)
Treatment of gastrinoma
In patients without MEN1 (75%) (sporadic ZES)
Surgery cures 60% in short term (<1 year); 30–35% at 5 years postresection
At surgery, routine duodenotomy and intraoperative ultrasound needed, with tumor enucleation used most frequently
Aggressive resection (Whipple procedure) rarely needed
In patients with MEN1/ZES (25%)
Tumor enucleation rarely (<1%) curative given that 60–80% of gastrinomas are duodenal in location, multiple, and/or with lymph node metastases (50–70%)
Surgery generally recommended if tumor is imaged ≥2–3 cm

Note. PPIs, proton pump inhibitors.

it is not controlled both acutely and for the long term. PPIs (e.g., omeprazole, lansoprazole, pantoprazole, rabeprazole, esomeprazole) are now the mainstay of treatment, starting at a dose equivalent to 60 mg per day of omeprazole. Patients with MEN1, severe GERD, or previous gastric acid-reducing surgery should be started on this dose given twice a day. With time, the dose can be decreased in most patients. During periods when parenteral antisecretory drugs are needed, intravenous PPIs (intravenous pantoprazole in the United States) are now the drugs of choice. Oral or intravenous histamine H<sub>2</sub> receptors can be used, but high frequent oral dosing is needed and continuous intravenous infusions are recommended. Total gastrectomy is rarely used today and is reserved for the rare patient who cannot take oral medication (Table III).

## Treatment of the Gastrinoma

### **Surgical**

In all patients with sporadic ZES (without MEN1), without liver or distant metastatic disease, and without concomitant illnesses that shorten life expectancy or increase the risk of surgery, exploratory laparotomy by a surgeon experienced in the treatment of this disease should be performed. Enucleation of pancreatic head/

body tumors should be performed. If needed, distal pancreatectomy for pancreatic tail lesions should be performed. Whipple resection is generally not recommended. In 5 to 15% of patients with advanced disease with liver metastases, cytoreductive surgery should be considered if more than 90% of the visible tumor can be removed.

In patients with MEN1/ZES, the type of surgery and when it should be performed are controversial because these patients cannot be cured generally but still have an excellent long-term prognosis. We recommend exploration when any lesion more than 2 cm in diameter is imaged because numerous studies demonstrate that tumor size is a strong prognostic factor for metastatic spread to the liver, which is associated with decreased survival. Lymph node metastases are frequent, even with small duodenal gastrinomas, but have not been shown to be associated with decreased survival.

### **Medical**

Patients with metastatic gastrinoma in the liver have a decreased survival rate. Studies show that survival is especially decreased in the 40% of patients with a rapidly growing metastatic tumor. In this subset of patients, antitumor treatment is indicated. Long-acting somatostatin analogues, such as octreotide-LAR (Fig. 1) (given once monthly) alone or in combination with  $\alpha$ -interferon (given three times/week), are now the recommended initial treatments. These agents cause a decrease in tumor size in only 15 to 20% of patients; however, they cause a cessation of growth in 50 to 70% of patients and continue to remain effective for years in many patients. Chemotherapy (e.g., streptozotocin, doxorubicin  $\pm$  5-fluorouracil) causes a decrease in tumor size in 5 to 50% of patients but has significant toxicity and has not been shown to extend life in patients with gastrinomas.

## CONCLUSION

Gastrinomas are an uncommon cause of PUD, but if diagnosed, the gastric acid hypersecretion can be controlled in every patient. The natural history of the gastrinoma is becoming an increasingly important determinant of long-term survival, so the typical 5-year delay in diagnosis remains a major challenge to instituting treatment early in the disease course in many patients. Furthermore, studies suggest that the widespread use of PPIs may further delay the diagnosis of ZES.

## See Also the Following Articles

Gastrin • Gastrin-Releasing Peptide • GI Hormones in Cancer • GI Tract, General Pathology of Endocrine Growths

## Further Reading

- Corleto, V. D., Annibale, B., Gibril, F., Angeletti, S., Serrano, J., Venzon, D. J., Delle Fave, G., and Jensen, R. T. (2001). Does the widespread use of proton pump inhibitors mask, complicate, and/or delay the diagnosis of Zollinger–Ellison syndrome? *Alimentary Pharmacol. Therapeut.* **15**, 1555–1561.
- Corleto, V. D., Delle Fave, G., and Jensen, R. T. (2002). Molecular insights into gastrointestinal neuroendocrine tumors: Importance and recent advances. *Digest. Liver Dis.* **34**, 668–680.
- DelValle, J., and Scheiman, J. M. (1999). Zollinger–Ellison syndrome. In “Textbook of Gastroenterology” (T. Yamada, D. H. Alpers, L. Laine, C. Owyang, and D. W. Powell, eds.), 3rd ed., pp. 1445–1462. Lippincott Williams & Wilkins, Philadelphia.
- Jensen, R. T. (1996). Gastrinoma. *Bailliere’s Clin. Gastroenterol.* **10**, 555–766.
- Jensen, R. T. (1998). Management of the Zollinger–Ellison syndrome in patients with multiple endocrine neoplasia type 1. *J. Int. Med.* **243**, 477–488.
- Jensen, R. T. (1999). Use of omeprazole and other proton pump inhibitors in the Zollinger–Ellison syndrome. In “Milestones in Drug Therapy” (L. Olbe, ed.), pp. 205–221. Birkhauser-Verlag, Basel, Switzerland.
- Jensen, R. T. (2001). Zollinger–Ellison syndrome. In “Surgical Endocrinology: Clinical Syndromes” (G. M. Doherty and B. Skogseid, eds.), pp. 291–344. Lippincott Williams & Wilkins, Philadelphia.
- Norton, J. A., Alexander, H. A., Fraker, D. L., Venzon, D. J., Gibril, F., and Jensen, R. T. (2003). Possible primary lymph node gastrinomas: Occurrence, natural history, and predictive factors—A prospective study. *Ann. Surg.* **237**, 650–659.
- Norton, J. A., Fraker, D. L., Alexander, H. R., Venzon, D. J., Doppman, J. L., Serrano, J., Goebel, S. U., Peghini, P., Roy, P. K., Gibril, F., and Jensen, R. T. (1999). Surgery to cure the Zollinger–Ellison syndrome. *N. Engl. J. Med.* **341**, 635–644.
- Roy, P., Venzon, D. J., Shojamanesh, H., Abou-Saif, A., Peghini, P., Doppman, J. L., Gibril, F., and Jensen, R. T. (2000). Zollinger–Ellison syndrome: Clinical presentation in 261 patients. *Medicine* **79**, 379–411.
- Shojamanesh, H., Gibril, F., Louie, A., Ojeaburu, J. V., Bashir, S., Abou-Saif, A., and Jensen, R. T. (2002). Prospective study of the anti-tumor efficacy of long-term octreotide treatment in patients with progressive metastatic gastrinomas. *Cancer* **94**, 331–343.

**Table I** General Characteristics of Gastrointestinal Neuroendocrine Tumors (NETs) (Pancreatic Endocrine Tumors and Carcinoids)

Share general neuroendocrine cell markers
Chromogranins (Cgs) (A, B, C) are acidic monomeric-soluble proteins found in the large secretory granules. CgA is most widely used.
Neuron-specific enolase is the $\gamma$ - $\gamma$ dimer of the enzyme enolase and is a cytosolic marker of neuroendocrine differentiation.
Synaptophysin is an integral membrane glycoprotein (38 kDa) found in small vesicles of neurons and neuroendocrine tumors.
Pathologic similarities
All NETs show amine precursor uptake and decarboxylation (APUDomas).
Ultrastructurally, they have dense-core secretory granules (> 80 nm).
Histologically, they appear similar with few mitoses and uniform nuclei.
Frequently, they synthesize multiple peptides/amines, which can be detected immunocytochemically but may not be secreted.
The presence or absence of clinical syndrome or type cannot be predicted by immunocytochemical studies.
Histological classifications do not predict biologic behavior. Only invasion or metastases establish malignancy.
Similarities of biologic behavior
They are generally slow growing, but a proportion are aggressive.
They secrete biologically active peptides/amines, which can cause clinical symptoms.
They generally have high densities of somatostatin receptors, which are used for both localization and treatment.

distinguish between benign and malignant NETs unless metastases or invasion are present.

PETs are classified according to the specific clinical syndrome they cause or, if no clinical syndrome occurs, as nonfunctional (Table II). Nonfunctional PET in the strict sense is a misnomer because these tumors frequently secrete multiple peptides (pancreatic polypeptide, CgA, and neurotensin); however, they do not cause a specific clinical syndrome. The symptoms caused by nonfunctional PETs are due to the tumor per se.

Carcinoid tumors are frequently classified by location of origin (foregut, midgut, or hindgut) because they share histochemical, functional, and biologic characteristics within these areas (Table III). Foregut carcinoids generally have low serotonin (5-HT) content, are argentaffin negative on silver staining, synthesize several hormones, and occasionally secrete ACTH or 5-hydroxytryptophane (5-HTP) causing an atypical carcinoid syndrome. Midgut carcinoids are generally argentaffin positive, have high serotonin content, frequently multihormonal, rarely secrete 5-HTP or ACTH, and most frequently cause carcinoid syndrome when they metastasize (Table III). Hindgut carcinoids (transverse colon to rectum) are argentaffin negative; rarely contain serotonin, 5-HTP, and ACTH or cause carcinoid syndrome; contain numerous peptides; and frequently metastasize to bone. Most carcinoids (70%) occur in one of four sites—bronchus, jejunum/ileum, rectum, or appendix—although they can occur in any tissue (Table III). Overall, the GI tract is the most common site for carcinoids (74% of

cases), and the respiratory tract the second most common site (25% of cases).

The frequency of NETs varies according to whether they are symptomatic. The incidence of clinically significant PETs is 10 cases/1 million population/year. Their relative incidence is insulinoma, gastrinoma, and nonfunctional (0.5–2.5 cases/1 million/year) > VIPomas (2–8 times less common) > glucagonomas (17–30 times less common) > somatostatinomas. In autopsy studies, 0.5–1.5% of all cases have PETs, with less than 1 in 1000 symptomatic. Clinically significant carcinoids occur in 7–13 cases/1 million/year and malignant cases at autopsy in 21–84 cases/1 million/year.

Both PETs and carcinoids can show malignant behavior. With PETs, 50–100% are malignant, except for insulinomas, in which < 10% are malignant (Table II). With carcinoid tumors the percentage that are malignant varies with different locations (Table III). A number of factors are predictive of tumor aggressiveness and/or survival for NETs (Table IV). The presence of liver metastases is the most important prognostic factor. Primary tumor size is also an important predictor. The factors in Table IV need to be considered when determining the aggressiveness of treatment of NETs.

Until recently, the molecular pathogenesis of NETs was largely unknown because common oncogenes (*ras*, *fos*, etc.) and common tumor suppressor genes (retinoblastoma gene, *p53*) were usually not altered. Recent studies provide evidence that alteration in the *HER2/neu* oncogene, *MEN-1* gene, and *p16INK4a*

Table II Gastrointestinal Neuroendocrine Tumor Syndromes

Syndrome	Biologically active peptide(s) secreted	Incidence (new cases/10 <sup>6</sup> population/year)	Tumor location	Malignant (%)	Main symptoms/signs
<b>Established specific functional syndrome</b>					
<b>Pancreatic endocrine tumor</b>					
Zollinger–Ellison syndrome (gastrinoma)	Gastrin	0.5–1.5	Duodenum (70%) Pancreas (25%) Other sites (5%)	60–90	Pain (79–100%) Diarrhea (30–75%) Esophageal symptoms (31–56%)
Insulinoma	Insulin	1–2	Pancreas (> 99%)	< 10	Hypoglycemic symptoms (100%)
VIPoma (Verner–Morrison syndrome, pancreatic cholera, WDHA)	Vasoactive intestinal peptide	0.05–0.2	Pancreas (90%; adult) Other (10%; neural, adrenal, periganglionic)	40–70	Diarrhea (90–100%) Hypokalemic (80–100%) Dehydration (83%)
Glucagonoma	Glucagon	0.01–0.1	Pancreas (100%)	50–80	Rash (67–90%) Glucose intolerance (38–87%) Weight loss (66–96%)
Somatostatinoma	Somatostatin	Rare	Pancreas (55%) Duodenum/jejunum (44%)	> 70	Diabetes mellitus (63–90%) Cholelithiasis (65–90%) Diarrhea (35–90%)
GRFoma	Growth hormone-releasing hormone	Unknown	Pancreas (30%) Lung (54%) Jejunum (7%) Other (13%)	> 60	Acromegaly (100%)
ACTHoma	ACTH	Rare	Pancreas (4–16% all ectopic Cushing's)	> 95	Cushing's syndrome (100%)
PET <sup>a</sup> causing carcinoid syndrome	Serotonin ? Tachykinins	Rare (43 cases)	Pancreas (< 1% all carcinoids)	60–80	Same as carcinoid syndrome above
PET causing hypercalcemia	PTHrP Others unknown	Rare	Pancreas (rare cause of hypercalcemia)	84	Abdominal pain due to hepatic metastases
<b>Carcinoid tumor</b>					
Carcinoid syndrome	Serotonin, possibly tachykinins, motilin, prostaglandins	0.5–2	Midgut (75–87%) Foregut (2–33%) Hindgut (1–8%) Unknown (2–15%)	95–100	Diarrhea (32–84%) Flushing (63–75%) Pain (10–34%) Asthma (4–18%) Heart disease (11–41%)
<b>Possible specific functional PET syndrome</b>					
PET secreting calcitonin	Calcitonin	Rare	Pancreas (rare cause of hypercalcitonemia)	> 80%	Diarrhea (50%)
PET secreting renin	Renin	Rare	Pancreas	Unknown	Hypertension
<b>No functional syndrome</b>					
PPoma/nonfunctional	None	1–2	Pancreas (100%)	> 60	Weight loss (30–90%) Abdominal mass (10–30%) Pain (30–95%)

<sup>a</sup>PET, pancreatic endocrine tumor.

Table III Carcinoid Tumor Location, Frequency of Metastases, and Association with Carcinoid Syndrome

	Location (% of total) <sup>a</sup>	Incidence of metastases <sup>a</sup>	Incidence of carcinoid syndrome <sup>b</sup>
<b>Foregut</b>			
Esophagus	< 1	67	—
Stomach	3.8	31	9.5
Duodenum	2.1	—	3.4
Pancreas	< 1	76	20
Gallbladder	< 1	56	5
Bronchus, lung	32.5	27	13
<b>Midgut</b>			
Jejunum	2.3	70	9
Ileum	17.6	70	9
Meckel's diverticulum	0.4	—	13
Appendix	7.6	35	< 1
Colon	6.3	71	5
Liver	< 1	29	—
Ovary	< 1	32	50
Testis	< 1	—	50
Cervix	< 1	67	3
<b>Hindgut</b>			
Rectum	10	14	—

<sup>a</sup>From 5468 cases studied from 1973 to 1991 reported by I. M. Modlin *et al.* (1997). *Cancer* 79, 813.

<sup>b</sup>From 4349 cases studied from 1950 to 1971 reported by J. D. Godwin (1975). *Cancer* 36, 560.

and overexpression of growth factors (epidermal growth factor, vascular endothelial growth factor, insulin-like growth factor-1, and transforming growth factor-2) and their receptors may be important in the pathogenesis (Table IV). An increased understanding of their molecular pathogenesis may lead to newer treatments as well as have prognostic significance.

## SPECIFIC PANCREATIC ENDOCRINE TUMOR SYNDROMES

Functional PETs almost always present with symptoms due to the specific hormone ectopically secreted. Non-functional PETs present with symptoms due to the tumor per se and usually present late in the disease course with advanced metastatic disease. Treatment of functional PETs requires strategies directed at controlling the hormone-excess state as well as at the PETs per se because they are frequently malignant. Unfortunately, these tumors are not curable by surgery due to the extent of the disease; therefore, both treatment aspects cannot adequately be dealt with by curative resection.

Table IV Molecular Abnormalities and Prognostic Factors in NETs<sup>a</sup>

Molecular abnormalities	
PETs	
<i>HER2/neu (erbB-2)</i> expressed in 100% GAS	
<i>MEN1</i> gene LOH, 46%; mutation, 42% GAS/10% INS	
p16 <sup>INK4a</sup> , 30% gene abnormality	
TGF- $\alpha$ /EGFR, 72% IHC+	
IGF-1/IGFR-1, 70% PCR+	
Carcinoids	
<i>MEN-1</i> gene LOH, 36%; mutation, 10%	
p16 <sup>INK4a</sup> , 33% gene abnormality	
TGF- $\alpha$ /EGFR, 85% IHC+	
IGF-1/IGFR-1, 40%–80% IHC+	
VEGF, 79% IHC+	
Both carcinoids and PETs	
Alterations in common oncogenes ( <i>ras</i> , <i>fos</i> , etc.) or common tumor suppressor (p53, Rb) uncommon	
<b>Prognostic factors</b>	
PETs	
<i>Ha-Ras</i> oncogene or p53 overexpression	
Female gender	
MEN-1 syndrome absent	
Laboratory findings (increased chromogranin A in some studies; gastrinomas, increased gastrin level)	
Carcinoid tumors	
The presence of carcinoid syndrome	
Laboratory results: urinary 5-HIAA level ( $p < 0.01$ ), plasma neuropeptide K ( $p < 0.05$ ), serum chromogranin A ( $p < 0.01$ )	
The presence of a second malignancy	
Male gender ( $p < 0.001$ )	
Older age ( $p < 0.01$ )	
Mode of discovery (incidental > symptomatic)	
Both carcinoid tumors and PETs	
The presence of liver metastases ( $p < 0.001$ )	
Extent of liver metastases ( $p < 0.001$ )	
The presence of lymph node metastases ( $p < 0.001$ )	
Depth of invasion ( $p < 0.001$ )	
Primary tumor site ( $p < 0.001$ )	
Primary tumor size ( $p < 0.005$ )	
Various histologic features	
Tumor differentiation ( $p < 0.001$ )	
High growth indices (high Ki-67 index, PCNA expression)	
High mitotic counts ( $p < 0.001$ )	
Vascular or perineural invasion	
Flow cytometric features (i.e., aneuploidy)	

<sup>a</sup>Abbreviations used: TGF- $\alpha$ , transforming growth factor- $\alpha$ ; EGFR, receptor for epidermal growth factor; IGF/IGFR, insulin-like growth factor and its receptor; IHC, immunohistochemistry; MEN-1, multiple endocrine neoplasia type 1; PCR, polymerase chain reaction; PET, pancreatic endocrine tumor; VEGF, vascular endothelial growth factor; MEN, multiple endocrine neoplasia; NET, neuroendocrine tumor; PCNA, proliferating cell nuclear antigen; Ki-67, proliferation-associated nuclear antigen recognized by Ki-67 monoclonal antibody.

## Zollinger–Ellison Syndrome

Zollinger–Ellison syndrome (ZES) is only briefly discussed here because it is included in a separate article in this encyclopedia.

### Clinical Features

ZES is a clinical syndrome caused by ectopic secretion of gastrin by a PET (gastrinoma) that results in gastric hypersecretion. The acid hypersecretion characteristically causes peptic ulcer disease (often severe and refractory), diarrhea, or esophageal reflux disease (Table II).

### Pathology/Pathogenesis

Almost all the presenting symptoms are due to the effects of gastric hypersecretion because they disappear when the acid secretion is treated. Sixty to 90% of gastrinomas are malignant. Most gastrinomas (50–70%) are found in the duodenum, followed by the pancreas (20–40%). Other intraabdominal sites include mesentery, lymph nodes, and ovary. Approximately 20–25% of ZES patients have multiple endocrine neoplasia type 1 (MEN-1).

### Treatment

Acid hypersecretion can be controlled medically in almost every patient. Proton pump inhibitors (PPIs) are the drugs of choice because of their long duration of action. Histamine H<sub>2</sub> receptor antagonists are also effective but frequent (every 4–6 h) high doses are usually needed. Long-term surgical cure is possible in 30% of patients with ZES without MEN-1 but is rare (< 1%) in patients with ZES with MEN-1.

## Insulinomas

Insulinomas are neuroendocrine tumors of the pancreas thought to be derived from the  $\beta$  cells of the islets that ectopically secrete insulin, which results in hypoglycemia.

### Clinical Features

The most common symptoms are due to the effects of hypoglycemia on the central nervous system (neuroglycemic symptoms), including confusion, headache, altered consciousness, and even seizure (Table V). Many patients also have symptoms due to excess catecholamine release, including sweating or tremulousness. Attacks characteristically occur during fasting.

### Pathology/Pathophysiology

Almost all insulinomas (> 98%) occur in the pancreas. Insulinomas are usually small (40% < 1 cm), solitary,

**Table V** Frequency of Clinical Symptoms, Signs, and Laboratory Findings in Patients with Insulinoma

Clinical symptoms signs	Frequency (%)
<b>Occurrence anytime during clinical course</b>	
Neuropsychiatric (loss of consciousness, confusion, dizziness, diplopia)	92
Confusion or abnormal behavior	80
Obesity	52
Amnesia or coma	47
Convulsions (grand mal)	12
Cardiovascular symptoms, palpitations, tachycardia	17
<b>Occurrence during first attack</b>	
Neuroglycopenic	
Visual disturbances (diplopia, blurred vision)	59
Confusion	51
Altered consciousness	38
Weakness	32
Transient motor defects, hemiplegia	29
Dizziness	28
Fatigue	27
Inappropriate behavior	27
Speech difficulty	24
Headache	23
Seizure	23
Syncope	21
Adrenergic	
Sweating	43
Tremulousness	23
Hunger, nausea	12
Palpitations	10

and distributed evenly throughout the pancreas. In adults with hyperinsulinism with islet disease, 86% have an insulinoma, 5–15% have adenomatosis, nesio-blastosis occurs in 4%, and hyperplasia occurs in 1%. Insulin is normally synthesized and stored in islet  $\beta$  cells. It is derived from proinsulin, which consists of a 21-amino acid  $\alpha$  chain and a 30-amino acid  $\beta$  chain connected by a 33-amino acid connecting peptide (C-peptide). In normal subjects, < 25% of serum insulin is proinsulin, whereas > 90% of patients with an insulinoma have an elevated serum proinsulin.

### Diagnosis/Differential Diagnosis

Whipple's triad, published in 1933, was long used as the diagnostic criteria for insulinoma and consists of the occurrence of hypoglycemic symptoms, hypoglycemia (blood glucose < 50 mg/dl), and relief of symptoms following glucose injection. Unfortunately, this

**Table VI Causes of Spontaneous Hypoglycemia**

<b>Postprandial (reactive) hypoglycemia</b>	
Functional: recognizable anatomic lesion	
Alimentary hyperinsulinoma, usually secondary to previous gastric surgery such as Billroth gastrectomy	
Secondary to mild diabetes	
Idiopathic	
Due to specific hepatic enzyme deficiencies	
Hereditary fructose intolerance (infants, children)	
Galactosemia (infants, children)	
Familial fructose and galactose intolerance (rare)	
<b>Fasting hypoglycemia</b>	
Organic hyperinsulinism: specific anatomic lesion present	
Pancreatic islet disease	
Insulinoma, single or multiple	
Microadenomatosis with or without macroscopic adenomas	
Carcinoma	
Hyperplasia	
Nesidioblastosis	
Nonpancreatic tumors	
Severe congestive heart failure	
Severe renal insufficiency in non-insulin-dependent diabetes	
Due to hepatic enzyme deficiencies or decreased hepatic glucose output (primarily in infants, children)	
Glucogen storage diseases	
Glucogen synthetase deficiencies	
Other enzyme deficiencies (fructose-1,6-diphosphate deficiencies)	
Endocrine hypofunction	
Anterior pituitary (usually in infants, children)	
Adrenocortical (Addison's disease)	
Diffuse acquired liver disease	
Ethanol	
Severe malnutrition, sepsis	
Due to exogenous agents (factitious)	
Sulfonylureas, biguanides	
Insulin administration	
Ingestion of ackee fruits (hypoglycine)	
Other drugs (aspirin, pentamidine)	
Functional hypoglycemia with no persistent anatomical defect	
Autoantibodies to insulin receptor	
Spontaneous autoimmune anti-insulin antibody syndrome	
Transient hypoglycemia on infancy	

triad is not specific for insulinoma (Table VI). The diagnosis requires the establishment of hypoglycemia (< 40 mg/dl) at the time of an elevated fasting plasma insulin level. The usual method of diagnosing insulinoma is to perform a fast of up to 72 h, with assessment of serum glucose, insulin, and C-peptide every 4–8 h. If at any time the patient becomes symptomatic, the test should be terminated by giving

glucose after obtaining the previously mentioned serum measurements. During the first 24 h, 70–80% of patients with insulinomas develop symptoms, and 98% of patients develop symptoms by 48 h. In non-obese normal subjects, blood insulin values should decrease to < 6  $\mu$ M/ml if blood glucose decreases to  $\leq$  40 mg/dl and the ratio of insulin to glucose (mg/dl) is < 0.3. Some investigators also require elevated serum C-peptide and proinsulin levels and/or insulin:glucose ratio > 0.3 for an insulinoma diagnosis. There are numerous causes of hypoglycemia besides insulinoma, including pancreatic islet disease and factitious use of insulin or hypoglycemic agents (Table VI). Combinations of proinsulin levels, C-peptide levels, assessment of antibodies to insulin, and sulfonylurea levels may be required to make the correct diagnosis.

### Treatment

Only 5–15% of insulinomas are malignant; therefore, after tumor localization studies, surgery should be performed. Between 75 and 95% of patients are cured by surgery. Prior to surgery, hypoglycemia is usually controlled by frequent, small meals and the use of diazoxide (150–800 mg/day). Diazoxide controls symptoms in 50–60% of patients. Other drugs used to control the hypoglycemia include verapamil and diphenylhydantoin. Long-acting somatostatin analogues such as octreotide are effective in 40–50% of patients. However, in some patients octreotide may worsen the hypoglycemia. The preferred tumor localization studies include computed tomography (CT) scan to exclude liver metastases and endoscopic ultrasound. Somatostatin receptor scintigraphy (SRS), which is the preferred imaging modality in all other NETs, is frequently negative (40–70%) in patients

**Table VII Clinical Characteristics of Patients with VIPoma Syndrome**

	Frequency (%)
<b>Symptoms/signs</b>	
Secretory diarrhea	89–100
Dehydration	44–100
Weight loss	36–100
Abdominal cramps, colic	10–63
Flushing	14–28
<b>Laboratory findings</b>	
Hypokalemia	67–100
Hypochlorhydria	34–72
Hypercalcemia	41–50
Hyperglycemia	18–50

with insulinomas due to low numbers of somatostatin receptors.

### VIPomas

VIPomas are neuroendocrine tumors that occur primarily in the pancreas, secrete excessive amounts of vasoactive intestinal peptide (VIP), and cause a distinct syndrome characterized by large-volume diarrhea, hypokalemia, and dehydration (Tables II and VII). This syndrome is also called the Verner–Morrison syndrome, pancreatic cholera, or WDHA syndrome (watery diarrhea, hypokalemia, achlorhydria; Table II).

#### Clinical Features

The mean age of occurrence is 49 years; however, the syndrome can occur in children. The principal symptoms are large-volume diarrhea leading to dehydration, acidosis, and hypokalemia. Other symptoms/signs due to the actions of VIP include flushing, hypochlorhydria, hyperglycemia, and hypercalcemia. Steatorrhea is uncommon (16%). The diarrhea is characteristically secretory in nature (persists during fasting).

#### Pathology/Pathophysiology

VIP is a 28-amino acid neuropeptide normally present in the GI tract and central nervous system. It causes stimulation of small intestinal chloride secretion, has effects on smooth muscle contractility, inhibits acid secretion, and has vasodilatory effects, which explain many of the symptoms of the VIPoma syndrome. In adults, 80–90% of VIPomas are pancreatic, whereas in children the syndrome is usually caused by gangliogliomas/ganglioblastomas. VIPomas are usually solitary, with 50–75% in the pancreatic tail, and 40–70% of patients have liver metastases at diagnosis.

#### Diagnosis/Differential Diagnosis

The diagnosis requires the demonstration of an elevated plasma VIP level and the presence of large-volume diarrhea. Almost all patients have a stool volume > 1 liter/day (most have a volume > 3 liters/day), and it is proposed that the diagnosis be excluded if the volume is not > 700 ml/day. A number of other causes of large-volume diarrhea can be excluded by fasting the patient. Other diseases that can cause fasting large volume diarrhea include Zollinger–Ellison syndrome, chronic laxative abuse, systemic mastocytosis, diabetic diarrhea, AIDS, carcinoid syndrome, and, rarely, medullary thyroid cancer.

#### Treatment

The most important first step in treatment is to control the diarrhea and correct the dehydration, hypokalemia, and acidosis. Patients may require > 5 liters/day of fluid and > 350 mm/day of potassium when the diarrhea is not controlled. Octreotide will control the diarrhea in 85–90% of VIPoma patients. In non-responders or patients who have become refractory, combinations of octreotide and glucocorticoids may be helpful. Other drugs that may help in individual patients include clonidine, indomethacin, prednisone, lithium, phenothiazine, loperamide, lidamide, propranolol, and metoclopramide.

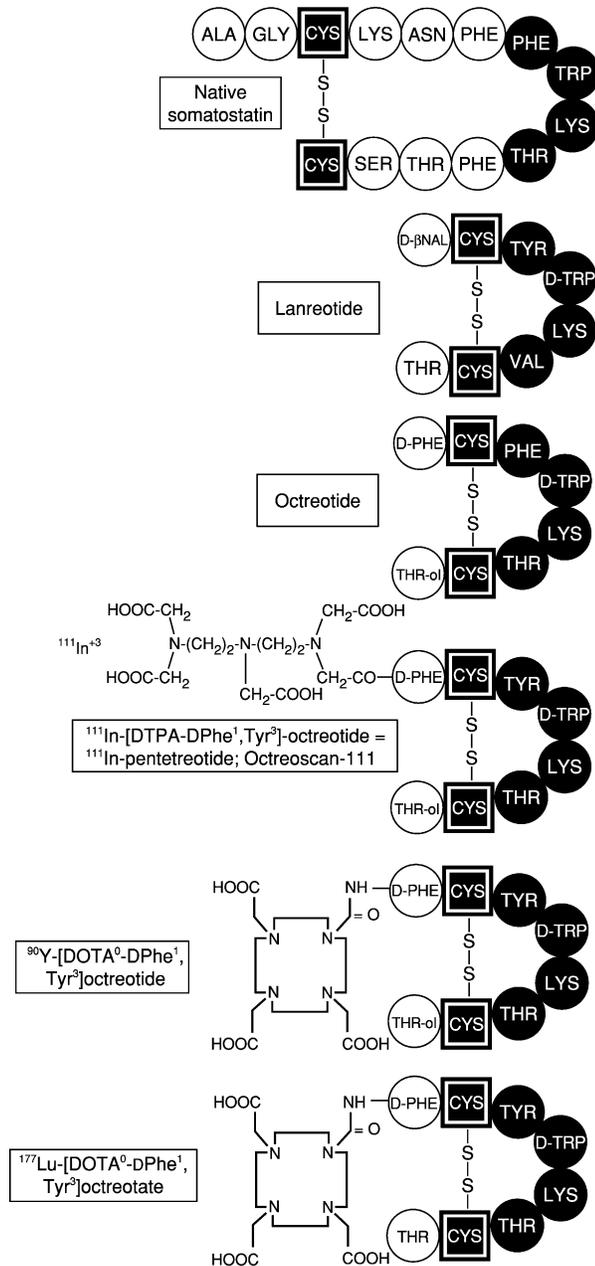
In 40–70% of adult patients with VIPoma, diffuse metastatic disease in the liver is present initially; therefore, surgical cure is not possible. For these patients, long-acting somatostatin analogues such as octreotide or lanreotide (Fig. 1) are the drugs of choice for treatment. If these fail or tumor growth continues, treatment with chemoembolization, hepatic embolization, or chemotherapy or radiolabeled somatostatin analogues (Fig. 1) may be helpful. If the majority of the metastatic disease in the liver can be safely resected, cytoreductive surgery may be of value in helping to control the symptoms of the hormone-excess state.

### Glucagonomas

Glucagonomas are neuroendocrine tumors of the pancreas that ectopically secrete excessive amounts of glucagon and cause a syndrome characterized by a dermatitis (migratory necrolytic erythema), glucose intolerance or diabetes, and weight loss (Table VIII).

#### Clinical Features

Glucagonomas usually occur in middle-aged and elderly populations. Migratory necrolytic erythema usually starts as an erythematous area typically at periaxial or intertriginous areas, such as the groin, buttocks, or thighs. It subsequently becomes raised and bullae form and break, resulting in eroded areas. The skin lesions may wax and wane and frequently precede the diagnosis of glucagonoma (mean, 6–8 years). Hypoaminoacidemia is a characteristic feature of glucagonomas, with plasma amino acid levels decreasing to < 25% of normal, especially glycogenic amino acids such as alanine and glycine. Weight loss is another prominent feature of this syndrome and is not seen early in other PETs unless malabsorption is present, suggesting that it is an intrinsic feature of this syndrome. Experimental studies support



**Figure 1** Structure of somatostatin and synthetic analogues used for diagnostic (tumor localization) or therapeutic indications in patients with neuroendocrine tumors. The indium ( $^{111}\text{In}$ ), yttrium ( $^{90}\text{Y}$ ), or lutetium ( $^{177}\text{Lu}$ ) compounds are used for therapeutic purposes. DOTA-1,4,7,10-tetra-azacyclododecane- $N,N',N'',N'''$ -tetra-acetic acid.

the conclusion that a novel anorectic substance, independent of glucagon, is released by these tumors and is responsible for the weight loss. Other prominent symptoms/clinical findings include glossitis

**Table VIII** Frequency of Clinical Symptoms and Laboratory Findings in Patients with Glucagonoma

	Frequency (%)
<b>Clinical symptoms</b>	
Dermatitis	64–90
Diabetes/glucose intolerance	38–96
Weight loss	56–96
Glossitis/stomatitis/cheilitis	29–40
Diarrhea	14–15
Abdominal pain	12
Thromboembolic disease	12–35
Venous thrombosis	24
Pulmonary emboli	11
Psychiatric disturbance	0–17
<b>Laboratory abnormality</b>	
Anemia	33–90
Hypoaminoacidemia	26–100
Hypocholesterolemia	80

or stomatitis, glucose intolerance or diabetes that may precede the diagnosis by years, and anemia (Table VIII).

#### Pathology/Pathophysiology

Glucagonomas are characteristically diagnosed late in their course and are usually large (average, 5–10 cm), and 50–80% have evidence of metastatic spread at diagnosis, usually to the liver (43–80%). Glucagonomas are usually single tumors and 50–80% occur in the pancreatic tail.

Glucagon is a naturally occurring, 29-amino acid peptide characteristically released by the pancreatic A cells. Most of the findings of the syndrome are compatible with the known actions of glucagon stimulating glycogenolysis and gluconeogenesis and affecting gut secretion and motility. The exact pathogenesis of the migratory necrolytic erythema remains unclear and has been attributed by some to hypoaminoacidemia or to nutritional deficiencies such as zinc.

#### Diagnosis/Differential Diagnosis

The diagnosis is established by demonstrating an elevated plasma glucagon level (normal level is usually < 150 ng/liter). Plasma glucagon levels usually exceed 1000 ng/liter in glucagonoma patients (90%), and in patients with symptoms/laboratory findings of glucagonoma (Table VIII), a level > 1000 ng/liter is diagnostic. Other conditions can also cause elevated plasma glucagon levels, including pancreatitis, hepatic failure disease, renal failure, Cushing's syndrome, prolonged fast, or familial hyperglucagonemia. Except

for cirrhosis, these disorders usually do not cause increases in plasma glucagon levels > 500 ng/liter.

### Treatment

Diffuse hepatic metastases are usually present at diagnosis (in up to 80% of patients), so curative surgical resection is usually not possible. Surgical debulking or other antitumor treatments may be of palliative benefit. The drugs of choice are the long-acting somatostatin analogues (octreotide and lanreotide) (Fig. 1), which improve the rash in 75–80% of patients and may improve the weight loss, pain, and diarrhea but usually do not improve the diabetes/glucose intolerance.

### Somatostatinoma Syndrome

The somatostatinoma syndrome is caused by a neuroendocrine tumor of the GI tract that ectopically secretes excess amounts of somatostatin (Fig. 1), which frequently causes diabetes mellitus, gallbladder disease, diarrhea, and steatorrhea (Table IX). There is no general agreement on what constitutes a somatostatinoma, and there is no distinction in the literature regarding the presence of a somatostatinoma and/or the somatostatinoma syndrome. In the literature, the term somatostatinoma is generally used to mean a GI neuroendocrine tumor possessing somatostatin-like immunoreactivity, and it may (11–45%) or may not (55–89%) be associated with clinical symptoms due to ectopic release of somatostatin (Table IX). Because of this confusion, the term somatostatinoma syndrome is used here to indicate a PET-releasing

somatostatinoma, which causes clinical symptoms, and the term somatostatinoma is used to indicate a PET containing somatostatin immunoreactivity.

### Clinical Features

The mean age of onset is 51–53 years. In one large review ( $N = 173$ ), only 11% of all somatostatinomas in the literature were associated with the somatostatinoma syndrome. The principal clinical features of the somatostatinoma syndrome are gallbladder disease, diabetes mellitus, diarrhea, weight loss, and steatorrhea. The frequency of these symptoms depends on the location of the somatostatinoma (Table IX). Each symptom characteristic of the syndrome is reported more frequently with pancreatic than intestinal somatostatinomas.

### Pathology/Pathophysiology

Somatostatin is a naturally occurring tetradecapeptide (Fig. 1) found widely in the central nervous system and GI tract, where it functions as a neurotransmitter or has paracrine or autocrine actions. In general, it is a potent inhibitor of processes, including the release of numerous hormones, gastric and intestinal pancreatic secretion, and absorption. The hypochlorhydria, diabetes, steatorrhea, and gallbladder diseases that occur in somatostatinoma syndrome are caused by the known inhibitor effects of somatostatin.

Somatostatinomas occur in the pancreas in 56–74% of cases and are mainly found in the pancreatic head. In most of the remaining cases, they are found in the intestine, with 90% in the duodenum, particularly in the periampullary area. They are usually solitary, large tumors (mean, 3.6–4.9 cm), and 53–84% have metastatic spread at diagnosis (usually to the liver).

### Diagnosis/Differential Diagnosis

In most cases in the literature, somatostatinomas have been found incidentally either at the time of cholecystectomy or during endoscopy. The presence of psammoma bodies in a duodenal neuroendocrine tumor should raise this possibility. Duodenal somatostatinomas are associated with von Recklinghausen's disease. Most of these patients do not develop the somatostatinoma syndrome and have normal plasma somatostatin levels. The diagnosis of the somatostatinoma syndrome requires the demonstration of an elevated plasma somatostatin level.

Table IX Clinical and Laboratory Findings in Patients with Somatostatinomas with or without Somatostatinoma Syndrome

Clinical finding	Somatostatinoma <sup>a</sup>		Overall frequency in somatostatinoma syndrome (%) <sup>a</sup>
	Pancreatic (%)	Intestinal (%)	
Diabetes mellitus	95	21	95
Gallbladder disease	94	43	68
Diarrhea	66–97	11–36	37
Weight loss	32–90	20–44	68
Laboratory finding			
Steatorrhea	83	12	47
Hypochlorhydria	86	17	26

<sup>a</sup>Somatostatinoma is the occurrence of a pancreatic endocrine tumor containing somatostatin by immunocytochemistry that can occur with (11%) or without (89%) the somatostatinoma syndrome, which is due to ectopically released somatostatin.

## Treatment

Symptoms of the somatostatinoma syndrome are improved by administration of long-acting synthetic somatostatin analogues such as octreotide (Fig. 1). Pancreatic tumors (70–92%) and intestinal tumors (30–69%) are associated with liver metastases at presentation and antitumor treatment is frequently needed.

## GRFomas AND OTHER RARER FUNCTIONAL PET SYNDROMES

GRFomas are neuroendocrine tumors that ectopically secrete growth hormone-releasing factor (GRF). GRF is a 44-amino acid peptide that stimulates growth hormone release. GRFomas occur in lung (47–54%), PETs (29–30%), small intestinal carcinoids (8–10%), and other sites (12%). The symptoms are those of acromegaly and the mean age of patients is 38 years. The acromegaly is indistinguishable from classical acromegaly due to a pituitary adenoma. Pancreatic GRFomas are usually large (mean, > 6 cm) and liver metastases are present in 39% of cases. They should be suspected in a patient with acromegaly with an abdominal tumor, MEN-1, or hyperprolactinemia, which occurs in 70% of GRFomas. The diagnosis is established by performing plasma assays for GRF and growth hormone. Surgery is the treatment of choice if possible. Symptoms can be controlled in >75% of patients by long-acting somatostatin analogues, such as octreotide or lanreotide (Fig. 1).

Cushing's syndrome due to a PET occurs in 4–16% of all patients with ectopic Cushing's syndrome. It occurs in 5% of patients with sporadic gastrinomas and is associated with hepatic metastases and a poor prognosis. Paraneoplastic hypercalcemia due to a PET releasing PTH-related peptide is rare. These tumors are usually large and malignant. PETs secreting calcitonin may cause a specific syndrome associated with diarrhea (Table II). PETs also cause the carcinoid syndrome (Table II). These are characteristically large and malignant (68–88%) and may cause an atypical carcinoid syndrome because they lack DOPA decarboxylase. A renin-producing PET was described in a patient presenting with hypertension (Table II). Ghrelin is a 28-amino acid peptide with a number of metabolic functions. Although expression has been demonstrated in most PETs, in a one study only 1 of 24 patients (4%) with a PET had an elevated plasma ghrelin level. This patient was asymptomatic, suggesting that no specific syndrome is associated with ectopic release of ghrelin by a PET.

## NONFUNCTIONAL PETs

Nonfunctional PETs are neuroendocrine tumors of the pancreas that secrete no products or the secreted products do not cause a distinct functional syndrome. The symptoms from these tumors are therefore due entirely to the tumor per se (i.e., pain, jaundice, weight loss, etc.). Nonfunctional PETs frequently secrete chromogranin A (CgA) (80–100%), CgB (90–100%), pancreatic polypeptide (PP) (58%), and the  $\alpha$ -subunit of hCG (40%), and many secrete other hormones such as neurotensin. These tumors characteristically present late in the disease course. They are generally large (72% > 5 cm) and invasive, and hepatic metastases are usually present (64–92%). The most common symptoms are abdominal pain (30–50%), jaundice (20–35%), weight loss, fatigue, and bleeding. In 10–15% of cases, they are found incidentally. The diagnosis is established by histology with appropriate neuroendocrine tumor immunohistochemistry and by assessing plasma hormone levels/clinical symptoms. Plasma PP is increased in 22–71% of patients, CgA levels are increased in 80–100%, and in patients with a pancreatic mass without a functional syndrome, this finding suggests that a nonfunctional PET is present. Curative surgical resection is rarely possible, and treatment needs to be directed against the malignant PET.

## FUNCTIONAL SYNDROMES DUE TO CARCINOID TUMORS

Carcinoid tumors can cause a specific functional syndrome, the carcinoid syndrome, or occasionally can release biologically active peptides that cause the specific PET syndromes discussed previously. Because carcinoids are also malignant (Table III), specific treatments need to be directed at both the carcinoid tumor and the functional syndrome it produces.

### Carcinoid Syndrome

The carcinoid syndrome is caused by a neuroendocrine tumor, usually present in the gastrointestinal tract, ectopically secreting bioactive amines/peptides, which results in a clinical syndrome characterized by diarrhea, flushing, asthma/wheezing, and, occasionally, heart disease (Table X).

### Clinical Features

The mean age at presentation is 57 years, but it occurs over a wide age range (9–91 years) (Table X). The principal symptoms are diarrhea and flushing, which occur in up to 73% of cases initially and 84% during

**Table X** Clinical Features in Patients with Carcinoid Syndrome

	At presentation	During course of disease
<b>Symptoms/signs</b>		
Diarrhea (%)	32–73	68–84
Flushing (%)	23–65	63–74
Pain (%)	10	34
Asthma/wheezing (%)	4–8	3–18
Pellagra (%)	2	5
None (%)	12	22
Carcinoid heart disease (%)	11	14–41
<b>Demographics</b>		
Male (%)	46–59	46–61
Mean age, years (range)	57 (25–79)	53 (9–91)
<b>Tumor location</b>		
Foregut (%)	5–9	2–33
Midgut (%)	78–87	60–87
Hindgut (%)	1–5	1–8
Unknown (%)	2–11	2–15

the course of the disease. The flush is usually of sudden onset, associated with a deep red to violaceous erythema of the upper body, and often associated with a feeling of warmth and occasionally with lacrimation, pruritis, or diarrhea. It may be precipitated by food, exercise, alcohol, or drugs, particularly serotonin reuptake inhibitors. Flushing is usually caused by metastatic midgut tumors. Flushing with midgut tumors and bronchial or gastric carcinoids may differ in duration, associated symptoms (salivation and lacrimation), and skin color. Diarrhea usually occurs with flushing (85% of cases) and is usually watery and of small volume (60% < 1 liter/day). Steatorrhea is present in 67%, and in 50% of patients it is > 15 g/day of fecal fat. Cardiac manifestations occur in 11% initially and 14–41% during the disease course (Table X), and they are due to endocardial fibrosis, primarily on the right side (tricuspid > pulmonary), but they can also occur on the left side. Fibrosis results in valve constriction and up to 80% of patients develop heart failure. Pellagra-like symptoms (2–25%) and symptoms due to increased fibrotic tissue (i.e., retroperitoneal fibrosis, Peyronie's disease, and intraabdominal fibrosis) are unusual features of this disease.

A life-threatening complication of the carcinoid syndrome is the development of a carcinoid crisis. This is most frequently seen in patients with high 5-hydroxyindoleacetic acid (5-HIAA) levels and may be provoked by anesthesia, endoscopy, stress, surgery, a

radiological procedure, or a biopsy. Patients develop intense flushing, diarrhea, abdominal pain, hypotension, and cardiac abnormalities. If not adequately treated, it can be fatal.

### **Pathology/Pathobiology**

Carcinoid symptoms occurred in 8% of 8876 patients with carcinoid tumors. Carcinoid syndrome occurs only when tumor-secreted products reach the systemic circulation in sufficient concentrations. In 91% of cases, this occurs due to liver metastases, and in the remainder it occurs due to retroperitoneal invasion by gut or pancreatic tumors or due to lung or ovary carcinoids with direct access to the systemic circulation. Midgut carcinoids account for 60–67% of cases of the carcinoid syndrome, foregut tumors account for 2–33%, hindgut accounts for 1–8%, and an unknown primary accounts for 2–15%.

One of the main secretory products of carcinoid tumors is serotonin (5-HT), and overproduction occurs in 90–100% of patients with carcinoid syndrome. Serotonin is thought to primarily mediate diarrhea by its effects on gut motility and intestinal secretion. However, prostaglandins and tachykinins (substance P, K; neuropeptide K) may also be important in causing diarrhea in some patients. Flushing is not relieved by serotonin antagonists, and in patients with gastric carcinoids it can be due to histamine secretion. Tachykinins are released during flushing and may be important in its mediation. Both histamine and serotonin may be responsible for the wheezing/asthma as well as the fibrotic reactions characteristic of this disease. The heart disease is likely due to the actions of serotonin because it is similar to that seen with appetite-suppressant drugs (e.g., fenfluramine) that have high affinity for serotonin receptors.

Patients may develop a typical or atypical carcinoid syndrome. The typical syndrome is due to a midgut tumor oversynthesizing serotonin from tryptophan. The atypical syndrome occurs when there is a deficiency in the enzyme required to convert 5-HTP to 5-HT and there is overproduction of 5-HTP. The atypical syndrome is more likely to occur with foregut carcinoids.

### **Diagnosis/Differential Diagnosis**

The diagnosis of carcinoid syndrome requires measurement of urinary or plasma serotonin or its metabolites. The measurement of urinary 5-HIAA is most frequently performed. False positives may occur if a patient eats serotonin-rich foods (bananas, walnuts, pecans, and pineapple) or takes certain medications (L-Dopa, cough syrups with guaniforesin, and

salicylates). If an atypical carcinoid syndrome is suspected, urinary 5-HIAA may be only slightly elevated, and other metabolites of tryptophan such as 5-HTP should be considered. Serum cgA levels are elevated in 50–100% of patients with carcinoid tumors, and the level correlates with tumor bulk. However, serum CgA levels are not specific for carcinoids because they can occur with other NETs, including PETs.

Flushing can be caused by other conditions, including systemic mastocytosis, reactions to alcohol or glutamate, effects of drugs, and menopause. These need to be excluded.

### Treatment

Treatment consists of avoiding conditions that precipitate attacks and treatment of wheezing with bronchodilators, treatment of heart failure with diuretics, and treatment of diarrhea with antidiarrheal agents, such as loperamide or diphenoxylate. If symptoms present, somatostatin analogues (octreotide and lanreotide) (Fig. 1) or serotonin receptor antagonists are the drugs of choice. Somatostatin analogues control symptoms in 80% of patients with flushing or diarrhea, and 70% show a > 50% decrease in urinary 5-HIAA. Approximately 40% of patients show some resistance after 4–6 months and doses may have to be increased. Sustained-release preparations, such as octreotide-LAR (long-acting release) and lanreotide-PR (prolonged release), are widely used because they can be given less frequently. Octreotide-LAR is usually given monthly and lanreotide-PR every 10–14 days as opposed to the usual preparation of octreotide/lanreotide that is given subcutaneously every 6–12 h. Somatostatin analogues can be given to both treat and prevent carcinoid crises. Prior to a possible precipitating event, such as surgery, anesthesia, or stress, it is recommended that octreotide be administered.

Side effects from the somatostatin analogues occur in 40–60% of patients with subcutaneous analogues. Pain at the injection site and GI side effects (discomfort, nausea, diarrhea, and cramping) are the most common. They are usually short-lived and do not interrupt treatment. Long-term complications include an increased incidence of gallstones/biliary sludge (52%), steatorrhea, and worsening glucose tolerance. Interferon- $\alpha$  may also control symptoms in some patients.

Surgery should be performed if possible; however, almost all patients with carcinoid syndrome have metastatic disease in the liver and curative resection is not possible. Treatment directed against the tumor may be needed.

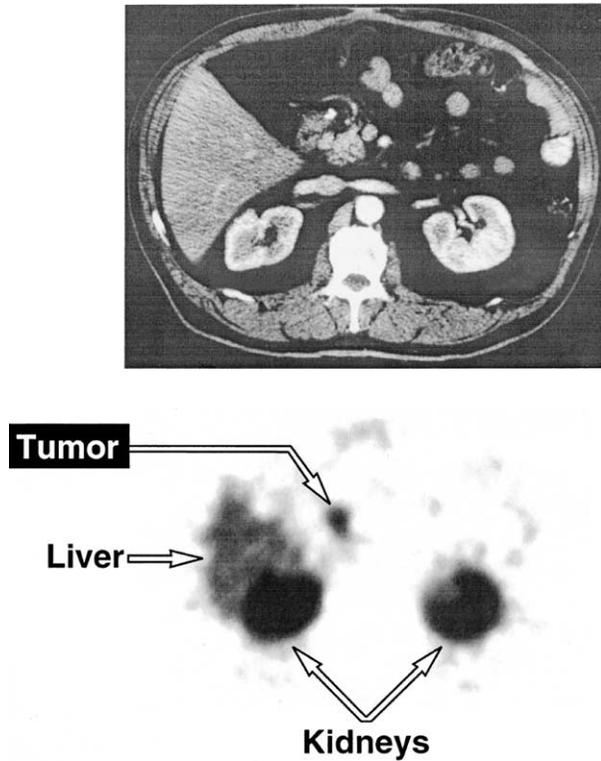
## NET TUMOR LOCALIZATION

Tumor localization is needed for all management phases of both PETs and carcinoid tumors. Both the localization of the primary tumor and the determination of the location and extent of metastatic disease are required to appropriately manage these patients. Conventional imaging studies [CT, magnetic resonance imaging (MRI), ultrasound, and angiography] and SRS are widely used. For PETs, endoscopic ultrasound is also widely used. Bronchial carcinoids are usually detected by chest X-ray and assessed by CT. Rectal, duodenal, colonic, and gastric carcinoids are usually detected by GI endoscopy.

Both PETs and carcinoids frequently (90–100%) overexpress somatostatin receptors, which have a high affinity for radiolabeled somatostatin analogues (Figs. 1–3) that can be used to localize them. Because of its greater sensitivity compared to that of conventional imaging studies and its ability to localize a number of tumors throughout the body simultaneously, SRS is the imaging modality of choice for localizing all primary and metastatic NET tumors except insulinomas. Insulinomas are usually small and have a low density of somatostatin receptors, with the result that SRS is positive in only 12–50% of patients with insulinomas. In contrast, SRS is positive in 73–89% of patients with carcinoids and 60–100% of patients with PETs other than insulinomas. Figures 2 and 3 show two examples of the ability of SRS to image a primary PET and metastatic disease in the liver when conventional imaging studies were negative. For PETs localized in the pancreas, endoscopic ultrasound is highly sensitive, localizing 73–100% of insulinomas, which occur almost exclusively within the pancreas. SRS occasionally gives false positives (12% in one study) because normal and abnormal cells can have increased somatostatin receptors such as granulomas, thyroid disease, and activated lymphocytes (abscess and infection). Furthermore, SRS does not provide information on tumor size or the exact location of metastases, and a CT scan or MRI are frequently used to provide this information.

## TREATMENT OF ADVANCED DIFFUSE METASTATIC DISEASE IN PATIENTS WITH MALIGNANT NETs

Of the numerous prognostic factors identified for NETs (Table IV), the presence and the extent of the hepatic metastases are the most important in almost every study. For patients with gastrinomas, 5-year

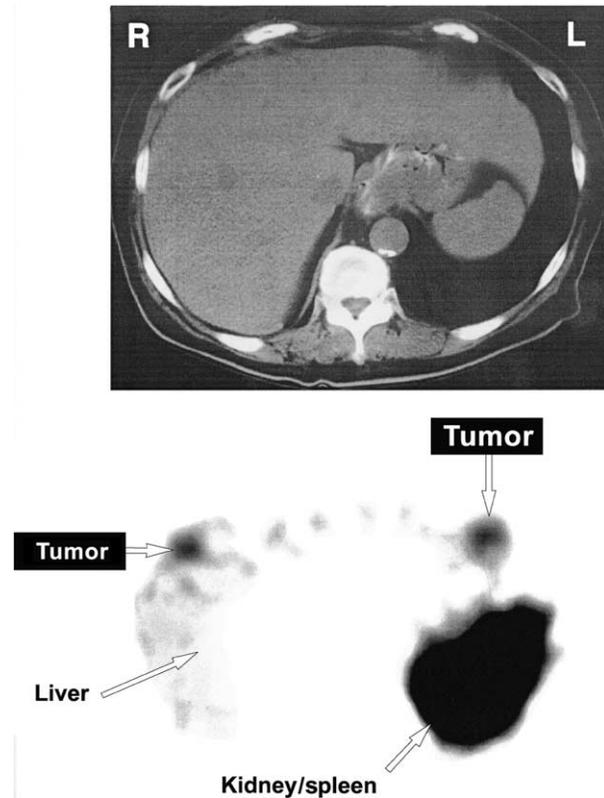


**Figure 2** Computed tomography (CT) scan (top) and somatostatin receptor scintigraphy (SRS) (bottom) of a patient with a pancreatic endocrine tumor (gastrinoma). This patient previously had a duodenal tumor removed. The CT scan was negative, whereas the SRS showed a tumor medial and anterior to the right kidney. At surgery, a PET was found in a metastatic lymph node in the peri-duodenal area. This case illustrates the greater sensitivity of SRS compared to conventional imaging studies.

survival is 98–100% without liver metastases, 78% with limited metastases in one lobe, and 16% with diffuse metastases. For carcinoid tumors without liver metastases, 5-year survival is 80–90%, and with diffuse metastases it is 50%. A number of treatments are reported to be effective, including embolization, chemoembolization, chemotherapy, cytoreductive surgery (removal of all visible tumor), somatostatin analogues,  $\alpha$ -interferon, radiotherapy using radiolabeled somatostatin analogues to target the tumor (Fig. 1), and liver transplantation.

### Specific Antitumor Treatments

Unfortunately, cytoreductive surgery is only possible in 10–20% of patients who have limited hepatic metastases, allowing surgical removal of at least 90% of visible tumor. Although it is reported to provide palliative treatment, there are no control studies to



**Figure 3** Computed tomography (CT) scan and somatostatin receptor scintigraphy (SRS) of a patient with a metastatic PET to the liver. The CT scan showed no liver lesions, whereas the SRS showed two hepatic metastases (one in the right lobe and one in the left lobe). This case illustrates the increased sensitivity of SRS compared to conventional imaging studies for localizing hepatic metastases.

support this conclusion. However, studies suggest that it may increase survival; therefore, cytoreductive surgery is recommended, if possible.

Chemotherapy for metastatic carcinoids is generally disappointing, with response rates of 0–40% with various two–three drug combinations. With PETs, chemotherapy has been more successful, with response rates of 30–70%. The regimen of choice is streptozotocin and doxorubicin. Unfortunately, chemotherapy is almost never curative and has substantial toxicity. Therefore, it is generally reserved for patients with rapidly growing PETs who fail other treatments.

Long-acting somatostatin analogues (Fig. 1) and  $\alpha$ -interferon rarely decrease tumor size (i.e., 0–15%); however, these drugs have prominent tumoristatic effects, stopping growth in 50–90% of patients with NETs. Studies suggest that they are particularly effective in slower growing tumors. It has not been proven that these agents extend survival; however, because of

the availability of long-acting formulations (e.g., one injection of somatostatin/month), their low toxicity, and their long-term effectiveness in some patients, they are the antitumor agents of choice.

Hepatic embolization or chemoembolization (i.e., chemotherapy with embolization) can decrease tumor bulk and help control the symptoms of the hormone-excess state. This approach is usually reserved for patients who have disease largely confined to the liver, who have a patent portal vein, and who fail treatment with other modalities.

Somatostatin receptor-directed cytotoxicity using radiolabeled somatostatin analogues that are internalized by somatostatin receptors (Fig. 1) overexpressed on the NET are being widely investigated. The following are being evaluated for treatment:  $^{111}\text{In}$  ( $^{111}\text{In}$ )-labeled compounds, which emit gamma rays, internal conversion, and Auger electrons;  $^{90}\text{Y}$  yttrium-coupled analogues, which emit high-energy  $\beta$  particles; and  $^{177}\text{Lu}$  lutetium ( $^{177}\text{Lu}$ )-coupled analogues, which emit both. The  $^{177}\text{Lu}$  and  $^{111}\text{In}$  compounds have respectively been shown to stabilize disease in 41 and 40% and decrease tumor size in 38 and 30% of patients with advanced metastatic NETs.

Liver transplantation, although largely abandoned for most metastatic tumors, is still a consideration for patients with metastatic NETs because of their slower growth. In 103 cases of malignant NETs (43 carcinoid, 48 PETs), liver transplantation achieved 2- and 5-year survival rates of 60 and 47%, respectively. Liver transplantation has been suggested for younger patients with metastatic NETs limited to the liver.

### See Also the Following Articles

Gastrinomas • Ghrelin • GI Hormones in Cancer • GI Tract, General Anatomy (Cells) • GI Tract, General Pathology of Endocrine Growths • Substance P

### Further Reading

- Alexander, R. A., and Jensen, R. T. (2001). Pancreatic endocrine tumors. In "Cancer: Principles and Practice of Oncology" (V. T. DeVita, S. Hellman, and S. A. Rosenberg, eds.), 6th ed., pp. 1788–1813. Lippincott Williams & Wilkins, Philadelphia.
- Arnold, R., Simon, B., and Wied, M. (2000). Treatment of neuroendocrine GEP tumours with somatostatin analogues: A review. *Digestion* **62**, 84–91.
- Caplin, M. E., Buscombe, J. R., Hilsen, A. J., Jones, A. L., Watkinson, A. F., and Burroughs, A. K. (1998). Carcinoid tumour. *Lancet* **352**, 799–805.
- Corleto, V. D., Delle Fave, G., and Jensen, R. T. (2002). Molecular insights into gastrointestinal neuroendocrine tumors: Importance and recent advances. *Digestive Liver Dis.* **34**, 668–680.
- Eriksson, B. (2001). Systemic therapy for neuroendocrine tumors of the pancreas. In "Surgical Endocrinology" (G. M. Doherty and B. Skogseid, eds.), pp. 393–403. Lippincott Williams & Wilkins, Philadelphia.
- Gibril, F., and Jensen, R. T. (2003). Diagnostic use of radiolabeled somatostatin analogues in patients with gastroenteropancreatic endocrine tumors. *Digestive Liver Dis.* (in press).
- Jensen, R. T., and Doherty, G. M. (2001). Carcinoid tumors and the carcinoid syndrome. In "Cancer: Principles and Practice of Oncology" (V. T. DeVita, Jr., S. Hellman, and S. A. Rosenberg, eds.), 6th ed., pp. 1813–1833. Lippincott Williams & Wilkins, Philadelphia.
- Jensen, R. T., and Norton, J. A. (2002). Pancreatic endocrine tumors. In "Sleisenger and Fordtran's Gastrointestinal and Liver Disease: Pathophysiology, Diagnosis, and Management" (M. Feldman, L. S. Friedman, and M. H. Sleisenger, eds.), 7th ed., pp. 988–1016. Saunders, Philadelphia.
- Kulke, M. H., and Mayer, R. J. (1999). Carcinoid tumors. *N. Engl. J. Med.* **340**, 858–868.
- Kwekkeboom, D., Krenning, E. P., and deJong, M. (2000). Peptide receptor imaging and therapy. *J. Nucl. Med.* **41**, 1704–1713.
- Lehnert, T. (1998). Liver transplantation for metastatic neuroendocrine carcinoma. *Transplantation* **66**, 1307–1312.
- Metz, D. C., and Jensen, R. T. (2003). Endocrine tumors of the gastrointestinal tract and pancreas. In "Gastrointestinal Cancers" (A. K. Rustgi, ed.), pp. 681–720. Saunders, New York.
- Mignon, M. and Jensen, R. T. (eds.) (1995). "Endocrine Tumors of the Pancreas: Recent Advances in Research and Management," Vol. 23. Karger, Basel, Switzerland.
- O'shea, D., and Bloom, S. R. (guest eds.) (1996). Gastrointestinal endocrine. *Balliere's Clin. Gastroenterol.* **10**, 555–766.



## Gastrointestinal Hormone

see GI Hormone entries

of the twentieth century. It was generally assumed that, if gender assignment had been carried out correctly, that is, in agreement with the true sex, the final outcome would be a heterosexual man or a heterosexual woman, in agreement with the assigned gender. There were problems with the true-sex policy, however. The presumed definitive criterion of the true sex did not always agree with a person's somatic physical appearance as a man or woman, and also did not always agree with the person's gender identity, and could therefore lead to dramatically adverse social and legal consequences for the intersexed individual.

### The Optimal-Gender Policy

On the basis of a critical examination of the true-sex policy and studies of the overall psychosocial outcome and quality of life of intersex patients, John Money and the Hampsons at Johns Hopkins Hospital in Baltimore, Maryland, formulated an optimal-gender policy during the 1950s. This policy had a number of underlying assumptions. (1) There is no single biological criterion that determines the development of psychological gender; rather, there is a cascade of biological processes that culminate in the development of gender identity. (2) Socialization factors have the decisive role in psychological gender development. (3) If one minimizes the barriers to socialization in the assigned gender, the outcome will be a heterosexual man or a heterosexual woman, in agreement with the assigned gender. (4) The condition of the body and particularly the genitalia limits the psychosocial and especially psychosexual functioning of the intersexed person in later years. Therefore, the newborn with intersexuality should be assigned to that gender that permits the optimal psychosexual and psychosocial functioning when all available medical treatment options are taken into account. So, in contrast to the question of the true-sex policy—"Is this a boy or a girl?"—the optimal-gender policy asked "Will this newborn have a better function later in life as a male or a female?"

From the central role of socialization followed the concept of the birth of an intersex newborn as a psychosocial emergency, with the implication that the decision time for gender assignment should be kept as short as possible. Another consequence was the recommendation of early feminizing or masculinizing surgery of the external genitalia so that their appearance would be as similar to the gender norm as possible and would therefore not interfere with gender-typical rearing conditions and body-image

development. A third consequence was the instruction to keep the intersex status of the child secret from all people who do not belong to the core family and to educate the child gradually, in line with his or her cognitive development. These guidelines were intended to prevent the parents from developing chronic doubts about the sex of their child and to protect the child from stigmatization by other people, but have frequently led to parental attempts of preventing the disclosure of details of the medical history to the intersex children themselves, sometimes even after they attained adulthood.

### The True-Brain-Sex Policy

Increasingly over the past decade, the optimal-gender policy has come under criticism. One reason is that also under this policy some individuals will turn out dissatisfied with their assigned gender and may seek gender reassignment, which is made more difficult if the external genitalia have been operated on to be more compatible with the originally assigned gender. A second argument is that, even without later gender change, genital surgery carries a risk of damage to sexual functioning in adolescence and adulthood. Intersex activists, therefore, demand that psychosocially indicated genital surgery be performed only with the informed consent of the mature individual. Some critics advocate a moratorium on genital surgery, unless necessary for strictly medical indications, and claim—without data—that psychological counseling can take care of all attending psychosocial problems that children with marked genital ambiguity and their families may face. Third, the optimal-gender policy is blamed (although unjustly; see above) for keeping the medical facts secret, especially from the patient him- or herself, which contributes to the maintenance of the social stigma of intersexuality and thereby to the negative self-image (especially shame) of the intersexed individual. Many intersex activists, therefore, argue for early comprehensive disclosure of their medical history to intersex patients. A fourth criticism of the optimal-gender policy derives from the extensive data, especially data generated by animal research, regarding the influence of sex hormones on the developing brain and long-term sex-dimorphic behavior. Because of these data, some biological determinists suggest that the decisive factor for gender identity formation is the prenatal androgenization of the brain and that psychosocial factors have only a secondary role. If such a "true-brain-sex policy" is valid, gender assignment decisions should be based on the

degree of androgenization/masculinization of the brain and minimization of barriers to socialization is unimportant. However, because brain imaging techniques are not yet capable of rendering such assessments, these theorists must fall back on the status of the genitalia as a—rather uncertain—indicator of brain androgenization/masculinization.

### Policy Effects on Gender Assignment

The three major policies outlined above may lead to quite varied decisions on gender assignment. Two cases will illustrate this: (1) a chromosomally female newborn with extreme genital masculinization (Prader stage V) due to prenatal androgen excess associated with classical congenital adrenal hyperplasia (CAH) and (2) a chromosomally male newborn with penile agenesis. According to the true-sex policy, case (1) should be raised female because of the clearly female histology of the gonads (ovaries) and female chromosomes (see Table I). By contrast, a 46,XY newborn with penile agenesis would be assigned to the male gender, because of the normal-male histology of the testicular gonads and the clearly male karyotype. Under the optimal-gender policy, case (1) would be assigned to the female gender, because the external genitalia can be feminized so as to permit coitus and to thereby retain the option of conception and pregnancy, because the internal female reproductive organs are intact. The 46,XY infant with penile agenesis would also be assigned to the female gender, because the surgical *de novo* construction of a functioning penis is not yet possible and therefore sexual functioning as a male would be compromised. Advocates of the true-brain-sex policy would recommend assigning the CAH case to the male gender because of the putative effects of prenatal androgens on the brain and recommend the same for the case of penile agenesis. Some activists recommend assigning the intersex child to the gender that seems to offer the more promising outcome, but to do so provisionally and to consider from the outset the possibility of later gender change and therefore not to operate on the genitals before the age of consent, unless medically necessary.

Table I Gender Assignment as a Function of Policy

Disorder	True-sex policy	Optimal-gender policy	True-brain-sex policy
Penile agenesis	Male	Female	Male
CAH/Prader stage V	Female	Female	Male

### The Status of the Evidence

The major objections to the optimal-gender policy are important to consider, but a consensus about a new management policy has yet to emerge, especially for 46,XY individuals with intermediate degrees of genital ambiguity. The major problem is insufficient data concerning the long-term psychological outcome. Patient-initiated later gender change, for instance, can be observed in intersex individuals in both directions, from male to female and from female to male (and, in some cases, from the assigned gender to “intersex”). This is not surprising, because such gender change is also well known—as transsexualism—in non-intersex individuals, but the relative frequency of later gender change is increased in intersex patients. Given the psychological, social, and medical problems that are associated with later gender change, clinicians usually favor a policy that minimizes its occurrence. Thus, long-term follow-up data are needed that will allow clinicians to state the relative frequency of gender change of patients with a given syndrome who were managed under a defined policy and in a particular cultural context. Because gender change can take place as late as in midlife, follow-up studies need to reach at least into that age range. Given the relative rarity of intersex patients, it is not surprising, therefore, that the data thus far available are clearly insufficient.

As an extreme example, consider the question of gender reassignment in chromosomal males with traumatic loss of the penis in infancy. One such case, in which reassignment to female at 17–21 months of age was followed by patient-initiated re-reassignment to male in adolescence, has drawn enormous attention among care providers as well as in the media, once his unfortunate story was investigated and published, when he was in his mid-thirties. However, only one other case with a history of traumatic loss of the penis in infancy and reassignment to female has been followed into adulthood and published and this person continues to live as a woman without gender dysphoria. Two cases with strikingly different outcomes are not sufficient for evidence-based policy decisions. Even for intersex syndromes that are more prevalent and less shrouded in secrecy than the cases of traumatic loss of the penis, long-term outcome data are hard to come by and those that have become available all suffer from problems that limit their validity: loss to follow-up of patients who have moved and cannot be found, who have died (especially if for other than medical reasons), or who have refused to participate in research (especially if for reasons of disappointment with their medical treatment).

Outcome data are also insufficient with regard to the question of impairment of sexual functioning after gender-confirming genital surgery. From a neuroanatomic standpoint, it is plausible that excision of the clitorophallus—widely practiced as a feminizing technique in the 1950s and 1960s—would lead to a diminution of erotic sensitivity and orgasmic capacity, although the limited outcome data available show considerable variability. But surgical techniques have undergone many changes since then and reduction of the clitoropenis rather than excision has become the norm. To evaluate sexual functioning in patients who were operated on with these newer surgical techniques, researchers need to wait at least until the patients are sexually active, sufficiently sexually experienced, and able to talk about these intimate details of their lives, and this in a population that is known to be relatively delayed in attaining psychosexual milestones. The same applies to patients with a similarly ambiguous genital status at birth who for one reason or another have not undergone genital surgery and could therefore serve as comparisons. Most surgical outcome reports are limited to anatomical data and the few that include data on sexual functioning leave much to be desired in terms of sample representativeness, assessment methods, and study details provided, quite apart from the fact that patients who have undergone surgeries with the newer techniques are not yet old enough for sexual functioning studies.

The true-brain-sex policy appears very plausible at first glance, but becomes more problematic on closer scrutiny. The assumption that the effects of prenatal hormones on the brain are the decisive biological factor in the development of gender and therefore the best criterion for the assignment of gender is primarily derived from research on the sexual differentiation of brain and behavior in nonhuman mammals, especially rodents, which has yielded several neuroendocrine models of the sexual differentiation of brain and behavior. In humans, most pertinent studies have been limited to androgens. In general, the resulting data allow the preliminary conclusion that effective prenatal androgens are indeed associated with the masculinization of gender-role behavior. However, the core gender identity, for instance, as a girl or as a woman, is compatible with large variability in gender-role behavior, as is vividly demonstrated in girls and women with CAH, thus making a close linkage to prenatal hormones less likely. It also must be kept in mind that in the limited literature available, the association between prenatal androgens and later gender change takes place predominantly in the context of further postnatal and pubertal cross-gender

hormonalization and the associated increased somatic sexual ambiguity. Therefore, in all likelihood, it is not the prenatal hormonal milieu by itself that determines such later gender change.

### Inferring Brain Masculinization from Genital Masculinization

Adherents of the true-brain-sex policy recommend a male assignment for a newborn with intersexuality if the child's brain was prenatally strongly masculinized, independent of the status of the genitalia. However, how can one determine in a newborn whether and to what extent the brain was prenatally masculinized and/or defeminized? Appropriate brain imaging techniques are not yet available, and even if they were available, it is rather uncertain that the brain is, at the time of birth, already sufficiently sexually dimorphic to show such clear-cut gender dimorphic structures. Some advocates of the true-brain-sex policy therefore suggest using the genital status at birth (staged according to Prader for excess masculinization in chromosomal females and according to Quigley for undermasculinization in chromosomal males) as an indicator of brain androgenization/masculinization. There are several problems with this approach. (1) Genital staging at birth is not suitable for all syndromes of interest. For instance, Prader staging is problematic for many prenatally dexamethasone-treated cases of 46,XX CAH and Quigley staging is problematic for 46,XY newborns with 5 $\alpha$ -reductase deficiency, and Quigley staging is not at all suitable for other conditions of interest, such as 46,XY penile agenesis and 46,XY cloacal exstrophy of the bladder. (2) The growth of the penis does not depend exclusively on sex hormones. (3) The hormone model of genital differentiation is not identical with the hormone models of brain differentiation. As far as is known, the prenatal development of the genitalia is determined by testosterone, dihydrotestosterone, and anti-Müllerian hormone, whereas the development of the brain depends on androgens and, possibly, estradiol. (4) Attempts to correlate in the 46,XX CAH syndrome the Prader stage with gender-role behavior have had only very limited success. (5) Available (scanty) data on patient-initiated gender change in adults with a given intersex syndrome show little relationship to genital status at birth. (6) It also must be kept in mind that gender identity incongruence occurs in non-intersexed persons as well, although prior to surgery the genitals of transsexuals are compatible with their original gender assignment. (7) Research has made it

likely that contributions to the sexual differentiation of brain and behavior by hormone-independent genetic factors and brain-tissue-specific hormonal factors need to be taken into account. (8) There may be early effects of the social environment on central nervous system structures, as is already known from animal research. If any of these mechanisms participate in the development of persons with gender identity incongruence without somatic intersexuality, they are likely to also contribute to the variability and development of intersex patients. It can be concluded from all of these considerations that the genital status at the time of birth cannot be interpreted as a clear-cut indicator of brain masculinization.

### Gender Assignment by Diagnostic Category

Intersex children and their families cannot be left in gender limbo. Despite controversies and uncertainties, clinicians must continue making decisions in this area. The degree of uncertainty varies with the syndrome. The situation is relatively clear-cut for the most frequent syndrome, CAH in 46,XX newborns, for which there is a general consensus that the assignment should be to the female gender, because the internal reproductive structures (ovaries, fallopian tubes, distal vagina) permit reproduction, provided that external genitalia and vaginal introitus are surgically corrected where necessary. Although many girls with CAH show variable degrees of behavioral masculinization, gender identity typically is female. Some genitally highly masculinized 46,XX children with CAH have been inadvertently raised as males and maintain a male gender identity when correctly diagnosed in adolescence, and a few CAH females change to male in adolescence or adulthood, apparently mostly in cases of protracted marked genital ambiguity and inconsistent hormonal control. However, the evidence is not sufficient to support the routine assignment of the most masculinized (Prader stage V) newborns to the male gender (as has been recommended).

Also, among 46,XY intersex syndromes not all gender assignment decisions are problematic. There is a consensus that 46,XY newborns with complete androgen insensitivity should be raised as female, because they will never respond to male sex hormones, neither somatically nor psychologically, and, given that it is (mostly) prenatal rather than postnatal androgens that affect gender-related behavior, the argument can be extended to the rare syndrome of complete gonadal dysgenesis. Even in the syndromes of partial androgen

insensitivity and/or testosterone biosynthesis defect, there is a consensus that the least undermasculinized cases should be raised male and the most undermasculinized cases should be raised female. However, given the scarcity of long-term follow-up reports on such syndromes, the best cutoff point on the Quigley scale for assignment to the male or female gender cannot be arrived at on an empirical basis and what is known about the variability of gender outcome in individuals with a given molecular genotype and endocrine or genital phenotype lets one expect that any cutoff point on the Quigley scale will be associated with occasional later patient-initiated gender reassignment, however rare. Similar uncertainties also affect the decision-making in cases of micropenis [i.e., a fully differentiated penis with the urethral meatus at the tip, but a (stretched) length  $2\frac{1}{2}$  standard deviations or more below the norm]. Individual cases of satisfactory long-term outcome have been shown among female-assigned as well as male-assigned individuals, but unsatisfactory outcomes have also been documented in both types of assignments, and evidence-based decision-making requires substantially larger samples and more comprehensive assessments to weigh the advantages and disadvantages of either assignment. Also problematic are the relatively rare pubertal change syndromes such as  $5\alpha$ -reductase deficiency and  $17\beta$ -hydroxysteroid-dehydrogenase deficiency in 46,XY. At birth, such children appear more female than male and they are typically assigned to the female gender, unless diagnosed correctly. There are a number of clinical-medical reports from resource-poor countries showing that, in the absence of medical intervention, the majority of such individuals start virilizing dramatically during spontaneous puberty and later change to the male gender, but very few data are available on the long-term outcome of such cases when medical intervention is introduced at newborn age or at the beginning of puberty. Similarly, the long-term outcome of those few cases who reportedly have had early hormone replacement therapy to masculinize the genitalia is not yet known.

Most controversial is the gender assignment of 46,XY infants who presumably had a normal-male sex-hormonal environment *in utero* so that the brain was normally masculinized, but who, for nonhormonal reasons, had genital abnormalities, for instance, in cases of penile agenesis, cloacal exstrophy of the bladder, or traumatic loss of the penis due to a circumcision accident. Although there are some well-known individual cases of later patient-initiated gender reassignment to male, there are other cases where this has not occurred, and in general, the

long-term outcome data are much too limited for an evidence-based decision.

In the syndrome of true hermaphroditism, the gender assignment decision is based on considerations of the preponderance of masculine or feminine structures in both the external and internal genitalia and the level of anti-Müllerian hormone, but again there is uncertainty regarding the best cutoff point for male versus female assignment, and more long-term follow-up data are needed for the formulation of a consensus policy. In the case of “XX males” secondary to a translocation of the SRY locus onto an X chromosome and consequent differentiation of a male gonad, there is no doubt that such infants should be raised as boys. Although the rapidly expanding knowledge on multiple genes participating in the sexual differentiation of the gonads and the brain complicates the picture and leads to the discovery of additional genotypes associated with intersex syndromes, it can be expected that over the next decade these discoveries will help refine both the diagnostic picture and the prognostic picture, especially when complemented by data about sexual brain dimorphism from brain imaging.

## Conclusion

Overall, the available data from both earlier and newer studies can be tentatively summarized in the statement that, with the exception of the pubertal gender-change syndromes, the majority of intersex patients develop a gender identity commensurate with the assigned gender. Later initiation of gender change by an intersex person appears to be the more likely the more the prenatal and postnatal biological factors and the postnatal psychosocial factors push in the same cross-gender direction. It follows that the gender assignment of the newborn should be based on the best prognosis for the future psychosocial and psychosexual functioning of the patient in the given cultural context, taking into account everything that is known about the likely steroid effects on the brain in the syndrome in question and what is known from follow-up investigations.

## PSYCHOSOCIAL MANAGEMENT

### Introduction

The issues of gender assignment and genital surgery and the potential implications for the development of gender, sexuality, and parenthood—three highly salient, emotional, moral, and frequently controversial social issues—make intersex conditions a challenge not only for patients and their families, but also for

physicians and other professionals involved in their clinical care and psychosocial management. Moreover, the psychosocial management itself is in a period of flux, since the optimal-gender policy of the second half of the twentieth century has come under severe criticism, whereas the evidence on which to base systematic improvements where warranted leaves much to be desired and no systematic randomized trials of psychosocial management procedures for intersex patients have been conducted. In these circumstances, recommendations for specific treatment options need to be more tentative, and not only the providers, but also the parents and, when old enough, the patients themselves will be burdened by more uncertainties in their decision-making than desirable.

The professionals involved in the medical and psychosocial management of intersex patients, especially when gender assignment is problematic, may involve multiple disciplines and subspecialties, e.g., specialists in neonatology, pediatric endocrinology, pediatric urology, gynecology, genetics, genetic counseling, and mental health. To minimize the burden on intersex patients and their families, these professionals should form a smoothly operating team and follow a common policy, also with regard to criteria for decisions on gender. Parents of intersex newborns are typically in a state of high stress while the gender of their child remains uncertain, and most are not in a position to decide between conflicting professional opinions. Those team members—both medical and mental health staff—who have the most influence on decisions regarding gender assignment and genital surgery and those who most likely will be involved with the parents and patients over the coming years must keep up with medical advances, emerging new treatment options, the growing research on psychosocial outcomes, and the prognostic alternatives for both health and psychological development in the intersex field. To be fully effective, they need to acquire up-to-date information on all psychosocially relevant aspects of the patient’s syndrome and specific symptoms. For instance, some intersex syndromes may be associated with short stature or neuropsychological impairments, both of which have psychosocial consequences of their own.

### Gender Assignment and Reassignment

#### **Gender Assignment at Birth**

The identification of ambiguous genitalia in a newborn only occasionally constitutes a medical emergency. Yet, having a newborn of undetermined

gender is highly stressful for most parents and thus presents a psychosocial emergency and speedy processing of necessary medical tests and rapid decision-making with regard to gender assignment are important. Prolonged periods of nondecision are thought to run the risk of chronically ambiguous or inconsistent sex typing by the family, or of rejection of the child altogether.

Given the diverse factors that influence gender identity development in intersex patients, the decision on gender assignment required in a patient with highly ambiguous genitalia is not only based on medical criteria, but also needs to take into account the differential prognosis for psychological (including psychosexual and reproductive) functioning in either gender role as well as the prognosis for how well the family will be able to cope with either decision. The prognostic considerations take into account the various options of genital surgery and sex-hormone treatment and their respective implications for future functioning, especially in terms of gender-role fit, sexuality, reproduction, and overall quality of life. As gender-related values differ in various societies, cultural factors have significant impact on gender assignment decisions. In many Asian countries, for example, gender assignment to male is more strongly favored, and infertility in the female more stigmatized, than in the West.

Intersex counseling at this stage should involve both parents and the inclusion of other family members should be considered. Legally, gender assignment is the parents' decision, but they are likely to rely on expert medical advice. Along with medical education about the nature and origin of the problem and the medical tests involved, the parents must be adequately informed by their clinicians about the diversity of long-range outcomes and not given overly optimistic assurances. Commensurate with their cognitive capacity, the parents need to be made aware of significant gaps in scientists' knowledge about intersex disorders or major controversies about management decisions. Such parent counseling requires a careful balance between providing gross oversimplification on the one hand and too much detail on the other, with a resulting paralysis of decision-making. If either parent is not convinced that the gender decision was correct, appropriate child rearing may be in jeopardy.

### **Monitoring Behavioral Development**

The psychosocial management needs of intersex patients vary with developmental stage. Once the diagnosis is established and the decision on gender assignment made, an approximate timetable for future

medical procedures and preventive psychosocial measures can be projected and visits planned accordingly. Again, a team approach in which medical procedures and psychosocial management are closely integrated is highly recommended. Active outreach by the mental health professional may facilitate prevention of psychosocial problems and adherence to medical treatment.

As for other congenital disorders, regular monitoring of behavioral development is also recommended for intersex children. This includes monitoring for medical problems and treatment compliance, as needed, and monitoring for cognitive and behavioral problems, with a special emphasis on atypical gender-role behavior. Some parents (and even some children) need help with handling atypical gender-role behavior. If symptoms of gender dysphoria emerge, parents and the child may need some help in dealing with gender issues and in persistent extreme cases the option of gender reassignment needs to be considered.

### **Gender Reassignment**

Sometimes, intersex children are initially misdiagnosed, especially when born outside of intersex-experienced medical settings, and receive their correct diagnosis later from medical specialists. The definitive diagnosis may alter the prognostic considerations and make a gender reassignment desirable. Such reassignment, if well justified, is usually not a problem during infancy, provided the parents are adequately counseled.

After infancy, during the toddler and preschool years, many intersex children develop some degree of gender-atypical behavior, but these rarely indicate a problem of gender dysphoria. Parents who are anxious about their child's gender atypicality may need reassurance and counseling. Gender reassignment decisions after infancy, which are quite uncommon, should never be based on purely medical considerations, but require careful psychological evaluation (over a prolonged period of time) of the child's overall behavioral development, with particular attention to the child's gender-role behavior and to any symptoms of gender dysphoria, as well as to the behavioral niche the child has occupied in the family system.

Gender dysphoria and wishes to change one's gender may also emerge later in adolescence or adult life, usually as the result of a long and gradual process. Even if the problem is of long standing, through counseling the patient can consider various options, and gender change and the attendant hormone treatment and surgical procedures are not the invariable outcome.

## Genital Surgery

Some children with intersexuality require genital surgery for medical reasons. In the majority of cases, however, genital surgery has been performed for psychosocial reasons in order to confirm the assigned gender by genital appearance and to thereby facilitate gender-appropriate rearing, help develop a gender-typical body image, and avoid social stigmatization and also to facilitate later peno-vaginal intercourse. Because genital surgery is associated with some risk to sexual tissues and erotic innervation and thereby sexual functioning, a vigorous debate has ensued as to whether such nonmedically necessitated surgery should be delayed until the intersex child is old enough to give informed consent. Surgeons argue that modern surgical techniques are much improved, but it may still require years until the patients so treated are old enough to provide data on sexual functioning. Experienced clinicians are concerned that obviously ambiguous genitalia put an unoperated intersex child at risk of undue attention and stigmatization. Moreover, in the absence of sufficient sexual experience, even an older adolescent or young adult may not be capable of giving appropriately informed consent, and the very fact of genital ambiguity may contribute to the delay of sexual initiation and some patients may never reach that developmental stage. Thus, there is as yet no consensus regarding the issue of early genital surgery, except to say that milder cases of genital ambiguity are less likely to be operated on than they would have been one or two decades ago and that there is a growing consensus that genital surgery should be confined to intersex-experienced centers of excellence.

If the decision for genital surgery has been made, both medical and psychological considerations determine the choice of time. From a psychological perspective, genital surgery is performed more easily in infancy, when counseling for the child is not an issue, and in adolescence, when cognitive maturation facilitates counseling and the patient has achieved a degree of autonomy, than in early and middle childhood. The older the child, the more he or she should be empowered to have the decisive vote in the decision for or against genital surgery and in choosing the time when it should take place.

Before and after genital surgery, genital examinations of intersex children and adolescents are frequent. Given the low prevalence rates of the various syndromes of intersexuality, it has been common medical practice to have many different physicians, especially medical students and residents, perform genital examinations on the same patient. However,

considerable evidence has accumulated showing that this practice has negative and often severe psychological aftereffects. In medical training, this practice should be replaced by the use of photographs, videotapes, and physical models. Also, older children and teenagers should have a say in whether someone, and, if yes, who, should accompany them during the examination.

In adolescence and adulthood, psychosocial monitoring needs to include concerns about romantic relationships and sexual functioning, because of the increased problems associated with genital ambiguity and related surgery.

## Sex-Hormone Treatment

During infancy, androgen treatment is sometimes used for chromosomally male children with underdeveloped male genitals to test for an androgen-receptor defect or, if the assignment decision was for male, to enhance penile size. Occasionally, such children show behavioral change that normalizes with the end of the treatment and the parents may need reassurance.

Intersex children who have no gonads or underfunctioning gonads will need sex-hormone replacement therapy to initiate or support pubertal maturation. This is best done at the age when their classmates experience endogenous puberty. Undue delay of pubertal development may add to difficulties in the later development of romantic and sexual relationships.

## Information Management

The medical and sexual education of intersex children and their parents and related counseling (including information about how to deal with relatives and friends) is a crucially important part of psychosocial management. The most difficult issue is the disclosure of their medical status and history to the intersex patients themselves, especially to those assigned to a gender discrepant with sex chromosomes and/or gonadal structure. Professionals generally support the right of patients to have access to their medical information and many experienced professionals agree that by the time of graduation from high school the patient should be fully informed, commensurate with the level of cognitive capacity. However, many parents are anxious and would like to postpone full disclosure, sometimes permanently. Appropriate timing of disclosure is important. Early disclosure runs into the problem of cognitive limitations. Late disclosure implies cover-up and deception and may lead to the

patient's long-term distrust of and anger at medical professionals and even endanger appropriate future utilization of medical services. Disclosure is best conceptualized as a long-term process starting in late preschool age, with many installments, preferably in conjunction with meaningful events. Substantively, there are the issues of past genital surgery and their tell-tale signs, namely, scars on the abdomen, possibly persisting abnormalities in genital appearance, and plans for future surgery; the need for the induction of puberty by sex-hormone treatment and the maintenance of secondary sex characteristics by its lifelong continuation; the issue of infertility or, better, the potential realization of parenthood by means other than pregnancy or insemination; the question of variability in gender-role behavior among individuals of the same core gender identity; and the potential for romance, sexuality, and sexual pair-bonding—regardless of sexual orientation. All of these issues of functional potential are more important to the average developing intersex person than the details of the molecular structure of genes, the variants of hormone synthesis, receptor defects, enzyme abnormalities, and the organization and activation of neural networks—although the occasional adolescent or adult patient may have specialized interests or even choose a career path in the respective biological sciences and may seek as complete an understanding of his or her biological condition as is possible.

For the developing child and adolescent, the medical information must be carefully tailored to the cognitive maturity level. The fact that the meaning and connotations of terms differ vastly between medical personnel and lay persons makes a deliberate and cautious choice of terminology necessary. The intended medical “truth” may be dramatically at variance with what the patient or parents perceive. Visual aids are often very helpful.

For both the patients and the parents, discussing the medical aspects of intersexuality tends to be highly emotional, so adequate retention is problematic. In the author's unit, the ideal procedure for the medical education of a newly referred family is that the physician in charge of the overall coordination of care, usually a pediatric endocrinologist or urologist, gives a carefully worded summary of the medical information to the parents, in the presence of the mental health professional, who then takes over. Typically, the parents are asked to recount in detail how they understand the medical condition, its origin, and its prognosis. Then, the instruction is repeated, the

parents' misunderstandings are corrected, and their particular concerns are addressed. At subsequent visits, the procedure is repeated as necessary, keeping in mind that new misunderstandings and different concerns develop over time. Similar procedures are used with intersex children, once they are old enough, except that, in addition, the clinician lets them explain to their parents (in the clinician's presence) what they have learned from the instruction, to open the channels of communication.

## Support Groups

One of the recurrent problems of persons suffering from rare disorders and their families is the feeling of isolation and having no one to talk to who really understands what they must cope with. Many support groups have sprung up, including groups specific for intersex persons and/or their parents, and the internet has greatly aided in this development. Generally, such groups have been a great help to many patients and their families. However, the quality of the groups is highly variable and the medical and other information that comes from them is sometimes quite misleading. Thus, frequent exchanges between support groups and professionals are recommended and group participants should be strongly encouraged to seek second opinions from their physicians and other care providers regarding the information they receive.

## See Also the Following Articles

Androgen Insensitivity Syndrome • Congenital Adrenal Hyperplasia, Prenatal Diagnosis and Therapy • Endocrine Disruptors and Male Sexual Differentiation • Genes and Gene Defects Affecting Gonadal Development and Sex Determination • Sexual Maturation, Female • Sexual Maturation, Male

## Further Reading

- Cohen-Kettenis, P. T., and Pfäfflin, F. (2003). Atypical sexual differentiation. *In* “Transgenderism and Intersexuality in Childhood and Adolescence: Making Choices” (P. T. Cohen-Kettenis and F. Pfäfflin, eds.), Chap. 3, pp. 23–49. SAGE Publications, Thousand Oaks, CA.
- Cohen-Kettenis, P. T., and Pfäfflin, F. (2003). Clinical management of intersex conditions. *In* “Transgenderism and Intersexuality in Childhood and Adolescence: Making Choices” (P. T. Cohen-Kettenis and F. Pfäfflin, eds.), Chap. 5, pp. 85–104. SAGE Publications, Thousand Oaks, CA.
- Daaboul, J., and Frader, J. (2001). Ethics and the management of the patient with intersex: A middle way. *J. Pediatr. Endocrinol. Metab.* **14**, 1575–1583.

- Meyer-Bahlburg, H. F. L. (1999). Variants of gender differentiation. *In* "Risks and Outcomes in Developmental Psychopathology" (H.-C. Steinhausen and F. C. Verhulst, eds.), pp. 298–313. Oxford University Press, Oxford, UK.
- Meyer-Bahlburg, H. F. L. (2002). Gender assignment and reassignment in intersexuality: Controversies, data, and guidelines for research. *In* "Pediatric Gender Assignment: A Critical Reappraisal" (S. A. Zderic, D. A. Canning, M. C. Carr, and H. M. Snyder, eds.), pp. 199–223. Kluwer/Plenum, New York.
- Money, J. (1994). "Sex Errors of the Body and Related Syndromes: A Guide to Counseling Children, Adolescents, and Their Families," 2nd ed. Paul H. Brookes, Baltimore, MD.
- Ruble, D. N., and Martin, C. L. (1998). Gender development. *In* "Handbook of Child Psychology" (W. Damon, ed.) 5th ed., Vol. 3, pp. 933–1016. Wiley, New York.
- Wilson, J. D. (1999). The role of androgens in male gender role behavior. *Endocr. Rev.* **20**, 726–737.
- Zucker, K. J. (1999). Intersexuality and gender identity differentiation. *Annu. Rev. Sex Res.* **10**, 1–69.

Table 1 Genes Involved in Mammalian Gonadal Development and Sex Determination

Gene	Chromosomal localization		Gene function	Mutation phenotype <sup>a</sup>
	Human	Mouse		
<i>WT1</i>	11p13	2	Transcription factor	XY male-to-female sex reversal Denys-Drash and Frasier syndromes
<i>SF1/AD4BP/NR5A1</i>	9q33	2	Transcription factor	XY male-to-female sex reversal Gonadal dysgenesis (mouse KO)
<i>LIM1/LHX1</i>	11p13	11	Transcription factor	Gonadal dysgenesis (mouse KO)
<i>LHX9</i>	1q31	1	Transcription factor	Gonadal dysgenesis (mouse KO)
<i>EMX2</i>	10q26	19	Transcription factor	Gonadal dysgenesis (mouse KO)
<i>SRY</i>	Yp11	Y	Transcription factor	XY gonadal dysgenesis
<i>SOX9</i>	17q24.3–25.1	11	Transcription factor	XY male-to-female sex reversal Campomelic dysplasia
<i>DMRT1</i>	9p24	19	Transcription factor	XY male-to-female sex reversal Multigene deletion in chromosome 9 encompassing <i>DMRT1</i> gene
<i>GATA4</i>	8p23.1–22	14	Transcription factor	XY gonadal dysgenesis (mouse mutation)
<i>FOG2</i>	8q23	15	Transcription factor	XY gonadal dysgenesis (mouse KO)
<i>DAX1</i>	Xp21.3–p21.2	X	Transcription factor	Gene duplication: XY gonadal dysgenesis Abnormal testis cord formation (mouse KO)
<i>M33</i>	17q25	19	Transcription factor	Delay in genital ridge formation and XY male-to-female sex reversal (mouse KO)
<i>FGF9</i>	13q11–q12	14	Signaling molecule	XY male-to-female sex reversal (mouse KO)
<i>WNT4</i>	1p35	4	Signaling molecule	XY gonadal dysgenesis in individuals carrying a duplication of a portion of chromosome 1 encompassing the <i>WNT4</i> gene XX masculinization (mouse KO)
<i>AMH/MIS</i>	19p13.3–p13.2	10	Signaling molecule/hormone	Persistent Müllerian duct syndrome in XY individuals

<sup>a</sup>Human data unless indicated otherwise.

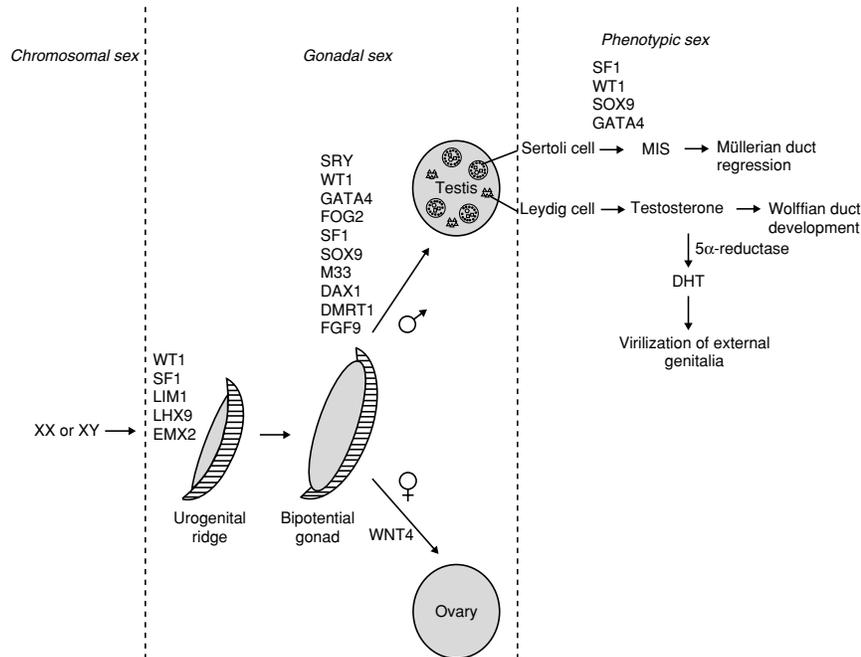


Figure 1 Molecular determinants of mammalian gonadal development and sex determination. DHT, dihydrotestosterone.

## GENES INVOLVED IN EARLY GONADAL DEVELOPMENT

A number of genes have been shown to be crucial for the earliest steps of gonadal morphogenesis. These include *WT1*, *SF1*, *LIM1/LHX1*, *LHX9*, and *EMX2*. Since these genes act prior to sex determination, they are required for early gonadal development in both sexes. Mice deficient in these genes either fail to develop genital ridges or have an early block in genital ridge development. Human mutations involving the *LIM1*, *LHX9*, and *EMX2* genes have not yet been linked to cases of human gonadal agenesis or dysgenesis. The *WT1* and *SF1* genes merit further discussion since these genes are known to be required for both early gonadal development and sex differentiation.

### *WT1*

The *WT1* gene, encoding a DNA-binding protein containing four zinc-fingers, was first identified as a tumor suppressor gene via genetic studies of Wilms' tumor patients in humans. Mutations in the *WT1* gene in humans were shown to be responsible for both Denys-Drash syndrome and Frasier syndrome, characterized by anomalies in kidney development and gonad formation including sex reversal. Mice homozygous for a null mutation in the *Wt1* gene show a phenotype of renal and gonadal agenesis and adrenal dysgenesis, demonstrating that *Wt1* is required for genital ridge formation prior to the events of sex determination. It is evident that the *WT1* gene can result in multiple protein isoforms via posttranscriptional processes including alternative splicing; these isoforms can function not only as transcription factors but also as RNA-processing factors. Important among these isoforms is the presence of a 3-amino-acid insertion (+KTS) between the third and fourth zinc-fingers or the absence of the insertion (-KTS). The *WT1* (-KTS) isoform functions as a transcription factor that can activate the *AMH* and *DAX1* promoters as well as the *SRY* promoter *in vitro*. The *WT1* (+KTS) isoform shows little transactivation activity but has been implicated in RNA-processing events. Isoform-specific null allele mouse studies have been reported and they show that the *Wt1* (+KTS) null allele results in a phenotype of XY sex reversal. This indicates that *WT1* (+KTS) isoforms as well as RNA processing are important contributors to the molecular mechanisms leading to sex determination.

### *SF1*

Steroidogenic factor 1 (SF1; also known as Ad4BP and NR5A1) is a transcription factor belonging to the nuclear receptor superfamily. SF1 was first identified as an essential regulator of the P450 hydroxylases in the adrenals and gonads. Later studies showed that SF1 is also present in the pituitary and ventromedial nucleus of the hypothalamus (VMH). In agreement with its expression pattern, *Sf1* knockout mice lack adrenal glands and gonads, exhibit VMH and pituitary gonadotrope abnormalities, and display female internal genitalia. At the transcriptional level, SF1 regulates the expression of many genes involved in reproduction, steroidogenesis, and sexual differentiation. An important SF1 target for sexual differentiation is the anti-Müllerian hormone gene (*AMH/MIS*) involved in Müllerian duct regression in the developing male embryo. In humans, the role of SF1 in male sex differentiation is supported by the identification of *SF1* gene mutations in a subpopulation of 46,XY sex-reversed patients: a dominant *de novo* heterozygous G35E mutation, a recessive homozygous R92Q mutation, and a heterozygous deletion of eight nucleotides causing a frameshift mutation and C-terminal truncation of the SF1 protein. For two of these *SF1* mutations (G35E and R92Q), the presence of normal Müllerian structures and the retention of a uterus in the affected individuals is indicative of insufficient AMH production. This suggests that the mutated SF1 proteins might interfere with the expression of a *Sf1* target gene(s) involved in the male sexual differentiation pathway, including *AMH*.

## GENES INVOLVED IN SEX DETERMINATION AND DIFFERENTIATION

The era of molecular genetics of mammalian sex determination was launched more than a decade ago with the identification and cloning of the elusive Y chromosome-linked testis determining gene, termed *SRY*. Since then, a multitude of additional genes have been described as being involved in testis differentiation; these include *SOX9*, *DMRT1*, *GATA4*, *FOG2*, *DAX1*, and *FGF9*.

### *SRY*

The first major advance in the molecular characterization of mammalian sex determination was reported

in 1990 with the positional cloning of the *SRY* gene, a candidate for the testis-determining factor (TDF) located on the Y chromosome. The equivalence of *SRY* and TDF was subsequently demonstrated via gain-of-function studies in the transgenic mouse, whereby a transgene coding for Sry caused an XX animal to develop with a male phenotype. Biochemical studies showed that *SRY* can bind to DNA via a structural motif termed an HMG-box and mutations in the HMG-box were shown to be of clinical relevance in cases of XY females in the human population. Structural comparisons of the *SRY* gene between mammalian species revealed a notable lack of sequence conservation. Although this is unremarkable for a gene located on the Y chromosome, it was not anticipated for such a key player in the mammalian sex determination process. Also surprising is the fact that after 13 years of study, no direct target genes of *SRY* have yet been identified, and the mechanisms by which *SRY* causes the activation of male-specific gene expression, cell proliferation, and cell migration within the male genital ridge remain obscure. However, some progress has been made in deciphering *SRY* promoter function, with *in vitro* evidence that WT1, SF1, SOX9, and GATA4 can activate *SRY* gene transcription. Genetic evidence suggests that DAX1 competes with *SRY* activity and additional *in vitro* evidence suggests that DAX1 inhibits the function of proteins that activate *SRY* activity.

### SOX9

The description of the HMG-box of *SRY* allowed the cloning of a family of related HMG-box-containing genes, termed SOX (SRY-related HMG-box) genes. One of these new SOX family members, *SOX9*, was associated via positional cloning with the musculo-skeletal disease campomelic dysplasia and also with autosomal sex reversal in humans. Demonstration of *SOX9* gene expression within the male genital ridge at approximately the time of sex determination in mammals as well as in birds suggested a conserved role for *SOX9* in sex determination and/or sex differentiation. Transgenic expression of *SOX9* within the genital ridge in the absence of *SRY* resulted in a male phenotype within XX animals, suggesting that *SOX9* expression is sufficient for sex determination and can account for downstream functions of *SRY* expression. Like *SRY*, *SOX9* contains a DNA-binding HMG-box domain, but unlike *SRY*, it also contains a recognizable transactivation domain. A key event in sex differentiation is the activation of the *AMH* gene, which codes for a secreted protein of the TGF- $\beta$  family.

*Sox9* has been shown by both *in vitro* and *in vivo* studies to transactivate the *Amb* promoter. *In vitro* evidence suggests that at least in some species, *SOX9* can also transactivate the *SRY* promoter, placing it both upstream and downstream of sex determination. Whether *SOX9* is necessary for sex determination per se, or can be compensated for by other SOX genes known to be expressed within the genital ridge, awaits further experimentation.

### DMRT1

DMRT1 was initially identified and cloned via a conserved DNA-binding domain sequence (DM domain) identified within the *Drosophila melanogaster* Doublesex gene, the *Caenorhabditis elegans* Mab-3 sex regulator gene, and human expressed sequences derived from a testis cDNA library. The chromosomal location of human *DMRT1* (9p) makes it a candidate gene responsible for XY gonadal dysgenesis seen with 9p deletions in humans. This region is syntenic with the chicken Z chromosome, making chicken *DMRT1* a candidate for a dosage-sensitive sex determination system in birds. Unlike most other genes involved in sex determination and differentiation that show fairly wide tissue expression patterns, *DMRT1* expression is restricted to the gonads. Furthermore, *DMRT1* expression is up-regulated concurrently with testis development, not just in humans and mice but also in other vertebrates. Male mice that are deficient for the *Dmrt1* gene display problems with cell differentiation and cell survival in the postnatal testis, but these animals undergo normal sex determination. These results could indicate that DMRT1 is simply not involved in sex determination or that functional redundancy can compensate for the absence of DMRT1; there are seven *DMRT* genes in mouse and humans. A DMRT family member (*Dmrt1bY*) was proposed as the sex-determining factor in the Medaka fish; however, this gene was found to be absent in closely related fish species. Thus, although DMRT1 sequences are widely distributed within the animal kingdom and are unquestionably implicated in gonadal development, a universal involvement of DMRT1 sequences in sex determination does not appear likely. Where mammalian *DMRT1* integrates with other genes known to be involved in mammalian sex determination and differentiation pathway also remains to be determined.

### GATA4

GATA4 is one of six proteins belonging to the GATA family of transcription factors. GATA factors are

named for the nucleotide sequence (WGATAR) that they bind to in the promoter regions of target genes. They were originally identified as crucial regulators of cardiac development and hematopoietic cell differentiation. *GATA* expression, however, is not limited to these two systems. In the mouse, the *Gata4* gene is abundantly expressed in the somatic cell population of the developing genital ridge prior to and during the time of *Sry* expression. Thus, based on its expression pattern, GATA4 was proposed to play a central role in early gonadal development and sex determination. This hypothesis has been confirmed in the mouse, where *in vivo* disruption of GATA4 function via a mutation of the *Gata4* gene leads to a block in testis development and a marked down-regulation of *Sry* expression. Thus, GATA4 appears to function as a direct upstream regulator of *SRY* expression in the developing testis. Although the latter has yet to be conclusively demonstrated, the presence of multiple GATA regulatory motifs in the mouse, human, and pig *SRY* promoters strongly supports this notion. After gonadal differentiation, GATA4 has been shown to be recruited as an important regulator of the *AMH* gene involved in Müllerian duct regression and, consequently, male sexual differentiation. GATA4 has been shown to regulate both the mouse and human *AMH* promoters through direct transcriptional cooperation with SF1. Although no human *GATA4* gene mutations have yet been linked to gonadal defects, evidence suggests that disruption of GATA4/SF1 synergism may account for some cases of human male-to-female sex reversal involving insufficient *AMH* expression.

## FOG2

GATA factors regulate the expression of target genes through cooperative interactions with other transcription factors. These include the Friend of GATA proteins (FOG1 and FOG2), which were cloned as GATA-specific cofactors. Although the FOG proteins do not directly bind to DNA, they have been shown to act as either enhancers or repressors of GATA transcriptional activity depending on the cell and promoter context being studied. In this regard, it has been suggested that the FOG proteins act as bridging molecules that link GATA proteins with other factors involved in either activation or repression. In the mouse, *Fog2* is coexpressed with *Gata4* in the developing genital ridge. Like *Gata4*, *Fog2* is likely expressed upstream of *Sry* in the sex determination pathway since *Fog2* knockout mice have markedly diminished *Sry* expression and do not develop a normal testis.

## DAX1

The X chromosome-linked *DAX1* [dosage-sensitive sex reversal-adrenal hypoplasia congenital (AHC) critical region on the X chromosome, gene 1; NR0B1] gene encodes an atypical member of the nuclear receptor superfamily. *DAX1* is expressed in several endocrine tissues, including the gonads, where it is present in the Sertoli and Leydig cells of the testis and the granulosa and theca cells of the ovary. *DAX1* plays a critical role in adrenal development and gonadal function; it was originally identified as the causative gene for AHC associated with hypogonadotropic hypogonadism (HHG), a recessive disorder affecting males. Indeed, several different groups have characterized numerous *DAX1* mutations (nucleotide substitutions, insertions, and deletions) in individuals affected with AHC and HHG. Females carrying heterozygous mutations in the *DAX1* gene are normal and there has been no report of homozygous females since males needed to transmit the nonfunctional allele are infertile. Since *Dax1* expression is down-regulated in the developing male gonad after overt testis differentiation in the mouse, the *Dax1* gene was initially believed to be an important ovarian determinant. However, female *Dax1* knockout mice are phenotypically normal. In contrast, male *Dax1* knockout mice are infertile due to a loss of germ cells. Thus, *Dax1* is apparently not required for ovarian differentiation but rather is essential for the maintenance of the integrity of the seminiferous epithelium of the testis. Closer examination of *Dax1*-deficient mice has revealed that *Dax1* is also essential for testis cord formation and, consequently, testis determination. In addition to *DAX1* mutations, another human disorder has been associated with the *DAX1* gene. In this case, duplication of a small region on the X chromosome that encompasses the *DAX1* gene leads to dosage-sensitive sex reversal in XY males. Overexpression of *Dax1* in transgenic mice has also been shown to induce sex reversal in males carrying a weakened *Sry* allele. Thus, adequate *Dax1* levels are required for normal testis development and function, but too much *Dax1* appears to have an “anti-testis” effect. In the female developing gonad, *Dax1* transcription appears to be up-regulated by *Wnt4*, a member of the *Wnt* family of signaling molecules. *Wnt4* signaling has been shown to function via activation of  $\beta$ -catenin and its subsequent interaction with SF1 on the *Dax1* promoter. In the developing male gonad, the action of *Wnt4* has been suggested to be inhibited by *Sry*.

## AMH

In mammals, male sexual differentiation is regulated by two hormones produced by the fetal testis: testosterone, secreted by Leydig cells, and AMH, produced by Sertoli cells. In the male embryo, AMH induces regression of the Müllerian ducts (the anlagen of the internal female reproductive tract). AMH is the earliest marker of testis formation; it is found in Sertoli cells of the developing human testis starting at approximately 8 weeks of gestation. In the mouse, Amh is first detected in Sertoli cells on embryonic day 12.5; its expression remains high throughout fetal life before declining markedly after birth. Consistent with its role in male sexual differentiation, the absence of *AMH* expression in humans causes persistent Müllerian duct syndrome, a form of pseudo-hermaphroditism characterized by the retention of Müllerian duct structures. The transcriptional regulation of the *AMH* gene has been intensely studied. Interestingly, several transcription factors involved in primary sex determination (SF1, WT1, SOX9, and GATA4) are later recruited as important regulators of *AMH* expression. Although many of these factors have been shown to activate *AMH* transcription on their own by directly binding to the *AMH* promoter, functional cooperation between these factors appears to be of paramount importance in directing the proper spatiotemporal expression of the *AMH* gene.

## OTHER GENES INVOLVED IN GONADAL DEVELOPMENT AND SEX DETERMINATION

Ever since the discovery of *SRY*, the majority of newly identified genes involved in gonadal development and sex determination have been transcription factors. Signaling molecules such as FGF9 and WNT4 have also been implicated in these processes. In the mouse, Fgf9 was shown to be involved in the induction of mesonephric cell migration into the XY gonad, an essential step required for seminiferous cord formation and, hence, proper testis organization. *Wnt4* expression is detected in the genital ridge of both sexes but its expression becomes largely restricted to the ovary after overt gonadal differentiation. *Wnt4*-deficient female mice are masculinized due to the expression of genes involved in testosterone biosynthesis that are normally not expressed in the fetal ovary. Much like *DAX1*, overexpression of WNT4 in humans has been reported to cause male-to-female sex reversal; this

anti-testis effect of too much WNT4 is thought to occur via up-regulation of *DAX1*.

## PERSPECTIVES

Although significant progress has been made in elucidating how the different genes in the gonadal development and sex determination pathway come together to form a complex regulatory network, it remains evident that many pieces of the genetic puzzle are still missing. Indeed, the majority of the cases of human XY sex reversal and approximately one-quarter of the cases of XX maleness have not yet been defined at the genetic level. Technologies for large-scale gene identification based on differential expression have the potential for making important inroads into finding new genes in this fundamental developmental process. For example, expression-based strategies such as gene microarray technology have begun to be successfully applied to the discovery of genes involved in testis and ovary development. The *vanin 1* gene, which encodes a glycosylphosphatidylinositol-anchored cell surface protein, is an example of a gene identified through the sex-specific screening process. Clearly, other genes are likely to be identified in coming years. This is especially true for genes that might control ovarian development, an area that has been long overlooked by the drive to study testis differentiation.

## See Also the Following Articles

Agonadism, Male and Female • Androgens, Gender and Brain Differentiation • Anti-Müllerian Hormone • Delayed Puberty and Hypogonadism, Male • Endocrine Disruptors and Male Sexual Differentiation • Sexual Maturation, Female • Sexual Maturation, Male

## Further Reading

- Cotinot, C., Pailhoux, E., Jaubert, F., and Fellous, M. (2002). Molecular genetics of sex determination. *Semin. Reprod. Med.* **20**, 157–167.
- Jameson, J. L., Achermann, J. C., Ozisik, G., and Meeks, J. J. (2003). Battle of the sexes: New insights into genetic pathways of gonadal development. *Trans. Am. Clin. Climatol. Assoc.* **114**, 51–63.
- Novartis Foundation Symposium (2002). “The Genetics and Biology of Sex Determination” (D. Chadwick and J. Goode, eds.). Wiley, West Sussex, UK.
- Parker, K. L., and Schimmer, B. P. (2002). Genes essential for early events in gonadal development. *Ann. Med.* **34**, 171–178.
- Tilmann, C., and Capel, B. (2002). Cellular and molecular pathways regulating mammalian sex determination. *Rec. Prog. Horm. Res.* **57**, 1–18.

unclear because autosomal dominant disorders usually occur as a result of either haploinsufficiency or secondary to dominant negative activity. IGHD-3 may be associated with hypogammaglobulinemia and is inherited as an X-linked disorder. No mutations in *GHI* or other candidate genes have been identified for this type.

Mutations causing MPHD have been identified in the genes encoding pituitary transcription factors that direct the embryonic development of the anterior pituitary gland, including *POU1F1* (formerly referred to as *Pit-1*), *PROPI*, *HESX1*, and *LHX3* and *LHX4*. Autosomal recessive and dominant mutations of *POU1F1* (POU domain, class 1, transcription factor 1), which is located on chromosome 3p11, result in deficiency of GH, prolactin, and thyroid-stimulating hormone (TSH); adrenocorticotropin hormone (ACTH) and the gonadotropins are spared. Autosomal recessive mutations of *PROPI* (*Prophet of Pit-1*), located on chromosome 5, cause deficiency of GH, prolactin, and TSH, as well as the gonadotropins. ACTH is spared initially, but there is a tendency toward the development of deficiency with increasing age. Mutations of *HESX1* (also referred to as *RPX*), a pituitary transcription factor that also plays a role in the development of the optic nerves, have been implicated in septo-optic dysplasia. This is a heterogeneous disorder characterized by midline neurological abnormalities associated with pituitary hypoplasia and optic nerve hypoplasia. Mutations of *LHX3* have been identified in humans with abnormal neck and cervical spine development in whom there is a deficiency of all anterior pituitary hormones except ACTH. Mutations have been identified in *LHX4* in a family with GH, TSH, and ACTH deficiency in combination with cerebellar and skull-base malformations.

Potentially, inactivating mutations of SST, or one of the five specific G protein-coupled receptor subtypes (SSTR1–5) to which somatostatin binds, could lead to loss of inhibition and subsequent excess GH production. Although a mutation of the *sst5* gene has been demonstrated in a patient with a GH-secreting pituitary adenoma, the simultaneous presence of a *gsp* mutation makes the role of the *SSTR5* mutation unclear.

#### **Activating Mutations**

GHRHR signals through a  $G_{s\alpha}$ -containing G protein. Activating somatic mutations of the gene encoding the  $G_{s\alpha}$  subunit give rise to the *gsp* oncogene. These *gsp* mutations have been identified as the molecular cause of 30–40% of GH-producing pituitary tumors. They have also been found in GH-producing tumors, which occur in patients with McCune–Albright

syndrome. The mutations result in continuous and markedly elevated levels of intracellular cAMP, with subsequent increases in the secretion of GH and the proliferation of somatotrophs.

#### **Thyroid-Stimulating Hormone**

TSH is a glycoprotein hormone that consists of  $\alpha$  and  $\beta$  subunits. The  $\alpha$  subunit is common to TSH, luteinizing hormone (LH), follicle-stimulating hormone (FSH), and human chorionic gonadotropin, whereas the  $\beta$  subunit in each of these hormones is unique and confers specificity of action. Production of TSH is stimulated by thyrotropin-releasing hormone (TRH) and inhibited by triiodothyronine.

#### **Inactivating Mutations**

Central hypothyroidism, which is a rare disorder characterized by insufficient production of TSH by the anterior pituitary, occurs most commonly as part of MPHD. Rare cases of isolated TSH deficiency have been described, typically resulting from missense, nonsense, or frameshift mutations of the TSH- $\beta$  gene on chromosome 1; a nonsense mutation of the *TRHR* gene has also been identified in a patient with isolated central hypothyroidism. Central hypothyroidism secondary to the production of a mutant TSH molecule has been demonstrated in two families in which the affected members demonstrated thyroid hormone levels consistent with hypothyroidism but TSH levels that were inappropriately in the normal range.

#### **Activating Mutations**

TSH-secreting pituitary adenomas are a rare cause of central hyperthyroidism. Potential molecular causes that have been investigated, such as activating mutations of the  $G_{s\alpha}$  subunit, *TRHR* and *POU1F1*, have not been identified.

#### **Adrenocorticotropin Hormone**

ACTH is a peptide hormone derived from a precursor polypeptide, pro-opiomelanocortin (POMC). The POMC gene encodes several peptides, including an N-terminal peptide,  $\beta$ -lipotropin, and ACTH. Synthesis of ACTH is primarily stimulated by hypothalamic corticotropin-releasing factor (CRF) through its interaction with the CRF receptor, another G protein-coupled receptor.

#### **Inactivating Mutations**

Although rare cases of isolated ACTH deficiency have been described, deficiency of this hormone usually

occurs as part of MPHD. ACTH deficiency can be simulated by isolated glucocorticoid deficiency because there is resistance to the action of ACTH at the level of the adrenal cortex. In approximately half of cases identified with this disorder, there is a demonstrable mutation in the ACTH receptor (also known as the melanocortin-2 receptor). This condition, referred to as GCCD1, is inherited in an autosomal recessive manner. Subjects are biochemically characterized by low serum cortisol and high ACTH levels, with normal mineralocorticoid activity.

### Activating Mutations

The prevalence of *gsp* mutations in ACTH-secreting pituitary adenomas has been demonstrated to be less than 10%.

## Luteinizing Hormone and Follicle-Stimulating Hormone

The pituitary gonadotropins LH and FSH, which belong to the family of glycoprotein hormones, are produced by the anterior pituitary in response to stimulation by the hypothalamic peptide gonadotropin-releasing hormone (GnRH). The  $\alpha$  subunit, common to both LH and FSH, is encoded by a gene on chromosome 6q12.21. The hormone-specific LH  $\beta$  subunit is encoded by the *LH $\beta$*  gene on chromosome 19q13.32, whereas the FSH  $\beta$  subunit is encoded by the *FSH $\beta$*  gene on chromosome 11p13.

### Inactivating Mutations

Five subjects with isolated FSH deficiency have been reported with inactivating mutations of *FSH $\beta$* . Two different mutations (a 2-base pair deletion and a missense mutation) were described in three women who presented with primary amenorrhea, poorly developed secondary sexual characteristics, and infertility. Mutations of *FSH $\beta$*  have also been reported in two men, one of whom presented with azoospermia but normal puberty, and the other with azoospermia and delayed puberty. These mutations consisted of the same 2-base pair deletion detected in the females plus a new missense mutation. Only one homozygous missense mutation causing inactivation of *LH $\beta$*  has been reported in a male, who presented with delayed puberty. No mutations of the human  $\alpha$  subunit gene have been described in the literature.

Hypogonadotropic hypogonadism (HH), which results from deficiency of both gonadotropins, can occur in isolation, in association with anosmia such

as in Kallmann's syndrome (KS), or with adrenal insufficiency such as in adrenal hypoplasia congenita. Inactivating mutations of the gene encoding the GnRH receptor (*GnRHR*), a member of the G protein-coupled receptor family located on chromosome 4q21.2, have been reported in isolated HH.

The defects in KS and adrenal hypoplasia congenita both occur at the hypothalamic rather than pituitary level. Briefly, KS is a form of HH that is associated with anosmia/hyposmia. X-linked recessive KS results from mutations of the *KAL1* gene that is located in the pseudo-autosomal region of Xp. This gene encodes the KAL protein (also known as anosmin), which plays a central role in the migration of both GnRH and olfactory neurons during embryonic development. However, the majority of KS is not X-linked and has been shown to have either autosomal recessive or dominant patterns of inheritance in familial cases. Adrenal hypoplasia congenita (AHC) is a rare X-linked disorder characterized by primary adrenal insufficiency and HH. It is caused by missense, nonsense, and frameshift mutations of the *DAX-1* gene (dosage-sensitive sex reversal, AHC-critical region of the X chromosome, gene 1), located on the short arm of the X chromosome Xp21. More than 60 different mutations of *DAX-1*, which is expressed in the hypothalamus, pituitary, adrenals, and gonads, have been reported in X-linked AHC.

### Activating Mutations

These mutations have not been identified in the *LH $\beta$* , *FSH $\beta$* , or *GnRHR* genes.

## Prolactin

Prolactin (PRL) is a polypeptide hormone encoded by the *PRL* gene on chromosome 6. Its release is stimulated by a number of prolactin-releasing factors, including vasoactive intestinal peptide, TRH, and PRL-releasing peptide. It is inhibited by prolactin-inhibiting factors, predominantly dopamine but possibly also by the 56-amino acid portion of the precursor to GnRH, known as GnRH associated peptide.

### Inactivating Mutations

Isolated PRL deficiency is extremely rare, usually occurring as a component of MPHD.

### Activating Mutations

No such mutations of genes encoding receptors for the hypothalamic factors or of *PRL* have been identified.

## Pituitary Disease Associated with Other Disorders

### Multiple Endocrine Neoplasia

This disorder is characterized by the presence of tumors involving two or more endocrine glands in a single patient. The two major forms, multiple endocrine neoplasia 1 (MEN1) and 2 (MEN2, which comprises 2A and 2B), are inherited as autosomal dominant conditions, with penetrance that increases with age.

MEN1 is characterized by tumor development in the parathyroid glands, pancreatic islet cells, and the anterior pituitary. Biochemically, approximately 60% of MEN1-associated pituitary tumors secrete prolactin, less than 25% secrete GH, and 5% secrete ACTH. The remainder appear to be nonfunctioning tumors.

There is variability in the phenotype, although some manifestations tend to be more common. This variability also occurs within family members of an affected kindred. Since penetrance is incomplete (especially early in life), MEN should always be considered in the diagnosis of “isolated” endocrine tumors, especially when familial.

MEN1 tumors are caused by inactivation of a tumor suppressor gene, *MEN1*. This gene is located on chromosome 11q13 and consists of 10 exons that encode a 610-amino acid protein, referred to as menin. Approximately 140 germ-line mutations of *MEN1* have been identified: approximately 25% are nonsense mutations, 45% deletions, 15% small insertions, 10% missense mutations, and less than 5% donor splice mutations.

### Carney Complex

Carney complex is an autosomal dominant condition that may be considered a multiple endocrine neoplasia syndrome. Endocrine tumors in Carney complex may involve the adrenal gland, the thyroid, the gonads, and the anterior pituitary gland. The majority of the pituitary tumors secrete GH and prolactin. The disorder is caused by mutations in one of two genes—the tumor suppressor gene *PRKAR1* (the type 1 $\alpha$  regulatory subunit of cAMP-dependent protein kinase A), located on chromosome 17q, or an unknown gene on chromosome 2p16.

## POSTERIOR PITUITARY

### Arginine Vasopressin

Arginine vasopressin (AVP) is synthesized in the supraoptic nuclei of the hypothalamus and transported within axons to the posterior lobe of the

pituitary for storage. The gene encoding AVP is arginine vasopressin–neurophysin II (*AVP–NPII*), located on chromosome 20p13. It encodes the precursor protein of AVP, consisting of AVP (i.e., antidiuretic hormone) and neurophysin II, the carrier protein for AVP. Diabetes insipidus may result from a mutation of the AVP portion or the neurophysin portion of the gene. Deficiency of AVP results in central or neurogenic diabetes insipidus. There are a number of genetic forms, such as familial neurohypophysial diabetes insipidus (FNDI) and Wolfram’s syndrome.

FNDI is a rare autosomal dominant syndrome in which more than 35 different germ-line mutations of *AVP–NPII* have been reported. Wolfram’s syndrome, also known as DIDMOAD (diabetes insipidus, diabetes mellitus, optic atrophy, and deafness), is an autosomal disorder that is caused by inactivating mutations of the gene encoding wolframin (*WFS1*) on chromosome 4p16.1.

### Acknowledgement

We thank Brian Betensky for his editorial assistance in the preparation of the manuscript.

### See Also the Following Articles

ACTH (Adrenocorticotrophic Hormone) • FSH (Follicle-Stimulating Hormone) • Gigantism: Excess of Growth Hormone • Growth Hormone Deficiency, Genetic • Growth Hormone (GH) • LH (Luteinizing Hormone) • Multiple Endocrine Neoplasia (MEN) Type 2 • Pituitary Gland Anatomy and Embryology • Pituitary Tumors, Molecular Pathogenesis • Prolactin (PRL) • TSH (Thyroid-Stimulating Hormone; Thyrotropin)

### Further Reading

- Achermann, J. C., and Jameson, J. L. (2001). Advances in the molecular genetics of hypogonadotropic hypogonadism. *J. Pediatr. Endocrinol. Metab.* **14**(1), 3–15.
- Achermann, J. C., Ozisik, G., Meeks, J. J., and Jameson, J. L. (2002). Genetic causes of human reproductive disease. *J. Clin. Endocrinol. Metab.* **87**(6), 2447–2454.
- Brandi, M. L., Gagel, R. F., Angeli, A., Bilezikian, J. P., Beck-Peccoz, P., Bordi, C., Conte-Devolx, B., Falchetti, A., Gheri, R. G., Libroia, A., Lips, C. J., Lombardi, G., Mannelli, M., Pacini, F., Ponder, B. A., Raue, F., Skogseid, B., Tamburrano, G., Thakker, R. V., Thompson, N. W., Tomassetti, P., Tonelli, F., Wells, S. A., Jr., and Marx, S. J. (2001). Guidelines for diagnosis and therapy of MEN type 1 and type 2. *J. Clin. Endocrinol. Metab.* **86**(12), 5658–5671.
- Cogan, J. D., and Phillips, J. A. (1998). Growth disorders caused by genetic defects in the growth hormone pathway. *Adv. Pediatr.* **45**, 337–361.
- Dattani, M., Martinez-Barbera, J., Thomas, P., Brickman, J., Gupta, R., Martensson, I., Toresson, H., Fox, M., Wales, J., Hindmarsh,

- P., Krauss, S., Beddington, R., and Robinson, I. (1998). Mutations in the homeobox gene HESX1/Hesx1 associated with septo-optic dysplasia in human and mouse. *Nature Genet.* **19**, 125–133.
- Gertner, J. M., Wajnrajch, M. P., and Leibel, R. L. (1998). Genetic defects in the control of growth hormone secretion. *Horm. Res.* **49**(Suppl. 1), 9–14.
- Layman, L. C. (1999). Genetics of human hypogonadotropic hypogonadism. *Am. J. Med. Genet.* **89**(4), 240–248.
- Machinis, K., Pantel, J., Netchine, I., Leger, J., Camand, O., Sobrier, M., Dastot-Le, M. F., Duquesnoy, P., Abitbol, M., Czernichow, P., and Amsellem, S. (2001). Syndromic short stature in patients with a germline mutation in the LIM homeobox LHX4. *Am. J. Hum. Genet.* **69**, 961–968.
- Marx, S., Spiegel, A. M., Skarulis, M. C., Doppman, J. L., Collins, F. S., and Liotta, L. A. (1998). Multiple endocrine neoplasia type 1: Clinical and genetic topics. *Ann. Internal Med.* **129**(6), 484–494.
- Netchine, I., Sobrier, M., Krude, H., Schnabel, D., Maghnie, M., Marcos, E., Duriez, B., Cacheux, V., Moers, A., Goossens, M., Gruters, A., and Amsellem, S. (2000). Mutations in LHX3 result in a new syndrome revealed by combined pituitary hormone deficiency. *Nature Genet.* **25**, 182–186.
- Parks, J. S., Brown, M. R., Hurley, D. L., Phelps, C. J., and Wajnrajch, M. P. (1999). Heritable disorders of pituitary development. *J. Clin. Endocrinol. Metab.* **84**(12), 4362–4370.
- Rittig, S., Robertson, G. L., Siggaard, C., Kovacs, L., Gregersen, N., Nyborg, J., and Pedersen, E. B. (1996). Identification of 13 new mutations in the vasopressin-neurophysin II gene in 17 kindreds with familial autosomal dominant neurohypophyseal diabetes insipidus. *Am. J. Hum. Genet.* **58**, 107–117.
- Robinson, A. G., and Verbalis, J. G. (2003). Posterior pituitary. In “Larsen Williams Textbook of Endocrinology” (J. Verbalis, ed.), 10th ed., pp. 291–292. Saunders, Philadelphia.
- Sandrini, F., and Stratakis, C. (2003). Clinical and molecular genetics of Carney complex. *Mol. Genet. Metab.* **78**(2), 83–92.
- Spiegel, A. M. (1997). Inborn errors of signal transduction: Mutations in G proteins and G protein-coupled receptors as a cause of disease. *J. Inherited Metab. Dis.* **20**(2), 113–121.
- Spiegel, A. M. (1997). The molecular basis of disorders caused by defects in G proteins. *Horm. Res.* **47**(3), 89–96.
- Spiegel, A. M. (2000). G protein defects in signal transduction. *Horm. Res.* **53**(Suppl. 3), 17–22.
- Stratakis, C. A. (2001). Clinical genetics of multiple endocrine neoplasias, Carney complex and related syndromes. *J. Endocrinol. Invest.* **24**(5), 370–383.
- Stratakis, C. A., Kirschner, L. S., and Carney, J. A. (2001). Clinical and molecular features of the Carney complex: Diagnostic criteria and recommendations for patient evaluation. *J. Clin. Endocrinol. Metab.* **86**(9), 4041–4046.
- Thakker, R. V. (2001). Multiple endocrine neoplasia. *Horm. Res.* **56**(Suppl. 1), 67–72.
- Themmen, A. P. N., and Huhtaniemi, I. T. (2000). Mutations of gonadotropins and gonadotropin receptors: Elucidating the physiology and pathophysiology of pituitary-gonadal function. *Endocr. Rev.* **21**(5), 551–583.
- Tsigos, C. (1999). Isolated glucocorticoid deficiency and ACTH receptor mutations. *Arch. Med. Res.* **30**(6), 475–480.

## CELL SIGNALING DURING SPERMATOGENESIS

Differentiation of male germ cells within the seminiferous epithelium proceeds in a complex series of stages, including the multiplication of spermatogonia, the successive phases of meiosis, the extensive remodeling of the nucleus and cellular structures during the haploid phase, and the final release of sperm cells. The whole process depends on a coordinated network of endocrine, paracrine, and cell-to-cell communication involving a variety of somatic partners. The kinds of extracellular molecules that are involved in signaling to the diverse types of spermatogenic cells are thought to be numerous, and not all have been identified. The most common are the gonadotropins luteinizing hormone (LH) and follicle-stimulating hormone (FSH), which have their target cells in the testicular somatic Leydig cells and Sertoli cells, respectively. Under hormonal signaling, Leydig cells secrete the male hormone testosterone, whereas Sertoli cells, under stimulation by FSH and testosterone, produce factors that induce germ cell proliferation, meiosis, and spermiogenesis.

Among the well-established Sertoli factors are inhibins, activins, and follistatin. In addition to the endocrine action of these peptides, which are generally considered endocrine regulators of FSH secretion, experimental evidence shows that activins, inhibins, and follistatin act locally within the testis as paracrine and autocrine factors involved in the regulation of spermatogenesis. Some studies suggest that inhibin can inhibit spermatogonial proliferation, whereas activin A seems to stimulate spermatogonial proliferation and follistatin may neutralize the action of activin. Other Sertoli factors reported to have a possible role in male germ cell differentiation are specific neurotrophins and cytokines and also putative “contacting” molecules—that is, the membrane components with selective expression and exposure on the Sertoli cell’s surface that exert a specific stage- or age-related action on spermatogenesis.

How is the message carried by so many different types of signaling molecules transduced in male germ cells? The answer is not known because of the inability to culture male germ cells for long periods of time. In fact, in contrast to cultured cells, no male germ cell lines are available for transfection experiments or to assay the signaling pathway evoked by a growth factor, an activator or inhibitor, a mutagen, and so on. Only recently have successful conditions for culturing primary spermatocytes and spermatogonia been reported; these may provide the groundwork

for future physiological or transfection studies. However, no cell lines are able to differentiate—that is, to undergo the spermiogenetic process to yield spermatozoa. Therefore, the only way to study spermatogenesis physiologically is to test *in vivo*-specific constructs by creating individual transgenic mice or mutant mice. The question is, what kind of mutant animal? To find an answer to this question, one has to ask the following: What are the major mechanisms regulating cell signaling?

## PROTEIN PHOSPHORYLATION

Protein phosphorylation is generally accepted as the universal tool that cells have developed to switch on or off dynamic processes. Phosphorylation can regulate the activity of transcription factors at multiple levels, including nuclear transport, dimerization, DNA binding and transcriptional activation, as well as a series of enzyme-catalyzed reactions, including enzyme-linked membrane receptor signaling, G protein-linked receptor signaling, and nutritional functions such as glycogen breakdown. It can also regulate complex phenomena, such as learning and memory. Consequently, protein phosphorylation may play a key role in spermatogenesis; this has been recognized for at least 40 years. How can the hierarchy of phosphorylation events that constitute a signaling pathway be identified?

### c-kit/SCF Signaling

Under cAMP and FSH stimulation, Sertoli cells express a growth factor that exists in both a soluble form and a transmembrane form and that is known as steel factor or stem cell factor (SCF). The SCF receptor is the c-kit transmembrane tyrosine kinase that is expressed by spermatogonia in the testis. In the whole organism, the c-kit/SCF system is required for normal hematopoiesis, melanogenesis, and gametogenesis; regarding spermatogenesis, c-kit/SCF signaling has been shown to be essential for proliferation and survival of the only proliferating germ cells, the spermatogonia. Which transduction pathway elicited by c-kit/SCF is able to support these effects? Among the c-kit/SCF-mediated pathways, there is direct activation of phosphatidylinositol (PI) 3'-kinase, which in turn leads to activation of Akt, a serine–treonin kinase with a key role in the growth factor-dependent proliferative and survival responses. Since PI 3'-kinase is active when it binds to a specific tyrosine phosphorylated residue of c-kit (Tyr<sup>719</sup> in the mouse),

Blume-Jensen and coworkers generated homozygous mutant mice carrying a Phe<sup>719</sup> instead of Tyr<sup>719</sup> in c-kit (Y719F/Y719F mutant mice) to search for c-kit/SCF signaling via PI 3'-kinase. No hematopoietic or pigmentation defects were found in homozygous mutant mice, but males were sterile due to a block in spermatogenesis, with a decrease in proliferation and then extensive apoptosis of spermatogonia. In contrast, female homozygotes were fully fertile. With such an approach, the researchers demonstrated the role of an individual signaling pathway downstream of c-kit/SCF in intact animals. Also, they demonstrated the cell line specificity of cell signaling because melanogenesis and hematopoiesis were not impaired by the inhibition of the transduction pathway acting through PI 3'-kinase.

## UBIQUITIN SYSTEM

The importance of post-translational protein modification for cell signaling within the cell is well recognized. In addition to protein phosphorylation, post-translational glycosylation also plays an essential role in the control of biological activities as the ligand receptor recognizing crossstalk; however, its dynamism and reversibility of action are not comparable to those of protein phosphorylation/dephosphorylation. Less canonical in comparison to protein phosphorylation and glycosylation is another highly dynamic mechanism, protein ubiquitination, involved in regulation of cell signaling. The ubiquitin system determines the half-life, stabilization, refolding, and translocation of proteins crucial for cell physiology. Ubiquitination is crucial for the downregulation of plasma membrane receptors, steroid hormone receptors, plasma membrane transporters, and ion channels. The ubiquitin machinery is complex; in fact, although the signaling is in the ubiquitin moiety covalently ligated to a well specified protein, other indispensable components are the ubiquitin-activating enzymes, the ubiquitin-conjugating enzymes, and the ubiquitin protein ligases that, through the cooperation of members of molecular chaperones, deliver the ubiquitinated protein to the proteasome or endocytotic vacuolar compartment. In addition to the enzymes involved in linking ubiquitin to proteins, there are a large number of deubiquitinating enzymes, which remove ubiquitin from the ubiquitinated proteins, allowing the ubiquitin moiety to recycle. Therefore, as phosphatases switch out the signal elicited by protein kinases, deubiquitinases counterbalance the signaling by ubiquitin ligases. The ubiquitin pathway is undoubtedly important for spermatogenesis. Using mouse transgene and knockout

models, it has been shown that certain components of the ubiquitin system are required for successful spermatogenesis. Different phases of mammalian spermatogenesis require different specialized activities of the ubiquitin machinery.

### HR23B Knockout

The *HR23B* gene encodes a mammalian homolog of *Saccharomyces cerevisiae* RAD23, an ubiquitin-like fusion protein involved in nucleotide excision repair (NER). To study its biological relevance, Ng and coworkers generated *HR23B*<sup>-/-</sup> mice. Unexpectedly, *HR23B* deficiency does not result in a NER defect, but *HR23B*<sup>-/-</sup> mice show impaired embryonic development and a high rate of intrauterine or neonatal death. When they survive, animals display abnormalities such as retarded growth and male sterility. Considering that *HR23B* is expressed in all mouse tissues and organs, why does male sterility occur? Because disruption of *HR23B* causes a failure of spermatogenesis, yielding a phenotype like that of the Sertoli cell-only syndrome. This finding indicates that *HR23B*, and consequently the protein stability via the ubiquitin/proteasome pathway, may be required for the postnatal initiation phase of spermatogenesis.

### *Siah1a* Knockout

Experimental evidence indicates a strict link among DNA repair, the ubiquitin system, the heat shock proteins/chaperones machinery, and the control of the meiotic cycle in spermatocytes. Dickins and coworkers used gene targeting to analyze the function of *Siah1a*, a component of E3 ubiquitin ligase complexes, during mammalian development. Mutant mice have normal weights at term but are postnatally growth retarded, despite normal levels of pituitary growth hormone. Also, serum gonadotropin levels are normal in mutant animals; however, females are subfertile, whereas males are sterile. Male sterility is due to a block in spermatogenesis. Although spermatocytes display normal meiosis I spindle formation, they accumulate at metaphase and then undergo apoptosis. Consequently, *Siah1a*, as a component of a male germ cell-specific ubiquitin ligase complex, is necessary for normal progression of meiosis.

### HR6B Knockout

It is during the postmeiotic phase, spermiogenesis, that ubiquitin signaling plays more roles. This cyto-differentiative process implies a massive remodeling

of cell structure and loss of most of the cytoplasm and a large fraction of cellular proteins. Some of this reorganization occurs due to condensation of cytoplasm into the residual body, phagocytosed by Sertoli cells while the residue is eliminated as cytoplasmic droplets by epididymal spermatozoa. However, this is not sufficient. In 1996, the first mammalian model of a defect in an enzyme involved in the ubiquitin pathway indicated the essential role of protein ubiquitination during spermatogenesis. To acquire its stream-like structure and simultaneously protect the genetic material during the long trip to the egg, the spermatozoon has evolved a very peculiar chromatin organization. Nucleosomes are disassembled, somatic and testis-specific histones are replaced by transition proteins, which in turn are replaced by protamines, and a highly condensed and inaccessible organization of chromatin is obtained. Roest and co-workers developed *HR6B*-deficient mice by gene targeting. The mouse *HR6B* gene is an autosomal homolog of the *S. cerevisiae Rad6* gene that codifies an ubiquitin-conjugating enzyme. This gene is highly conserved from yeast to mammals. The phenotype of *HR6B* knockout mice is remarkable and, at first glance, surprising: Although the *HR6B* gene is expressed throughout the body, the only pronounced defect of *HR6B*<sup>-/-</sup> mice is male infertility. This defect is due to impairment of the process of postmeiotic chromatin remodeling, indicating the involvement of the ubiquitin pathway in chromatin dynamism. The mouse *HR6B* knockout does not cause a complete and uniform block of spermatogenesis at a given point, but spermatozoa of knockout mice are morphologically abnormal and not able to complete fertilization. This finding advanced the study of the ubiquitin system and spermatogenesis, and since 1996 numerous laboratories have been investigating the signaling supported by the ubiquitin system during male germ cell differentiation. Not surprisingly, it has been found that the ubiquitin machinery during spermiogenesis is involved not only in chromatin remodeling but also in the targeted destruction of certain proteins. It appears that whereas the removal of residual cytoplasm as residual body displays the features of a rough phenomenon, the signal-mediated degradation/elimination or rescue of specific proteins by the ubiquitin/proteasome system plays a key role in yielding a fully fertile spermatozoon.

## CONCLUSION

Cell signaling is a complex network of various interacting objects. Paradoxically, in some cases, data

obtained in the past two decades have complicated our understanding because it has been found that most molecular signal transducers operate in the context of networks in which there is a considerable overlap of functions, some with a synergizing and some with an antagonizing effect. On the other hand, male germ cell signaling is the Cinderella of cell signaling. Historically, these cells have been the model for morphological, caryological, and comparative studies, but because of the lack of a cell culture technology they have attracted the interest of very few researchers engaged in signal transduction.

In this article, I have tried to provide a glimpse of the underexplored and fascinating isle of male germ cell signaling. Of course, due to space limitations and my intention to focus onto two fundamental mechanisms of control of cell signaling (i.e., the traditional protein phosphorylation and the emerging protein ubiquitination), I have not provided a comprehensive review of the topic. However, I hope to communicate to the scientific community that male germ cell signaling is a matter worthy of attention and study. A better understanding of the mechanisms governing spermatogenesis is highly desirable.

## See Also the Following Articles

FSH (Follicle-Stimulating Hormone) • LH (Luteinizing Hormone) • Spermatogenesis, Endocrine Control of

## Further Reading

- Blume-Jensen, P., Jiang, G., Hyman, R., Lee, K.-F., O’Gorman, S., and Hunter, T. (2000). Kit/SCF receptor-induced activation of PI 3'-kinase is essential for male fertility. *Nature Genet.* **24**, 157–162.
- de Krester, D. M., Loveland, K. L., Meehan, T., O’Bryan, M. K., Phillips, D. J., and Wreford, N. G. (2001). Inhibins, activins and follistatin: Actions on the testis. *Mol. Cell. Endocrinol.* **180**, 87–92.
- Dickins, R. A., Frew, I., House, C. M., O’Bryan, M. K., Holloway, A. J., Haviv, I., Traficante, N., de Krester, D. M., and Bowtell, D. D. (2002). The ubiquitin ligase component Siah1a is required for completion of meiosis I in male mice. *Mol. Cell. Biol.* **22**, 2294–2303.
- Hicke, L. (1999). Getting down with ubiquitin: Turning off cell-surface receptors, transporters and channels. *Trends Cell. Biol.* **9**, 107–112.
- Ng, J. M., et al. (2002). Developmental defects and male sterility in mice lacking the ubiquitin-like DNA repair gene mHR23B. *Mol. Cell. Biol.* **22**, 1233–1245.
- Roest, H. P., et al. (1996). Inactivation of the HR6B ubiquitin-conjugating DNA repair enzyme in mice causes male sterility associated with chromatin modification. *Cell* **86**, 799–810.
- Sette, C., Dolci, S., Geremia, R., and Rossi, P. (2000). The role of stem cell factor and of alternative c-kit gene products in the establishment, maintenance and function of germ cells. *Int. J. Dev. Biol.* **44**, 599–608.

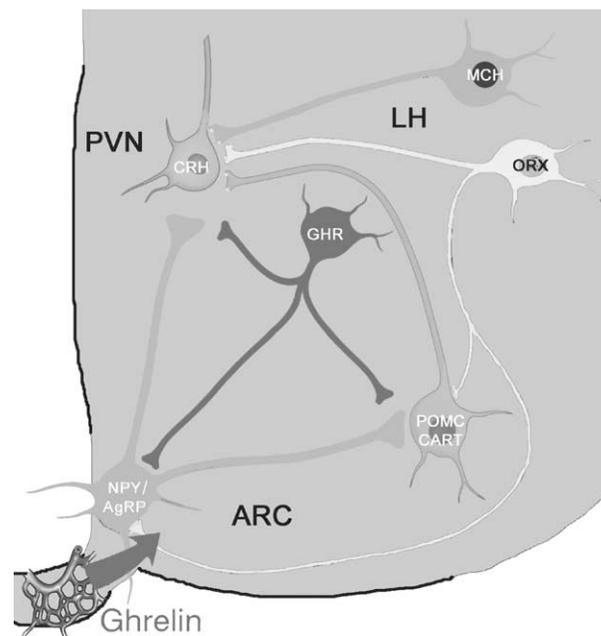
extent, thereby depriving it of its ability to bind and activate the ghrelin receptor, GHS-R1a. It is curious, however, that a peptide encoded obviously without a fatty acid side chain, in which form it also circulates in large quantities, would have no biological activity, and some findings suggest a possible biological role for unacylated ghrelin, i.e., in cell proliferation processes. Future research will shed light on this apparent paradox.

## DETERMINANTS OF GHRELIN ACTION

Whereas an enormous amount of data on ghrelin biology and physiology is emerging, the possibility that additional ghrelin receptors (in addition to GHS-R1a/b), as well as other endogenous ligands, might exist and could play a relevant role should not be overlooked. The vast majority of data either focus on GHS-R1a or extrapolate from changes in the overall concentration of circulating (bioactive and biologically inactive) ghrelin peptide. Whereas the detection of relative differences in total circulating ghrelin levels (predominantly representing bio-inactive peptide) between disease states or in response to physiological challenges can be regarded as useful, it is important that these data not be overinterpreted. When more sophisticated methods become available for monitoring the plasma concentrations of active ghrelin and the expression and activation levels of specific ghrelin receptor subtypes, substantial parts of the view on ghrelin physiology might have to be readjusted. The extent and magnitude of ghrelin action most likely involve multiple regulatory levels that may sometimes be independent of one another. Relevant mechanisms include the following: the regulation of transcription and translation of the ghrelin gene, the level of enzymatic activity of the putative acyltransferase that is responsible for the posttranslational octanoylation of the ghrelin molecule, secretion rates of the bioactive ghrelin molecule, putative enzymatic processes that deactivate circulating ghrelin, the possible influence of ghrelin-binding proteins on the hormone's bioactivity (e.g., binding of HDL), the variable accessibility of target tissue (i.e., blood-brain barrier transport), the clearance or degradation of ghrelin by passage through the kidney or liver, the circulating concentration of additional endogenous ligands or other possibly cross-reacting hormones, the level of expression of the ghrelin receptor(s) in target tissues, and the sensitivity of the target tissues at the level of intracellular signaling mechanisms.

## GASTRIC AND HYPOTHALAMIC GHRELIN

Ghrelin is primarily expressed in the stomach and upper intestinal tract. However, studies using polymerase chain reaction (PCR) amplification techniques have revealed the potential localization of ghrelin mRNA in several other tissues, such as the kidneys, immunocompetent blood cells, placenta, testes, ovaries, pancreas, pituitary, and hypothalamus. With respect to the putative role of ghrelin in the regulation of energy homeostasis, as well as other homeostatic mechanisms, localization in the hypothalamus seems to be particularly interesting (Fig. 1). In a study using antibodies as well as reverse transcription-PCR (RT-PCR), a uniquely distributed hypothalamic group of mostly bipolar neurons has been shown to produce small amounts of ghrelin. These neurons are not



**Figure 1** Schematic illustration of the relationship between ghrelin and hypothalamic peptidergic circuits. Ghrelin may reach the hypothalamus via the circulation and affect neuronal activity within the arcuate nucleus (ARC), where the NPY/AgRP- and POMC-producing neurons are located. Ghrelin is also produced in a subset of hypothalamic neurons located in the periventricular area. Ghrelin axon terminals innervate arcuate and paraventricular neurons (PVN), including those producing CRH. It is also likely that the central ghrelin system affects the lateral hypothalamic orexin (ORX) and MCH neuronal systems. Note that the hypothalamus operates with many more peptidergic and classical neuromodulators, such as  $\gamma$ -aminobutyric acid and glutamate, and this part of the brain is only one component in the central regulation of homeostasis. Therefore, it is reasonable to suggest that ghrelin's effect in the central nervous system will involve a variety of neurotransmitters and regions not depicted in this figure.

colocalized with any known centrally expressed hormone or neuropeptide, but, intriguingly, they do project directly to several previously identified hypothalamic appetite control centers. Ghrelin receptor expression and binding can furthermore be localized in multiple hypothalamic areas that directly neighbor cells containing neuropeptide Y (NPY), agouti-related protein (AgRP), proopiomelanocortin (POMC), orexin, and other neuropeptides and neurotransmitters that are substantially involved in appetite control. These neuroanatomical findings, complemented by electrophysiological studies, provide evidence for the existence of a central circuit that involves ghrelin as a key modulator. The full understanding of the role of hypothalamic ghrelin in endocrine and autonomic regulation remains elusive.

The main source of circulating endogenous ghrelin is the stomach. Circulating ghrelin derived from the gastrointestinal tract could be comodulating central networks regulating energy balance together with ghrelin derived from the hypothalamus after crossing the blood–brain barrier or by neural projection from areas that are not protected by the blood–brain barrier (circumventricular organs, e.g., the median eminence). Gastrointestinally derived circulating ghrelin could be responsible for the peripheral effects of ghrelin, including direct effects on the endocrine axes at the pituitary level and cardiovascular, anti-proliferative, or adipocyte-specific effects, whereas centrally derived ghrelin modulates mainly energy balance control circuits. Gastrointestinally derived ghrelin comodulates the central regulation of energy balance at the gastric level via the vagal nervous system and the brainstem. There is no reason that ghrelin could not act concurrently via paracrine mechanisms in the brain, via endocrine mechanisms by way of the circumventricular organs or after crossing the blood–brain barrier, and via the parasympathetic nervous system at the gastrointestinal level to regulate energy balance.

Ghrelin administration induces adiposity, raises the respiratory quotient (reflecting reduced fat utilization), and suppresses spontaneous locomotor activity in rodents. Neutralization studies with polyclonal ghrelin antibodies yielded encouraging results, showing decreased food intake in rodents after intracerebroventricular injection. These data were confirmed by a transgenic rat model that overexpresses antisense oligonucleotides against the ghrelin receptor, GHS-R1a, and decreased food intake and lower body fat were observed as a consequence. There is clear evidence that, despite its relatively short half-life, administration of ghrelin in physiologically relevant doses

induces a positive energy balance. Ghrelin affects body weight and food intake more than 1000-fold more potently following central administration, strongly supporting the hypothesis that ghrelin influences energy homeostasis predominantly by the modulation of central mechanisms. Therefore, whatever the derivation of the decisive endogenous amount of ghrelin that regulates energy balance, neutralizing or blocking its endogenous actions acutely as well as chronically will, at the very least, provide a valuable physiology lesson, yet has the potential to pave the way for the development of a new drug.

## REGULATION OF GASTRIC GHRELIN EXPRESSION AND SECRETION

Ghrelin is secreted mainly by gastric A/X-like cells within the oxyntic glands; the half life of ghrelin is relatively short (5–15 min) and less than 20% of the circulating immunoreactive ghrelin appears to be octanoylated and therefore bioactive. Gastrointestinal X/A-like cells represent approximately one-quarter of all endocrine cells in the oxyntic mucosa, and other cells within these glands, such as histamine-rich enterochromaffin-like cells (ca. 70%) and D-(somatostatin) cells (10%), are ghrelin-negative. From the stomach to the colon, ghrelin is found with caudally decreasing expression. Most ghrelin-containing enteroendocrine cells have no continuity with the lumen, probably respond to physical and/or chemical stimuli from the basolateral side, and are closely associated with the capillary network running through the lamina propria. Ghrelin-secreting cells occur as open- and closed-type cells (open or closed toward the stomach lumen), with the number of open-type cells gradually increasing in the direction from the stomach to the lower gastrointestinal tract. A closer look at the structural and functional relationship between ghrelin and its receptor, as well as the structure of motilin and its receptor, suggests that a larger family of peptide hormones is comodulating gastrointestinal motility, appetite, secretion of pituitary hormones, and other physiological processes. This peripheral endocrine network most likely also includes gastrointestinal hormones such as cholecystikinin, peptide YY (PYY1-36, PYY3-36), glucagon-like peptide 1, and gastric inhibitory peptide. Both ghrelin peptide secretion and ghrelin mRNA expression are regulated according to metabolic challenges. Acute and chronic periods of food deprivation increasing from mild to severe (e.g., fasting) raise ghrelin peptide levels and ghrelin mRNA concentrations, whereas refeeding reduces both.

## OTHER SOURCES OF GHRELIN

It is evident that ghrelin plays a classical endocrine role as a peptide hormone that is secreted into a capillary network, but local, paracrine activities of ghrelin might play an additional role. Removal of the stomach or of the acid-producing part of the stomach in rodents reduces serum ghrelin concentration by approximately 80%, further supporting the notion that the stomach is the main source of this endogenous GHS receptor ligand. However, in another study, plasma levels of ghrelin after total gastrectomy gradually increased again, suggesting that other tissues can compensate for the loss of ghrelin production after gastrectomy. The fact that total plasma ghrelin is barely detectable following gastric bypass surgery was interpreted by Cummings and co-workers as a “shutdown” of gastric ghrelin secretion due to a complete lack of contact with ingested nutrients.

Ghrelin mRNA and ghrelin peptide have also been detected in rat and human placenta. Here, ghrelin is expressed predominantly in cytotrophoblast cells and very sporadically in syncytiotrophoblast cells. A pregnancy related time-course, represented by an early increase in ghrelin expression in the third week, decreasing levels in the latest stages of gestation, and still detectable amounts of ghrelin at term, was found in rats. In human placenta, ghrelin is expressed mainly in the first half of pregnancy and is not detectable at term, whereas a putative involvement of ghrelin in fetal-maternal interactions via autocrine, paracrine, or endocrine mechanisms still remains to be shown.

Small concentrations of ghrelin are found in the pancreas, where ghrelin immunoreactivity has been localized in a subgroup of endocrine cells that are also immunopositive for pancreostatin. Ongoing, partially contradictory studies suggest that ghrelin-positive cells must be either the pancreatic alpha or beta cells.

In normal pituitary cells as well as in pituitary tumors, ghrelin mRNA expression and ghrelin-immunopositive cells were detected. This is in addition to the known presence of GHS receptors in pituitary cells. This suggests a possible autocrine or paracrine role for hypophyseal ghrelin, although only approximately 5% of the detected ghrelin peptide derived from the pituitary has been found to be octanoylated.

Using real-time PCR methodology, small amounts of ghrelin were detected in the adrenal glands, esophagus, adipocytes, gallbladder, muscle, myocardium, ovary, prostate, skin, spleen, thyroid, blood vessels, and liver. Preproghrelin production was shown in rat mesangial cells and mouse podocytes,

indicating the production of ghrelin in kidney, glomerulus, and renal cells and suggesting possible paracrine roles for ghrelin in the kidney. Human ghrelin expression and GHS receptor mRNA expression were shown by real-time PCR and confirmed by DNA sequencing in human T lymphocytes, B lymphocytes, and neutrophils from venous blood of healthy volunteers. Cell type and the maturity of the cells did not seem to have an influence on ghrelin production in immune cells. Interestingly, it has been shown that the small molecule GHS has a considerable immune-enhancing effect. In summary, ghrelin is predominantly expressed by the stomach and is expressed at decreasing levels as one moves caudally through the gastrointestinal tract. Although its physiological significance as a paracrine factor in extra-gastrointestinal tissue is the subject of ongoing studies, a classical endocrine role for extra-gastrointestinal ghrelin appears to be unlikely since ghrelin expression levels in other organs are relatively low in comparison. Thus, published studies on the regulation of ghrelin expression primarily focused on gastric ghrelin. However, studies on ghrelin expression or secretion in rodents are not necessarily relevant to the physiological regulation of ghrelin in humans.

## REGULATION OF CIRCULATING GHRELIN LEVELS

A very intriguing series of clinical studies by Cummings *et al.* indicates that each daily meal is followed by decreases in circulating ghrelin levels, most likely reflecting acutely reduced ghrelin secretion from the gastrointestinal tract. The authors speculate further that an observed premeal rise in circulating human ghrelin levels might reveal a role for ghrelin in meal initiation, which fits well with the observation that ghrelin administration in healthy volunteers causes hunger sensations. Ghrelin levels might also reflect the acute state of energy balance, signaling to the central nervous system in times of food deprivation that increased energy intake and an energy-preserving metabolic state are desirable.

Only a few determinants of circulating ghrelin concentration have been identified; these include insulin, glucose, somatostatin, and possibly growth hormone, leptin, melatonin, and the parasympathetic nervous system tone. In several species (e.g., mice, rats, cows, and humans), ghrelin mRNA expression levels or circulating ghrelin levels have been shown to be increased by food deprivation and to be decreased postprandially. This phenomenon further supports

the concept of ghrelin as an endogenous regulator of energy homeostasis that has apparently been preserved throughout evolution in all species. Rat ghrelin expression can also be stimulated by insulin-induced hypoglycemia, leptin administration, and central leptin gene therapy. Ingestion of sugar suppresses ghrelin secretion in rats *in vivo*, indicating a direct inhibitory effect of glucose/caloric intake on ghrelin-containing X/A-like cells in the oxyntic mucosa of the rat stomach rather than an exclusively insulin-mediated effect. The fact that insulin is an independent determinant of the circulating ghrelin concentration has been shown by several research groups using hyperinsulinemic–euglycemic clamp studies in humans. These findings add further evidence that ghrelin provides the link between mechanisms governing energy balance and the regulation of glucose homeostasis. However, it remains to be shown whether postprandially occurring insulin peaks are sufficient to decrease circulating ghrelin levels, since hyperinsulinemic–euglycemic clamp studies causing decreased ghrelin secretion involve either supraphysiological or markedly prolonged (e.g., > 120 min) periods of hyperinsulinemia. Further insight into the complex mechanisms regulating ghrelin secretion is based on studies showing an increase in circulating ghrelin levels in rats following surgical interventions such as vagotomy and hypophysectomy. Human growth hormone (GH) deficiency, however, does not seem to lead to increased plasma ghrelin levels. On the other hand, administration of synthetic GH in rats decreases circulating ghrelin levels and therapeutic intervention causing normalization of GH levels in patients with acromegaly increases endogenous ghrelin levels. Somewhat contradictory observations could be caused by species-specific differences between rodents and humans or could indicate that an acute, but not a chronic, change in GH levels modulates ghrelin concentrations. A previously neglected pathophysiological factor that might increase circulating ghrelin levels is the production of ghrelin by tumors of the stomach and the intestine, such as carcinoids.

In summary, ghrelin expression and ghrelin secretion are predominantly influenced by changes in energy balance and glucose homeostasis and influenced to a somewhat lesser degree by alterations in the endocrine axes (e.g., increasing GH concentrations). Based on the available data, ghrelin seems to represent a molecular regulatory interface between energy homeostasis, glucose metabolism, and physiological processes regulated by the classical endocrine axes, such as growth and reproduction. One particular biological purpose of these multiple roles of ghrelin

might be to ensure the provision of calories that GH requires for growth and repair.

## GHRELIN AND OBESITY

In contrast to earlier models that expected the endogenous ligand of the growth hormone secretagogue receptor to exclusively govern growth hormone secretion, ghrelin is believed to play its main physiological role in the regulation of energy balance. As the only peripherally circulating orexigenic agent known, ghrelin triggers appetite and nutrient intake. Ghrelin might even represent the first known “meal initiation factor.” However, conclusive evidence that meal-related circadian changes in plasma ghrelin concentrations are responsible for the initiation of nutrient intake rather than representing an epiphenomenon of trained meal patterns is still missing. Based on clinical investigations of meal-related changes in plasma ghrelin levels and data generated by insulin- and glucose-clamp studies, plasma insulin and blood glucose levels are very likely to be involved in the general regulation of ghrelin secretion. Although hyperinsulinemic–euglycemic clamps have been repeatedly shown to decrease circulating ghrelin levels, it remains unclear whether experimental conditions during clamp studies are comparable with the lower maximum peaks and the shorter duration of postprandial insulin levels. The possibility cannot be excluded that additional blood-derived factors may be responsible for meal-related changes in ghrelin concentrations or that gastrointestinal nutrient sensors may modulate ghrelin expression and secretion rates.

Although counterintuitive, the finding that circulating ghrelin concentrations are low in obesity not only mirrors earlier observations of hyperleptinemia in obesity, but may be explained by compensatory mechanisms aiming to communicate to the central regulatory centers that energy stores are full. Though it is unclear which signal communicates increased adipocyte size to ghrelin-secreting cells (leptin, interleukin-6, and adiponectin would be candidates), other phenomena and symptoms that commonly occur during obesity (such as a frequently filled stomach or insulin resistance) should be carefully investigated to determine their contribution to hypoghrelinemia. Ghrelin gene polymorphisms have been described by several groups; linkage analysis studies, however, failed to prove a solid association between ghrelin and obesity. Although diet-induced human obesity and polygenic (e.g., Pima Indians) or monogenic (e.g., MC4-R defect) causes of human obesity all present with low plasma ghrelin levels, there is one group

of severely obese patients in whom markedly increased plasma ghrelin concentrations have been observed. Prader-Willi syndrome, an impressive hunger syndrome in which patients exhibit morbid obesity and numerous other symptoms, is caused by a defect in the short arm of chromosome 15 and is accompanied by circulating ghrelin levels that are three- to fivefold higher than in healthy controls. Although the overlap between symptoms of Prader-Willi syndrome and the effects of ghrelin administration is impressive, only treatment with a potent, but safe ghrelin antagonist compound will reveal whether ghrelin is part of the pathogenesis in Prader-Willi syndrome. Discussion is ongoing as to whether increased plasma ghrelin is only a consequence of the severe caloric restrictions that are a central part of the treatment strategies for patients with Prader-Willi syndrome in an attempt to control their energy balance. The only other population in which comparably high ghrelin levels have been reported are patients with cachexia or anorexia nervosa, in whom high ghrelin levels are believed to reflect a physiological compensation effort in response to either a chronically empty stomach or a markedly decreased fat mass. Though circulating ghrelin levels are significantly lower in obese individuals, these levels are still very substantial when compared to nearly undetectable ghrelin concentrations in patients after gastric bypass surgery. The superior effectiveness of this bariatric procedure is considered to be partially due to a “knock-down” of endogenous ghrelin secretion caused by the lack of stimulation of gastric cells by incoming nutrients. On the other hand, an increase in endogenous ghrelin in response to diet-induced weight loss could contribute to the very high likelihood of recurrence of obesity. Carefully conducted clinical studies are imperative to discover the answer to this important question.

## SUMMARY

In summary, ghrelin represents a gastric hormone that induces hunger and increases fat deposition via central and possibly peripheral mechanisms in response to a negative energy balance. Ghrelin is one of the most potent orexigenic agents and the only known peripherally circulating orexigenic agent; it may also represent the first meal initiation factor. Although, normally, plasma ghrelin concentrations are negatively correlated with fat mass, substantial levels are still secreted in the vast majority of obese individuals, whereas weight loss occurs along with the loss of circulating ghrelin in patients with a gastric bypass.

Apart from the possible effectiveness of a ghrelin antagonist for the general prophylaxis and treatment of adiposity, the blockade of ghrelin could be the first specific pharmacotherapeutic approach to successfully treat patients with Prader-Willi syndrome. However, although the various additional effects of ghrelin on physiological processes and organ systems suggest other possible therapeutic uses, they also make unwanted cardiovascular, gastrointestinal, or proliferative effects caused by the blockade of ghrelin action a likely occurrence.

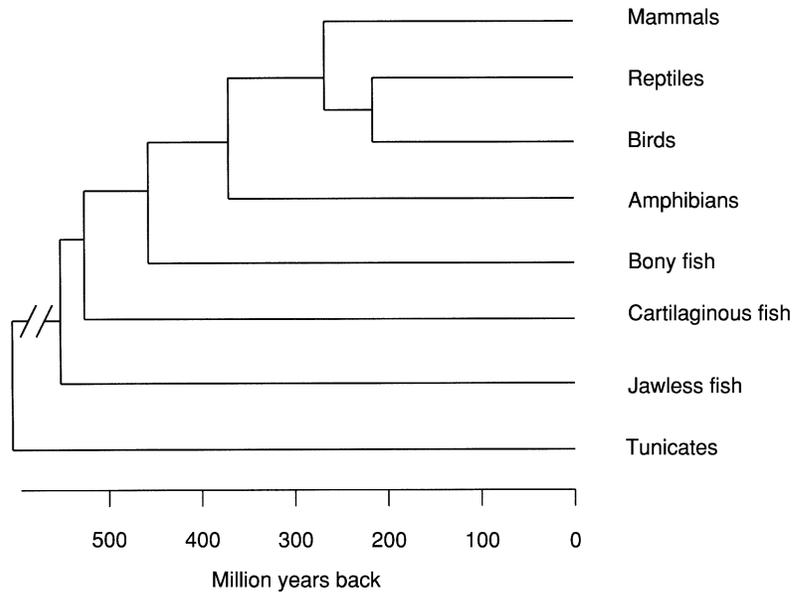
## See Also the Following Articles

CCK (Cholecystokinin) • GI Hormone Development (Families and Phylogeny) • GI Hormones Outside the Gut: Central and Peripheral Nervous System • GIP (Gastric Inhibitory Polypeptide) • Hunger and Satiation • Leptin • Motilin • Natural and Synthetic Growth Hormone Secretagogues • Neuropeptide Y • Obesity and Diabetes, Regulation of Food Intake • Peptide YY (PYY)

## Further Reading

- Ariyasu, H., Takaya, K., Tagami, T., Ogawa, Y., Hosoda, K., Akamizu, T., Suda, M., Koh, T., Natsui, K., Toyooka, S., Shirakami, G., Usui, T., Shimatsu, A., Doi, K., Hosoda, H., Kojima, M., Kangawa, K., and Nakao, K. (2001). Stomach is a major source of circulating ghrelin, and feeding state determines plasma ghrelin-like immunoreactivity levels in humans. *J. Clin. Endocrinol. Metab.* **86**, 4753–4758.
- Baldanzi, G., Filigheddu, N., Cutrupi, S., Catapano, F., Bonisconi, S., Fubini, A., Malan, D., Baj, G., Granata, R., Broglio, F., Papotti, M., Surico, N., Bussolino, F., Isgaard, J., Deghenghi, R., Sinigaglia, F., Prat, M., Muccioli, G., Ghigo, E., and Graziani, A. (2002). Ghrelin and des-acyl ghrelin inhibit cell death in cardiomyocytes and endothelial cells through ERK1/2 and PI 3-kinase/AKT. *J. Cell Biol.* **159**, 1029–1037.
- Baumont, N. J., Skinner, V. O., Tan, T. M., Ramesh, B. S., Byrne, D. J., MacColl, G. S., Keen, J. N., Bouloux, P. M., Mikhailidis, D. P., Bruckdorfer, K. R., Vanderpump, M. P., and Srai, K. S. (2003). Ghrelin can bind to a species of high density lipoprotein associated with paraoxonase. *J. Biol. Chem.* **278**, 8877–8880.
- Bednarek, M. A., Feighner, S. D., Pong, S. S., McKee, K. K., Hreniuk, D. L., Silva, M. V., Warren, V. A., Howard, A. D., Van der Ploeg, L. H., and Heck, J. V. (2000). Structure–function studies on the new growth hormone-releasing peptide, ghrelin: Minimal sequence of ghrelin necessary for activation of growth hormone secretagogue receptor 1a. *J. Med. Chem.* **43**, 4370–4376.
- Chanoine, J. P., Yeung, L. P., Wong, A. C., and Birmingham, C. L. (2002). Immunoreactive ghrelin in human cord blood: Relation to anthropometry, leptin, and growth hormone. *J. Pediatr. Gastroenterol. Nutr.* **35**, 282–286.
- Date, Y., Murakami, N., Toshinai, K., Matsukura, S., Nijijima, A., Matsuo, H., Kangawa, K., and Nakazato, M. (2002). The role of the gastric afferent vagal nerve in ghrelin-induced feeding and growth hormone secretion in rats. *Gastroenterology* **123**, 1120–1128.

- Havel, P. J. (2001). Peripheral signals conveying metabolic information to the brain: Short-term and long-term regulation of food intake and energy homeostasis. *Exp. Biol. Med.* **226**, 963–977.
- Horvath, T. L., Castaneda, T., Tang-Christensen, M., Pagotto, U., and Tschöp, M. H. (2003). Ghrelin as a potential anti-obesity target. *Curr. Pharm. Des.* **9**, 1383–1395.
- Horvath, T. L., Diano, S., Sotonyi, P., Heiman, M., and Tschöp, M. (2001). Minireview: Ghrelin and the regulation of energy balance—A hypothalamic perspective. *Endocrinology* **142**, 4163–4169.
- Hosoda, H., Kojima, M., Mizushima, T., Shimizu, S., and Kangawa, K. (2003). Structural divergence of human ghrelin: Identification of multiple ghrelin-derived molecules produced by post-translational processing. *J. Biol. Chem.* **278**, 64–70.
- Kojima, M., Hosoda, H., Date, Y., Nakazato, M., Matsuo, H., and Kangawa, K. (1999). Ghrelin is a growth-hormone-releasing acylated peptide from stomach. *Nature* **402**, 656–660.
- Muccioli, G., Tschöp, M., Papotti, M., Deghenghi, R., Heiman, M., and Ghigo, E. (2002). Neuroendocrine and peripheral activities of ghrelin: Implications in metabolism and obesity. *Eur. J. Pharmacol.* **440**, 235–254.
- Tschöp, M., Smiley, D. L., and Heiman, M. L. (2000). Ghrelin induces adiposity in rodents. *Nature* **407**, 908–913.
- Van der Lely, A. J., Tschöp, M., Heiman, M. L., and Ghigo, E. Biological physiological pathophysiological and pharmacological aspects of ghrelin. *Endocr. Rev.*, in press.
- Zhang, Y., Proenca, R., Maffei, M., Barone, M., Leopold, L., and Friedman, J. M. (1994). Positional cloning of the mouse obese gene and its human homologue. *Nature* **372**, 425–432.



**Figure 1** Evolutionary relationship of the chordate classes. Apart from the tunicates (or urochordates), the remaining classes are vertebrate animals. The lengths of the horizontal lines indicate the period when the respective classes are assumed to have segregated.

Thus, observations in one species based on antibodies raised against peptides from another species must be evaluated with caution, since both false-positive and false-negative results can occur.

During evolution, duplication of genes and exons followed by substitutions, insertions, and deletions has been an important mechanism for the creation of new peptides with new functions. Two major rounds of duplications occurred during the evolution of vertebrates, one at the beginning of the “Cambrian explosion” (approximately 500 million years ago) and one in the early Devonian (approximately 400 million years ago), to produce up to four copies of the original genome. In fish, a third duplication appears to have occurred in the late Devonian. Many of the duplicated gene products developed into new proteins with new functions, thus giving rise to distinct families of proteins (e.g., regulatory peptides), with family members having related or even very distant functions.

Many of the gastrointestinal hormones can clearly be classified as belonging to families originating from a common ancestor. Others appear as singular peptides although they can be traced to most vertebrate classes. The following discussion of the relationships between specific peptides and peptide families considers mainly solid structural information and to some extent function and anatomical origin. Also, only hormones from the Chordata (including tunicates) are considered.

## HORMONE FAMILIES

### The PACAP/Glucagon Superfamily

The PACAP/glucagon superfamily includes nine structurally related hormones in human, namely, PACAP, GHRH (also known as growth hormone-releasing factor), glucose-dependent insulinotropic polypeptide (GIP), peptide histidine–methionine (PHM) or peptide histidine–isoleucine (PHI), secretin, vasoactive intestinal peptide (VIP), glucagon, and glucagon-like peptide-1 and-2 (GLP-1 and-2). In mammals, these peptides are encoded by six genes: VIP and PHM are encoded by the same gene, glucagon, GLP-1, and GLP-2 are encoded by a single gene, and the remaining peptides come from separate genes.

#### **PACAP/GHRH**

The PACAP/GHRH superfamily can be traced back to tunicates, in which PACAP and GHRH have been identified; actually, there are two different precursors in tunicates, each coding for both peptides. Throughout evolution, PACAP and GHRH have remained together on the same precursor until the emergence of mammals, in which they appear on separate genes. This could be due to a late gene duplication in a common ancestor of mammals. Alternatively, it has been suggested that a second

gene remains undetected in nonmammalian vertebrates and that one has evolved into the mammalian PACAP preceded by a related flanking peptide and the other has evolved into the mammalian GHRH followed by another related flanking peptide. Interestingly, PACAP has been extremely well conserved from tunicate to human, whereas the structure of GHRH is highly variable and even between different mammals there is a low degree of similarity.

### **The Glucagon Family**

Glucagon, GLP-1, and GLP-2 have been detected in all vertebrate classes, but not (yet) in tunicates. The three peptides are derived from a common precursor and this organization has been conserved from jawless fish to mammals. Glucagon is highly conserved in all vertebrates but both GLP-1 and GLP-2 appear to be considerably more variable.

### **GIP, PHM/PHI, Secretin, and VIP**

GIP has been identified only in mammals and secretin and PHM/PHI have been identified only in mammals and birds. Whether they are represented in the other vertebrate classes remains an open question. In contrast, VIP has been detected in all vertebrates except jawless fish. Furthermore, the structure of VIP is highly conserved with only minor variations allowed at nine positions of the 28-residue peptide.

### **The PP-fold Family**

The PP-fold family of peptides derives its name from a structural characteristic of all the members: a hair-pin-like fold with an extended proline helix containing three proline residues, a turn, and an  $\alpha$ -helix with two tyrosine residues that fit in between the three proline residues. The family consists of neuropeptide Y (NPY) and peptide YY (PYY), present in all vertebrate classes, pancreatic polypeptide (PP), found only in tetrapods, and a pancreatic peptide Y (PY), found in some fish. NPY and PYY most likely originate from duplication of a common ancestor before the development of the jawless fish. NPY displays a high degree of sequence conservation, whereas PYY is more varied. The occurrence of PP only in tetrapods suggests that it is the result of yet another gene duplication that occurred before the emergence of the amphibians. The sequence variability of PP is quite high, with only 50% identity between mammals, birds, and amphibians. PY constitutes a special case and probably arose from a duplication of the PYY gene unrelated to the duplication event that generated PP.

### **The Cholecystokinin/Gastrin Family**

Cholecystokinin (CCK) and gastrin constitute a small family comprising only these two members. They are characterized by a common amidated C-terminal tetrapeptide sequence, which also constitutes the minimal structure necessary for biological activity of both peptides. Hence, it appears most likely that CCK and gastrin have evolved from a common ancestor. Cionin, isolated from *Ciona intestinalis*, a representative of the tunicates, which occupy a key position at the transition to vertebrates, also contains the characteristic tetrapeptide sequence and thus represents the oldest genuine member of the CCK/gastrin family thus far known, dating the emergence of these peptides back to at least 500 million years ago. The CCK/gastrin family is represented throughout the entire chordate phylum, including cartilaginous and bony fish, amphibians, reptiles, birds, and mammals. A duplication of the ancestral gene appears to have occurred at the level of cartilaginous fish, giving rise to two peptides most likely homologous to CCK and gastrin. At the amphibian level, the two separate peptide systems have been shown to exert distinct physiological gastrin and CCK actions. Interestingly, though CCK is well conserved in all vertebrate species, the gastrins vary more and the mammalian gastrins at first sight appear as a distinct group with little similarity to the nonmammalian gastrins outside the invariant C-terminal tetrapeptide. However, closer examination reveals that even if a major structural change was introduced at the transition to mammals, there exists a clear evolutionary relationship between mammalian and nonmammalian gastrins.

### **The Tachykinin Family**

The tachykinins share a C-terminally amidated pentapeptide, -Phe-Xaa-Gly-Leu-Met-amide, where Xaa can vary among a few possible amino acids. In mammals, the family comprises four peptides encoded by two genes: preprotachykinin A, encoding substance P (SP), neurokinin A (NKA), and neuropeptide  $\gamma$  (NP $\gamma$ , which is an N-terminally extended form of neurokinin A); and preprotachykinin B, encoding neurokinin B. SP, NKA, and NP $\gamma$  have been identified in most vertebrate classes, whereas NKB has been identified only in mammals and a frog (Amphibia). NKA (a decapeptide) has been highly conserved and SP (an undecapeptide) somewhat less so. Interestingly, SP from amphibians shows less identity to the mammalian form than SP from the phylogenetically more distant fish. However, the peptides are rather short

and even if many tachykinins from nonmammalian vertebrates have been identified, almost no genes for these tachykinins have been discovered. Thus, solid conclusions regarding the evolution of the tachykinins are difficult to make, but there is no doubt that the tachykinins constitute a family, that they must have arisen from a common ancestor, and that their diversity must have been brought about by exon duplication, gene duplication, and point mutations.

### The Somatostatin Family

In mammals, the somatostatin family contains two members, somatostatin (known since 1973) and cortistatin, which shows a high degree of similarity to somatostatin. The bioactive forms of somatostatin are somatostatin-14 and the N-terminally extended somatostatin-28. Somatostatin-14 is highly conserved, being identical in all vertebrates from jawless fish to mammals. In addition to genuine somatostatin, many fish and amphibian species express a second gene very similar to preprosomatostatin-I. It has been suggested that the duplication of the original somatostatin gene occurred early in evolution, predating or concomitant with the development of the chordates. Then, a second duplication may have occurred in the bony fish after divergence from the line leading to tetrapods.

### The Gastrin-Releasing-Peptide/Bombesin/Neuromedin B Family

Gastrin-releasing peptide (GRP) has been identified in all major vertebrate classes except jawless fish. It consists of 22 (goldfish) to 29 (rat) amino acid residues and has an invariant C-terminal octapeptide sequence in all species. Neuromedin B, which shows similarity to GRP, has been identified only in mammals and a toad (Amphibia) and is encoded by a separate gene. The bombesins constitute an interesting subgroup, all being isolated from the skin of amphibians, where they may serve protective purposes, and also arising from independent genes. Since only GRP is known to exist in many vertebrate classes, it is not possible to further elaborate on the phylogeny of these peptides.

## HORMONES WITH NO OBVIOUS FAMILY RELATIONSHIPS

### Galanin

Galanin is a widespread neuropeptide and is also common in the gut. Since its discovery in 1983,

galanin has been identified in a number of species representing all major vertebrate classes. It comprises 29 (or 30, in human) amino acid residues, of which the 14 N-terminal residues are identical in all species investigated. Despite repeated efforts to discover related peptides, none were found until 1999, when a new peptide, galanin-like peptide (GALP), was isolated from hypothalamus based on its stimulation of galanin type 2 receptors. Thirteen residues (positions 9–21) of GALP are identical to the N terminus of galanin. GALP has been identified only in mammals (human, pig, and rat). Thus, it still remains to be seen whether a galanin family can be identified.

### Motilin and Ghrelin

Motilin is produced in endocrine cells of the upper intestine. Since the identification of porcine motilin in 1972, motilin has been identified in a number of mammals but in only one nonmammalian species, the chicken. The mature form of motilin consists of 22 amino acid residues and between human and chicken there are seven substitutions, although most are conservative. Thus, little can be said about the evolution of motilin. In 1999–2000, a new GHRH was identified independently by two groups and was designated ghrelin and prepromotilin-related peptide, respectively. The latter name refers to a limited structural similarity with motilin. Identification of the two hormones from more vertebrate classes will be needed before possible relationships can be evaluated.

### See Also the Following Articles

ACTH,  $\alpha$ -MSH, and POMC, Evolution of • Angiotensin, Evolution of • Insulin and Insulin-like Growth Factors, Evolution of • Natriuretic Peptide System, Evolution of • Neuropeptide Y, Evolution of • Pituitary Adenylate Cyclase-Activating Polypeptide (PACAP)/Glucagon Superfamily • Prolactin, Evolution of • Somatostatin, Evolution of • Steroid Receptors, Evolution of

### Further Reading

- Cerda-Reverter, J. M., and Larhammar, D. (2000). Neuropeptide Y family of peptides: Structure, anatomical expression, function, and molecular evolution. *Biochem. Cell. Biol.* **78**, 371–392.
- Feng, D.-F., Cho, G., and Doolittle, R. F. (1997). Determining divergence times with a protein clock: Update and reevaluation. *Proc. Natl. Acad. Sci. USA* **94**, 13028–13033.
- Holmgren, S., and Jensen, J. (2001). Evolution of vertebrate neuropeptides. *Brain Res. Bull.* **55**, 723–735.
- Hoyle, C. H. (1998). Neuropeptide families: Evolutionary perspectives. *Regul. Pept.* **73**, 1–33.
- Irwin, D. M. (2001). Molecular evolution of proglucagon. *Regul. Pept.* **98**, 1–12.

- Johnsen, A. H. (1998). Phylogeny of the cholecystokinin/gastrin family. *Front. Neuroendocrinol.* **19**, 73–99.
- Lin, X., Otto, C. J., Cardenas, R., and Peter, R. E. (2000). Somatostatin family of peptides and its receptors in fish. *Can J. Physiol. Pharmacol.* **78**, 1053–1066.
- Montero, M., Yon, L., Kikuyama, S., Dufour, S., and Vaudry, H. (2000). Molecular evolution of the growth hormone-releasing hormone/pituitary adenylate cyclase-activating polypeptide gene family: Functional implication in the regulation of growth hormone secretion. *J. Mol. Endocrinol.* **25**, 157–168.
- Sherwood, N. M., Krueckl, S. L., and McRory, J. E. (2000). The origin and function of the pituitary adenylate cyclase-activating polypeptide (PACAP)/glucagon superfamily. *Endocr. Rev.* **21**, 619–670.

GI hormone cholecystokinin (CCK) is expressed in the postnatal rat pancreas. Transgenic overexpression of gastrin and transforming growth factor- $\alpha$  has been found to result in islet hyperplasia, perhaps reflecting a trophic role of gastrin during early islet development.

### Secretin

Secretin is the main product of endocrine S cells of the duodenum and jejunum. S cells occur on villi, but never in crypts, documenting their identity with postmitotic cells. Secretin is transiently expressed in insulin cells of the developing pancreatic islets.

### Preprotachykinin A

Expression of substance P was first noted in endocrine, enterochromaffin (EC) cells of the intestinal mucosa. Subsequently, preprotachykinin A (PPTA) was found to be transiently expressed in both insulin- and non-insulin-producing endocrine cells of the fetal and neonatal pancreas. Postnatally, the number of PPTA-positive cells declines and no such cells can be detected in adults. It has been suggested that substance P and neurokinin A, which both are encoded by the PPTA gene, could serve as growth factors during islet development.

### Vasoactive Intestinal Polypeptide

VIP is normally expressed by nerve cells of the GI tract and pancreas but is also expressed by pancreatic tumors causing the WDHA (watery diarrhea, hypokalemia, achlorhydria) syndrome, showing that neoplastic pancreatic endocrine cells have the capacity for VIP production.

### Serotonin

Serotonin (5-hydroxytryptamine) is produced by endocrine EC cells of the GI tract and pancreas and is also detected in pancreatic insulin cells of certain species, such as the guinea pig. Early maturing GI endocrine cells coexpress serotonin with other hormones.

### Insulin

Transient insulin expression occurs in a few neuron-like cells present at the juxtaduodenal myenteric plexus at embryonic days 18–19 in the rat.

### Glucagon

The glucagon gene is expressed by both pancreatic A cells and intestinal L cells. Differences in proteolytic processing account for the production of classical glucagon in A cells and of glucagon-like peptide-1 and -2 and other fragments in L cells.

### Somatostatin

Somatostatin is produced by D cells present in the GI tract and pancreas and is expressed by both the adult and the developing gut and pancreas. It is believed to act primarily as a local paracrine regulator of hormone release. In some species, both pancreatic and GI somatostatin cells emit long neuron-like processes with which they contact their target cells.

### PP Family

The PP family includes three structurally related peptides: pancreatic polypeptide (PP), neuropeptide Y (NPY), and peptide tyrosine tyrosine (PYY). PP was originally localized to the fourth (PP) cell type of the pancreas. Cat and dog PP (F) cells are characterized by large granules. In other species, such as human and rat, PP cells are small granulated cells, distinct from the fifth ( $D_1$ ) islet cell type, believed to produce an as yet unidentified hormone. Reverse transcription-polymerase chain reaction data show that pancreatic PP expression occurs before PP cells can be immunocytochemically identified and transgenic data indicate that insulin cell progenitors express PP. Only limited GI expression of PP occurs. PYY is expressed by both gut and pancreatic endocrine cells (including antropyloric gastrin cells) and may be a marker for the earliest appearing endocrine pancreatic cells. NPY is expressed in a subpopulation of insulin cells in rat pancreas and by nerves in the GI tract.

### Islet Amyloid Polypeptide (Amylin)

Amylin was originally discovered as the chief constituent of amyloid present in insulinoma tissue and in islets of patients with non-insulin-dependent diabetes. Amylin occurs in islet insulin and somatostatin cells as well as in a majority population of somatostatin cells and a minority population of gastrin cells of the antropyloric mucosa of the stomach.

## HORMONE COEXPRESSION IN THE GASTRODUODENOPANCREATIC REGION

Many studies have shown that mature endocrine cell types frequently coexpress multiple hormones. Although these observations challenged well-guarded concepts, such as “the one cell–one hormone principle,” they have stood the test of time better than the original concepts. It is widely accepted that many endocrine cells secrete a host of molecules that act in concert to elicit coordinated biological responses.

Also, developing or maturing endocrine cells frequently express multiple hormones. Observations of such coexpression led to the finding that the three main antropyloric endocrine cell types (gastrin, serotonin, and somatostatin cells) develop from common multihormone-expressing precursors in the isthmus of gastric glands. In addition, duodenal endocrine cells develop from multihormonal precursors present in the crypts of Lieberkühns. Cells coexpressing insulin and glucagon have been detected in the developing pancreas and it has been suggested that they represent islet cell precursors. However, transgenic studies have documented that insulin and glucagon cells are derived from cell lineages that do not coexpress these two hormones. Instead, mature insulin cells may develop from cells transiently expressing PP. Moreover, the large increases in insulin cell numbers that occur during rat pancreatic development suggest that most of these cells are derived from insulin-negative precursors.

## TRANSCRIPTION FACTORS AND ENDOCRINE DIFFERENTIATION

Gene knockout studies have demonstrated that a number of transactivating factors are important for the differentiation of both pancreatic and gastroduodenal endocrine cells. This is not surprising since the distal stomach, duodenum, and pancreas develop from closely adjacent parts of the primitive foregut. Furthermore, overlaps in the distribution of these factors may explain the transient and permanent overlaps in the expression of gut hormones in the pancreas and of pancreatic hormones in the gut.

Such factors include pancreatic(-antropyloric)-duodenal homeobox-1 (Pdx-1),  $\beta$ 2/NeuroD, Pax4, and Pax6. Additionally, indirect data suggest that factors such as islet-1 (Isl-1) and Nkx6.1 may share this dual importance.

Table 1 Effects of Gene Knockouts of Pdx-1, Pax4, Pax6, and  $\beta$ 2/NeuroD on Hormone Expression in the Gastroduodenopancreatic Region

Deficiency	Pancreas	Duodenum	Stomach
<b>Pdx-1</b>	No pancreas	Neurotensin ↓ GIP ↓ Secretin ↓	Gastrin ↓ Serotonin ↑
<b>Pax4</b>	Insulin ↓ Somatostatin ↓ Glucagon ↑	Serotonin ↓ CCK ↓ GIP ↓ PYY ↓ Secretin ↓	Serotonin ↓ Somatostatin ↓
<b>Pax6</b>	Glucagon ↓	GIP ↓	Gastrin ↓ Somatostatin ↓
<b><math>\beta</math>2/NeuroD</b>	Insulin ↓ Glucagon ↓ Somatostatin ↓	CCK ↓ Secretin ↓	

Pdx-1 was identified as a factor that interacts with the insulin and somatostatin promoters. In Pdx-1-deficient mice, the pancreas fails to develop, the rostral duodenum shows anomalies, and in the antropyloric mucosa, gastrin cells fail to develop, whereas somatostatin cells occur in normal numbers and serotonin cells occur in increased numbers (Table 1). These data show that Pdx-1 is not required for gastric somatostatin expression, but is needed for the differentiation of gastrin cells, probably from a common gastrin–somatostatin precursor cell. Additionally, Pdx-1<sup>-/-</sup> mice show normal gastric expression of amylin, probably reflecting normal expression in somatostatin cells and reduced expression of PYY, reflecting the decrease in gastrin cells. The pancreas of Pax4<sup>-/-</sup> mice is deficient in insulin and somatostatin cells, but contains increased numbers of glucagon cells. In the antropyloric mucosa, Pax4 deficiency eliminates somatostatin cells and severely reduces the number of serotonin cells, while leaving gastrin cells unaffected (Table 1). The duodenum of Pax4<sup>-/-</sup> mice shows near total elimination of serotonin, secretin, CCK, GIP (gastric inhibitory polypeptide), and PYY cells. Mice deficient in Pax6 show abnormal pancreatic development with no glucagon cells. In the antropyloric mucosa, Pax6<sup>-/-</sup> mice are deficient in gastrin and somatostatin cells, but show normal numbers of serotonin cells, whereas in the duodenum severe reductions in CCK and GIP, but not in serotonin, secretin, or PYY cells, are observed. These data show that the two Pax genes are needed for the differentiation of virtually all endocrine cells of the gastroduodenopancreatic region. Pax4 is transiently

expressed during pancreatic development and may constitute a transcriptional repressor. A fourth factor of proven importance is  $\beta 2$ /NeuroD. Mice deficient in this factor show reductions in the number of secretin and CCK cells, while expressing apparently normal numbers of PYY, GIP, and somatostatin cells. In the pancreas,  $\beta 2$ /NeuroD $^{-/-}$  mice show a 75% reduction in insulin cells, a 40% reduction in glucagon cells, and a 20% reduction in somatostatin cells. A number of additional signaling or transactivating factors have been shown to be important for cell-specific hormone expression in the pancreas but have not yet been tested for their effects on gut endocrine development. Primary candidates include transforming growth factor- $\beta$  family cytokines, sonic hedgehog, neurogenin-3, Nkx6.1, and Isl-1.

## CONCLUSIONS

The distal stomach, duodenum, and pancreas develop from closely adjoining areas of the primitive foregut that show very similar hormone and transcription factor expression profiles. Targeted deletion of genes shows that several factors, such as Pdx-1, Pax4, Pax6, and  $\beta 2$ /NeuroD, are important in the development of both pancreatic and gastroduodenal endocrine cell types. As documented by gene knockout experiments, these factors are required for the expression of multiple hormones (Table I). Thus, they probably act together in different combinations to direct the cell-specific expression of different hormones. This helps to explain why certain gut hormones are transiently expressed in the pancreas and vice versa. Moreover, these observations may help to explain why certain hormones, such as gastrin, are expressed in pancreatic endocrine tumors. Interestingly, many of the factors identified are also important in the development of specific subsets of neurons. This adds to the long list of factors that are shared by endocrine cells and neurons, but does not indicate a common embryological origin. Thus, many observations solidly

document that GI and pancreatic endocrine cells are of endodermal origin. It appears that different combinations of factors, acting in a strict spatial and temporal hierarchy as permanent or transient transcriptional activators and repressors, drive cell-specific hormone gene expression in the gastroduodenopancreatic region. The quest for new factors and/or combinations that may be important for directing insulin expression and tumor-specific hormone expression is ongoing.

## Acknowledgments

Grant support for this work was provided by the Danish Medical Research Council and Cancer Society.

## See Also the Following Articles

CCK (Cholecystokinin) • Gastrin • GI Hormones and Endocrine Pancreas: Growth • GI Hormones Outside the Gut: Central and Peripheral Nervous System • GI Hormones Outside the Gut: Other Tissues • Pancreatic Cancer

## Further Reading

- Alpert, S., Hanahan, D., and Teitelman, G. (1988). Hybrid insulin genes reveal a developmental lineage for pancreatic endocrine cells and imply a relationship with neurons. *Cell* **53**, 295–308.
- Grappin-Botton, A., Majithia, A. R., and Melton, D. A. (2001). Key events of pancreas formation are triggered in gut endoderm by ectopic expression of pancreatic regulatory genes. *Genes Dev.* **15**, 444–454.
- Herrera, P. L. (2000). Adult insulin- and glucagon-producing cells differentiate from two independent cell lineages. *Development* **127**, 2317–2322.
- Jonsson, J., Carlsson, L., Edlund, T., and Edlund, H. (1994). Insulin-promoter-factor 1 is required for pancreas development in mice. *Nature* **371**, 606–609.
- Larsson, L.-I. (1988). Regulatory peptides and amines during ontogeny and in non-endocrine cancers: Occurrence and possible functional significance. *Prog. Histochem. Cytochem.* **17**, 1–222.
- Larsson, L.-I. (2000). Developmental biology of gastrin and somatostatin in the antropyloric mucosa of the stomach. *Microsc. Res. Tech.* **48**, 272–281.

demonstrated the expression of heparin-binding epidermal growth factor-like growth factor (HB-EGF) during pancreas development. This factor, expressed abundantly in ductal cells and immature endocrine cells, can be potentially regulated by PDX-1 as both molecules colocalize during development. It is tempting to suggest that HB-EGF could mediate the effects of PDX-1 along with betacellulin and TGF- $\beta$ 2 via the EGF receptor expressed in the endocrine and exocrine cells of the fetal pancreas. It would seem that EGF itself is not involved since it is not expressed in these cells during development. Furthermore, Sarvetnick clearly demonstrated significant immunoreactivity of heregulin isoforms in fetal pancreatic ducts; these factors bind to the Erb $\beta$  receptors, of which Erb $\beta$ 1 is the EGF receptor. It is strongly suggested that this receptor family operates through receptor heterodimerization between family members, with Erb $\beta$ 2 being the preferred partner. The association of these receptor-ligands with endocrine cells and islet formation strongly supports their involvement in mediating islet growth during pancreatic development.

Hayek's group established that hepatocyte growth factor (HGF), an expressed fetal mesenchyme-derived factor, can regulate beta cell growth and differentiation as the HGF receptor is preferentially expressed in developing beta cells. When used alone in monolayer culture, this factor was associated with a tremendous increase in the number of fetal pancreatic epithelial cells and marked down-regulation of insulin and glucagon gene expression. However, when cells were combined into three-dimensional aggregates, hormone gene expression increased, suggesting the importance of cell-cell contact for the complete biological effect of HGF. The implication of vascular endothelial growth factor (VEGF) in endocrine pancreas development was suggested when coculture of isolated endoderm with tissue fragments containing vascular endothelium induced the expression of both PDX-1 and insulin in the endodermal cells; however, tissues without vascular endothelium did not support pancreatic differentiation. Furthermore, mice expressing VEGF under the control of the PDX-1 promoter to target the factor to the developing pancreas exhibited hypervascular islets with islet hyperplasia. Although VEGF increased mitogenesis, insulin content, and insulin release in response to glucose from fetal islets, it did not stimulate direct differentiation of ductal epithelial cells into endocrine islet cells. Instead, VEGF increased the number of blood vessels or the production of an unknown factor from the vascular endothelial cells. These few examples

clearly indicate the importance of specific growth factors in the physiological processes leading to the development and differentiation of the fetal pancreatic islets.

## Hormones

Experimental data that clearly show the involvement of hormones and more specifically GI hormones during fetal pancreas development are rather scattered. Studies by Swenne indicate that growth hormone (GH) can stimulate the *in vitro* replication of fetal rat beta cells, an effect mimicked by prolactin and placental lactogen. However, because of the doses used, the physiological significance of the observed effects was called into question. Later, Swenne proposed that part of GH's mitogenicity in beta cells could be attributed to insulin-like growth factor-I (IGF-I). Rhodes described IGF-I and GH signal transduction pathways and suggested that each molecule operates via its own specific route to activate different mitogenic signals.

Although cholecystokinin (CCK), a gastrointestinal peptide hormone, can stimulate proliferation of the exocrine pancreas after birth, it remains unlikely that it would be involved during fetal development since its expression in fetal life is negligible. However, by immunochemistry and immunofluorescence methods, Sarvetnick detected CCK in pancreatic cells located in the acinar region of the pancreas on embryonic day 16. These data suggest that CCK signaling could be established early in development but its exact function remains unknown. However, treatments throughout pregnancy with caerulein, a CCK analogue, led Morisset to show that it induced pancreas aplasia in the mature fetus; unfortunately, the endocrine pancreas was not examined in this study. Another GI hormone that could influence the growth of the endocrine pancreas is gastrin, as it is transiently expressed in the islets, its major source in the fetus. Although the role of pancreatic gastrin in islet development remains undefined, it may influence islet growth since fetal pancreas has high levels of CCK<sub>B</sub>/gastrin receptor mRNA transcripts and high concentrations of gastrin. Although once again endocrine pancreas was not investigated, Morisset clearly demonstrated that pentagastrin treatment throughout pregnancy resulted in fetal pancreas hypertrophy, whereas treatment with L365,260, a specific CCK<sub>B</sub> receptor antagonist, caused pancreas atrophy. It may be hypothesized that the endocrine pancreas could also have been affected.

## POSTNATAL ENDOCRINE PANCREAS GROWTH

### Normal Growth

Discovery of the mechanisms of islet cell development and regeneration remains an important medical issue because of the steady increase in the incidence of diabetes among populations. For this reason, research has mostly focused on beta cell proliferation. As indicated by Bonner-Weir, an increase in islet mass in the adult can occur by neogenesis (islet formation from ductal precursor cells), replication of existing islet cells, or beta cell hypertrophy. Although the adult beta cell population has a limited proliferative potential, a considerable number of endocrine cells can regenerate. Such a process involves an orchestration of hormones and growth factor stimuli.

### Response to Growth Factors

Kim showed that a mutation in the activin receptors led to hypoplastic pancreatic islets with a normal exocrine pancreas. Along this line, Korc established the presence of TGF- $\beta$  isoforms in the adult human pancreas. TGF- $\beta$ 2 and TGF- $\beta$ 3 in particular are present and coexpressed with insulin. These observations led Korc to suggest that these factors may participate in the regulation of the biological functions of the endocrine pancreas. In perinatal rats fed a low-protein diet, reduced expression of IGF-I led to hypoplastic islets. With the HGF receptor preferentially expressed in beta cells, adult beta cells proliferate in response to HGF stimulation. These few experimental data indicate the importance of certain growth factors in controlling the normal functions of the endocrine pancreas, including turnover of their representative cell components.

### Response to Hormones and GI Hormones

GH, prolactin, and placental lactogen are implicated in regulating marked beta cell hyperplasia and increased islet proliferation during gestation in mice and humans. From studies performed on the pancreatic beta cell line INS-1, it seems that the effect of GH on DNA synthesis would be direct and dependent on glucose concentration. Among the gastrointestinal hormones, pituitary adenylate cyclase-activating protein (PACAP), a member of the secretin/glucagon/VIP family, was first shown to be a potent insulin secretagogue and its distribution in pancreatic nervous fibers surrounding the islets suggests a paracrine effect. However, studies by Petruzzo raised the possibility of PACAP being a growth factor. Indeed,

incubation of the peptide with purified islets resulted in the appearance of PACAP in nuclei of 80% of the islet cells; such an observation may suggest a genomic action of the internalized PACAP. Intestinal glucagon-like peptide-1, known to modulate insulin, glucagon, and somatostatin release, has been shown to stimulate PDX-1 expression and to increase islet size in mouse pancreas. Even though it has been reported that in the rat pancreas, free CCK<sub>A</sub> receptors lost their ability to secrete insulin and glucagon, the fact remains that chronic CCK treatment alone or with secretin in normal rats did not affect total insulin content while inducing exocrine pancreas growth. Similarly, as shown by Morisset, caerulein, a CCK analogue, did not modify the labeling index of the endocrine cells after 4 days of treatment, whereas all the other cell populations had their labeling index significantly increased, by even as much as 26% in acinar cells. Finally, data presented by Brand convincingly indicated that gastrin does not stimulate islet cell growth. Indeed, the pancreas of transgenic mice overexpressing gastrin had normal histology with an islet mass comparable to that of the controls.

## Endocrine Pancreas Regeneration

The aims of most studies related to endocrine pancreas regeneration have been to restore the beta cell mass to maintain euglycemia. After the neonatal period, the replication rate of the beta cells is low although it is known that its mass continues to grow well into adulthood. In order to study islet regeneration, different experimental animal models were used, including partial pancreatectomy, duct ligation, cellophane wrapping, and overexpression of specific genes.

### Pancreatectomy

In 40% pancreatectomized rats, Leahy observed regrowth of much of the excised islet mass after 3 weeks. Similarly, in 60% pancreatectomized rats, beta cell regeneration was still observed but remained incomplete, with the islet non-beta-cell mass remaining unchanged. In large animals, such as the dog, there is a critical threshold of resection at approximately 88 to 92% that causes the immediate onset of diabetes. Development of diabetes is also a major setback for regeneration; indeed, after 74–92% pancreatectomy, the regeneration rate, evaluated as changes in pancreas size, was more than 40% in the nondiabetic dogs but only 15 to 23% in the diabetic animals. Treatment with insulin enhanced DNA synthesis in the remnant

pancreas, thus promoting regeneration. Unfortunately, the factors other than insulin that are involved in islet regeneration are still unknown.

### **Pancreatic Duct Ligation**

Previous studies indicated that pancreatic duct ligation induced islet cell neogenesis from duct cells in the adult rat pancreas. Klöppel found that, in response to this procedure, gastrin mRNA was strongly expressed in the newly developed duct-like cell structures along with gastrin. Expression of three isoforms of the transforming growth factor- $\alpha$  (TGF- $\alpha$ ) protein was also observed. The appearance of these two growth factors preceded the peak of DNA synthesis in beta cells. These data strongly suggest that gastrin and TGF- $\alpha$  act as growth factors during islet neogenesis. Using a similar duct ligation model in rats, Bouwens found that infusion of gastrin for 3 days increased the beta cell mass. However, there was no mitogenic effect of gastrin on the beta cells, which is not surprising since they do not express CCK<sub>B</sub> receptors. It was then suggested that gastrin extends the process of neogenesis already under way by the duct ligation procedure with an up-regulation of gastrin and the CCK<sub>B</sub> receptors in the ductal complexes of the ligated part of the pancreas. Under these conditions, gastrin would have a paracrine effect on the responsive ductal cells through its newly expressed CCK<sub>B</sub> receptor.

### **Cellophane Wrapping**

A new factor was discovered in response to cellophane wrapping of the head of the pancreas in Syrian golden hamsters. This surgical procedure led to the induction of new islet formation from ductal cells within 2 weeks through a paracrine or autocrine mechanism responsible for islet neogenesis. The responsible factor was cloned and designated INGAP (islet neogenesis-associated protein) and is a member of the Reg family. Its administration resulted in a two- to threefold increase in DNA synthesis in ductal epithelial cells, the potential precursor cells; the peptide had no effect on existing beta cells. The INGAP site of origin would be the acinar cells and its action on ductal cells would likely be paracrine, leading to islet neogenesis, as proposed by Vinik. The phenomenon was also observed in monkeys, with the same wrapping procedure being used.

### **Overexpression of Growth Factors**

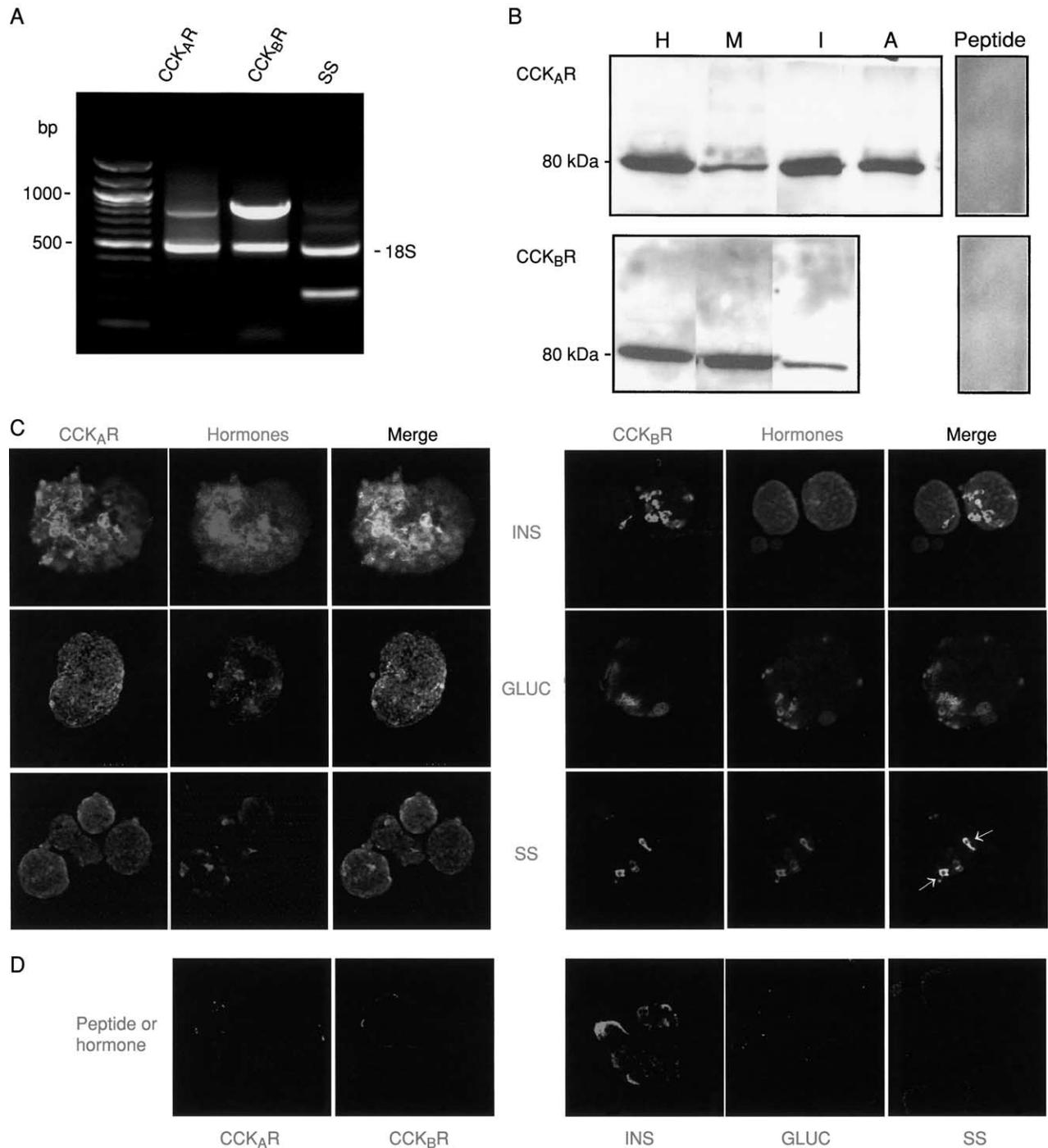
From studies by Brand, it was shown that pancreatic coexpression of gastrin and TGF- $\alpha$  resulted in significant increases in islet mass in mice with both

transgenes. These studies indicate that both factors can act in synergy to stimulate islet growth, whereas each peptide alone failed to do so. In transgenic mice expressing interferon- $\gamma$  in their islets, pancreatic duct cell proliferation and new islet formation were observed. These specific modifications were associated with the expression in the islets of the Erb- $\beta$  receptors and their specific agonists, the heregulin isoforms, suggesting their involvement in islet neogenesis. Finally, the overexpression of insulin-like growth factor-II in mouse pancreas led to oversized islets with no change in the number of islets per area. In pancreas, the area occupied by insulin cells was decreased, the area occupied by glucagon cells was increased, and the area occupied by somatostatin cells was unchanged. However, given the islet cell hyperplasia, the total number of beta cells per islet was increased in these transgenic animals. Overexpression of this growth factor also resulted in increased cell replication and reduced apoptosis.

## **AN APPROACH TO ENDOCRINE PANCREAS GROWTH CONTROL BY GI HORMONES**

From data presented in this article, it is clear that the role played by the GI hormones in the development and regeneration of the pancreatic islets has remained underinvestigated over the years. Although the involvement of the various GI hormones in secretion of the different islets' hormones has been well investigated, results on this specific subject remain controversial. As an example, ambiguity can come from the location of CCK in the pancreas. Indeed, the hormone colocalized either with insulin in rat or with glucagon in mice. Similar divergent data regarding the location of the CCK<sub>A</sub> and CCK<sub>B</sub> receptor subtypes exist. The research groups of Fourmy and Bouwens established the presence of the CCK<sub>B</sub> receptor on glucagon cells in human and rat pancreas, respectively, and Amselgruber identified the CCK<sub>A</sub> receptor on glucagon cells in the pig pancreas. Neither Fourmy nor Bouwens was able to localize the CCK<sub>B</sub> receptor on pancreatic somatostatin cells; however, Morisset and co-workers clearly showed their colocalization in human, rat, mouse, pig, dog, and calf pancreas.

The authors' approach to evaluating the growth potential of the GI hormones on pancreatic endocrine cells is systematic. Indeed, before any experimental assays on islet cell proliferation are performed, the rat purified islets were adopted as a model and the GI



**Figure 1** CCK<sub>A</sub> and CCK<sub>B</sub> receptor mRNA (A) and protein expression (B) in total rat pancreas and purified islet. (A) RT-PCR was performed using the TITANIUM One-Step RT-PCR Kit (Clontech Laboratories, Palo Alto, CA) and 1  $\mu$ g of total RNA from purified islets. Amplification of CCK<sub>A</sub>R, CCK<sub>B</sub>R, and somatostatin (SS) mRNA is observed. The amplified product sizes are 812, 669, and 231 bp for CCK<sub>A</sub>R, CCK<sub>B</sub>R, and SS, respectively. (B) Western blots of total pancreas homogenate (H), purified membranes (M), purified islets (I), and acini (A) represent 25  $\mu$ g protein fractionated on a 12% sodium dodecyl sulfate-acrylamide gel transferred to a polyvinylidene difluoride membrane and probed with the polyclonal rabbit anti-CCK<sub>A</sub>R and anti-CCK<sub>B</sub>R described in Table I. Specificity was established by preincubation of each antibody with its specific peptide. Bands were visualized by enhanced chemiluminescence. Localization of CCK<sub>A</sub>R, CCK<sub>B</sub>R, insulin (INS), glucagon (GLUC), and somatostatin by confocal microscopy of purified rat islets (C) and specificity of the reactions (D). The primary and secondary antibodies were incubated in 1 $\times$  phosphate-buffered saline, 0.2% Triton X-100, and 1.4% normal donkey serum overnight at 4  $^{\circ}$ C and 1 h at room temperature. For specificity (D), the peptide antigens and hormones were preincubated at 40  $\mu$ g/ml with their respective antibodies.

Table I Origin and Characteristics of the Antibodies

Primary antibody	Dilution and source	Secondary antibody
Mouse anti-insulin	1:50 (Santa Cruz Biotechnologies)	Rhodamine-conjugated donkey anti-mouse IgG (Santa Cruz Biotechnologies)
Mouse anti-glucagon	1:200 (Sigma Chemical Co.)	Rhodamine-conjugated donkey anti-mouse IgG (Santa Cruz Biotechnologies)
Goat anti-somatostatin	1:50 (generous gift from P. Brazeau)	Alexa Fluor 546 donkey anti-goat (Molecular Probes)
Rabbit anti-CCK <sub>A</sub> R 1122	1:1500 (generous gift from M. L. Kruze)	Alexa Fluor 488 donkey anti-rabbit (Molecular Probes)
Rabbit anti-CCK <sub>B</sub> R 9262	1:200 (generous gift from J. Walsh)	Alexa Fluor 488 donkey anti-rabbit (Molecular Probes)

hormones CCK and gastrin were chosen because their receptors are well characterized and because specific antibodies, raised against parts of these molecules, are also readily available. On these islets, the authors have thus far evaluated (1) the expression of the CCK<sub>A</sub> and CCK<sub>B</sub> receptor mRNA, (2) the presence of these two receptor proteins by Western blot, and (3) the specific localization of each receptor protein subtype among the islet cells by confocal microscopy.

As shown in Fig. 1, the data clearly show that rat purified islets express both CCK<sub>A</sub> and CCK<sub>B</sub> receptor subtypes as indicated by the presence of their respective mRNA (Fig. 1A) and protein (Fig. 1B). Note that the specificity of the two CCK receptor antibodies used was established by preincubation with their specific antigens (Fig. 1B, Peptide; see also Table I). Furthermore, by confocal microscopy, the CCK<sub>A</sub> receptors were specifically localized on insulin and glucagon cells, whereas the CCK<sub>B</sub> receptors appeared exclusively on the somatostatin cells (Fig. 1C). As shown in Fig. 1D, preabsorption of each antibody with its corresponding peptide or hormone confirms the specificity of the data. With this CCK receptor distribution in mind, it becomes possible to design more accurate protocols (1) to study the anti-diabetogenic action of CCK<sub>8</sub> in type 2 diabetics as shown by Ahren, (2) to explain the increased pancreatic contents of somatostatin in diabetic animals and its potential role in the development of type 2 diabetes, and (3) to determine the growth potential of CCK and gastrin on specific islet cells using hormone concentrations corresponding to the affinity of each CCK receptor subtype. It will also become easier to test and eventually use specific CCK receptor antagonists to target and treat pathologies of

the endocrine pancreas. It is sincerely believed that characterization of the receptors and determination of their specific cellular location are critical data needed to study the role of GI hormones in any physiological response attributed to these hormones.

### See Also the Following Articles

CCK (Cholecystokinin) • Gastrin • GI Hormones and Endocrine Pancreas: Expressional Regulation • GI Hormones as Growth Factors • Growth Hormone (GH) • Pancreatic Cancer • Pancreatic Polypeptide (PP)

### Further Reading

- Bonner-Weir, S. (2000). Islet growth and development in the adult. *J. Mol. Endocrinol.* **24**, 297–302.
- Morisset, J., and Grondin, G. (1989). Dynamics of pancreatic tissue cells in the rat exposed to long-term caerulein treatment. 2. Comparative analysis of the various cell types and their growth. *Biol. Cell* **66**, 279–290.
- Morisset, J., Lainé, J., and Mimeau-Worthington, T. (1999). Hormonal control of rat fetal pancreas development. *Biol. Neonate* **75**, 327–336.
- Rhodes, C. J. (2000). IGF-I and GH post-receptor signaling mechanisms for pancreatic  $\beta$ -cell replication. *J. Mol. Endocrinol.* **24**, 303–311.
- Rooman, I., Lardon, J., and Bouwens, L. (2002). Gastrin stimulates  $\beta$ -cell neogenesis and increases islet mass from transdifferentiated but not from normal exocrine pancreas tissue. *Diabetes* **51**, 686–690.
- Sumi, S., and Tamura, K. (2000). Frontiers in pancreas regeneration. *J. Hepatobiliary Pancreatic Surg.* **7**, 286–294.
- Wang, R. N., Rehfeld, J. F., Nielsen, F. C., and Kloppel, G. (1997). Expression of gastrin and transforming growth factor  $\alpha$  during duct to islet cell differentiation in the pancreas of duct-ligated adult rats. *Diabetologia* **40**, 887–893.

number of gastric inhibitory peptide-, substance P-, somatostatin-, and serotonin-producing enteroendocrine cells was also observed in these mice, suggesting that several small intestinal endocrine cell lineages are developmentally related. Finally, expression of a dominant negative Pax6 allele in SEY<sup>Neu</sup> mice results in reduced levels of proglucagon mRNA transcripts in the small and large intestine and an almost complete absence of enteroendocrine cells containing glucagon-like peptide-1 (GLP-1) or glucagon-like peptide-2 (GLP-2) immunoreactivity.

### Enteric Nervous System

The ENS consists of autonomic ganglia located in the submucosal and myenteric plexus and associated connecting neural structures in the bowel wall. Afferent sensory neurons detect both chemical and mechanical stimuli arising from the intestinal wall and lumen and efferent motor neurons interact with effector cells including smooth muscle cells, pacemaker cells, mucosal glands, blood vessels, epithelia, and cells that regulate both immune response and intestinal endocrine function.

The neurons and glial cells of the ENS are derived from the neural crest. During early fetal development, a wave of migrating neural crest (NC) cells (somites 1–5 in the mouse) enters the foregut mesenchyme and colonizes the entire length of the gastrointestinal tract in a rostral-to-caudal manner. Truncal-derived NC cells (somites 6 and 7 in the mouse) also contribute to the development of the ENS in the foregut (esophagus and stomach) and sacral-derived NC cells contribute to the development of the hindgut. As NC cells colonize the gut, these precursor cells undergo cellular proliferation and differentiation into glial cells and other cell types that constitute the ENS. Little is known about the various factors that regulate cellular differentiation; however, genetic loss-of-function experiments in rodents have identified several genes that are essential for regulating the spatial development of the ENS. For example, the transcription factor MASH1 is required for the development of the ENS in the esophagus; glial-derived neurotrophic factor and its receptors RET and GDNF family receptor  $\alpha$ -1 as well as the transcription factor Phox2b are essential for the development of the ENS distal to the stomach; whereas endothelin 3 and its receptor Ednrb/edn3/ece1 as well as the transcription factors sox10 and Hoxa-4 are essential for the development of the ENS in the hindgut. Several members of the *hedgehog* gene family are also important for developmental formation of the ENS.

## ENTEROENDOCRINE CELL-DERIVED PEPTIDES

### Gastrin

In the gastrointestinal tract, gastrin is produced predominantly in G cells located in the gastric antrum and duodenal bulb. The 101-amino-acid progastrin precursor polypeptide is posttranslationally processed into several biologically active molecular forms, including G-34, G-17, and G-14, which can be either amidated or glycine-extended. The amidated forms of gastrin are essential for regulating gastric acid secretion and the progastrin-derived peptides exhibit multiple trophic effects on the stomach, pancreas, and colon.

### *Oncofetal Relationships*

#### *Stomach*

Gastrin content increases gradually from birth and reaches a peak at 5 weeks of age, after which levels decrease in adult rats. Although glycine-extended gastrin (G-Gly) precursors increase in the first day of life, the process of weaning results in a further increase in gastric gastrin content and enhanced gastrin amidation. The amidated forms of gastrin (gastrin-17 and -34) stimulate proliferation of gastric stem cells and enterochromaffin-like (ECL) cells in the oxyntic mucosa, resulting in an increased parietal and ECL mass and ensuring adequate gastric acid production in the rodent stomach during the weaning process.

Mice with a null mutation in the gastrin gene exhibit decreased numbers of parietal and ECL cells and reduced gastric acid production. Gastric carcinoid tumors occur in approximately 30% of patients with hypergastrinemia due to gastrin-producing tumors as part of the multiple endocrine neoplasia type 1 syndrome, whereas approximately 5% of patients with pernicious anemia and hypergastrinemia secondary to reduced parietal cell mass will develop ECL cell carcinoid tumors, which may resolve following surgical resection of the antrum. Long-term treatment of animals with proton pump inhibitors is also associated with hypergastrinemia, ECL hyperplasia, and the development of carcinoid tumors.

#### *Pancreas*

The biological effects of the progastrin-derived peptides on the endocrine pancreas have yet to be clearly defined. The fetal pancreas of rodents contains a large amount of amidated gastrin that is first detected by E15 and persists until just after weaning. This time period is associated with considerable fetal islet

growth and differentiation and suggests a possible role for gastrin in pancreatic endocrine cell development. Transgenic mice expressing amidated gastrin in the islet adult beta cell exhibit a minimal phenotype; however, a doubling of islet cell mass is detected when these mice are crossed with transgenic mice expressing transforming growth factor- $\alpha$  (TGF- $\alpha$ ). Gastrin is not essential for islet development, as gastrin-deficient mice do not demonstrate abnormalities in pancreatic islet morphology.

### Colon

Progastrin and the progastrin-derived peptides can be detected in the rat colon during fetal development. Following birth, amidated gastrins are no longer detected in the colon and the levels of glycine-extended gastrins gradually decrease such that only progastrin can be detected in the adult rat colon. The mechanisms underlying the trophic effects of progastrin and G-Gly on the colonic epithelium remain incompletely defined. Transgenic mice expressing progastrin or G-Gly exhibit increased colono-cyte proliferation and mucosal thickness and are more prone to form aberrant crypt foci following the administration of azoxymethane. Consistent with these observations, targeted inactivation of the gastrin gene results in reduced rates of colonic epithelial proliferation. Gastrin stimulates the proliferation of several colon cancer cell lines expressing the gastrin receptor (CCK<sub>2</sub>); however, the majority of colon cancer cell lines and normal colonic epithelium do not normally express CCK<sub>2</sub>. A truncated gastrin-binding receptor and a constitutively active CCK<sub>2</sub> receptor mutant have been identified in some human colorectal cancers. These findings form the basis for the potential use of gastrin-neutralizing antisera for the treatment of subsets of patients with colon cancer.

### Cholecystokinin

CCK was initially described as a factor that stimulates gallbladder contraction. In the gastrointestinal tract, CCK is expressed in "open-type" enteroendocrine I cells located in the proximal small intestine and in nerve fibers branching to the gastric and colonic myenteric plexus and submucosal plexus, where it acts as a neurotransmitter. The CCK gene encodes a 94-amino-acid prohormone that is posttranslationally processed in a tissue-specific fashion into CCK<sub>83</sub>, CCK<sub>58</sub>, CCK<sub>39</sub>, CCK<sub>33</sub>, CCK<sub>22</sub>, CCK<sub>8</sub>, and CCK<sub>5</sub>, all sharing a common C terminus. The major active form of CCK is an octapeptide containing a sulfated

tyrosine residue and an amidated C-terminal phenylalanine residue.

### Oncofetal Relationships

CCK cannot be detected in fetal and neonatal pancreatic islet cells but is expressed in islets during and after weaning, suggesting that CCK is probably not important for development of the endocrine pancreas. CCK stimulates pancreatic enzyme secretion and exhibits trophic effects on pancreatic acini. Chronically elevated levels of CCK are associated with pancreatic hyperplasia and enhanced tumor formation. Rats fed a soybean trypsin inhibitor that enhances the release of CCK develop preneoplastic pancreatic lesions that eventually progress to carcinoma (Table I). Similarly, long-term pancreatobiliary diversion in the rat causes elevated levels of plasma CCK, pancreatic growth, and premalignant changes. Elevated circulating levels of CCK also enhance the development of preneoplastic acinar lesions induced by azaserine, a potent pancreatic carcinogen in rats. In contrast, loss of CCK signaling produces variable effects on the pancreas. The Otsuka Long-Evans Tokushima Fatty rat fails to express the CCK-A receptor and exhibits reduced pancreatic size, whereas the CCK-A receptor knockout mouse exhibits normal pancreatic morphology. CCK may also play a role in stimulating the growth and invasiveness of human pancreatic cancer cell lines.

### GLP-1 and GLP-2

The proglucagon gene is expressed in pancreatic A cells, in enteroendocrine L cells, and in the brainstem and hypothalamus. In mammals, both glucagon and the glucagon-like peptides are encoded within a single proglucagon precursor prohormone. Post-translational processing of proglucagon occurs in a tissue-specific manner, resulting in the liberation of the proglucagon-derived peptides, which possess remarkably diverse biological actions. In the pancreatic A cell, posttranslational processing liberates glucagon and the major proglucagon fragment, whereas in enteroendocrine L cells and brain, posttranslational processing liberates glicentin, oxyntomodulin, GLP-1, GLP-2, and several spacer or intervening peptides.

### Oncofetal Relationships

Proglucagon-immunoreactive cells and proglucagon mRNA transcripts are first detected in the rat intestine by E14; however, an intestinal profile of

Table 1 Intestinal-Derived Factors and Their Growth-Promoting Effects

Growth factor	Site of intestinal expression	Trophic effects	Oncogenic effects
Gastrin	G cells	Stomach oxyntic mucosa, colonic epithelium, pancreas (?)	ECL cellular hyperplasia, carcinoid tumors, colonic adenocarcinoma
CCK	I cells and ENS	Pancreatic acini	Pancreatic hyperplasia/carcinoma, tumor invasion (?)
GLP-1	L cells	Pancreatic beta cell differentiation and neogenesis	?
GLP-2	L cells	Stomach, small intestine, colonic epithelium	?
NT	N cells and ENS	Small intestinal and colonic epithelium, pancreas (?)	Colonic and pancreatic adenocarcinoma
PYY	L cells	Small intestinal and colonic epithelium	?
TRH	G cells and ENS	Pancreas	?
VIP	ENS	Stimulates the release of neurotrophic factors important in development and functioning of ENS	Promotes growth of non-small-cell lung cancer and pancreatic adenocarcinoma but inhibits growth of gastric adenocarcinoma
Bombesin and related peptides	ENS	Pancreas	Pancreatic adenocarcinoma, gastrinoma, colonic adenocarcinoma, invasion and metastasis of colon and prostate cancer (?)

Note. ECL, enterochromaffin-like; ENS, enteric nervous system; CCK, cholecystokinin; GLP-1, glucagon-like peptide-1; GLP-2, glucagon-like peptide-2; NT, neurotensin; PYY, peptide YY; TRH, thyrotropin-releasing hormone; VIP, vasoactive intestinal peptide.

glucagon-like immunoreactive peptides cannot be detected until E17–19 when the molecular machinery associated with posttranslational processing of proglucagon has developed.

#### GLP-1

GLP-1 inhibits food intake, gastric emptying, and glucagon secretion and stimulates insulin secretion. Although the GLP-1 receptor is expressed in multiple tissues during murine development, a role for GLP-1 during embryonic development has not been established. GLP-1 stimulates beta cell proliferation, beta cell differentiation, and beta cell neogenesis; however, GLP-1 receptor<sup>-/-</sup> mice develop normally and GLP-1R<sup>-/-</sup> islets exhibit only minor abnormalities in islet cell formation, precluding a major role for GLP-1 in islet cell development. Administration of GLP-1 to aging rats leads to an increase in beta cell mass, whereas activation of GLP-1R signaling enhances islet proliferation in rats and mice with experimental diabetes.

#### GLP-2

GLP-2 is produced in the gut during late gestation, following which levels increase progressively in neonatal rats and then subsequently decrease to adult levels. Similarly, GLP-2 receptor expression can be detected along the entire length of the gastrointestinal

tract during late fetal development and during the neonatal period. Exogenously administered GLP-2 induces growth of the stomach, small intestine, and colon of neonatal rats. In contrast, administration of GLP-2 to fetal pigs does not produce an intestinotrophic response. Furthermore, although Pax6 mutant SEY<sup>Neu</sup> mice exhibit a marked reduction in the number of GLP-2-producing L cells, intestinal development appears relatively normal in these mice, strongly suggesting that GLP-2 is not essential for the development of the fetal murine intestine.

GLP-2 exerts trophic effects on the small intestine and colonic mucosa by stimulating crypt cell proliferation and inhibiting apoptosis within the crypt and villus compartments. GLP-2 also exerts multiple actions independent of intestinal growth including enhancement of intestinal epithelial barrier function and stimulation of intestinal hexose transport. The therapeutic utility of GLP-2 has been demonstrated in pilot studies of human subjects with short bowel syndrome and in rodent models of intestinal disease including major small bowel resection, total parenteral nutrition (TPN)-induced intestinal hypoplasia, small and large intestinal inflammation, chemotherapy-induced mucositis, and intestinal ischemia-reperfusion injury. Intriguingly, the GLP-2 receptor has been localized to human endocrine cells and to murine submucosal and myenteric neurons, highlighting the

interaction between the enteric endocrine and nervous systems in controlling intestinal epithelial growth.

### Neurotensin

Neurotensin (NT) is a 13-amino-acid peptide originally discovered in bovine hypothalamus. Peptides that share structural homology with NT include neuromedin N, xenin, and xenopsin. In the gastrointestinal tract, NT is synthesized by N cells located within the small intestinal mucosa, predominantly the ileum, and can also be detected within enteric neurons.

#### **Oncofetal Relationships**

In the human fetal colon, NT is transiently expressed at midgestation between 16 and 18 weeks and the extent of NT expression in the adult human colon is unclear. In the adult rat, NT administration augments the adaptive response to small bowel resection in the intestinal remnant and stimulates growth of the colonic epithelium *in vivo*. NT receptor expression has been detected in human colonic adenocarcinomas and pancreatic cancer cell lines. The importance of NT for pancreatic growth remains unclear as NT binding is detectable by autoradiography in pancreatic cancer cells but not in normal human pancreas.

### Peptide YY

PYY, neuropeptide Y, and pancreatic polypeptide are members of the pancreatic polypeptide superfamily. All three peptides contain 36 amino acids, share considerable amino acid homology, demonstrate amidated C-terminal ends, and possess several tyrosine residues (single-letter abbreviation Y) including tyrosines at both the amino and carboxy termini. Only PYY has been implicated as a putative intestinal epithelial growth factor.

#### **Oncofetal Relationships**

PYY is one of the first hormones to be expressed in the developing fetal gastrointestinal tract and colocalizes with GLP-1 in a subset of enteroendocrine L cells present in the ileum, colon, and rectum. PYY coexpression is detected in the majority of colonic enteroendocrine cells throughout fetal and postnatal development, with the exception of serotonin-producing cells, suggesting a possible common PYY<sup>+</sup> progenitor cell for enteroendocrine cells in the intestine. Immunoreactive PYY has also been detected

in the developing endocrine pancreas and in a subpopulation of glucagon-producing A cells in mature islets, although a role for PYY in the endocrine pancreas has yet to be elucidated.

In the intestine, PYY regulates small and large bowel motility and inhibits both water and chloride secretion. PYY expression is first detected in the rat colon at E17 and levels remain relatively elevated until after weaning (postnatal day 21) when PYY mRNA transcripts decline to adult levels. Administration of PYY to nursing rats increased the weight and DNA content of the duodenum, suggesting a potential trophic role for PYY during intestinal development and dietary adaptation.

In the adult mouse, administration of PYY increased the weight and DNA content of the duodenum, ileum, and colon, whereas co-infusion of PYY in TPN-fed rats increased jejunal mass and protein content. In the pancreas, PYY inhibits pancreatic exocrine secretion but does not exhibit direct trophic effects on either the exocrine or the endocrine pancreas.

### Thyrotropin-Releasing Hormone

Thyrotropin-releasing hormone (TRH) is expressed throughout the gastrointestinal tract in gastric G cells, in pancreatic islet beta cells, and in neurons constituting the myenteric plexus of the esophagus, stomach, and intestine. In the stomach, TRH modulates pentagastrin-stimulated gastric acid secretion and may attenuate acid secretion in subjects with acid secretory disorders.

In the pancreas, TRH is expressed during perinatal development and TRH administration in the adult rodent induces pancreatic hyperplasia and inhibits amylase release. TRH also reduces CCK-induced gallbladder smooth muscle contraction and inhibits cholesterol synthesis within the intestinal mucosa but is not trophic to the intestinal mucosa.

## ENS-DERIVED PEPTIDES

### Vasoactive Intestinal Peptide

Vasoactive intestinal peptide (VIP) is a 28-amino-acid neuropeptide originally isolated from porcine small intestine. VIP belongs to a family of neuropeptides that includes pituitary adenylate cyclase-activating peptide, peptide histidine isoleucine, and peptide histidine methionine, which function as neurotransmitters and neuromodulators of the enteric nervous system.

### **Oncofetal Relationships**

In the human digestive tract, immunoreactive VIP can be detected as early as 10 weeks gestation and in the pancreas as early as 9 weeks gestation. VIP stimulates the release of neurotrophic factors from glial cells, promotes the growth of astrocytes and several non-small-cell lung cancers, stimulates the growth of VIP-1 receptor-bearing pancreatic adenocarcinoma-derived cells *in vitro*, and inhibits the growth of colon adenocarcinoma cells *in vitro*.

### **Bombesin and Related Peptides**

The bombesin family of peptides was originally isolated from frog skin extracts and includes bombesin, gastrin-releasing peptide (GRP; the mammalian homologue of bombesin), neuromedin B, and neuromedin C. Bombesin is a tetradecapeptide, both neuromedin B and neuromedin C are decapeptides, and GRP is a 27-amino-acid peptide.

### **Oncofetal Relationships**

Administration of bombesin to 7-day-old rats resulted in increased stomach, intestinal, and pancreatic weights. GRP is also expressed in human cancer cells, including neuroendocrine tumors of the lung (carcinoids and small-cell lung cancer), breast cancer, neoplastic human thyroid C cells, pancreatic cancer, stomach cancer, and colon cancer. Bombesin stimulates the growth of several human small-cell lung cancers and pancreatic cancer cell lines and the detection of GRP and GRP RNA in neoplastic cells suggests that GRP may potentially function as an autocrine growth factor. Bombesin stimulates enhanced growth of xenografted tumors *in vivo* and GRP receptor mRNA has been detected in colonic tumors but not in normal colonic epithelium.

### **SUMMARY**

Trophic peptides exert effects on the intestinal epithelium via stimulation of crypt cell proliferation and/or inhibition of apoptosis in the crypt and villus compartments. These biological actions may be important for intestinal development and for ensuring adequate epithelial regeneration following intestinal injury. Although the enteric endocrine and nervous systems secrete peptides important for gut growth, intestine-derived growth factors, such as epidermal growth factor, TGF- $\alpha$ , hepatocyte growth factor,

and keratinocyte growth factor, along with growth hormone, insulin-like growth factor-I, interleukin-11, and interleukin-15, also exhibit trophic effects on the gastrointestinal epithelium. These factors function in a coordinated interdependent manner to regulate fetal organ maturation and mucosal healing and repair following epithelial injury. Understanding how these peptides interact to exert their cytoprotective and regenerative effects may be important for realizing the clinical potential of one or more gut peptides and growth factors for the treatment of human intestinal injury.

### **See Also the Following Articles**

CCK (Cholecystokinin) • Gastrin • Gastrin-Releasing Peptide • GI Hormone Development (Families and Phylogeny) • GI Hormones and Endocrine Pancreas: Growth • GI Hormones in Cancer • Glucagon-like Peptide 2 (GLP-2) • Neurotensin • Peptide YY (PYY) • Thyrotropin-Releasing Hormone (TRH)

### **Further Reading**

- Boushey, R. P., Yusta, B., and Drucker, D. J. (2001). Glucagon-like peptide (GLP)-2 reduces chemotherapy-associated mortality and enhances cell survival in cells expressing a transfected GLP-2 receptor. *Cancer Res.* **61**, 687–693.
- Chave, H. S., Gough, A. C., Palmer, K., Preston, S. R., and Primrose, J. N. (2000). Bombesin family receptor and ligand gene expression in human colorectal cancer and normal mucosa. *Br. J. Cancer* **82**, 124–130.
- Drucker, D. J., Ehrlich, P., Asa, S. L., and Brubaker, P. L. (1996). Induction of intestinal epithelial proliferation by glucagon-like peptide 2. *Proc. Natl. Acad. Sci. USA* **93**, 7911–7916.
- Hellmich, M. R., Rui, X. L., Hellmich, H. L., Fleming, R. Y., Evers, B. M., and Townsend, C. M., Jr. (2000). Human colorectal cancers express a constitutively active cholecystokinin-B/gastrin receptor that stimulates cell growth. *J. Biol. Chem.* **275**, 32122–32128.
- Jensen, J., Pedersen, E. E., Galante, P., Hald, J., Heller, R. S., Ishibashi, M., Kageyama, R., Guillemot, F., Serup, P., and Madsen, O. D. (2000). Control of endodermal endocrine development by Hes-1. *Nat. Genet.* **24**, 36–44.
- Naya, F. J., Huang, H., Qiu, Y., Mutoh, H., DeMayo, F., Leiter, A. B., and Tsai, M.-J. (1997). Diabetes, defective pancreatic morphogenesis, and abnormal enteroendocrine differentiation in  $\beta 2$ /NeuroD-deficient mice. *Genes Dev.* **11**, 2323–2334.
- Wang, T. C., Koh, T. J., Varro, A., Cahill, R. J., Dangler, C. A., Fox, J. G., and Dockray, G. J. (1996). Processing and proliferative effects of human progastrin in transgenic mice. *J. Clin. Invest.* **98**, 1918–1929.
- Yusta, B., Boushey, R. P., and Drucker, D. J. (2000). The glucagon-like peptide-2 receptor mediates direct inhibition of cellular apoptosis via a cAMP-dependent protein kinase-independent pathway. *J. Biol. Chem.* **275**, 35345–35352.

Elevation of serum gastrin levels is a common occurrence via a number of pathological, physiological, and pharmaceutical mechanisms. First, it occurs as a result of pernicious anemia, in which autoimmune antibodies directed against parietal cells result in reduced acid secretion. Second, it can occur in patients with Zollinger-Ellison syndrome, caused by islet cell tumors secreting gastrin. Finally, *Helicobacter pylori* infection and administration of proton pump inhibitors lead to an increase in mainly amidated, mature forms of gastrin. Colonic proliferation has been shown to be altered in these conditions of hypergastrinemia, with an increase in normal crypt cell proliferation. Preclinical models evaluating the effect of hypergastrinemia, induced by proton pump inhibitors (PPIs), on the development of carcinogen-induced colon cancer, show no deleterious effect of the hormone. However, in the APC<sup>Min</sup> mouse model of familial adenomatous polyposis and in a human colonic adenoma xenograft grown in nude mice, PPI-induced hypergastrinemia resulted in increased adenoma proliferation and reduced survival. A carefully controlled clinical study has consolidated these preclinical findings by showing that elevated gastrin levels (>90 pg/ml) were associated with a three- to fourfold increased risk of colorectal cancer development.

A number of studies have demonstrated cholecystokinin-2 (CCK<sub>2</sub>)/gastrin receptors on GI tumor cells and exogenously administered gastrin peptides were found to have proliferative effects both *in vitro* and *in vivo*. In addition to the endocrine mechanism of gastrin's action, activation of the gastrin gene within GI adenocarcinoma cells results in the production of immature gastrin peptides, mediating proliferation in an autocrine/paracrine manner. Specific gastrin peptides implicated in the promotion of GI tumor growth include amidated G17, GlyG17, glycine-extended G34, and progastrin.

Gastrin mediates its physiological effects through the G protein-coupled seven-transmembrane domain CCK<sub>2</sub> receptor. The receptors mediating the effects of trophic precursor forms of gastrin still require clarification and universal agreement. However, a number of CCK<sub>2</sub> receptor isoforms that result from alternative splicing have been described, including an amino-terminal truncation and retention of introns 2 and 4. Certain members, in particular those with retained introns, may be tumor-specific and have been implicated in binding precursor gastrin species.

The gastrin gene is a downstream target of  $\beta$ -catenin-mediated transcription, suggesting that

the gastrin autocrine pathway is operational early in the adenoma-carcinoma sequence and may play a role in CCK<sub>2</sub> receptor up-regulation and tumor progression.

Gastrin circumvents apoptosis in a number of experimental systems. Gastrin stimulated the phosphorylation and subsequent activation of the survival-inducing protein kinase B/Akt, activated in many cancers, promoting proliferation and resistance to chemotherapy and radiation. Furthermore, autocrine gastrin produced by the colorectal cancer cell line HCT116 reduced the activation of caspase-3 and up-regulated cytochrome c oxidase Vb. This resulted in a decreased sensitivity of the cells to pro-apoptotic stimuli by retaining cytochrome c within the mitochondria. Camptothecin activation of caspases 3 and 9 was also reduced by progastrin in the intestinal cell lines IEC18 and IEC6.

The incidence of mutated p53 was elevated in *Mastomys natalensis* carcinoid model mice with serum hypergastrinemia. *In vitro* gastrin stimulation of the AGS gastric cancer cell line increased the expression of p53. Such an effect on mutant p53 may also increase resistance to apoptotic stimuli.

The role of gastrin as a transcriptional activator has expanded to include many key targets involved in the establishment and maintenance of malignancy. Gastrin activates the transcription of heparin-binding epidermal growth factor, a potent angiogenic factor, and amphiregulin. This leads to a concomitant up-regulation of the EGF receptor. Gastrin peptides increase activated matrix metalloproteinase 2 secretion in the human colon cancer cell line LoVo, resulting in enhanced invasion, as well as up-regulated cyclooxygenase 2 (COX-2) gene and protein expression. The *Reg* family of genes, which are activated during regeneration of intestinal mucosa and which are positively correlated with colorectal cancer recurrence, have been shown to be up-regulated by gastrin. Long-term administration of the PPI lansoprazole strongly increased the expression of the *Reg* gene, which was reversed by inclusion of a CCK<sub>2</sub> receptor antagonist.

## OPTIONS FOR BLOCKADE OF GASTRIN-MEDIATED GROWTH PATHWAYS

### Gastrin Receptor Blockade

Higher affinity CCK<sub>2</sub> receptor antagonists have been developed, including the benzodiazepam derivative L-365,260, which, in preclinical studies, reverses gastrin-stimulated GI tumor growth. Another CCK<sub>2</sub>

receptor antagonist, Gastrazole (JB5008), has been shown to be highly potent in preclinical studies and has been used in an open-labeled pilot clinical trial in patients with pancreatic cancer. When compared to historical controls, a significant survival advantage was suggested.

A number of CCK<sub>2</sub> receptor antagonists block the effect of Gly-gastrin peptides, thus circumventing potential autocrine gastrin pathways. YM022 was shown to reverse both G17- and GlyG17-stimulated proliferation of the human colon cancer cell line LoVo, in addition to inhibiting basal growth. JMV1155 reduced the GlyG17-stimulated *in vivo* growth of the human colon tumor DLD1.

CCK<sub>2</sub> receptor antagonists that block endocrine and potential autocrine pathways exist. Studies delineating the importance of splice variants in terms of tumor expression and ligand specificity will direct development of the next generation of CCK<sub>2</sub> receptor antagonists.

### Gastrin Immunoneutralization

Anti-gastrin antibodies have been shown to be effective at reducing the basal growth of a gastrin-secreting colon cancer cell line *in vivo* and *in vitro*. A continuous infusion of anti-gastrin antibodies is difficult to administer, may be unstable, may cause anaphylaxis, and may fail to achieve consistent therapeutic serum levels sufficient to neutralize postprandial gastrin surges. Such limitations can be overcome by the use of active immunization against gastrin species. G17DT is a gastrin immunogen, composed of a nonapeptide derived from the amino-terminal of human gastrin-17 linked to diphtheria toxoid which acts as an immunogenic carrier. The G17 epitope is monovalent, thereby preventing complement fixation. Antibodies raised do not cross-react with other hormones, such as G34, smaller C-terminal fragments, or CCK. G17DT antibodies raised in rabbits have a high affinity for G17 and GlyG17 and functionality was confirmed by their ability to competitively displace gastrin ligands (<sup>125</sup>I-G17) from the CCK<sub>2</sub> receptor expressed by a tumor cell line even at dilutions of 1:100 of the original serum titer.

### Preclinical Efficacy of G17DT Antibodies

G17DT has been shown to exert therapeutic effects following passive immunization in nude mouse models of gastric, pancreatic, and colorectal cancer.

### Clinical Studies of G17DT

G17DT has been assessed in phase I/II trials in advanced colorectal cancer patients; increased survival of G17DT patients compared to a well-matched placebo group was observed.

A dose-finding phase II study of G17DT in 22 patients with pancreatic carcinoma demonstrated greater survival in patients who mounted an adequate antibody response than in nonresponders. G17DT is being assessed in phase III trials for gastric, pancreatic, and colorectal cancer.

### SOMATOSTATIN

Somatostatin exerts its actions through interaction with specific heptahelical G protein-coupled plasma membrane receptors. Five different somatostatin receptor subtypes have been cloned in humans. Different receptor subtypes are coupled to different intracellular transmission cascades in a cell type-dependent manner. Somatostatin can also exert cytostatic (G<sub>1</sub>-phase cell arrest) or cytotoxic (apoptosis induction) effects, depending on the receptor subtype expressed on the target cell. In gastroenteropancreatic neuroendocrine tumors, a predominance of sst1 and sst2 with a lesser extent of sst3 and sst5 subtype receptors has been demonstrated.

Since the short half-life of somatostatin makes continuous intravenous infusion mandatory, several long-acting analogues have been synthesized. Of these, octreotide (which binds mainly to somatostatin receptor subtypes sst2 and sst5) has been the most extensively investigated. These synthetic analogues have specific decreasing affinity for sst2 > sst5 > sst3 receptor subtypes and have been used as antiproliferative drugs in the treatment of gastroenteropancreatic tumors. Octreotide and lanreotide treatment resulted in a modest growth-inhibition activity, in functioning or nonfunctioning tumors. Longer-lasting formulations of somatostatin analogues have been developed to provide patients with the convenience of monthly administration and to ensure stable drug serum concentrations between injections. Side effects of these agents consist mainly of gastrointestinal complaints, cholelithiasis, and effects on glucose metabolism.

Inoperable liver tumors have an unfavorable natural course despite various therapeutic modalities. Octreotide, a somatostatin analogue, has shown considerable antitumoral activity in animal models of various hepatic tumors and in isolated cell culture lines. A randomized controlled trial of octreotide in the treatment of advanced hepatocellular carcinoma

has shown a significant survival benefit in the treated patients. Literature reports indicate a stimulatory effect of octreotide on Kupffer cells as a possible antitumor mechanism. Octreotide administration is the best available treatment for advanced inoperable hepatocellular carcinoma and better patient selection, based on receptor subtypes, might further improve the results in the future.

Somatostatin analogues have been used in numerous preclinical studies providing contradictory evidence on the growth inhibition of ductal pancreatic adenocarcinoma. Monotherapy did not result in a prolongation of survival; however, in 15–20% of patients, the progression of the process was halted for several months accompanied by a significant improvement of the clinical condition without notable side effects. Somatostatin analogues have also been used in combination with tamoxifen in which patients with unresectable and resected ductal adenocarcinoma of the pancreas had an apparently increased survival when compared to historical controls.

Octreotide has been used to treat hypersecretion in Zollinger-Ellison syndrome in a small study. It had an antitrophic effect on parietal cell mass.

Octreotide is extremely useful for palliative care. It has analgesic properties when administered by the spinal and intraventricular routes. Its actions in reducing gut motility and secretions make it a valuable adjunct in the management of inoperable bowel obstruction, fistulae, and intractable diarrhea. Octreotide, for example, has been extensively used to ameliorate the gut motor dysfunction that characterizes carcinoid diarrhea.

## SOMATOSTATIN AS A DIAGNOSTIC TOOL

Radiolabeled somatostatin analogues have been employed for the localization of primary and metastatic tumors expressing somatostatin receptors. The so-called “somatostatin receptor scintigraphy” is an important clinical diagnostic investigation for patients with suspected neuroendocrine tumors.

## SOMATOSTATIN TARGETED THERAPY

An extension of somatostatin scintigraphy is somatostatin receptor-targeted chemotherapy and radiotherapy (with conjugates of somatostatin peptides and cytotoxic drugs) and gene therapy (e.g., transferring the *ss2* gene into neoplastic cells). Having been successfully tested in experimental studies, these therapies are being evaluated in clinical trials.

## GASTRIN-RELEASING PEPTIDE

Gastrin-releasing peptide (GRP) and its seven-transmembrane-domain G protein-coupled receptor (GRPR) are frequently expressed by cancers of the gastrointestinal tract, breast, lung, and prostate. Most studies have found that GRP acts to increase tumor cell proliferation, leading to the hypothesis that it is an important mitogen for the growth of these cancers.

Gastrin-releasing peptide receptors are normally expressed on intestinal smooth muscle cells rather than epithelial cells of the GI tract. However, aberrant receptor expression has been shown on GI adenocarcinomas. Early studies revealed that 24–40% of human colon cancers overexpressed GRP receptors. However, a later study showed that 62% of the 50 colonic cancers examined aberrantly overexpressed both GRPR and the GRP hormone, unlike the normal adjuvant epithelium. The receptors were expressed equally across the different grades of tumor and appeared to be down-regulated in metastases, being expressed in only 1 of 37 metastases.

Poorly differentiated tumors were less likely to coexpress GRP and GRPR than well-differentiated tumors and it was concluded that the proteins act as morphogens rather than as mitogens.

GRPR mutations have also been described and result in the lack of production of functional receptor protein, indicating that studies exclusively examining gene expression of the GRPR may not reflect their biological significance in malignant progression. Such mutated GRPRs have also been shown in nonantral gastric adenocarcinomas.

Further evidence for the role of GRPR as a morphogen was provided in both wild-type and GRPR-deficient C57BL/J6 mice treated with the carcinogen, aoxymethane. Tumors that were induced in wild-type mice expressed GRPR and were well differentiated, whereas tumors in GRPR-deficient mice were poorly differentiated mucinous adenocarcinomas.

It was concluded that aberrant expression of GRPR and GRP did not result in larger tumors but due to GRP increasing focal adhesion kinase, the receptors promoted a well-differentiated phenotype.

The lack of a proliferative effect of GRP via GRPR was reinforced in a study of colorectal tumor specimens, where it was shown that 93% of tumor specimens expressed GRPR mRNA, which was expressed at higher levels in tumors with lymphatic vessel invasion but there was no relationship with p53 expression or proliferation index. Activation of GRPR results in stimulation of activation protein 1 expression, which has been shown to impact on

COX-2 expression. In a rat intestinal cell line over-expressing GRPR, bombesin was shown to stimulate COX-2 mRNA and protein expression in addition to prostaglandin E<sub>2</sub> production.

These data suggested that a COX-2-dependent pathway may be responsible for the effect of GRP on GRPR expressed by colon cancer cells.

A number of bombesin receptor antagonists have been described by Schally's group. The bombesin/GRP antagonist RC-3095 inhibited the growth of the human gastric MKN45 xenograft model and the HT29 human colon xenograft model in nude mice.

## FUTURE THERAPEUTIC APPLICATIONS

Radiolabeled agonists may be used for imaging and therapy, as they appear to be internalized, yielding a higher target:background ratio.

## See Also the Following Articles

CCK (Cholecystokinin) • Childhood Cancer, Endocrine Effects of • EGF and Related Growth Factors • Gastrin • Gastrin-Releasing Peptide • Gastrointestinal Neuroendocrine Tumor Syndromes (GI NETS) • GI Hormones as Growth Factors • GI Tract, General Pathology of Endocrine Growths • Pancreatic Cancer • Pancreatic Islet Cell

Tumors • Parathyroid Cancer • Prostate Cancer • Somatostatin Analogs • Thyroid Carcinoma

## Further Reading

- Carroll, R. E., Matkowskyj, K. A., Chakrabarti, S., McDonald, T. J., and Benya, R. V. (1999). Aberrant expression of gastrin-releasing peptide and its receptor by well-differentiated colon cancers in humans. *Am. J. Physiol.* **276**, G655–G665.
- Carroll, R. E., Ostrovskiy, D., Lee, S., Danilkovich, A., and Benya, R. V. (2000). Characterization of gastrin-releasing peptide and its receptor aberrantly expressed by human colon cancer cell lines. *Mol. Pharmacol.* **58**, 601–607.
- Hocker, M., and Wiedenmann, B. (1999). Therapeutic and diagnostic implications of the somatostatin system in gastroenteropancreatic neuroendocrine tumour disease. *Ital. J. Gastroenterol. Hepatol.* **31**, S139–S142.
- Jensen, J. A., Carroll, R. E., and Benya, R. V. (2001). The case for gastrin-releasing peptide acting as a morphogen when it and its receptor are aberrantly expressed in cancer. *Peptides* **22**, 689–699.
- O'Byrne, K. J., Schally, A. V., Thomas, A., Carney, D. N., and Steward, W. P. (2001). Somatostatin, its receptors and analogs, in lung cancer. *Chemotherapy* **47**, 150–161.
- Scarpignato, C., and Pelosini, I. (2001). Somatostatin analogs for cancer treatment and diagnosis: An overview. *Chemotherapy* **47**, 1–29.
- Vainas, I. G. (2001). Octreotide in the management of hormone-refractory prostate cancer. *Chemotherapy* **47**, 62–77.
- Watson, S. A., and Gilliam, A. D. (2001). G17DT—A new weapon in the therapeutic armoury for gastrointestinal malignancy. *Expert Opin. Biol. Ther.* **1**, 309–317.

(and tumors). The gene encoding a specific regulatory peptide may also be expressed by multiple cell types. These patterns may themselves change during evolution, giving rise to species differences in cellular expression.

## Receptors

Hormonal and neurotransmitter peptides typically act at G protein-coupled receptors (GPCRs). GPCRs constitute a large family of structurally related proteins that are thought to have evolved by the duplication and divergence of ancestral genes. Multiple GPCRs may exhibit differential affinity for the various members of regulatory peptide families. The receptors may themselves be expressed by multiple cell types including CNS or PNS neurons and nonneuronal cells, e.g., secretory cells or smooth muscle cells.

## CRITERIA FOR THE IDENTIFICATION OF GUT HORMONES IN PNS OR CNS NEURONS

Evidence for the presence of gut hormones in PNS and CNS neurons should be based on the following lines of evidence: (1) expression of the gene identified by techniques such as reverse transcription-polymerase chain reaction, Northern blot analysis, and *in situ* hybridization and confirmation by cloning and sequencing of the relevant cDNA; (2) the presence of the relevant peptide determined by radioimmunoassay, preferably using antibodies to several epitopes, coupled with high-performance liquid chromatography and confirmation by bioassay, chemical isolation, and amino acid sequencing; and (3) demonstration of cellular origins by immunocytochemistry, preferably using multiple antibodies to several epitopes. Localization of neuropeptides may be a useful way of distinguishing subsets of neurons already identified by the presence of classical transmitters and for this purpose double-labeling immunocytochemistry is useful. There may be problems with the specificity of particular antibodies. For these reasons, the interpretation of studies based on the use of a single experimental method, e.g., immunohistochemistry, using a single antibody should be approached with caution.

## CELLULAR ASPECTS

Regulatory peptides produced in gut endocrine cells, CNS neurons, or PNS neurons are, in each case, generated by mRNA translation at the rough

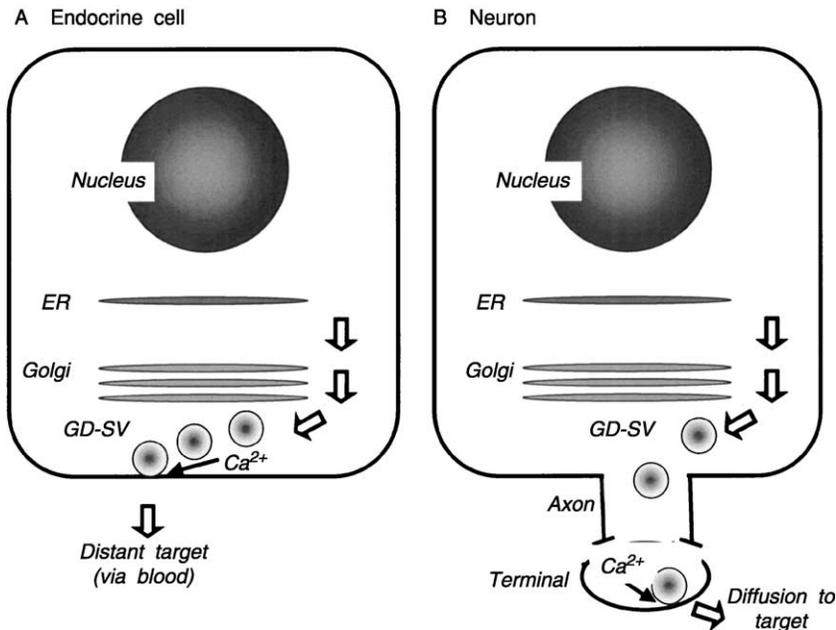
endoplasmic reticulum (Fig. 1). The product (prepro-peptide) is usually biologically inactive and is rapidly converted to a precursor peptide (propeptide) that progresses through the Golgi complex and is sequestered in secretory vesicles that bud from the *trans*-Golgi network. During transit through the Golgi complex, or in secretory vesicles, the propeptide is subject to any of a variety of different posttranslational modifications. These may include glycosylation, Tyr sulfation, and Ser phosphorylation (all of which occur in the Golgi complex), cleavage of the peptide chain, or COOH-terminal  $\alpha$ -amino amidation (which occurs in secretory vesicles). The patterns of posttranslational modification for a specific peptide may differ in neurons and endocrine cells. The Golgi-derived secretory vesicles often possess electron-dense cores and their morphology is characteristic of particular cell types. These vesicles are distinct from the small, clear, synaptic vesicles that recycle at nerve terminals and contain classical transmitters such as acetylcholine, monoamines,  $\gamma$ -aminobutyric acid (GABA), and glutamate. Expression of the genes encoding gut hormones is physiologically regulated, but different mechanisms may regulate the expression of these genes in PNS or CNS neurons.

## COTRANSMISSION

Increased intracellular  $\text{Ca}^{2+}$  leads to exocytosis of both Golgi-derived (neuropeptide-containing) and synaptic vesicle populations (Fig. 2). The actions of neuropeptides are exerted at GPCRs, which are not exclusively localized to postsynaptic membranes or even to neurons. As a consequence, peptides may act over longer distances and for longer times than classical neurotransmitters. Where both neuropeptide and classical transmitters act on the same postsynaptic cell, there may be interactions between them. For example, increased or decreased sensitivity to the classical transmitter may be mediated by neuropeptides.

## OVERVIEW OF GUT HORMONES EXPRESSED BY PNS NEURONS

Representatives of the gut hormone families are commonly expressed in the major divisions of the autonomic nervous system. In postganglionic parasympathetic neurons, the secretin-like peptides VIP, peptide histidine isoleucine amide (PHI), and pituitary adenylate cyclase-activating peptide (PACAP) are



**Figure 1** Schematic representation of the intracellular progression of hormonal peptides produced by gut endocrine cells (A) and by central or peripheral neurons (B). In both cases, mRNA is translated at the endoplasmic reticulum (ER) and precursor peptides progress through the Golgi complex; mature peptides are stored in Golgi-derived secretory vesicles (GD-SV) and released by calcium-dependent exocytosis. Peptide released at the basolateral membrane of endocrine cells is conveyed to target cells via the circulation. In neurons, secretory vesicles progress along the axon to nerve terminals for release; in this case, the secreted peptide typically diffuses to its target cells, although peptides released into the hypothalamo-hypophyseal portal vessels are an exception.

commonly expressed, whereas in postganglionic sympathetic neurons, neuropeptide tyrosine (NPY) is commonly expressed. The enteric nervous system, which is sometimes considered to be a third division of the autonomic nervous system, is an abundant source of neuropeptides, including VIP, CCK, and NPY. Many primary afferent neurons express neuropeptides; some of those belonging to the main families of gut hormones (VIP, NPY) are up-regulated after nerve damage. Gut hormonal peptides are not generally found in somatic efferent neurons.

## OVERVIEW OF GUT HORMONES EXPRESSED BY CNS NEURONS

The expression of neuropeptides is a property of very many CNS neurons. Gut hormones or related peptides, e.g., somatostatin, CCK, and NPY, are expressed in hypothalamic neurons, where they may act in local circuits or in the control of pituitary hormone release after secretion into hypothalamo-hypophyseal portal vessels. In addition, these and other gut hormones are found in many other CNS

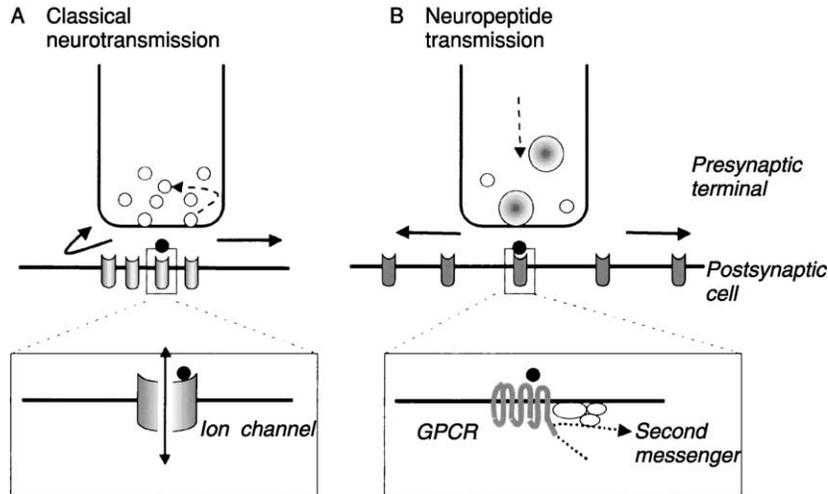
regions, including cerebral cortex, hippocampus, brainstem, and striatum.

## MAJOR GUT HORMONE FAMILIES AND THEIR EXPRESSION IN CNS AND PNS NEURONS

The following brief sketches are grouped on the basis of peptide families that include at least one member that is produced and released by gut endocrine cells and at least one that is produced and released by CNS or PNS neurons.

### The Cholecystokinin/Gastrin Family

The *CCK* gene is expressed in intestinal I-type endocrine cells and in many CNS neurons and some PNS neurons. There are differences in posttranslational processing that account for the reported variation between neurons and endocrine cells in the major form of CCK present. In cerebral cortical neurons, CCK may occur together with GABA and in some nigrostriatal neurons it occurs with dopamine.



**Figure 2** Schematic representation of some events involving neurotransmission by classical (A) and neuropeptide (B) transmitters. Classical transmitters are stored in small, clear synaptic vesicles that are recycled at the nerve terminal. There are reuptake mechanisms for the neurotransmitter, which is resequenced in these vesicles. Receptors for classical transmitters are concentrated in the postsynaptic membrane and are often ion channels. In contrast, neuropeptides are located in Golgi-derived dense-cored vesicles and after secretion they diffuse away from the release site. Neuropeptides typically act at G protein-coupled receptors (GPCR) to trigger second-messenger responses; the receptors may be diffusely distributed on postsynaptic cells and may be some distance from the release site. (●) Putative transmitter.

Gastrin originates from pyloric antral G cells and may be found in small amounts in hypothalamus. The  $CCK_A$  (also called  $CCK_1$ ) receptor has high affinity for CCK and low affinity for gastrin; it is expressed in pancreas, gallbladder, and some CNS and PNS neurons. The gastrin- $CCK_B$  (also called  $CCK_2$ ) receptor has high affinity for both gastrin and CCK and is found on parietal and enterochromaffin-like cells and many CNS neurons. The main function that has been ascribed to CCK in the brain is inhibition of food intake. In addition, CCK is thought by some to be associated with anxiety or panic attacks and it may modulate noxious sensations.

### The Secretin/Glucagon/VIP/PHI/PACAP/GIP Superfamily

The superfamily of structurally related peptides that includes the classical gut hormones secretin and glucose-dependent insulinotropic peptide (GIP) and the pancreatic hormone glucagon also includes major neuropeptide transmitters such as VIP, PHI, and PACAP. In the case of glucagon, there is good evidence that a single precursor molecule may yield several different products through alternative post-translational processing. In the pancreatic islet alpha cells, glucagon itself is a primary product; in intestinal L cells, glucagon-like peptide-1 (GLP-1) and GLP-2

are primary products (but not glucagon). CNS neurons resemble intestine more closely than pancreas in the processing of the glucagon precursor; one possible role for the CNS GLPs is in the control of food intake. Two members of the family that seem to be predominantly expressed in neurons, and not endocrine cells, are VIP (the precursor of which also gives rise to the related peptide PHI) and the closely related peptide PACAP.

### Somatostatin

Somatostatin-producing endocrine (D) cells are found throughout the gastrointestinal tract and in the islets of Langerhans. In these systems, somatostatin acts locally as a paracrine inhibitor of secretion from nearby cells, e.g., G cells in the pyloric antrum and beta cells in the pancreas. In addition, somatostatin is produced in hypothalamic neurons and functions as an inhibitor of growth hormone secretion from the anterior pituitary. It is also found in enteric neurons, sympathetic neurons, and some somatic afferent neurons.

### Neurotensin

Neurotensin was first isolated from the hypothalamus and then discovered to be produced in N-type

endocrine cells of the ileum and colon, where it is a putative mediator of the ileal brake in the gut. Neurotensin has been identified in multiple CNS neurons; interactions with dopamine have been identified and possible actions as an anti-psychotic agent reported.

### **The Peptide YY/NPY/Pancreatic Polypeptide Family**

Peptide YY is named for the presence of tyrosine residues (Y in the single-letter notation) at both the COOH- and NH<sub>2</sub>-terminal positions. It is normally produced in endocrine cells of the ileum and colon. Putative hormonal actions include the inhibition of pancreatic secretion, intestinal transit, and food intake. There is little evidence for neurotransmitter functions. But the closely related NPY is widely expressed in CNS and peripheral neurons. One of its most striking actions is the stimulation of food intake on injection into the hypothalamus. In addition, it is present in sympathetic postganglionic neurons and may modulate responses to noradrenaline. Another member of the family is pancreatic polypeptide, produced in the islets of Langerhans; its functions as either a hormone or a neurotransmitter remain uncertain.

### **Ghrelin and Motilin**

The related peptides ghrelin and motilin are produced in gastric and small intestinal endocrine cells,

respectively. They are distinct from other gut hormones in that they are released from these cells during fasting. Ghrelin is a naturally occurring ligand for an orphan receptor originally identified as a putative mediator of growth hormone secretion. Small amounts are thought to be produced in the hypothalamus. It is a powerful stimulant of food intake and it is thought that the peptide released from the stomach may stimulate food intake by delivery to the hypothalamus through the circulation.

### **See Also the Following Articles**

CCK (Cholecystokinin) • Gastrin • Ghrelin • GI Hormones Outside the Gut: Other Tissues • Motilin • Neuropeptide Y • Neurotensin • Peptide YY (PYY) • Pituitary Adenylate Cyclase-Activating Polypeptide (PACAP)/Glucagon Superfamily • Somatostatin Analogs

### **Further Reading**

- Dockray, G. J. (1999). The brain gut axis. In "Textbook of Gastroenterology" (T. Yamada, D. H. Alpers, L. Laine, C. Owyang, and D. W. Powell, eds.), 3rd ed. Vol. 1, pp. 67–81. Raven Press, New York.
- Furness, J. B. (2000). Types of neurons in the enteric nervous system. *J. Auton. Nervous Syst.* **81**, 87–96.
- Hokfelt, T., Broberger, C., Xu, Z. Q., Sergeev, V., Ubink, R., and Diez, M. (2000). Neuropeptides—An overview. *Neuropharmacology* **39**, 1337–1356.
- Walsh, J. H., and Dockray, G. J. (1994). "Gut Peptides: Biochemistry and Physiology." Raven Press, New York.

## Secretin Family

Several members of the secretin family [secretin, glucagon, vasoactive intestinal polypeptide (VIP), pituitary adenylate cyclase-activating peptide (PACAP), and growth hormone-releasing factor (GRF)] have been localized to Leydig cells, germ cells, or both. Additionally, PACAP, VIP, and GRF have been localized to the ovaries and GRF has also been localized to the placenta. Secretin expression has been detected in the pituitary and VIP expression has been noted in adrenal chromaffin cells, in mast cells, and in leukocytes. VIP suppresses T cell proliferation and production of interleukin-2 (IL-2), IL-4, and IL-10 and acts as an anti-inflammatory mediator. However, VIP may also enhance certain aspects of lymphocyte function and these opposing actions may be mediated by different subtypes of VIP receptors. The effects of VIP on lymphocytes also reflect the rather abundant innervation of lymphoid tissues by VIP nerves.

## Neurotensin

Neurotensin has been detected in adrenal chromaffin cells and in leukocytes. In addition, neurotensin is produced by SCLC cells.

## Somatostatin

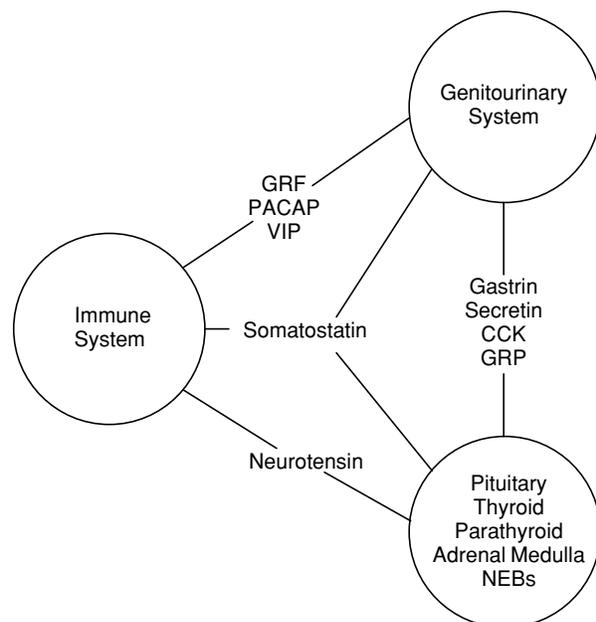
Somatostatin shows a widespread distribution and has been localized to neuroendocrine neurons ending on hypophyseal portal blood vessels in the median eminence, thyroid parafollicular C cells, parathyroid cells of some species, adrenal chromaffin cells, Langhans cells of the epidermis, endocrine-like cells of the prostate and toad urinary tract, and leukocytes. In all of these locations, it seems certain that somatostatin acts as a local paracrine mediator or as a short-range (portal system-delivered) hormone. In general, somatostatin acts as an inhibitor of secretory activities and also reduces cell proliferation, possibly through actions on specific phosphatases. Somatostatin has also been localized to several types of tumors, including breast cancers. It inhibits the proliferation of several normal and neoplastic cell types. Interestingly, antisense oligonucleotides blocking somatostatin expression stimulate the proliferation of rat splenocytes. This effect could be inhibited by the addition of exogenous somatostatin. Additionally, somatostatin inhibits the release of mediators (histamine, leukotriene D4) from human and leukemic basophils.

## Gastrin-Releasing Peptide/Bombesin

Gastrin-releasing peptide (GRP) has been localized to somatotrophs in the anterior lobe and to melanotrophs of the intermediate lobe of the pituitary and it exerts growth hormone-stimulating activity. GRP also occurs in NEBs of the bronchopulmonary tree, particularly during fetal and neonatal life, and has been localized to endocrine-like cells of the prostate. It is produced by a wide selection of human cancer cells of lung (SCLC), mammary, and prostate origin in which it may act as an autocrine growth factor.

## CONCLUSIONS

The above data clearly demonstrate that many regulatory peptides simultaneously are produced by cells derived from all three germ layers and that their designation as either gastrointestinal or hormonal is contextual rather than absolute. Many of the peptides may play roles in normal organs, in reproductive functions, and in the immune and cellular defense system of the body (Fig. 1). Additionally, some factors are transiently expressed during development and may serve as tumor growth factors. Indeed, an impressive array of peptides are expressed by lung, gastrointestinal, pancreatic, breast, and prostate cancers and pre-clinical data suggest that they are potential targets for therapy.



**Figure 1** Scheme summarizing some of the overlaps in the distribution of GI hormones and regulatory peptides.

## Acknowledgments

Grant support for this work was provided by the Danish Medical Research Council and Cancer Society.

## See Also the Following Articles

CCK (Cholecystokinin) • Gastrin • GI Hormones and Endocrine Pancreas: Expressional Regulation • GI Hormones Outside the Gut: Central and Peripheral Nervous System

## Further Reading

- Aguila, M. C., Rodriguez, A. M., Aguilla-Mansilla, H. N., and Lee, W. T. (1996). Somatostatin antisense oligodeoxynucleotide-mediated stimulation of lymphocyte proliferation in culture. *Endocrinology* **137**, 1585–1590.
- Bajo, A. M., Schally, A. V., Krupa, M., Hebert, F., Groot, K., and Szepeshazi, K. (2002). Bombesin antagonists inhibit growth of MDA-MB-435 estrogen-independent breast cancers and decrease the expression of the ErbB-2/HER-2 oncoprotein and c-jun and c-fos oncogenes. *Proc. Natl. Acad. Sci. USA* **99**, 3836–3841.
- Heasley, L. E. (2001). Autocrine and paracrine signaling through neuropeptide receptors in human cancer. *Oncogene* **20**, 1563–1569.
- Kimoto, Y. (1998). A single human cell expresses all messenger ribonucleic acids: The arrow of time in a cell. *Mol. Gen. Genet.* **258**, 233–239.
- Lambrecht, B. N. (2001). Immunologists getting nervous: Neuropeptides, dendritic cells and T cell activation. *Respir. Res.* **2**, 133–138.
- Larsson, L.-I. (1988). Regulatory peptides and amines during ontogeny and in non-endocrine cancers: Occurrence and possible functional significance. *Prog. Histochem. Cytochem.* **17**, 1–222.
- Schalling, M., Persson, H., Pelto-Huikko, M., Ødum, L., Ekman, P., Gottlieb, C., Høkfelt, T., and Rehfeld, J. F. (1990). Expression and localization of gastrin mRNA and peptide in spermatogenic cells. *J. Clin. Invest.* **86**, 660–669.
- Sherwood, N. M., Krueckl, S. L., and McRory, J. E. (2000). The origin and function of the pituitary adenylate cyclase-activating polypeptide (PACAP)/glucagon superfamily. *Endocr. Rev.* **21**, 619–670.

**Table I History of Gut Endocrine Cells**

Heidenhein (1870)	Chromaffin cells described
Nussbaum (1879),	Gritzner and Menzel (1879) Osmiophilic cells described
Kultschitzky (1897)	Basigranular acidophil cells described
Ciaccio (1907)	Term enterochromaffin cells coined
Kull (1913)	Nonargentaffin endocrine-like cells described
Masson (1914)	Argentaffin cells described
Vialli and Erspamer (1937)	Enteramine in EC cells described
Feyrter (1938)	Clear cells described
Erspamer (1939)	Argyrophilic cells (interpreted as EC cell progenitors) described
Dawson (1948)	Argentophilic (argyrophil) cells in the stomach described
Erspamer and Asero (1952)	Enteramine determined to be identical to 5-HT (serotonin)

formaldehyde vapors under certain conditions (the Falck-Hillarp technique). Using this technique, it was found that the endocrine cells in the GI tract had the capacity to form fluorogenic monoamines, such as dopamine and 5-HT, on administration of the corresponding amine precursor (L-DOPA and 5-hydroxytryptophan, respectively) and to store them in large amounts in their secretory granules. These properties of the cells have sometimes been referred to as APUD (amine precursor uptake and decarboxylation) and became an important means of collectively demonstrating gut endocrine cells, at the same time providing the possibility of discriminating the

EC cells (displaying a yellow fluorescence due to their high endogenous content of 5-HT) from other endocrine cells (displaying green fluorescence due to the newly formed dopamine after L-DOPA administration). These amine-handling properties of gut endocrine cells, as well as some other histochemical features, resemble those of neurons and it was long thought that the embryonic origin of the gut endocrine cells was neuroectodermal. However, it is generally agreed that the endocrine cells, like other epithelial cells lining the GI tract, are of endodermal origin.

During the past few decades, immunocytochemistry, with the use of antibodies against individual hormones or other specific cell markers, has become the most important technique for distinguishing the individual endocrine cell populations with respect to their cellular morphology, their regional and topographic distribution, and—when hormone antibodies are used—their function. However, there are still histochemically or ultrastructurally defined cell populations to which a hormone has not yet been linked. Conversely, there are cell populations that differ in their hormone content, but nevertheless share both light microscopic features and secretory granule appearance.

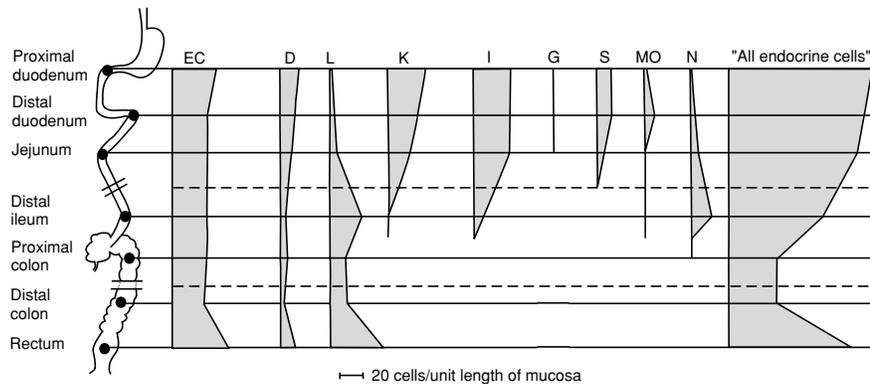
## REGIONAL DISTRIBUTION

The endocrine cells have as a rule a regional distribution along the GI tract that is characteristic for each

**Table II Endocrine Cells in the Mammalian Gastrointestinal Tract**

Cell type	Stomach		Small intestine		Large intestine	Hormone
	Oxyntic	Pyloric	Upper	Lower		
A	+ <sup>a</sup>	–	–	–	–	Glucagon
A-like/X	+	+	+	–	–	Ghrelin
D	+	+	+	+	+	Somatostatin
D <sub>1</sub>	+	+	+	+	+	Unknown (ghrelin in human)
EC	+	+	+	+	+	Serotonin + various peptides
ECL	+	–	–	–	–	Histamine
G	–	+	+	–	–	Gastrin
I	–	–	+	+	–	Cholecystokinin
K	–	–	+	+	–	GIP
L	–	–	+	+	+	Gut glucagon/GLP-1 and -2 + PYY
MO	–	–	+	+	–	Motilin
N	–	–	–	+	–	Neurotensin
P	+	+	+	–	–	Unknown
S	–	–	+	+	–	Secretin

<sup>a</sup>In certain species, e.g., carnivores.



**Figure 1** Schematic diagram illustrating the distribution and relative frequency (number of cells/unit length of mucosa) of several different populations of endocrine/paracrine cells in the human intestine. Data are based on immunocytochemical observations using hormone antibodies. (Left) Black dots on the outline of the intestine indicate tissue sampling sites. (Right) A compilation of the data. Note that some cell types are distributed throughout the intestine, whereas other cell types have a more restricted distribution. Note also that the rectum is almost as rich in endocrine/paracrine cells as the duodenum.

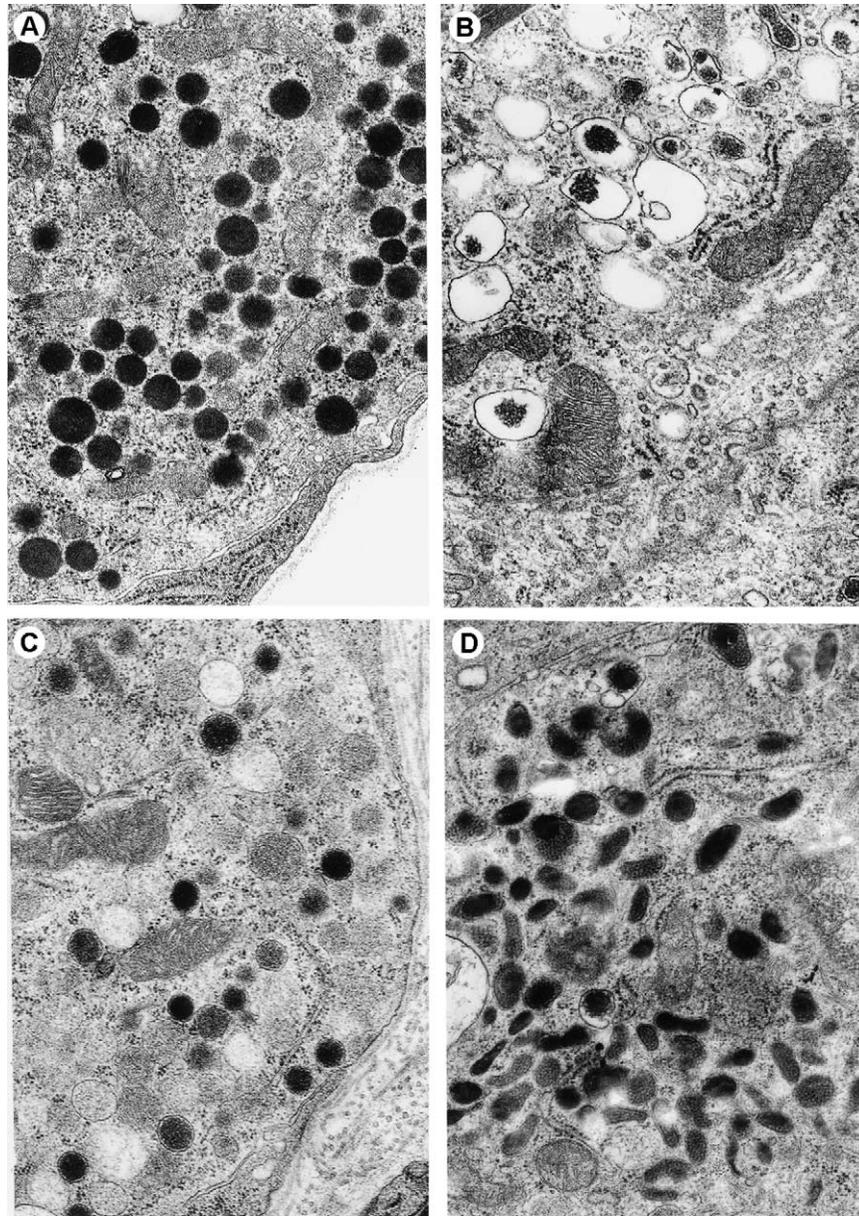
cell type (Table II, Fig. 1). Thus, certain cell types are restricted to, or greatly predominate, in the stomach. Others are restricted to the upper small intestine and still others predominate in the distal small intestine and/or the large intestine. A few cell types are distributed all along the GI tract. Interestingly, there are several indications that such widely distributed cells may serve to influence cells in their vicinity rather than to deliver their product(s) to the circulation. Such widely distributed cells are therefore often described as paracrine cells. However, since the possibility cannot be excluded that they may, in fact, have dual paracrine and endocrine functions, and since they may share several major morphological features with cells known to be genuinely endocrine, they are sometimes referred to as endocrine/paracrine. With respect to genuinely endocrine cells, it has been argued that each such population should have a restricted distribution along the GI tract in order for the target cells to be able to respond to a circulating messenger in a meaningful way. Nevertheless, an established endocrine function of a cell does not preclude an additional paracrine function.

## Stomach

Based on various histochemical techniques, including silver stains, the APUD properties, and immunohistochemistry, and, in particular electron microscopy, with special emphasis on the morphological features of the secretory granules (exemplified in Fig. 2), at least seven endocrine cell types have been described in the stomach. They comprise G cells, D cells, EC cells, enterochromaffin-like (ECL) cells, A-like cells (also referred to as X cells), and two types of cells with small

granules and designated P cells and D<sub>1</sub> cells (Table II). Additionally, in certain species, notably carnivores such as dogs and cats, “true” glucagon-producing A cells, being well-known constituents of the pancreatic islets, are also present in substantial numbers in the stomach. The ECL cells constitute one—and possibly the only—gastric endocrine cell type that is confined to the acid-producing, upper part (corpus fundus or body) of the stomach (Table III). This seems to be the case in all vertebrate species examined, from fish to human. Moreover, the ECL cells are the largest endocrine cell population in this area of the stomach. The ECL cells have a very characteristic secretory granule ultrastructure, which at least in part seems to be linked to the histamine production of these cells. Also, the A-like cells, and, when present, the A cells, are primarily found in this part of the stomach, but they may also occur in smaller numbers in the distal non-acid-producing part (antrum, pylorus) of the stomach.

The G cells are exclusively found in the antrum, where they are located within a zone at the upper part of the glands in most species. However, in rodents they predominate at the basal portion of the glands. The EC cells greatly predominate in the antrum in certain mammals, such as rodents, whereas in the stomach of human and many larger mammals they also occur in the acid-producing part of the stomach in quite high numbers. The D cells are diffusely distributed all over the stomach in all species. These cells very often have narrow cytoplasmic processes that can have a beaded, nerve fiber-like appearance, ending with a knob-like swelling filled with granules (Fig. 3). The small granule cells (P and D<sub>1</sub> cells), like the D cells, seem to occur all over the stomach. The

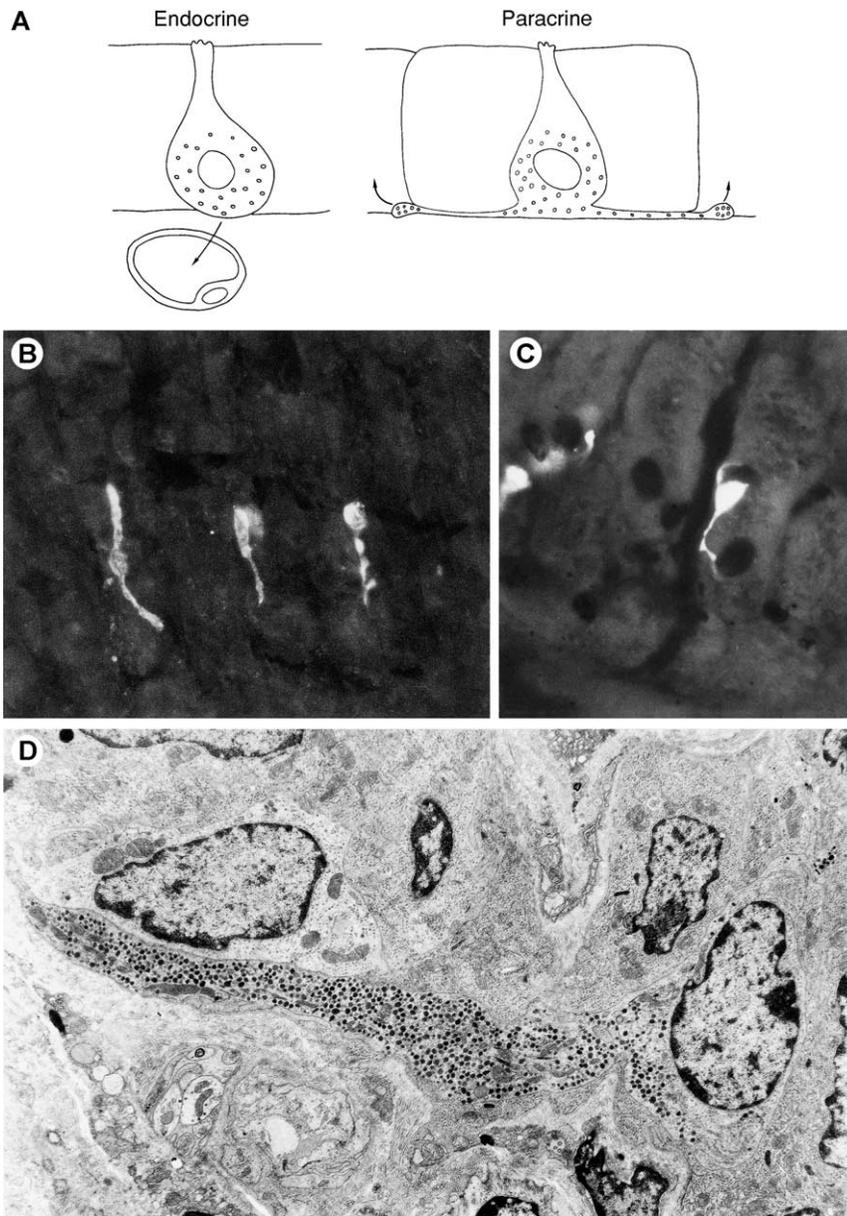


**Figure 2** Electron micrographs of secretory granules of endocrine/paracrine cells in the rat stomach. (A) A-like cell. The granules are round and highly electron-dense. (B) ECL cell. The granules are of the vesicular type with a flocculent electron-dense core surrounded by a wide electron-lucent halo. (C) G cell. Note the varying electron density of the granules. (D) EC cell. Note granule pleomorphism.

**Table III** Relative Frequency of Different Endocrine Cells in Gastric Oxyntic Mucosa of Rat and Human

Cell type	Rat (%)	Human (%)
ECL cells	65	35
A-like (X) + P + D <sub>1</sub> cells	24	14
Somatostatin cells	10	26
EC cells	0	25

D cells have been found to often make direct contact with G cells in the antrum and with parietal cells in the acid-producing part of the stomach. This contact is usually established via the cytoplasmic processes of the D cells. These features, together with the wide distribution of the cells both within the GI tract and outside of it (the pancreatic islets in particular), have contributed to the designation of the D cells as paracrine cells.



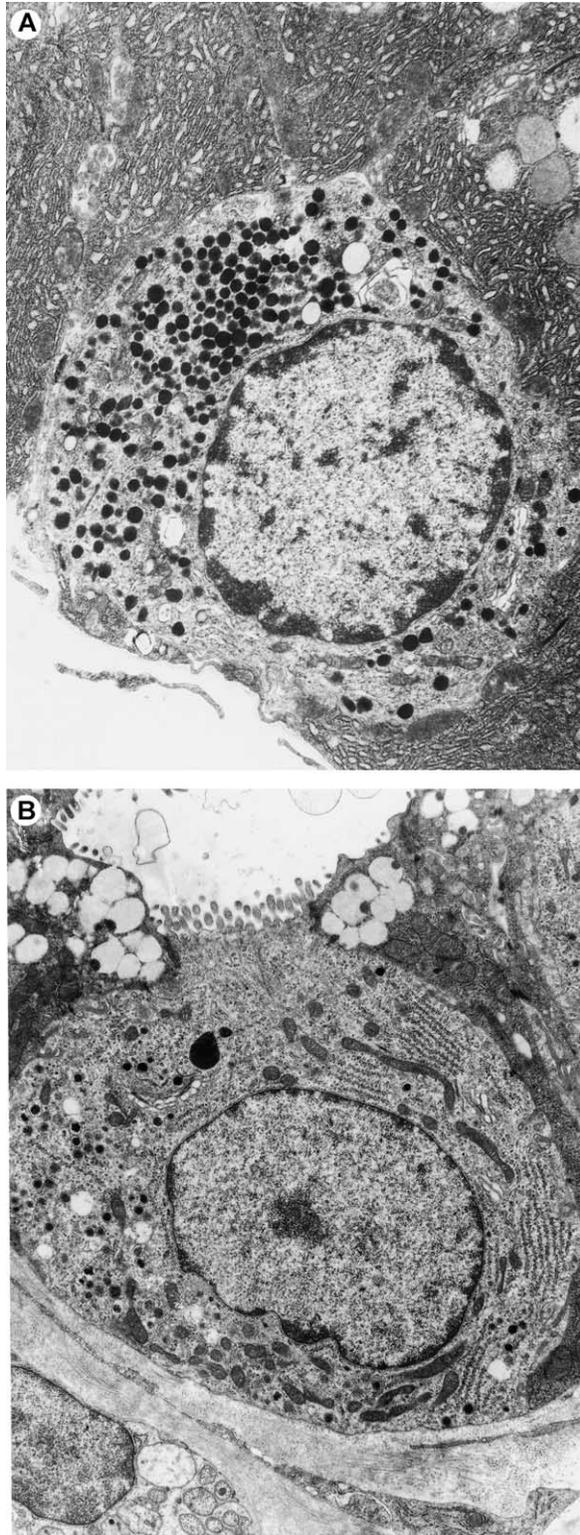
**Figure 3** (A) Schematic morphological distinction between endocrine and paracrine cells. (B and C) D cells of rat gastric mucosa (somatostatin immunofluorescence). (B) The cells issue long, slender cytoplasmic processes, sometimes ending in a knob-like swelling. (C) D cell process ending on a G cell (stained black by gastrin immunoperoxidase). (D) Electron micrograph showing a D cell process filled with secretory granules. The process runs along the base of the gland epithelium near the top of the figure.

A general morphological characteristic of the endocrine cells in the upper, acid-producing part of the stomach (i.e., the corpus or fundus) is that they are “closed” in that the apical end of the cell does not reach the gland lumen. In the antrum, on the other hand, the vast majority of cells do reach the lumen via a narrow neck (Fig. 4). One possible explanation for this morphological difference is that open cells respond to luminal stimuli (e.g., nutrients, pH

changes), whereas closed cells are sensitive to other types of stimuli (e.g., distension, temperature changes, neuronal and hormonal messengers).

### Small Intestine

Endocrine cells are distributed as single cells within the epithelium of both crypts and villi, but collectively they predominate in the crypts. They are



**Figure 4** Electron micrographs illustrating (A) an endocrine cell of closed type (A-like cell) in the acid-producing mucosa and (B) an open endocrine cell in the antral mucosa (G cell) of the rat stomach. Note numerous microvilli at the luminal surface.

characteristically flask-shaped with a narrow, often quite elongated neck reaching the lumen and thus of open type. The secretory granules accumulate at the base of the cell, but may sometimes be more diffusely distributed, although they are always confined to the cytoplasm and are never found in the nucleus. The classification and naming of the intestinal endocrine cells were for a long time based primarily on the size and morphology of the secretory granules. Accordingly, the cells were categorized as, e.g., S (small granules), I (intermediate-size granules), and L (large granules) cells ([Table II](#), [Fig. 1](#)). Subsequent immunohistochemical observations about the cellular localization of various hormones complicated matters and resulted in attempts to retain the letter classification, the inevitable renaming of some cells, and the addition of hormonally classified cells that did not fit into previous classifications. Thus, the classification of endocrine cells in the small intestine is often a mixture of letter naming and naming based on the hormone content of the cells. At least 11 different cell types have been identified ([Table I](#)). The EC cells are distributed throughout the small intestine. The D cells also occur throughout the small intestine, although in gradually smaller numbers distally. Some cell types are more restricted in their regional distribution. Thus, S cells, I [cholecystokinin (CCK)] cells, K cells, and MO (motilin) cells predominate in the duodenum and upper jejunum and are only rarely found in the ileum. The opposite is true for L cells and N cells, which predominate in the ileum, but are scarce in the duodenum. According to ultrastructural criteria, P cells are also present in the small intestine, with a distribution mainly in the upper part. One cell type with very large granules (VL cells) has been reported to occur throughout the human small intestine.

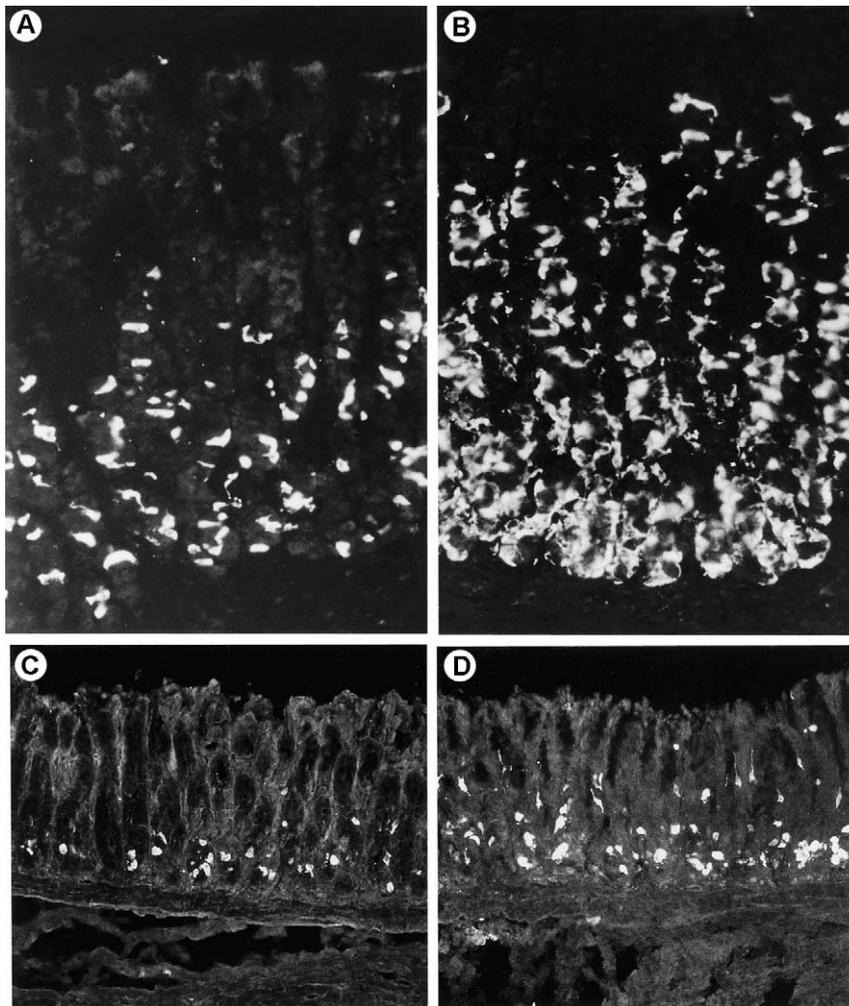
### Large Intestine

The large intestine also harbors endocrine cells of several different kinds, although it seems that the number of different populations is somewhat smaller than in the small intestine. Interestingly, the endocrine cells are collectively more numerous in the rectum than in the colon, as calculated in the human gut ([Fig. 1](#)). At least five different endocrine cell populations can be distinguished in the large intestine. It is notable that there is a marked overlap in the distribution of certain cell populations between the distal small intestine and the large intestine ([Table II](#)). Thus, EC cells and L cells are quite numerous also in the large intestine. N cells, on the

other hand, are only rarely found in the large intestine and D cells are also clearly fewer in number here than in the small intestine. As to morphology, many of the endocrine cells have a very characteristic appearance in that each cell at its base issues a long, tapering, sometimes beaded process running along the base of the epithelium. Furthermore, the process often seems to be directed toward the base of the crypt. This gives the cells a paracrine-like appearance. However, since cells thought to play a role as hormone producers, such as the L cells, also have this morphology, it is still open to speculation why the cells are equipped with such long processes. It is not inconceivable that the cells serve both endocrine and paracrine functions.

## CONCLUDING REMARKS

As is obvious from the data presented here, there is some confusion and remaining uncertainties as to the classification and naming of endocrine cells in the GI tract. This is due to the fact that endocrine cells sharing staining characteristics with cells occurring outside the GI tract, notably, the pancreatic islets, which are known by the “islet” name. This applies to islet A cells and D cells. Another problem arose when an ultrastructurally defined cell type, e.g., the L cell, was later found to comprise at least two functionally distinct cell populations, one producing proglucagon-derived peptides (retaining the L cell designation) and another producing neurotensin (renamed N cells).



**Figure 5** Endocrine cell plasticity illustrated by the marked hyperplasia of rat ECL cells (B) and G cells (D) on pharmacological blockade of acid secretion for 8 weeks. Acid blockade causes G cell activation and hyperplasia with hypergastrinemia, which in turn brings about activation of the ECL cells with increased histamine secretion and, with time, proliferation. ECL cells are visualized with histidine decarboxylase immunofluorescence and G cells are visualized with gastrin immunofluorescence.

Another problem is related to the I cell. Obviously, several functionally distinct endocrine cell populations fall into this somewhat vague category. Thus, CCK cells, K cells, and MO cells are all “I” cells, but with the current possibilities of a more detailed functional classification, the hormone name, or its short name, is gradually replacing the original letter name.

Furthermore, the largest population of endocrine cells, the EC cells, are probably not one single cell type, but rather several functionally distinct populations, unified merely because they produce large amounts of serotonin. This view is favored by available ultrastructural data showing different types of granules in EC cells in different regions of the gut, and sometimes even in different EC cells within the same region, together with the fact that certain regulatory peptides are restricted to subpopulations of EC cells only.

Finally, it ought to be mentioned that there is a remarkable potential for plasticity in gut endocrine cell systems. This is perhaps most readily observed in the stomach, where each of the endocrine cell populations has a restricted regional distribution and the cells are more frequent per unit area. Thus, marked hyperplasia of G cells and ECL cells (within their normal regional boundaries) evolves within some time (days to a few weeks) after, e.g., profound pharmacological blockade of acid secretion (Fig. 5) and ECL cell hypoplasia after, e.g., removal of the G cells (antrectomy).

### See Also the Following Articles

GI Hormone Development (Families and Phylogeny) • GI Hormones Outside the Gut: Central and Peripheral Nervous System • GI Hormones Outside the Gut: Other Tissues • GI Tract, General Pathology of Endocrine Growths

### Further Reading

- Bordi, C., and D’Adda, T. (1991). Ultrastructural morphometry of gastric endocrine cells. *In* “The Stomach as an Endocrine Organ” (R. Håkanson and F. Sundler, eds.), pp. 53–69. Elsevier, Amsterdam, The Netherlands.
- Buchan, A. M. J. (1999). Structure and function of gastrointestinal endocrine cells. *In* “Gastrointestinal Endocrinology” (G. H. Greeley, ed.), pp. 1–30. Humana Press, Totowa, NJ.
- Capella, C., and Solcia, E. (1972). The endocrine cells of the pig gastrointestinal mucosa and pancreas. *Arch. Histol. Jpn.* **35**, 1–29.
- Capella, C., Finzi, G., Cornaggia, M., Usellini, L., Luinetti, O., Buffa, R., and Solcia, E. (1991). Ultrastructural typing of gastric endocrine cells. *In* “The Stomach as an Endocrine Organ” (R. Håkanson and F. Sundler, eds.), pp. 27–51. Elsevier, Amsterdam, The Netherlands.
- Dayal, Y. (1999). Neuroendocrine cells of the gastrointestinal tract: Introduction and historical perspective. *In* “Endocrine Pathology of the Gut and Pancreas” (Y. Dayal, ed.), pp. 1–31. CRC Press, Boca Raton, FL.
- Feyrter, F. (1938). “Über diffuse endokrine epitheliale Organe.” J. A. Barth, Leipzig, Germany.
- Sjölund, K., Sandén, G., Håkanson, R., and Sundler, F. (1983). Endocrine cells in human intestine: An immunocytochemical study. *Gastroenterology* **85**, 1120–1130.
- Solcia, E., Capella, C., Buffa, R., and Frigerio, B. (1976). Histochemical and ultrastructural studies on the argentaffin and argyrophil cells of the gut. *In* “Chromaffin, Enterochromaffin and Related Cells” (R. E. Coupland and T. Fujita, eds.), pp. 209–225. Elsevier, Amsterdam, The Netherlands.
- Solcia, E., Capella, C., Vassallo, G., and Buffa, R. (1975). Endocrine cells of the gastric mucosa. *Int. Rev. Cytol.* **42**, 223–286.
- Sundler, F., and Håkanson, R. (1988). Peptide hormone-producing endocrine/paracrine cells in the gastro-entero-pancreatic region. *In* “Handbook of Chemical Neuroanatomy” (A. Björklund, T. Hökfeldt, and Ch. Owman, eds.), Vol. 6, pp. 219–295. Elsevier, Amsterdam, The Netherlands.
- Sundler, F., and Håkanson, R. (1991). Gastric endocrine cell typing at the light microscopic level. *In* “The Stomach as an Endocrine Organ” (R. Håkanson and F. Sundler, eds.), pp. 9–26. Elsevier, Amsterdam, The Netherlands.
- Sundler, F., Håkanson, R., Lorén, I., and Lundquist, I. (1980). Amine storage and function in peptide hormone-producing cells. *Invest. Cell Patol.* **3**, 87–103.

manipulation in mice. Proliferating or transformed gut endocrine cells may variably express the same antigens as their normal counterpart depending on their differentiation status. Expression of the SSR2 in endocrine tumor cells is important for both diagnostic and therapeutic applications.

## NONNEOPLASTIC GROWTHS OF GUT ENDOCRINE CELLS

Nonneoplastic growths of gut endocrine cells have been reported only in the stomach and occasionally in the small intestine.

### Stomach

Nonneoplastic growths are restricted to proliferation of histamine-producing enterochromaffin-like (ECL) cells in the corpus/fundus and of gastrin-producing (G) and somatostatin-producing (D) cells in the antrum/pylorus. Hypergastrinemia-promoted ECL cell hyperplasia is categorized as diffuse (an increase in endocrine cells of more than twice the standard deviation compared to age- and sex-matched controls), linear (sequences of five or more cells inside the basement membrane of the gastric gland), micronodular (clusters of five or more endocrine cells up to 150  $\mu\text{m}$  in diameter), and adenomatoid (collection of five or more micronodules adherent to one another but with interposition of basement membranes and thin strands of lamina propria). Dysplasia is characterized by 150 to 500  $\mu\text{m}$  lesions formed by moderately atypical endocrine cells and defined as

enlarged micronodules (nodules of  $\geq 150 \mu\text{m}$ ), adenomatous micronodules (collections of at least five micronodules), fused micronodules (disappearance of the intervening basal membrane between adjacent micronodules), and microinfiltrative lesions (microinfiltration of the lamina propria by endocrine cells filling the space in between glands).

G cell hyperplasia is defined as an increase in gastrin cell number (above 140) when counted per linear millimeter of mucosa in 5  $\mu\text{m}$  thick histological sections. It may be associated with reduced somatostatin (D) cell count, resulting in an elevated G/D cell ratio.

Long-standing hyperchlorhydria associated with duodenal G cell tumor may result in increased D cells of the antrum and a reduced G/D cell ratio.

### Intestine

Increased numbers of somatostatin D cells are observed in the small intestine of patients with celiac disease. Hyperplasia of unspecified argyrophil endocrine cells was reported in chronic inflammatory bowel disease.

## TUMORS

### General

Endocrine tumors of the gut are found at any level of the gastrointestinal tract (Table II). According to epidemiological data, in Western countries endocrine tumors more commonly display an age-standardized rate of approximately 1/100,000 and more frequently

Table II Tumor Type, Distribution, and Cell Features of the Endocrine Tumors of the Gastrointestinal Tract

Tumor type	Preferred site	Main cell type	Hormonal products
Well-differentiated endocrine/tumor/carcinoma	Stomach, body/fundus	ECL	Histamine 5-HT/5-HTP
	Duodenum, antrum, jejunum	G	Gastrin
	Duodenum	D	Somatostatin
	Appendix, ileum, jejunum, cecum	EC	Serotonin <sup>d</sup>
	Rectum, colon	L	Glicentin, PP, PYY
Gangliocytic paraganglioma	Duodenum	PP, D	PP, somatostatin
Poorly differentiated endocrine carcinoma	Stomach intestine	Proto-endocrine	None

Note. ECL, enterochromaffin-like cell; EC, enterochromaffin; 5-HT, 5-hydroxytryptamine (serotonin); 5-HTP, 5-hydroxytryptophan; PP, pancreatic polypeptide; PYY, peptide tyrosine tyrosine.

<sup>d</sup>Substance P, tachykinins, and other peptides.

occur in the colorectum, followed by the small intestine and stomach, occur more frequently in females, and increase in incidence with age.

The diagnosis of endocrine tumors is based on the identification of their characteristic morphology and of their antigenic asset, as defined by the expression of markers of endocrine differentiation (as described above) and hormones. According to the new World Health Organization classification, gut endocrine tumors are classified as follows: (1) well-differentiated endocrine tumors, including tumors with benign behavior and tumors with indefinite behavior at diagnosis; (2) well-differentiated endocrine carcinomas, low grade; and (3) poorly differentiated endocrine carcinomas, high grade.

### Well-Differentiated Neoplasms

Well-differentiated tumors are characterized by bland histological and cytological features, low cytological atypia, low mitotic index, and diffuse expression of general markers of endocrine differentiation. Cell-specific markers are in general observed in tumor cell subpopulations.

The definition of carcinoma is restricted to well-differentiated endocrine neoplasms with overt malignancy at diagnosis, i.e., synchronous metastasis, and/or deep wall invasion. In the absence of the above features, the behavior of well-differentiated tumors may be unpredictable, though independent of tumor cell typing and related to various clinicopathological variables. Predictors of poor outcome are as follows: tumor size (larger tumors usually are more aggressive); invasion of nearby tissue (appendix) or wall invasion beyond the submucosa; angioinvasion and invasion of perineural spaces; solid, nonorganoid structure; necrosis; overt cell atypia; more than two mitoses in 10 microscopic high-power fields (HPF); Ki67 proliferation index higher than 2% or of >100 in 10 HPF; reduction in or loss of chromogranin A immunoreactivity, argyrophilia, and/or nuclear p53 protein accumulation. The predictive value of most of these variables still needs confirmation from investigations on large series of endocrine tumors of the small and large intestine.

### Stomach

#### *ECL Cell Tumors*

By far the most frequent endocrine tumor of the stomach, ECL cell tumors display strong argyrophilia by silver impregnation techniques, immunoreactivity for chromogranin A and vesicular monoamine

transporter 2, and no or focal immunoreactivity for ghrelin, serotonin, gastrin, and somatostatin.

Three clinicopathologic subtypes are defined as follows: type I, associated with diffuse chronic atrophic gastritis of autoimmune or A type; type II, associated with hypertrophic gastropathy, usually in conjunction with MEN1 and Zollinger-Ellison syndrome (ZES); and type III or sporadic, not associated with a specific gastric pathology. Of the three subtypes, type I tumors are the most frequent (70–80% of published series), occur in elderly female patients, and, though multiple and multicentric, are small, limited to the mucosa and submucosa and demonstrate an excellent prognosis for survival. Type II tumors are rare (~6% of published series), multiple, and multicentric; they usually display a good prognosis, though metastases may be present and rare cases with an aggressive course have been described. In contrast, the solitary type III tumors are less frequent (~15% of published series), occur more frequently in male patients, are larger, and, in general, display a more aggressive behavior with frequent metastases and malignant course. Of note, type III tumors may associate with the so-called “atypical carcinoid syndrome” due to histamine and/or 5-hydroxytryptophan hypersecretion.

#### *Other Tumor Types*

Exceedingly rare gastrin G cell tumors have been reported in the antrum, similar to rare D cell tumors of both the oxyntic and antral mucosa. Such tumors usually are not associated with any hyperfunctional syndrome.

### Intestine

#### *Gastrin G Cell Tumors*

Gastrin G cell tumors occur in the duodenum or upper jejunum of male patients. When “functioning,” they can be defined as gastrinomas, are the cause of the ulcerogenic ZES, and frequently are associated with MEN1. Though small in size, G cell tumors, especially when functioning, frequently metastasize, especially to local lymph nodes.

#### *D Cell Tumors*

Somatostatin D cell tumors are rare tumors of the duodenum (preferentially at the ampulla of Vater) and upper jejunum in patients in the fifth decade of life. Often associating with type 1 neurofibromatosis and only rarely with diabetes and/or gallstones, they may be defined as functioning, though the complete somatostatinoma syndrome (diabetes mellitus, diarrhea, steatorrhea, hypo- or achlorhydria, anemia, and

gallstones) has been described for pancreatic D cell tumors only. Intestinal D cell tumors are often malignant and may metastasize to both local lymph nodes and the liver.

#### *Ganglioneuromatous Paraganglioma*

Rare tumors composed of endocrine cells, mature ganglion cells, and Schwann-like spindle cells develop in the submucosa of the periampullary duodenal region in middle-aged patients. In general, ganglioneuromatous paragangliomas display benign behavior, though local lymph node metastases of the endocrine component have been reported in occasional cases.

#### *Serotonin EC Cell Tumors*

Most serotonin-producing enterochromaffin (EC) cell tumors develop in the intestine, with decreasing frequency in the ileum, cecum, appendix, jejunum, duodenum, distal colon, and rectum. Intestinal EC cell tumors may display relatively aggressive behavior since often they are multiple (small intestine), are of large size (colon), and deeply infiltrate the intestinal wall, with frequent lymph node metastases. A typical carcinoid syndrome due to the unregulated release of serotonin and other active substances by tumor EC cells may occur in association with such tumors, depending on the establishment of liver metastases. In contrast, appendiceal EC cell tumors, despite the deep wall invasion often observed, in general run a benign course.

#### *Glicentin/PYY L Cell Tumors*

L cell tumors most frequently develop in the colon and rectum and rarely at other sites of the intestine and appendix. The majority of L cell tumors are small, do not associate with any hyperfunctional syndrome, and run a benign course.

### Poorly Differentiated Endocrine Carcinomas

Poorly differentiated endocrine carcinomas are aggressive carcinomas reported to occur at any site

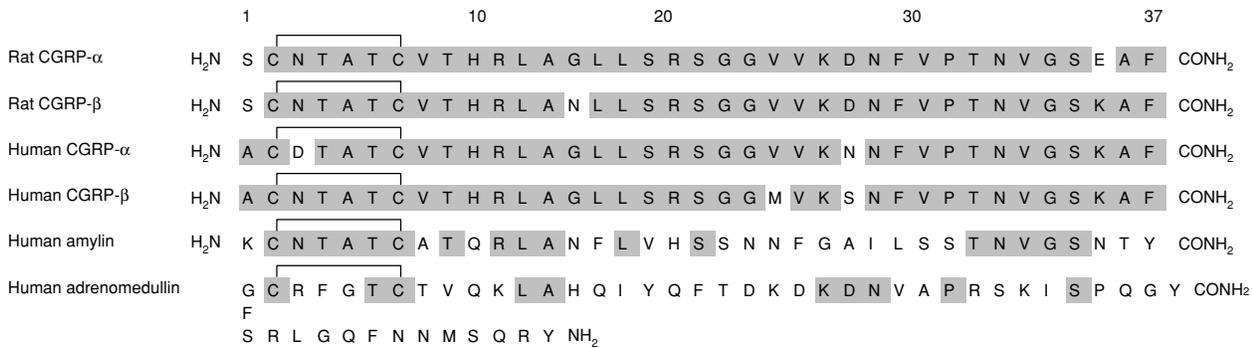
in the gut, except the appendix, of patients in the seventh decade of life. Poorly differentiated endocrine carcinomas display a solid structure, abundant necrosis, small cell cytology, severe cellular atypia, a high mitotic index, no or faint chromogranin A expression, and diffuse expression of NSE and synaptophysin. Additional features include a high Ki67 index and p53 accumulation as observed with immunohistochemistry, whereas cell-specific endocrine markers are absent. Similar to undifferentiated carcinomas, poorly differentiated endocrine carcinomas run a very aggressive course.

### See Also the Following Articles

GI Hormones in Cancer • GI Tract, General Anatomy (Cells)

### Further Reading

- Capella, C., Riva, C., Rindi, G., Sessa, F., Usellini, L., Chiaravalli, A., Carnevali, L., and Solcia, E. (1991). Histopathology, hormone products and clinico-pathologic profile of endocrine tumors of the upper small intestine: A study of 44 cases. *Endocr. Pathol.* **2**, 92–110.
- Fiocca, R., Capella, C., Buffa, R., Fontana, R., Solcia, E., Hage, E., Chance, R. E., and Moody, A. J. (1980). Glucagon-, glicentin-, and pancreatic polypeptide-like immunoreactivities in rectal carcinoids and related colorectal cells. *Am. J. Pathol.* **100**, 81–92.
- Moertel, C. G., Sauer, G., Dockerty, M. B., and Baggenstoss, A. H. (1961). Life history of the carcinoid tumor of the small intestine. *Cancer* **14**, 901–912.
- Rindi, G., Azzoni, C., La Rosa, S., Klersy, C., Paolotti, D., Rappel, S., Stolte, M., Capella, C., Bordi, C., and Solcia, E. (1999). ECL cell tumor and poorly differentiated endocrine carcinoma of the stomach: Prognostic evaluation by pathological analysis. *Gastroenterology* **116**, 532–542.
- Rindi, G., Luinetti, O., Cornaggia, M., Capella, C., and Solcia, E. (1993). Three subtypes of gastric argyrophil carcinoid and the gastric neuroendocrine carcinoma: A clinico-pathologic study. *Gastroenterology* **104**, 994–1006.
- Solcia, E., Capella, C., Fiocca, R., Sessa, F., LaRosa, S., and Rindi, G. (1998). Disorders of the endocrine system. In "Pathology of the Gastrointestinal Tract" (S. C. Ming and H. Goldman, eds.), 2nd ed., pp. 295–322. Williams & Wilkins, Philadelphia, PA.
- Solcia, E., Klöppel, G., and Sobin, L. H. (2000). "Histological Typing of Endocrine Tumors." World Health Organization International Histological Classification of Endocrine Tumors, Springer-Verlag, New York.



**Figure 1** Comparison of the amino acid sequences of human and rat CGRP- $\alpha$  and CGRP- $\beta$ , human amylin, and human adrenomedullin.

## CGRP RECEPTORS

The biological effects of CGRP are brought about by interaction with specific membrane receptors, two subtypes of which, the CGRP<sub>1</sub> and CGRP<sub>2</sub> receptors, have been proposed to exist on the basis of bioassay observations. Characteristic of the CGRP<sub>1</sub> receptors is their sensitivity to the antagonistic effect of the human CGRP<sub>8-37</sub> fragment. The molecular identification of functional CGRP receptors turned out to be very difficult because they are assembled from three different proteins (Fig. 2): the calcitonin receptor-like receptor (CRLR), the receptor-associated membrane protein 1 (RAMP-1), and the receptor component protein (RCP). Although CRLR is the CGRP-recognizing protein, the CGRP receptor becomes functional only if CRLR is associated with RAMP-1. This chaperone protein is important for the intracellular translocation of CRLR and its insertion into the plasma membrane and for conferring a CGRP<sub>1</sub> receptor-like binding profile on CRLR. RCP seems to be required for efficient coupling of the receptor to the G protein-adenylate cyclase signaling machinery (Fig. 2).

CRLR can associate not only with RAMP-1 and RCP to produce a CGRP<sub>1</sub> receptor, but also with RAMP-2 or RAMP-3, two other chaperone proteins sharing approximately 30% sequence homology with RAMP-1. These RAMPs dictate the pharmacological profile of the receptor complex inasmuch as CRLR associated with RAMP-2 behaves as an adrenomedullin receptor and human calcitonin receptor isotype 2 associated with RAMP-1 or RAMP-3 behaves as an amylin receptor.

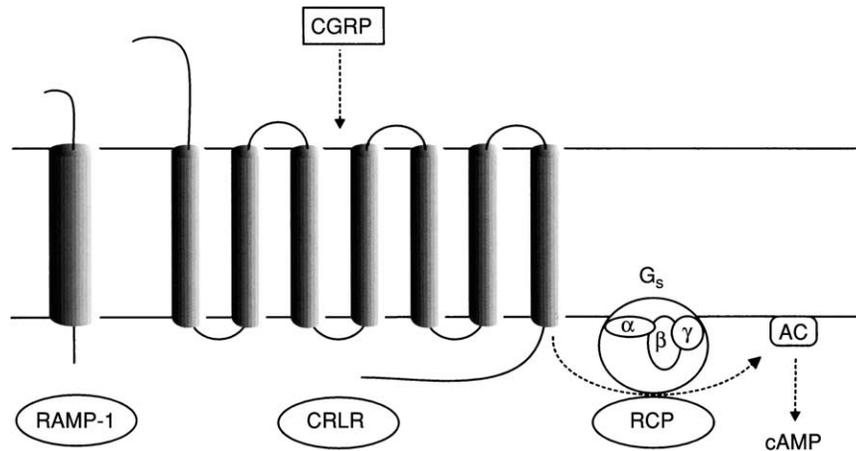
## EXPRESSION AND RELEASE OF CGRP IN THE GI TRACT

The principal sources of CGRP in the digestive system are extrinsic primary afferent nerve fibers and

intrinsic enteric neurons. Although the CGRP that is expressed in primary afferents of the rat is CGRP- $\alpha$ , the only form of CGRP in enteric neurons is CGRP- $\beta$ . The relative contribution of extrinsic afferent and intrinsic neurons to the overall CGRP content of the GI tract varies among different gut regions and different mammalian species. Most of the CGRP found in the esophagus and stomach of small rodents is derived from primary afferents, whereas the small intestine and large intestine contain a sizable number of CGRP-positive enteric neurons that issue specific projections to other enteric ganglia, muscle, and mucosa.

The majority of the CGRP-expressing extrinsic afferent neurons in the rodent gut originate from cell bodies in the dorsal root ganglia and reach the gut via sympathetic and sacral parasympathetic nerves. Within the wall of the GI tract, they innervate primarily arteries and arterioles but also project to the mucosa, enteric nerve plexuses, and muscle layers. CGRP-positive vagal afferents originating from the nodose ganglion supply the esophagus and proximal part of the stomach but make a relatively small contribution to the content of CGRP in the gastric corpus, antrum, and intestine.

As is expected for substances with a vesicular localization, CGRP is released from GI neurons in a calcium-dependent manner when these cells are stimulated. This is, for instance, the case when the mucosa of the stomach and duodenum is exposed to excess acid, which releases CGRP from primary afferent nerve fibers. Peptide release from spinal afferents in the gut can also be elicited by the vanilloid capsaicin because these nerve cells, but not enteric neurons, express vanilloid receptors of type 1. With the use of capsaicin, it has furthermore been found that CGRP in the general circulation represents primarily an overflow of peptide released from peri- and paravascular afferent nerve fibers.



**Figure 2** Diagram of the CGRP receptor complex consisting of the calcitonin receptor-like receptor (CRLR), the receptor-associated membrane protein 1 (RAMP-1), and the receptor component protein (RCP). Association of CRLR with RAMP-1 forms a high-affinity CGRP<sub>1</sub> receptor, which, through RCP, is coupled to the G protein G<sub>s</sub> and thereby can activate adenylylate cyclase (AC) to produce cAMP.

Apart from neurons, CGRP is also found in endocrine cells of the human GI mucosa and rat pancreas and in blood-derived or resident immune cells within the lamina propria of the rat gastric mucosa. However, the quantitative and functional significance of these sources is still not very well understood.

## EFFECTS, PHYSIOLOGICAL ROLES, AND PATHOLOGICAL IMPLICATIONS OF CGRP IN THE GI TRACT

### Motor Activity

The most prominent motor action of CGRP in the active gut is muscle relaxation via CGRP<sub>1</sub> receptors, which leads to retardation of gastric emptying and attenuation of motor activity throughout the digestive tract. However, CGRP is also able to excite cholinergic motor pathways, which is in keeping with the peptide's ability to depolarize intrinsic sensory neurons of the myenteric plexus, to enhance the release of acetylcholine from enteric neurons, and to cause contraction of the resting muscle.

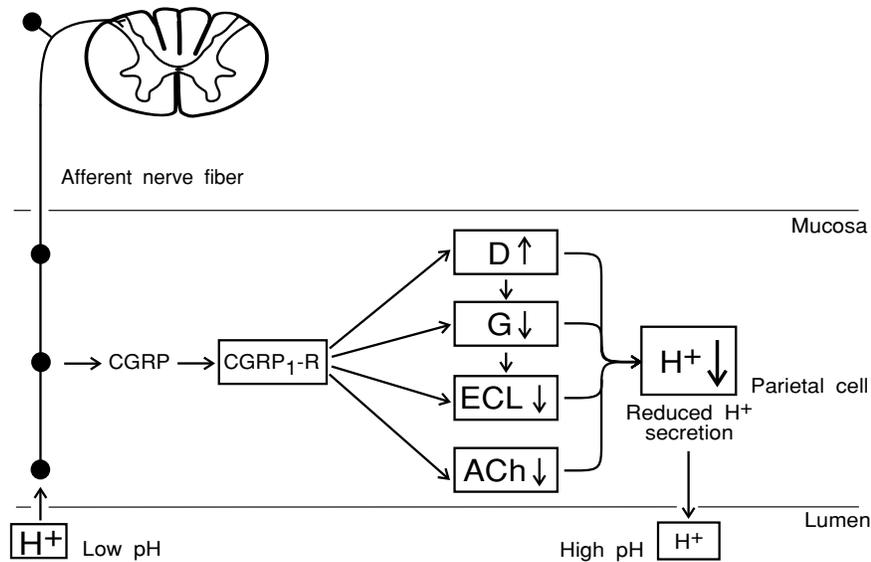
There is still scarce information as to whether CGRP released from intrinsic or extrinsic neurons of the gut plays a physiological role in the neural control of GI motility. The claim that CGRP released from sensory neurons contributes to distension-induced peristalsis is not universally accepted. There is, however, evidence that CGRP released from extrinsic afferent nerve fibers contributes to the pathological shutdown of GI motility in postoperative and peritonitis-induced ileus. This is consistent with the

observation that CGRP acting via CGRP<sub>1</sub> receptors contributes to the inhibition of intestinal peristalsis that ensues after sensory neuron stimulation.

### Secretory Processes

Electrolyte and fluid secretion in the small intestine of the dog and in the colon of the guinea pig and rat is stimulated by CGRP. Whereas the secretory effect of CGRP in the rat colon arises from a direct action on enterocytes, as is the case with human epithelial cell lines, CGRP's secretory action in the guinea pig colon is mediated by enteric neurons. Although a physiological role for CGRP in the control of intestinal secretory activity has not yet been determined, it seems as if the peptide contributes to the pathological fluid secretion that in the rat ileum is evoked by *Clostridium difficile* toxin A. The secretagogue-evoked secretion of enzyme, bicarbonate, and fluid from the pancreas of the dog and rat *in vivo* is blunted by CGRP through an action involving somatostatin, whereas amylase secretion from isolated acini of the rat and guinea pig pancreas is enhanced by the peptide.

There is good evidence that, in the stomach, CGRP contributes to the homeostatic regulation of endocrine and exocrine secretory processes. Thus, CGRP potently depresses basal and secretagogue-evoked output of acid and pepsin in the stomach of human, dog, rabbit, and rat, an action that is brought about by CGRP<sub>1</sub> receptors, depends on somatostatin as an essential mediator, and goes along with attenuation of the release of acetylcholine, gastrin, and



**Figure 3** Diagram illustrating the role of CGRP-releasing afferent nerve fibers in the feedback control of gastric acid secretion in the rat stomach. When the acidity in the lumen rises, afferent nerve fibers release CGRP, which, via activation of CGRP<sub>1</sub> receptors (CGRP<sub>1</sub>-R), stimulates the release of somatostatin and inhibits the release of gastrin, histamine, and acetylcholine. D, endocrine D cells releasing somatostatin; G, endocrine G cells releasing gastrin; ECL, enterochromaffin-like cells releasing histamine; ACh, neurons releasing acetylcholine.

histamine (Fig. 3). These effects are physiologically relevant, given that the CGRP<sub>1</sub> receptor antagonist CGRP<sub>8-37</sub> augments basal and stimulated acid secretion and acid accumulation in the gastric lumen releases CGRP from sensory nerve fibers. Through its effects on the release of somatostatin, gastrin, histamine, and acetylcholine, CGRP halts further secretion of acid and thus mediates feedback inhibition of gastric acid output (Fig. 3).

### Vascular Functions

The arteries and submucosal arterioles of the GI tract receive the densest innervation by extrinsic afferents containing CGRP. This finding, the expression of CGRP<sub>1</sub> receptors on the endothelium and smooth muscle of arteries and arterioles, and the peptide's vasodilator activity point to a vasoregulatory function of CGRP. Indeed, nonadrenergic noncholinergic dilation of the rat superior mesenteric artery is mediated by capsaicin-sensitive afferent nerve fibers releasing CGRP. In contrast, the physiological significance of CGRP in the microcirculation of the small and large intestine is little known. Although CGRP dilates submucosal arterioles in the guinea pig ileum, it fails to alter mucosal blood flow in the rat small and large intestine.

The situation is different in the rat stomach, where CGRP is highly potent in causing a CGRP<sub>1</sub>

receptor-mediated dilation of submucosal arterioles, but not venules. Vasodilation induced by low doses of CGRP is mediated through a mechanism that involves nitric oxide, whereas high doses of the peptide increase blood flow independent of nitric oxide. Although CGRP does not seem to regulate gastric blood flow in physiological circumstances, there is evidence that CGRP comes into play under pathological conditions. This role is best exemplified by the hyperemic response that ensues when the gastric mucosal barrier is disrupted by ethanol or bile salts so that acid can enter the tissue and damage the gastric mucosa. The CGRP-mediated rise in blood flow in response to acid influx serves a protective role in the gastric mucosa as it helps to neutralize and wash away intruding acid and delivers bicarbonate and other factors to defend and repair the mucosa. Another example relates to the *C. difficile* toxin A-evoked inflammation in the rat ileum, which involves CGRP as a pro-inflammatory mediator.

### Mucosal Homeostasis

There are several lines of evidence to indicate that GI mucosal integrity and repair are under the control of extrinsic afferent neurons releasing CGRP. First, CGRP protects the mucosa in a number of experimental models of gastric injury and colonic inflammation. The action of CGRP to reduce

ethanol-induced damage in the gastric mucosa is mediated by CGRP<sub>1</sub> receptors and involves nitric oxide. Second, CGRP mediates the gastroprotective effect of a number of factors and drugs that stimulate capsaicin-sensitive afferent nerve fibers. Thus, blockade of CGRP<sub>1</sub> receptors with CGRP<sub>8-37</sub> prevents the ability of intragastric capsaicin to attenuate experimentally imposed injury as does immunoneutralization of CGRP with polyclonal and monoclonal antibodies to the peptide. Third, CGRP<sub>8-37</sub> and active immunization of rats against CGRP exacerbate experimental injury in the stomach.

In summary, CGRP released from sensory nerve fibers strengthens gastric mucosal defense and facilitates repair of the wounded mucosa via vasodilation, hyperemia-dependent processes, such as appropriate delivery of bicarbonate, and hyperemia-independent mechanisms, such as secretion of mucus. Complementary evidence for such a homeostatic role for CGRP-releasing nerve fibers in the GI mucosa comes from the observation that sensory neuropathies weaken the resistance of the tissue to injury. This applies not only to the stomach, but also to the esophagus, small intestine, and colon, where experimentally induced inflammation and damage are aggravated.

### Gastrointestinal Sensitivity and Nociception

Since CGRP is a transmitter of nociceptive afferent neurons innervating the gut, it does not come as a surprise that CGRP can mediate GI pain and inflammatory hyperalgesia. Intraperitoneal administration of exogenous CGRP or acetic acid-induced release of endogenous CGRP in the rat peritoneum triggers abdominal muscle contractions, a reaction that is indicative of pain. Of particular importance is the finding that CGRP<sub>8-37</sub> prevents inflammation-induced hypersensitivity to colonic distension. Since, in this respect, intrathecal CGRP<sub>8-37</sub> is more potent than intravenous CGRP<sub>8-37</sub>, the site of CGRP-mediated hyperalgesia is primarily in the spinal cord.

### SUMMARY

CGRP is a transmitter candidate of intrinsic enteric neurons and extrinsic afferent nerve fibers in the GI tract. As such, this neuropeptide seems to be involved in the neural regulation of GI functions. Its major

actions include depression of GI motility, rise of gastric blood flow, inhibition of gastric acid secretion, enforcement of GI mucosal resistance, and mediation of inflammatory hyperalgesia. Furthermore, there is evidence that a disturbance of the GI CGRP system may contribute to a number of GI disorders and that a correction of these perturbations may be of therapeutic potential.

### Acknowledgment

Evelin Painsipp is greatly appreciated for drawing the figures.

### See Also the Following Articles

Calcitonin, Overview • CCK (Cholecystokinin) • Gastrin • Gastrin-Releasing Peptide • GIP (Gastric Inhibitory Polypeptide)

### Further Reading

- Foord, S. M., and Marshall, F. H. (1999). RAMPs: Accessory proteins for seven transmembrane domain receptors. *Trends Pharmacol. Sci.* **20**, 184–187.
- Gschossman, J. M., Coutinho, S. V., Miller, J. C., Huebel, K., Naliboff, B., Wong, H. C., Walsh, J. H., and Mayer, E. A. (2001). Involvement of spinal calcitonin gene-related peptide in the development of acute visceral hyperalgesia in the rat. *Neurogastroenterol. Motil.* **13**, 229–236.
- Holzer, P. (1994). Calcitonin gene-related peptide. In “Gut Peptides: Biochemistry and Physiology” (J. H. Walsh and G. J. Dockray, eds.), pp. 493–523. Raven Press, New York.
- Holzer, P. (1998). Neural emergency system in the stomach. *Gastroenterology* **114**, 823–839.
- Holzer, P. (2000). Calcitonin gene-related peptide in gastrointestinal homeostasis. In “The CGRP Family: Calcitonin Gene-Related Peptide (CGRP), Amylin, and Adrenomedullin” (D. Poyner, I. Marshall, and S. D. Brain, eds.), pp. 125–139. Landes Bioscience, Georgetown, TX.
- Ichikawa, T., Ishihara, K., Kusakabe, T., Hiruma, H., Kawakami, T., and Hotta, K. (2000). CGRP modulates mucin synthesis in surface mucus cells of rat gastric oxyntic mucosa. *Am. J. Physiol.* **279**, G82–G89.
- Juaneda, C., Dumont, Y., and Quirion, R. (2000). The molecular pharmacology of CGRP and related peptide receptor subtypes. *Trends Neurosci.* **21**, 432–438.
- Rosenfeld, M. G., Emeson, R. B., Yeakley, J. M., Merillat, N., Hedjran, F., Lenz, J., and Delsert, C. (1992). Calcitonin gene-related peptide: A neuropeptide generated as a consequence of tissue-specific, developmentally regulated alternative RNA processing events. *Ann. N. Y. Acad. Sci.* **657**, 1–17.
- Taché, Y. (1992). Inhibition of gastric acid secretion and ulcers by calcitonin gene-related peptide. *Ann. N. Y. Acad. Sci.* **657**, 240–247.
- Wimalawansa, S. J. (1996). Calcitonin gene-related peptide and its receptors: Molecular genetics, physiology, pathophysiology, and therapeutic potentials. *Endocr. Rev.* **17**, 533–585.

Table I Causes of Gigantism

Pituitary origin
$G_{s\alpha}$ gene mutation, with or without McCune-Albright syndrome
Pituitary adenoma due to other molecular causes
Excess GHRH
Eutopic—Hypothalamic dysfunction and tumors
Ectopic
Somatostatin deficiency

### Pituitary GH Excess Caused by Activation of the $G_{s\alpha}$ Subunit of the G Protein Receptor with or without McCune-Albright Syndrome

The G protein-coupled receptors are the largest family of membrane receptors. Peptides (vasopressin, glycoproteins, adrenocorticotrophic hormone [ACTH], GHRH, melanocyte-stimulating hormone [MSH]), neurotransmitters, and prostaglandins can bind to this receptor family. When a stimulatory ligand binds to the receptor, the  $\alpha$  subunit of the G protein ( $G_{s\alpha}$ ) is activated by replacement of GDP by GTP. This subsequently leads to activation of adenylate cyclase and cyclic AMP formation. It finally results in phosphorylation of substrates that control metabolism, gene transcription, secretion, and cell proliferation. Normally, intrinsic GTPase stops activation of  $G_{s\alpha}$ . A mutation in the gene encoding the  $\alpha$  subunit of the stimulatory G protein receptor results in constitutive activation of the protein, by inhibiting the intrinsic GTPase activity. This has a tumorigenic effect. Thirty to 50% of pituitary adenomas in gigantism are caused by this mechanism.

McCune-Albright syndrome is a rare disorder, classically defined by fibrous dysplasia, café au lait spots, precocious puberty, and other hyperfunctional endocrinopathies. In all patients with the McCune-Albright syndrome, a mutation in the gene for the  $\alpha$  subunit is found. It is assumed that the timing of the mutation in development determines which tissues are affected and in this way contributes to the heterogeneity of the clinical presentation. Approximately 20% of the cases of gigantism are associated with McCune-Albright syndrome. GH-producing tumors in these patients are the result of a mutation in the gene encoding the  $\alpha$  subunit of the G protein receptor of somatotroph cells, leading to constitutive activation of the  $G_{s\alpha}$  protein, resulting in hyperplasia and adenoma with GH overproduction.

Table II Histological Classification of Adenoma

Somatotroph adenomas
Sparsely granulated
Densely granulated
Plurihormonal tumors of the acidophil cell line
Mixed somatotroph-lactotroph adenomas
Mammomatotroph adenomas
Acidophil stem cell adenomas
Miscellaneous tumors
Plurihormonal adenomas

### Pituitary GH Excess Caused by an Adenoma by Other Molecular Causes

Pituitary adenomas are benign tumors consisting of adeno-hypophysial cells. They are often slowly growing, expansive tumors confined to the sella turcica, which are usually detected because of the clinical effects of the oversecreted hormones. Adenomas can be classified according to their size, microadenomas (<10 mm) and macroadenomas (>10 mm), but a morphological classification (Table II) is more useful.

In gigantism, mammomatotroph adenomas are the most frequent histological type. The mammomatotroph cells, producing prolactin and GH, observed in these adenomas resemble the mammomatotrophs identified in the fetal pituitary at 9.5 weeks gestation. It is thought that pituitary adenomas are derived from an intrinsic pituitary cell defect, leading to clonal expansion of a single transformed cell. The exact mechanisms by which these cells transform into tumorous cells is unknown. Inactivating mutations of tumor suppressor genes are associated with pituitary tumorigenesis, as shown in multiple endocrine neoplasia (MEN-I) syndrome, which is characterized by hyperfunction or tumor formation in the parathyroids, anterior pituitary, and pancreatic cells and rarely in the thyroid and adrenal cells. Pei and Melmed cloned a novel pituitary tumor-transforming gene. High-level expression of this gene in GH-secreting adenomas suggests that it plays a role in pituitary tumorigenesis.

### Excess of GHRH

#### Eutopic GHRH Excess

Secondary GH overproduction due to hypothalamic GHRH excess is an important cause of gigantism. GHRH excess usually results in mammomatotroph hyperplasia and rarely in adenoma formation.

In a case report of congenital gigantism due to GH and prolactin-producing cell hyperplasia, high serum levels of GHRH were found. There was no evidence of a hypothalamic GHRH-producing tumor, suggesting that a congenital hypothalamic regulatory defect resulting in early GHRH exposure leads to mammosomatotroph hyperplasia.

### **Ectopic GHRH Excess**

Ectopic GHRH-producing tumors are a well-known cause of acromegaly, but very rarely cause gigantism. Thus far, only two cases have been described. In one patient, a 15-year-old girl, the GHRH-producing primary tumor was located in the jejunum and resulted, together with GHRH-producing liver metastases, in severe GH hypersecretion causing gigantism. The other case is an 18-year-old boy with GHRH-like activity in a metastatic carcinoid tumor in the foregut.

Another cause of intracranial GHRH excess occurs in the setting of neural tumors, such as gangliocytoma or neurocytoma, within or in close proximity to the sella.

### **GH Excess by Somatostatin Deficiency**

A few cases have been documented in which GH excess is found in children with neurofibromatosis and optic gliomas or astrocytomas. It has been suggested that infiltration of the glioma in the medial temporal lobe causes a reduction in somatostatin, leading to increased GH levels and interference with the normal pulsatility of GH release.

## **DIAGNOSIS**

In all patients with an increased growth velocity and progressive upward deviation from the population reference curves, GH excess should be considered. In young children, body proportions are normal and usually the increased growth is the only sign of gigantism. In adolescents, the typical features of acromegaly, such as thickening of the skin, enlargement of the lower jaw, hands, and feet, coarsening of facial features, and excessive body sweating, may be present. Later, a eunochoid habitus can develop because of incomplete or absent puberty. Depending on the type and expansion of the adenoma, secretion of other pituitary hormones can be affected. As adenomas often secrete GH as well as prolactin, hyperprolactinemia may occur, manifesting as galactorrhea. Gonadotropin deficiency leads to hypogonadotropic hypogonadism and incomplete or absent puberty.

Less often, ACTH or TSH deficiency occurs. Because of compression of the optic chiasm, impaired vision and visual field abnormalities can be presenting symptoms.

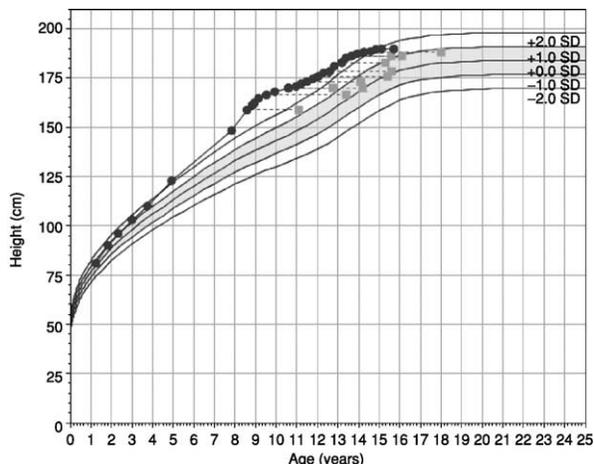
Serum GH levels are elevated in gigantism. Whereas in overnight GH profiles of healthy individuals, baseline levels do not exceed 0.5 ng/ml, in giants these levels are seldom less than 5 ng/ml. Serum levels of IGF-I and IGF-binding protein 3 (BP3) are elevated compared with age and gender references. However, interpretation of IGF-I levels is difficult during puberty because of the physiological IGF-I increase at that time.

Normally, oral glucose ingestion (1.75 g per kilogram of body weight) suppresses GH levels to less than 1 ng/ml, whereas in patients with GH excess, GH levels are not suppressed and may even show a paradoxical increase. However, in children and adolescents, the oral glucose tolerance test is less useful for diagnosing GH excess, as absent suppression is also common in healthy tall children and adolescents. Similarly, the presence of paradoxical GH response to thyrotropin-releasing hormone and luteinizing hormone-releasing hormone is of little use in this age group.

In the diagnostic workup, the other pituitary hormones should obviously also be assessed. Prolactin is elevated in approximately half the cases and may be associated with galactorrhea and contribute to hypogonadism. Gonadotropin secretion is often decreased because of compression of gonadotropin-secreting cells by the hyperplastic or adenomatous somatotroph cells. The thyroid-stimulating hormone and ACTH axes should be investigated, although these are mostly not affected. GHRH measurement can be helpful in differentiating GHRH excess from primary GH hypersecretion.

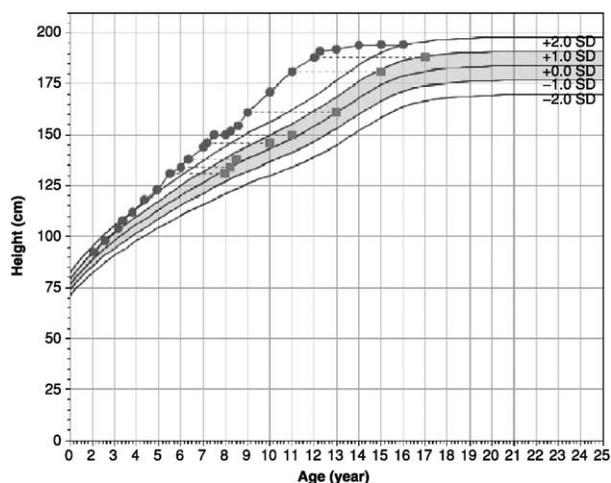
As insulin resistance can occur, the insulin response to glucose must be measured. Bone age should be assessed, but is usually not advanced, or only little advanced, during childhood. A skull X ray may reveal enlargement of the sella turcica. When GH excess is confirmed, neuroradiological studies including magnetic resonance imaging for tumor localization and extension are essential.

For illustrative purposes, two cases with gigantism are shown in [Figs. 1 and 2](#). Case 1 presented with tall stature at 8.5 years of age. He had serum GH levels fluctuating between 100 and 150 ng/ml, very elevated serum IGF-I levels, and a GH-producing adenoma with suprasellar extension. After transsphenoidal resection of the adenoma, a small tumor remnant remained. Bromocriptine treatment was sufficient for



**Figure 1** Growth curve of a patient with a growth hormone-producing adenoma.

a normalization of plasma GH levels and final height was limited by supraphysiologic testosterone treatment. His growth pattern is shown in Fig 1. Case 2 presented with pseudo-precocious puberty, a pigmented macula with irregular margins, resembling the coast of Maine, typical for McCune-Albright syndrome, and polyostotic fibrous dysplasia. Initially, he was treated with cyproterone acetate and ketoconazole for his pseudo-precocious puberty, but soon it became clear that he also had GH excess, with mean GH levels between 10 and 15 ng/ml. From 12.4 years of age, he was subsequently treated with bromocriptine, octreotide, and cabergoline. Bisphosphonates were prescribed to restrict the damage of the fibrodysplastic lesions. His growth curve is shown in Fig 2.



**Figure 2** Growth curve of a patient with McCune-Albright syndrome.

## DIFFERENTIAL DIAGNOSIS

Tall stature in childhood and adolescence can be classified in various ways. The authors prefer the classification summarized in Table III, in which three main groups are distinguished: constitutional tall stature, primary growth disorders (with a defect presumably in the growth plate), and secondary growth disorders (abnormalities in the milieu interieur).

By far the most frequent cause of tall stature is constitutional (familial) tall stature, which is a reflection of the fact that 50–90% of height variation is accounted for by genetic factors. In most cases, height standard deviation scores (SDS) lie within a range of 1.3 SDS above or below the target height SDS (the sex- and secular trend-corrected midparent height). Measurements of parental height and calculating target height are therefore of critical importance. Birth length is usually more than +0.7 SDS (75th percentile). In the first 3–4 years, height velocity is accelerated, followed by a growth pattern parallel to, but above, the +1.9 SDS (97th percentile). No abnormalities are present on physical examination.

Primary growth disorders include a number of genetic syndromes. There are a number of sex chromosome-related disorders, of which Klinefelter syndrome (47, XXY) is most frequent. Usually, boys with this syndrome are not exceptionally tall in childhood, but adolescents and adults can be tall and often have eunuchoid body proportions. It is assumed that the additional chromosomal material and inadequate pubertal development account for the tall stature in this syndrome.

Sotos syndrome, Weaver syndrome, and Marshall-Smith syndrome are characterized by increased birth length and head circumference, increased growth velocity during the first year of life, tall stature during childhood and adolescence, advanced bone age, and moderately increased final height. All these syndromes have psychomotor retardation and their own set of typical craniofacial features. The phenotype of fragile X syndrome is very similar to that of Sotos syndrome. In Beckwith-Wiedemann syndrome, birth weight and length are increased and infants have a large protruding tongue. Final height is often within the normal range due to advanced bone maturation.

Marfan syndrome is an autosomal dominant inherited disorder, in which tall stature is the most prominent feature. Associated symptoms are arachnodactyly, joint hyperlaxity, cardiovascular anomalies, and lens subluxation. Homocystinuria is an autosomal recessive disorder caused by an enzyme deficiency. Its

**Table III Causes of Tall Stature**

Constitutional (familial)
Primary growth disorders
Sex chromosome-related disorders: Klinefelter syndrome (XXY), XXX syndrome, XYY syndrome
Fragile X syndrome
Syndromes characterized by advanced growth and maturation: Sotos syndrome, Weaver Syndrome, Marshall-Smith syndrome
Marfan syndrome
Marfan-like phenotype: MEN type IIB, homocystinuria
Secondary growth disorders
GH excess
Hyperthyroidism
Precocious puberty
True (central) precocious puberty
Pseudo-precocious puberty: Leydig cell hyperplasia (testotoxicosis), McCune-Albright syndrome, congenital adrenal hyperplasia, gonadal tumors
Impaired estrogen effect (estrogen receptor defect or aromatase deficiency)

clinical features are similar to those of Marfan syndrome, with the exception of mental retardation, which is always present in homocystinuria. MEN-IIB syndrome mimics the signs of Marfan syndrome, but has additional signs, such as nodules at the tongue and lips.

Of the secondary growth disorders, GH excess is most important. Hyperthyroidism can cause an increased growth velocity and bone maturation, but in childhood tall stature is rare and final height is normal. In infants, hyperinsulinism, due to nesidioblastosis or as a result of diabetes mellitus of the mother, can result in overgrowth in terms of weight and length, but this usually normalizes shortly afterward. Pseudo-precocious puberty temporarily leads to increased growth and thus tall stature in childhood, but due to premature closure of the epiphyses final height is usually in the lower normal range. It has been observed that an absence of estrogen effect (in one case occurring due to a mutation in the estrogen receptor and in two cases occurring due to aromatase deficiency) results in normal growth in childhood, but brings about an absence of epiphyseal closure and thus failure to stop growing, leading to very tall stature in adulthood.

## COMPLICATIONS

In addition to the psychological and social problems that patients with extremely tall stature encounter, complications of gigantism include increased risk of

cardiovascular, neuromuscular, and pulmonary problems as in acromegaly. The risk of colorectal tumors and other malignancies seems to be correlated with the degree of GH control. The diabetogenic effects of GH cause insulin resistance in acromegaly. In young children with GH excess, diabetes is rare because of their increased islet reserve. Periods of stress, however, can exhaust these reserves and result in temporary diabetes and even diabetic ketoacidosis.

## TREATMENT

If an adenoma is present, transsphenoidal surgery is the treatment of choice. Irradiation can cause permanent damage to the pituitary and the central nervous system and is therefore not desirable in children. Primary or adjuvant pharmacotherapy is used pre- and postoperatively and in cases of pituitary hyperplasia.

Dopamine agonists (bromocriptine and cabergoline) release GH in normal individuals but paradoxically suppress GH hypersecretion in some patients with gigantism. The reaction to this therapy depends on the expression of dopamine receptor type D2 on the tumor cells. Often dopamine agonists alone fail to achieve complete biochemical control. In cases with pituitary hyperplasia or in cases with an insufficient response to transsphenoidal surgery, somatostatin analogues (octreotide) are often used. Somatostatin receptor subtypes 2 and 5 mediate the inhibitory effect of somatostatin. Complete biochemical control was reported in a case with an acidophilic stem cell adenoma using a combination of a long-acting dopamine agonist and a long-acting somatostatin analogue.

A GH receptor antagonist has shown to be effective in acromegaly. It is a mutated human GH molecule that binds to the GH receptor, but does not lead to signal transduction. Thus far, it has not been used in gigantism, but theoretically it should be as effective as in acromegaly. A competitive GHRH antagonist, blocking the GHRH receptor, has proven to be effective in the ectopic GHRH syndrome.

## Acknowledgments

The authors are grateful to Dr. M. Jansen and Dr. A. M. Pereira Arias for providing information on both patients.

## See Also the Following Articles

Acromegaly, Diagnosis of • Genetic Testing for Pituitary Disease • Growth Hormone (GH) • McCune-Albright Syndrome • Pituitary Tumors, Molecular Pathogenesis • Postnatal Non-Endocrine Overgrowth • Postnatal Normal

Growth and Its Endocrine Regulation • Somatostatin, Evolution of

### Further Reading

- Daughaday, W. H. (1992). Pituitary gigantism. *Endocrinol. Metab. Clin. N. Am.* **21**, 633–647.
- Drake, A. J., Lewis, S. P., Bouffet, E., and Crowne, E. C. (2000). Growth hormone hypersecretion in a girl with neurofibromatosis type 1 and an optic nerve glioma: Resolution following chemotherapy. *Horm. Res.* **53**, 305–308.
- Eugster, E. A., and Pescovitz, O. H. (1999). Gigantism. *J. Clin. Endocrinol. Metab.* **84**, 4379–4384.
- Holl, R. W., Bucher, P., Sorgo, W., Heinze, E., Homoki, U., and Debatin, K. M. (1999). Suppression of growth hormone by oral glucose in the evaluation of tall stature. *Horm. Res.* **51**, 20–24.
- Kahn, C. R., Smith, R. J., and Chin, W. W. (1998). Mechanism of action of hormones that act at the cell surface. In “Williams’ Textbook of Endocrinology” (J. D. Wilson, D. W. Foster, H. M. Kronenberg, and P. R. Larsen, eds.), 9th ed., pp. 95–143. Saunders, Philadelphia, PA.
- Maheshwari, H. G., Prezant, T. R., Herman-Bonert, V., Shahinian, H., Kovacs, K., and Melmed, S. (2000). Long-acting peptidomimergic control of gigantism caused by pituitary acidophilic stem cell adenoma. *J. Clin. Endocrinol. Metab.* **85**, 3409–3416.
- Pei, L., and Melmed, S. (1997). Isolation and characterization of a pituitary tumor-specific transforming gene. *Mol. Endocrinol.* **11**, 433–441.
- Sano, T., Asa, S. L., and Kovacs, K. (1988). Growth hormone-releasing hormone-producing tumors: Clinical, biochemical, and morphological manifestations. *Endocr. Rev.* **9**, 357–373.
- Shimon, I., and Melmed, S. (1997). Pituitary tumor pathogenesis. *J. Clin. Endocrinol. Metab.* **82**, 1675–1681.
- Zimmerman, D., Young, W. F., Ebersold, M. J., Scheithauer, B. W., Kovacs, K., Horvath, E., Whitaker, M. D., Eberhardt, N. L., Downs, T. R., and Frohman, L. A. (1993). Congenital gigantism due to growth hormone-releasing hormone excess and pituitary hyperplasia with adenomatous transformation. *J. Clin. Endocrinol. Metab.* **76**, 216–222.

Glucose stimulation of GIP release appears to be direct as it occurs *in vitro* from a GIP-secreting cell line. The GIP response to fat is slower to develop, greater in magnitude, and more prolonged. The delayed GIP response to fat may be in part due to delayed gastric emptying caused by fat. It should be noted that GIP released by fat would not be insulinotropic unless circulating glucose values were also elevated. In addition to carbohydrates and fat, there are reports of GIP release by ingested amino acids as well as peptone and protein hydrolysate.

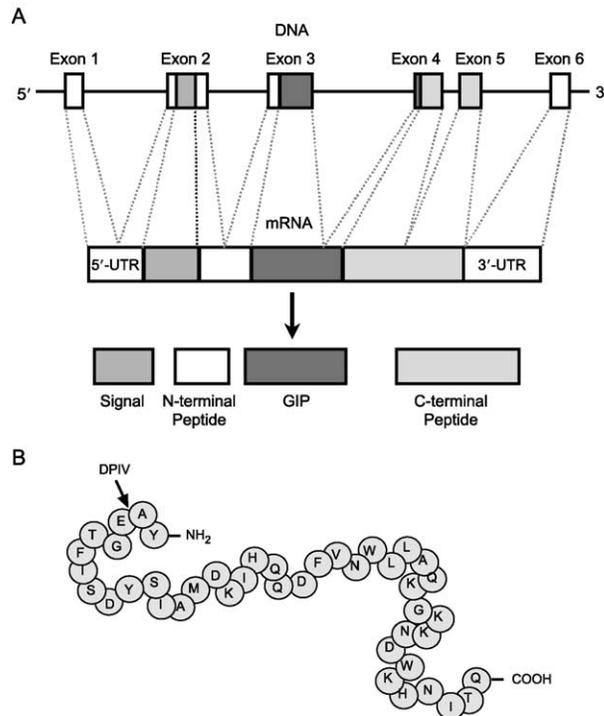
### GIP GENE STRUCTURE AND STRUCTURE–ACTIVITY RELATIONSHIPS

The secreted form of GIP is a 42-amino-acid peptide derived by proteolytic processing of precursors that are 144 and 153 amino acids in length in rat and human, respectively. The gene coding for the human GIP precursor spans approximately 10 kb, whereas that for the rat spans only 8.2 kb. Both consist of six exons, with exons 3 and 4 coding for GIP<sub>1–42</sub> (Fig. 1A). The promoter region of the human GIP gene contains potential binding sites for a number of transcription factors including Sp1, activator protein-1 (AP-1), and AP-2. The human GIP gene has been assigned to chromosome 17q21.3–q22.

A number of studies have been directed at determining the regions of the GIP molecule that are important for biological activity. Experimental evidence indicates that there are four dissociable domains in GIP<sub>1–42</sub>. Residues 6–30 in the GIP molecule constitutes a high-affinity binding domain. Truncation of 12 amino acids from the carboxyl terminus of GIP<sub>1–42</sub> was shown to result in a peptide with intact insulinotropic activity, but without the ability to stimulate gastric somatostatin secretion, indicating the existence of two bioactive domains in the native molecule. The insulinotropic domain of GIP has been further localized to residues 19–30 and Hinke and co-workers presented evidence suggesting that a third bioactive domain of GIP resides in residues 1–14.

### THE GIP RECEPTOR

The pancreatic islet GIP receptor is a glycoprotein with an estimated molecular weight of 59 kDa that belongs to the class B family of heptahelical G protein-coupled receptors. GIP receptor cDNAs have been



**Figure 1** (A) Diagrammatic representation of the GIP gene, propeptide, and the peptides produced by processing. (B) Primary structure of human GIP showing the cleavage site for dipeptidyl peptidase IV.

cloned and it has been shown that the amino acid sequences of the rat, human, and hamster receptors share 40–47% identity with the glucagon-like peptide-1 (GLP-1) and glucagon receptors. Splice variants of the GIP receptor exist in some tissues and it is possible that there is variability in hormone–receptor interaction.

Through the study of mutant and chimeric receptors expressed in different cell lines, it has been shown that the N-terminal domain and first extracellular loop are important for GIP binding and that intracellular (IC) loops are involved in G protein coupling, whereas threonine and serine residues in IC3 and the C-terminal tail are important for receptor internalization. Localization studies have confirmed the broad distribution of the GIP receptor in pancreatic islets, fat, stomach, brain, heart, lung, endothelium, adrenal cortex, and bone, in agreement with GIP's broad pleiotropic actions.

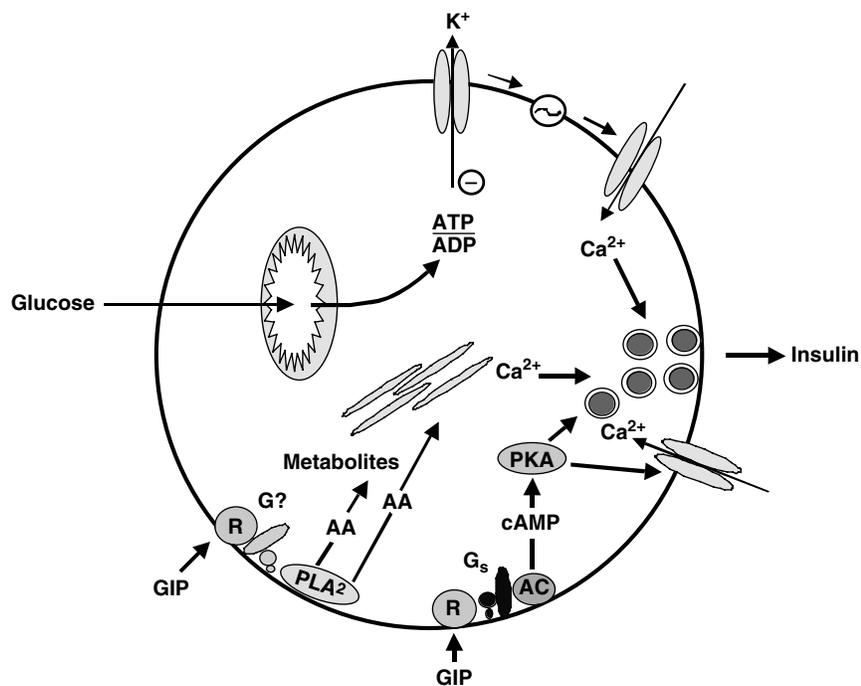
### The GIP Receptor Gene

The human receptor gene contains 14 exons and 12 introns, with a protein coding region of 12.5 kb, whereas that of the rat receptor gene spans ~10.2 kb

and contains 13 exons. The promoter region of the human gene has not yet been characterized, but 5'-flanking sequences of the rat gene contain several transcription factor-binding motifs, including a cyclic AMP (cAMP)-response element, an octamer-binding site, three Sp1 sites, and an initiator element. One potential Sp1-binding site has been shown to be important for transcriptional activity and distal negative regulatory sequences have been suggested to control cell-specific expression. Missense mutations in the receptor gene have been identified in Japanese and Danish populations. Cells expressing one of the mutated receptors (Gly<sup>198</sup>/Cys) exhibited elevated EC<sub>50</sub> values for cAMP production. However, allelic frequency studies have shown no association of the mutations with type 2 diabetes, although in the Danish group those homozygous for a Glu<sup>354</sup>/Gln variant had decreased fasting serum C-peptide levels, suggesting that GIP normally regulates beta cell secretion even in the fasting state. Lynn and co-workers showed that GIP receptor mRNA and protein levels are dramatically reduced in the Vancouver Diabetic Zucker fatty rat, thus partially explaining the resistance to GIP found in these animals.

### GIP-Activated Signal Transduction Pathways

The endocrine pancreas is the only target organ for which there is information on the mode of action of GIP. Glucose is the main nutrient stimulator of insulin secretion (Fig. 2). Glucose enters the pancreatic beta cell and its metabolism results in an increase in the ratio of cytosolic ATP:ADP. The resulting closure of ATP-dependent K<sup>+</sup> channels leads to membrane depolarization and opening of voltage-dependent Ca<sup>2+</sup> channels. The increase in cytosolic Ca<sup>2+</sup> is the major stimulus for insulin secretion. Agents such as GIP, GLP-1, and PACAP (pituitary adenylate cyclase-activating polypeptide) act as both competence factors for glucose-induced secretion and potentiators of the glucose-induced responses. GIP has been shown to stimulate adenylyl cyclase (AC) in pancreatic beta cells and the major pathway by which GIP potentiates insulin secretion is thought to be at the level of Ca<sup>2+</sup>-induced exocytosis via activation of protein kinase A (PKA). Although the pathways involved have not been completely identified, GIP is capable of increasing Ca<sup>2+</sup> influx through both L-type Ca<sup>2+</sup> channels and nonselective ion channels and releasing



**Figure 2** Diagram of a pancreatic islet beta cell showing K<sup>+</sup>-dependent ATP channel-associated glucose stimulation of insulin secretion and proposed mechanisms by which GIP activation of adenylyl cyclase and phospholipase A<sub>2</sub> can potentiate glucose-mediated responses.

Ca<sup>2+</sup> from intracellular stores. Eshes and co-workers have demonstrated that GIP can also activate a Ca<sup>2+</sup>-independent phospholipase A<sub>2</sub> (PLA<sub>2</sub>) in beta cells, resulting in the generation of arachidonic acid. Interestingly, in different cell types, GIP may stimulate PLA<sub>2</sub> via receptor coupling to G<sub>βγ</sub> dimers or via coupling of G<sub>α</sub> and adenylyl cyclase.

### GIP, Insulin Gene Transcription, and Beta Cell Proliferation

In addition to its effect on insulin secretion, glucose is the most important stimulator of proinsulin gene transcription. Its action involves stimulation of cytosolic/nuclear transport of the transcription factor PDX-1 and increased PDX-1 gene transcription. However, both GIP and GLP-1 stimulate the expression of several genes in the beta cell, including those of proinsulin, glucose transporters, and hexokinases. Additionally, these peptides appear to be involved in regulating the differentiation of beta cells from ductal progenitor cells and stimulating beta cell proliferation. Incretins are therefore important regulators of beta cell replication. Although it is not known exactly how GIP and GLP-1 exert their mitogenic actions, it is known that they can induce *c-fos* activity and activate mitogen-activated protein kinase pathways.

### METABOLISM

Through the use of assays of immunoreactive GIP levels, it was determined that the kidney is the major site of GIP clearance. By measuring the rate of disappearance of infused peptide, the circulating half-life in humans was calculated to be 20 min. These measurements can be called into question as it is known that postsecretory degradation of incretins results in products that may be immunoreactive but biologically inactive. Thus, RIAs may overestimate their biological half-lives. Dipeptidyl peptidase IV (DP IV) is a ubiquitously distributed enzyme with substrate specificity for small peptides with proline or alanine residues in the penultimate position from the N terminus. Both GIP and GLP-1 serve as substrates for this enzyme (penultimate alanine residues), yielding N-terminally truncated, biologically inactive products (GIP<sub>3-42</sub>, GLP-1<sub>9-36</sub>). It has been determined that DP IV degradation is the major route of GIP inactivation in the circulation and that the biological half-life of the hormone is approximately 2 min. The fact that most RIAs employ antibodies

directed against the C-terminal region of GIP calls much of the previously published data on circulating GIP levels into question since biologically inactive GIP contributes to immunoreactive GIP levels measured with these RIAs. The physiological role played by DP IV in the regulation of incretin activity, and thus the regulation of blood glucose, has been exploited by Pederson and co-workers in strategies used to enhance insulin secretion and improve glucose tolerance in type 2 diabetes. These strategies have taken the form of administration of DP IV inhibitors that increase the circulating half-life of incretins and the design and synthesis of DP IV-resistant GIP and GLP-1 analogues. An extensive body of evidence indicates that altering plasma incretin concentrations by inhibition of DP IV is a valid therapeutic approach to type 2 diabetes.

## BIOLOGICAL ACTIONS

### Acid Secretion—The Enterogastrone Concept

As discussed earlier, GIP was isolated on the basis of its acid inhibitory properties. GIP was pursued as a candidate for the inhibitor of gastric acid secretion: enterogastrone. This proposal gained considerable support when, in addition to its inhibitory actions on the stomach, it was determined that ingestion of fat was the most potent stimulus for the release of GIP into the circulation. Initial studies were carried out in the denervated canine stomach and subsequent studies in the intact innervated stomach of humans and rats demonstrated weaker inhibitory effects. In later studies, GIP was shown to be a potent stimulus for somatostatin release from the stomach and this release could be inhibited by the vagus or cholinomimetic agents. These studies suggested an indirect pathway for the action of GIP on the parietal cell, i.e., by GIP-induced somatostatin release with neuronal modulation by the parasympathetic nervous system. GIP is undoubtedly not the only enterogastrone and it is likely that it acts in concert with other peptides and/or nervous inhibitory mechanisms that would be initiated during the digestive and absorptive process in the upper small bowel.

### GIP Stimulation of Islet Hormones—The Incretin Concept

As outlined earlier, both enterogastrone and incretin actions of GIP were hypothesized as a result of studies

on impure (GIP-containing) preparations of CCK that demonstrated their ability to inhibit acid secretion and stimulate insulin release. Dupre and co-workers showed that concomitant infusion of porcine GIP and glucose in humans produced a pronounced increase in circulating insulin levels and improved glucose tolerance compared to glucose infusion alone. The insulin response was sustained for the duration of the GIP infusion and was not observed in fasted (euglycemic) individuals. This glucose dependency was subsequently verified in the isolated perfused rat pancreas and a threshold of 5.5 mM glucose for the insulinotropic action of GIP was established. Elegant studies by Anderson *et al.* established conclusively that endogenous GIP released by oral glucose could account for a major part of the enteric contribution to the insulin response to oral glucose in human. The glucose-dependent insulinotropic action of GIP has been demonstrated in many mammalian species as well as a variety of tumor-derived insulin secreting beta cell lines. The action of other insulin secretagogues, such as CCK, acetylcholine, GLP-1, and arginine, have also been demonstrated to be glucose-dependent. A common mechanism whereby glucose metabolism in the beta cell permits stimulation by nonglucose secretagogues has been proposed. GIP has been shown to be glucagonotropic in both the isolated rat pancreas and pancreatic islets. Several studies support a direct effect of GIP on glucagon release. Receptors for GIP are found on rat A cells and GIP increased whole-cell  $Ca^{2+}$  currents and potentiated exocytosis in purified rat pancreatic A cells.

## GIP and Fat Metabolism

In addition to its effect on insulin secretion, GIP has been implicated in the regulation of fat metabolism, an action that could also be of paramount importance in the normal regulation of insulin secretion. Ingestion of triglycerides increases circulating GIP by a pathway that involves both metabolism and absorption of fatty acids. Several lines of evidence support a role for GIP in the subsequent disposal of circulating triglycerides. Administration of GIP promotes chylomicron-associated triglyceride clearance from blood in dogs and reduces plasma triglyceride levels during intraduodenal fat infusion in rats. GIP has also been shown to stimulate lipoprotein lipase activity in cultured preadipocytes. Regarding the effects of GIP on adipocyte metabolism, it has been reported that GIP enhances the incorporation of

both fatty acids and glucose into lipids and inhibits glucagon-stimulated lipolysis and cAMP production. This seemed paradoxical, however, in that GIP signals via cAMP, which is known to stimulate hormone-sensitive lipase and thus would be expected to exert a lipolytic effect. It was hypothesized by McIntosh and co-workers that the response of the adipocyte to GIP may depend on the prevailing circulating insulin concentration and evidence that GIP exerts a lipolytic effect on adipocytes that is inhibited by insulin was provided. One can postulate a physiological role for GIP in fat metabolism in light of the demonstration that critical levels of circulating free fatty acids (FFAs) are required for optimal glucose stimulation of insulin secretion during fasting. GIP could be capable of stimulating lipolysis under conditions in which insulin levels are of insufficient magnitude to inhibit its action and this may ensure that levels of circulating FFAs are optimal for glucose- and GIP-stimulated insulin secretion.

## See Also the Following Articles

CCK (Cholecystokinin) • Gastrin • Gastrin-Releasing Peptide • GI-CGRP (Calcitonin Gene-Related Peptide) • GI Hormone Development (Families and Phylogeny) • Pancreatic Islet Cell Tumors

## Further Reading

- Andersen, D. K., Elahi, D., Brown, J. C., Tobin, J. D., and Andres, R. (1978). Oral glucose augmentation of insulin secretion: Interactions of gastric inhibitory polypeptide with ambient glucose and insulin levels. *J. Clin. Invest.* **62**, 152–161.
- Brown, J. C., Pederson, R. A., Jorpes, E., and Mutt, V. (1969). Preparation of highly active enterogastrone. *Can. J. Physiol. Pharmacol.* **47**, 113–114.
- Dupre, J., Ross, S. A., Watson, D., and Brown, J. C. (1973). Stimulation of insulin secretion by gastric inhibitory polypeptide in man. *J. Clin. Endocrinol. Metab.* **37**, 826–828.
- Ehses, J. A., Lee, S. T., Pederson, R. A., and McIntosh, C. H. S. (2001). A new pathway for glucose-dependent insulinotropic polypeptide (GIP) receptor signaling. *J. Biol. Chem.* **276**, 23667–23673.
- Hinke, S. A., Manhart, S., Pamir, N., Demuth, H.-U., Gelling, R. W., Pederson, R. A., and McIntosh, C. H. S. (2001). Identification of a bioactive domain in the amino-terminus of glucose-dependent insulinotropic polypeptide (GIP). *Biochim. Biophys. Acta* **364**, 1–13.
- Kosaka, T., and Lim, R. K. S. (1930). Demonstration of the humoral agent in fat inhibition of gastric secretion. *Proc. Soc. Exp. Biol. Med.* **27**, 890–891.
- Lynn, F. C., Pamir, N., Ng, E. H. C., McIntosh, C. H. S., and Pederson, R. A. (2001). Defective glucose-dependent insulinotropic polypeptide receptor expression in diabetic fatty Zucker rats. *Diabetes* **50**, 1004–1011.

McIntosh, C. H. S., Pederson, R. A., Koop, H., and Brown, J. C. (1981). Gastric inhibitory polypeptide stimulated secretion of somatostatin-like immunoreactivity from the stomach: Inhibition by acetylcholine or vagal stimulation. *Can. J. Physiol. Pharmacol.* **59**, 468–472.

Pederson, R. A., White, H., Schlezinig, D., Schmidt, J., McIntosh, C. H. S., and Demuth, H. U. (1998). Improved glucose tolerance in Zucker fatty rats by oral administration of the dipeptidyl peptidase IV (DP IV) inhibitor, isoleucine-thiazolidide. *Diabetes* **47**, 1253–1258.



## Gitelman's Syndrome

see Bartter's Syndrome

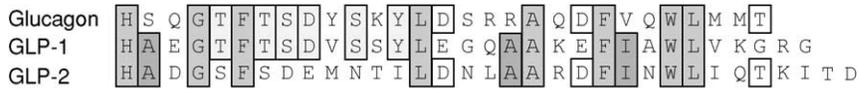


Figure 1 Homology of amino acid sequences of glucagon and the glucagon-like peptides.

of the gut GLP-1 and GLP-2 are the main gene products. The steps in the expression of the preproglucagon gene are outlined in Fig. 2. Several other potentially bioactive peptides are generated from proglucagon such as IP2 (intervening peptide 2) (as shown in Fig. 2) but this article focuses on glucagon and the glucagon-like peptides.

### HORMONE RECEPTORS AND CELL SIGNALING

The functional effects of glucagon and the GLPs are mediated by specific receptors located within the plasma membrane of target cells. Like the peptide hormones, the hormone-specific receptors share extensive amino acid sequence homology and form part of a closely related subfamily of receptors as part of the larger superfamily of G protein-coupled receptors. This superfamily is distinguished by the tertiary structure of its members, comprising seven transmembrane domains, and the coupling to trimeric G proteins. The coupling to the stimulatory G protein G<sub>s</sub>, the activation of adenylyl cyclase, and the subsequent increase in the intracellular production of cyclic AMP (cAMP) constitute the most established functional responses elicited by all the members of the

glucagon receptor subfamily. This uniform response in intracellular function suggests that the divergent physiological effects of glucagon, GLP-1, and GLP-2 are due primarily to tissue-specific expression of the receptors, whereby cAMP imparts distinct cellular functions due to the equally diverse array of gene expression patterns in different cell types. Furthermore, temporal and tissue-specific patterns of secretion of the hormones themselves play a critical role in initiating the appropriate functional response. Notably, in the pancreatic beta cell, in which GLP-1 and glucagon receptors are expressed, both glucagon and GLP-1 stimulate insulin secretion, mediated at least in part by cAMP. Paradoxically, however, insulin gene expression is inhibited by glucagon and stimulated by GLP-1. This contradiction suggests that more subtle distinctions in receptor signaling exist, leading to differential intracellular responses to the two hormones, glucagon and GLP-1.

### PHYSIOLOGY

#### Glucagon

The glucagon receptor is widely expressed and can be found in the liver, adipose tissue, heart, kidney,

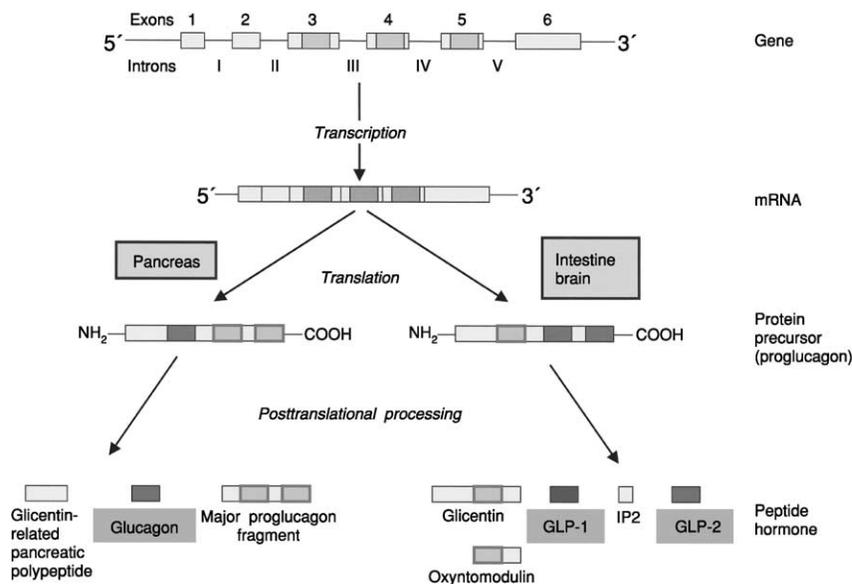


Figure 2 Tissue-specific processing of the preproglucagon gene.

pancreatic islets, stomach, small intestine, thyroid, and skeletal muscle. Although contradictions in the specific expression pattern exist, likely due to the sensitivity of detection techniques between laboratories, the less ambiguous targets with high receptor levels correspond to the most well-established physiological actions of glucagon. The hepatocyte is a primary target cell of glucagon to which it is exposed when the hormone is released in the portal vein following secretion from the pancreatic alpha cells. Glucagon increases glucose output from the liver, an effect that results from inhibition of glycogen synthesis and stimulation of both glycogenolysis and gluconeogenesis (Table I). Evidence suggests that these effects of glucagon are mediated by cAMP, although the glycogenolytic effect of glucagon may occur via a cAMP-independent mechanism. It is not surprising, therefore, that glucagon helps to maintain hepatic glucose output and thereby to sustain blood glucose levels. There exists a general consensus that the hyperglycemic activity of glucagon represents the first line of defense against hypoglycemia. Accordingly, stimulants of glucagon secretion from alpha cells include the presence and actions of factors at times when the body requires glucose for sustenance (during fasting) or fuel (during exercise).

The interplay between insulin and glucagon provides a tightly controlled equilibrium in blood glucose concentration. Insulin stimulates glucose uptake into fat and muscle cells and also inhibits glucagon expression and secretion from alpha cells. Thus, during periods when plasma insulin levels are low, such as during periods of fasting and exercising, plasma glucagon levels are elevated. The anatomical structure of the pancreatic islets of Langerhans further reflects this mechanistic action of insulin on glucagon secretion. Morphological studies have shown that in a given islet, the microcirculation goes from the core to the mantle. This circumstance suggests that *in vivo*, peripheral alpha cells are exposed to high concentrations of insulin released from the more centrally located beta cells in the islets. Thus, insulin and glucagon function inversely to regulate blood glucose levels.

**Table I** Actions of Glucagon

↑	Hyperglycemia
↑	Glycogenolysis
↑	Gluconeogenesis
↑	Glycerol and free fatty acid release
↑	Lypolysis
↓	Glycogen synthesis

Dysregulation of glucagon by a breakdown in the intra-islet insulin–glucagon relationship appears to be a substantial component in the pathogenesis of diabetes. As a result of insufficient insulin secretion, plasma glucagon levels are elevated in type 1 (insulin-dependent) diabetes and also to a lesser extent in type 2 (non-insulin-dependent) diabetes, thereby exacerbating the hyperglycemic state. This effect is further heightened when plasma glucagon levels fail to be suppressed following a meal and postprandial hyperglycemia ensues. It is compelling to speculate, therefore, that agents that are able to antagonize glucagon action or secretion may be of value in the treatment of patients with diabetes.

Paradoxically, a major feature of long-term diabetes is the impaired response to insulin-induced hypoglycemia, a condition of impairment that is almost universally present after 5 years of duration of type 2 diabetes. Studies have shown that hyperinsulinemia, which may be caused by insulin therapy, and insulin resistance can induce defective counterregulation in the manifestation of hypoglycemia, via defective glucagon release. These examples serve to emphasize the finely balanced hormonal equilibrium that characterizes the normal physiological state of glucose homeostasis.

Consistent with the role of glucagon in providing the body with usable energy, additional established targets of glucagon action are the adipocyte and skeletal myocytes (Table I). Although studies of the effects of glucagon on adipose tissue and other tissues are difficult to evaluate owing to the rapid breakdown of the peptide by proteolytic activity associated with these cells, substantial evidence demonstrates that glucagon induces lipolysis, stimulates the release of glycerol and free fatty acids in adipocytes, and stimulates glycogenolysis in myocytes. Therefore, glucagon could qualify as the “hormone of fuel need” rather than just a “hormone of glucose need.”

## GLP-1

The major secreted and bioactive form of GLP-1 from enteroendocrine L cells is a truncated isopeptide cleaved from the full 37-amino-acid peptide. Cleavage of the six N-terminal amino acids of GLP-1(1–37) to form GLP-1 (7–37) and removal of the glycine residue at position 37 followed by amidation of the exposed arginine at position 36 generate GLP-1(7–36)amide. The most significant physiological function of GLP-1 is the stimulation of the synthesis and secretion of insulin from pancreatic islet beta cells

Table II Actions of GLP-1

↑	Insulin biosynthesis and secretion
↑	Islet proliferation and neogenesis
↑	Islet cell differentiation
↑	Glucose uptake
↓	Gastric emptying
↓	Acid secretion
↓	Food intake (appetite)
↓	Glucagon secretion

(Table II). The mechanism of action of GLP-1 is a potentiation of glucose-induced insulin secretion. In the absence of glucose, GLP-1 fails to stimulate insulin secretion, whereas at elevated glucose concentrations, GLP-1 stimulates secretion at a rate that exceeds that induced by glucose alone and thus potentiates secretion in a glucose-dependent manner. Notably, the secretion of GLP-1 from L cells is stimulated by glucose (and other nutrients) following oral ingestion. Thus, the temporal release of GLP-1, coincident with enhanced blood glucose levels, cooperates synergistically to provide a highly sensitive insulin-releasing mechanism. GLP-1 is thus regarded as a “glucose competence factor.” As described above, insulin suppresses plasma glucagon levels and activates glucose uptake and metabolism, thus returning blood glucose levels to normal following a meal and maintaining glucose homeostasis in general. In this regard, insulin and GLP-1 play an important role in preventing hyperglycemia.

Evidence for additional mechanisms of action of GLP-1 in lowering blood glucose levels have emerged; GLP-1 has been shown to augment insulin-mediated glucose uptake in patients with diabetes. Extrapancreatic targets of GLP-1 actions, particularly those of adipose and skeletal muscle tissues, are under investigation. It seems possible that a novel GLP-1 receptor mediates these effects of GLP-1 on adipose tissue and skeletal muscle, as well as liver.

Additional functions of GLP-1 include the inhibition of gastric emptying (by inhibiting intestinal motor activity in response to nutrients in the distal gut) and of meal-stimulated gastric acid secretion. A role for GLP-1 in feeding behavior has also been identified. Centrally administered GLP-1 inhibited both feeding and drinking activity in rats, suggesting that GLP-1 may be a central neurotransmitter that modulates visceral functions. Finally, in addition to the detection of pluripotential stem cells within the pancreas, some evidence suggests that GLP-1 stimulates the proliferation and differentiation of these

stem/progenitor cells into beta cells and thus GLP-1 may play a developmental role in the endocrine pancreas.

As the characterization of GLP-1 action continues and a clearer picture of its multifunctional profile appears, it is evident that GLP-1 offers considerable potential as a therapeutic treatment for diabetes mellitus. By enhancing both the means to secrete insulin (by incretin action and differentiation of new beta cells) and the functional ability of insulin to induce glucose disposal, it may be possible to better control circulating blood sugar levels in diabetic patients.

## GLP-2

GLP-2 is synthesized in the enteroendocrine L cells (along with GLP-1) and in the brainstem and hypothalamus of the central nervous system. GLP-2 may act on the hypothalamus to curtail appetite and food intake. However, the regulation and function of GLP-2 in the intestinal system are better understood than those of GLP-2 in the central nervous system.

Initially, several lines of evidence suggested the importance of GLP-2 in maintaining the functional integrity of the gastrointestinal epithelium. Examples included the coincidence of intestinal injury and inflammatory bowel disease with elevated levels of circulating GLP-2 and patients with glucagon-secreting tumors who presented with small bowel villus hyperplasia (Table III). It was subsequently shown that administration of GLP-2 to mice leads to induction of intestinal epithelial proliferation. Furthermore, following treatment of mice and rats with exogenous GLP-2 for 7–10 days, a marked increase in villus height and small bowel mass and a smaller increment in small bowel length are observed. GLP-2 also appears to slow the ingestion and transit of food through the gastrointestinal tract, while increasing the absorption of nutrients from the small intestine.

As with glucagon and GLP-1, the tissue-specific effects of GLP-2 action are closely reflected by the expression profile of the GLP-2 receptor, which is predominantly found in the stomach, jejunum, ileum, and colon. Although GLP-2 presumably exerts direct effects on cells expressing its putative receptor, some physiological effects of GLP-2 in the intestine are thought to be exerted indirectly in an autocrine, paracrine, or endocrine manner by stimulating the release of as yet unknown factors from GLP-2 receptor-positive cells. This hypothesis arose due to conflicting data regarding the exact tissue distribution of the GLP-2 receptor. Whereas one group observed

Table III Actions of GLP-2

↑	Nutrient absorption
↑	Crypt cell proliferation
↑	Small and large bowel growth
↑	Villus height
↓	Food intake (appetite)
↓	Gastric motility
↓	Enterocyte and crypt apoptosis

that binding of labeled GLP-2 localized diffusely along the villous epithelium, others have reported the absence of the receptor among several intestinal epithelial cell lines. Thus, because the localization profile of the GLP-2 receptor is not known with certainty, the mechanism of the pleiotropic biological actions of GLP-2 in the gut, such as modulation of gastric motility, small bowel permeability, and both crypt cell proliferation and apoptosis, is confined to speculation. However, some evidence suggests that GLP-2 can directly mediate proliferative and anti-apoptotic effects through its receptor in cultured cells.

Apart from its intestinal functions, GLP-2 is also expressed and received in the brain. Receptors were located in the hypothalamus and in certain cell groups in the cerebellum, medulla, amygdala, and hippocampus. Although studies of GLP-2 action in the brain are just beginning, a functional role of GLP-2 appears to be in the regulation of food intake. Pharmacological and behavioral studies indicate that GLP-2 acts as a specific transmitter that may inhibit feeding behavior and has potential long-term effects on body weight homeostasis.

### INACTIVATION OF GLPs

The active half-life of both GLP-1 and GLP-2 is regulated by dipeptidyl peptidase IV (DPP IV/CD26), an extracellular soluble or membrane-anchored protease, which cleaves the two N-terminal amino acids and subsequently inactivates the hormones, with respect to their known functions. This potent inactivation restricts the circulating bioactive

half-lives to 2–3 min *in vivo*. Particularly high levels of expression of DPP IV are found in the kidney, lung, liver, and jejunum. Expression in endothelial cells of the blood vessels and the presence of small amounts of DPP IV as a soluble enzyme in the blood result in close contact with circulating substrate and added rapidity of peptide inactivation. Consequently, specific DPP IV inhibitors are of special interest for physiological investigations and for potential clinical applications, such as treatment of type 2 diabetes and improvement of mucosal regeneration. Studies have shown that prolonged activation of GLP-1 by DPP IV inhibitors results in a potentiation of the insulinotropic effect of GLP-1 *in vivo* and administration of GLP-2 to DPP IV-deficient rats markedly increased the bioactivity of GLP-2, resulting in a significant increase in small bowel weight. A second approach to enhance the bioactive half-life of the GLPs is the synthesis of DPP IV-resistant analogues of the GLPs for pharmacological administration. These approaches are in development.

### See Also the Following Articles

GI Hormone Development (Families and Phylogeny) • Insulin Secretion: Functional and Biochemical Aspects • Pituitary Adenylate Cyclase-Activating Polypeptide (PACAP)/Glucagon Superfamily

### Further Reading

- Ahren, B. (1998). Glucagon-like peptide-1 (GLP-1): A gut hormone of potential interest in the treatment of diabetes. *BioEssays* **20**, 642–651.
- Drucker, D. J. (1998). Glucagon-like peptides. *Diabetes* **47**, 159–169.
- Habener, J. F. (2001). Glucagonlike peptide-1 agonist stimulation of  $\beta$ -cell growth and differentiation. *Curr. Opin. Endocrinol. Diabetes* **8**, 74–81.
- Kieffer, T. J., and Habener, J. F. (1999). The glucagon-like peptides. *Endocr. Rev.* **20**, 876–913.
- Lefgbyre, P. J. (1995). Glucagon and its family revisited. *Diabetes Care* **18**, 715–730.
- Lovshin, J., and Drucker, D. J. (2000). New frontiers in the biology of GLP-2. *Regul. Pept.* **90**, 27–32.
- Mentlein, R. (1999). Dipeptidyl-peptidase IV (CD26)-role in the inactivation of regulatory peptides. *Regul. Pept.* **85**, 9–24.

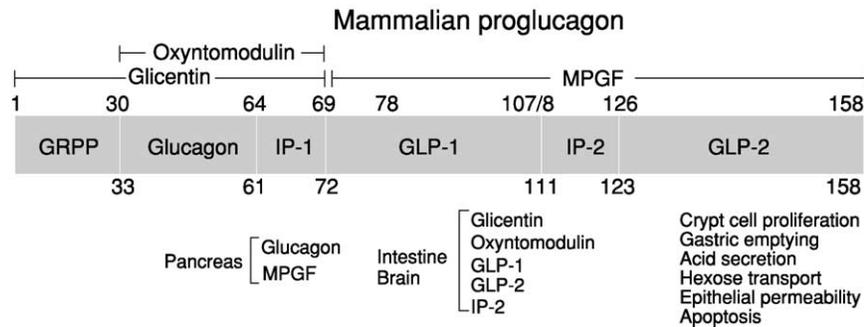


Figure 1 Structure of proglucagon and biological actions of GLP-2.

GLP-2 administration reduces epithelial permeability in the small intestine, effects noted within hours of peptide administration. The capacity of the GLP-2-treated bowel for nutrient absorption appears to be normal to enhanced, as assessed by nutrient challenge studies in normal mice following chronic GLP-2 administration.

### GLP-2 ACTIONS IN THE BRAIN

GLP-2 is synthesized in the brain, predominantly in the nucleus of the solitary tract in the brainstem and to a lesser extent in the hypothalamus. GLP-1 and GLP-2 are transported from the brainstem to distant central nervous system (CNS) nuclei via axonal transport. Within the hypothalamus, GLP-2 is distributed in a pattern similar to that of GLP-1; however, GLP-2 administration specifically activates neurons in the dorsomedial nucleus of the hypothalamus, where GLP-2 receptors have been localized by *in situ* hybridization. The GLP-2 receptor is also expressed in thalamic, cortical, cerebellar, hippocampal, and amygdaloid nuclei. Although the physiological role of GLP-2 in the CNS remains unclear, intracerebroventricular administration of pharmacological amounts of GLP-2 transiently inhibits food intake in mice and rats.

### GLP-2 AND GUT DEVELOPMENT

GLP-2 and GLP-2 receptors are expressed in the fetal and neonatal gut and GLP-2 stimulates intestinal epithelial proliferation in neonatal pigs and rats. The trophic effects of GLP-2 in the gastrointestinal tract may be restricted to postnatal life, as exogenous GLP-2 does not stimulate gut growth in late-gestational fetal pigs despite the presence of immunoreactive circulating GLP-2 in the fetal pig.

### THE GLP-2 RECEPTOR

GLP-2 exerts its actions through a distinct 550-amino-acid G protein-coupled GLP-2R (GLP-2 receptor). The cDNA encoding the GLP-2R was initially cloned from human and rat hypothalamic and intestinal cDNA libraries. Consistent with the known actions of GLP-2, GLP-2R expression is predominantly confined to the stomach, duodenum, jejunum, ileum, and colon. GLP-2R transcripts have also been detected in the brain and lung, although the biological significance of GLP-2R expression in lung remains unclear. Within the gastrointestinal tract, GLP-2 receptor-like immunopositivity has been localized to a subset of human endocrine cells using GLP-2R-specific antisera and GLP-2R expression was demonstrated in the murine enteric nervous system by *in situ* hybridization.

Although cell lines expressing an endogenous GLP-2 receptor have not yet been identified, GLP-2 increases intracellular cyclic AMP (cAMP) and activates protein kinase A but does not increase intracellular  $Ca^{2+}$  release in cells expressing a transfected human or rat GLP-2 receptor. GLP-2 directly activates immediate-early gene expression and cellular growth in serum-starved fibroblasts *in vitro*; however, the growth-promoting effects of GLP-2 in the gut are largely indirect *in vivo*. GLP-2 also prevents apoptosis following administration of cycloheximide (an inhibitor of protein synthesis) or chemotherapeutic agents to cells expressing a GLP-2R through cAMP-dependent mechanisms; hence, GLP-2R activation directly and indirectly signals intracellular pathways regulating both mitogenic and anti-apoptotic actions.

### THERAPEUTIC POTENTIAL OF GLP-2

GLP-2 infusion completely prevented the mucosal atrophy associated with withdrawal of enteral

nutrients in the small intestine but not the large intestine of the rat. Similarly, GLP-2 administration following small bowel resection in rats increased intestinal remnant adaptation and improved absorptive function. Reparative and protective effects of GLP-2 have also been observed in mice following ischemic injury induced by occlusion of the superior mesenteric artery.

GLP-2 exerts reparative and cytoprotective effects in rodents in the setting of intestinal inflammation. The protease-resistant GLP-2 analogue h[Gly2]-GLP-2 significantly reduced the severity of enteritis, bacterial sepsis, and mortality in mice following administration of the nonsteroidal anti-inflammatory agent indomethacin. Similarly, mice treated with h[Gly2]-GLP-2 exhibit reduced large bowel injury and significantly less weight loss in the setting of dextran sulfate colitis. The anti-apoptotic actions of GLP-2 are also detectable *in vivo* following chemotherapy administration. Administration of h[Gly2]-GLP-2 prior to irinotecan or 5-fluorouracil attenuated epithelial injury in the gut, decreased intestinal epithelial apoptosis, and reduced mortality in mice. Although GLP-2 promotes intestinal epithelial proliferation in the gastrointestinal tract, co-infusion of GLP-2 and parenteral nutrition does not affect tumor progression or tumor growth in tumor-bearing rats.

The reparative and proabsorptive effects of GLP-2 have been examined in a pilot trial in human subjects with short bowel syndrome. Eight patients with stable chronic intestinal insufficiency without a colon in continuity were treated with human GLP-2, 400  $\mu$ g twice daily by subcutaneous injection for 35 days. GLP-2-treated subjects exhibited significantly increased energy, wet weight, and nitrogen absorption, as measured in metabolic balance studies. Body weight and lean body mass increased, whereas fat mass decreased and the time to 50% gastric emptying was increased in GLP-2-treated subjects. Analysis of intestinal histology revealed increases in crypt depth and villus height following GLP-2 administration.

## SUMMARY

GLP-2 has been established as the elusive proglucagon-derived peptide with potent effects on gut growth and nutrient absorption. The essential physiological role(s) of GLP-2 has not yet been elucidated due to the lack of effective specific antagonists or immunoneutralizing antisera. The cytoprotective, proabsorptive, and regenerative properties of GLP-2 *in vivo*

appear to be indirect, mediated by a GLP-2 receptor expressed on endocrine cells and enteric neurons. The downstream effectors released following GLP-2R activation are unknown. Although intracerebroventricular administration of GLP-2 inhibits food intake in rodents, it seems likely that GLP-2 exerts additional as yet unknown actions in the central nervous system. The rapid degradation and clearance of GLP-2 suggest that protease-resistant analogues may exhibit advantages for long-term therapeutic administration *in vivo*. The diverse actions of GLP-2 in the gut are focused on maintaining the integrity of the epithelial mucosa and optimizing nutrient absorption, consistent with the physiological importance of proglucagon-derived peptides in the integrated control of energy homeostasis.

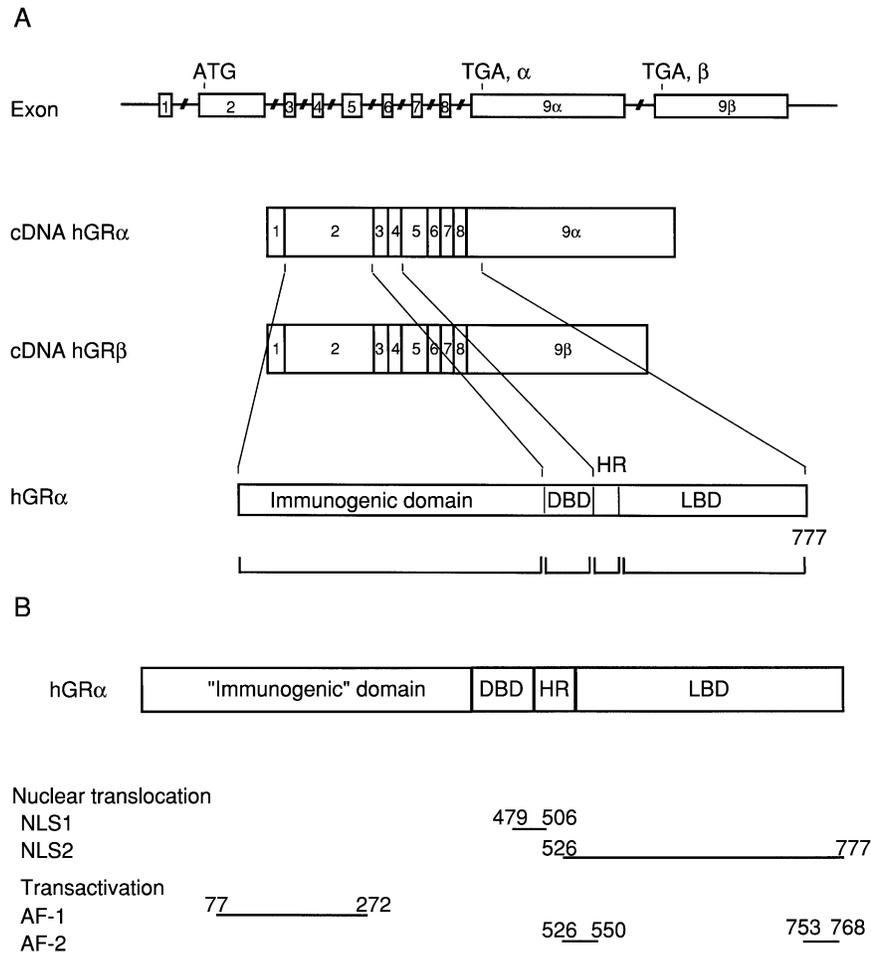
## See Also the Following Articles

GI Hormone Development (Families and Phylogeny) • GI Hormones as Growth Factors

## Further Reading

- Bjerknes, M., and Cheng, H. (2001). Modulation of specific intestinal epithelial progenitors by enteric neurons. *Proc. Natl. Acad. Sci. USA* **98**, 12497–12502.
- Boushey, R. P., Yusta, B., and Drucker, D. J. (1999). Glucagon-like peptide 2 decreases mortality and reduces the severity of indomethacin-induced murine enteritis. *Am. J. Physiol.* **277**, E937–E947.
- Boushey, R. P., Yusta, B., and Drucker, D. J. (2001). Glucagon-like peptide (GLP)-2 reduces chemotherapy-associated mortality and enhances cell survival in cells expressing a transfected GLP-2 receptor. *Cancer Res.* **61**, 687–693.
- Brubaker, P. L., Crivici, A., Izzo, A., Ehrlich, P., Tsai, C.-H., and Drucker, D. J. (1997). Circulating and tissue forms of the intestinal growth factor, glucagon-like peptide 2. *Endocrinology* **138**, 4837–4843.
- Burrin, D. G., Stoll, B., Jiang, R., Petersen, Y., Elnif, J., Buddington, R. K., Schmidt, M., Holst, J. J., Hartmann, B., and Sangild, P. T. (2000). GLP-2 stimulates intestinal growth in premature TPN-fed pigs by suppressing proteolysis and apoptosis. *Am. J. Physiol. Gastrointest. Liver Physiol.* **279**, G1249–G1256.
- Drucker, D. J. (2001). Glucagon-like peptide 2. *J. Clin. Endocrinol. Metab.* **86**, 1759–1764.
- Drucker, D. J., Ehrlich, P., Asa, S. L., and Brubaker, P. L. (1996). Induction of intestinal epithelial proliferation by glucagon-like peptide 2. *Proc. Natl. Acad. Sci. USA* **93**, 7911–7916.
- Drucker, D. J., Shi, Q., Crivici, A., Sumner-Smith, M., Tavares, W., Hill, M., Deforest, L., Cooper, S., and Brubaker, P. L. (1997). Regulation of the biological activity of glucagon-like peptide 2 by dipeptidyl peptidase IV. *Nat. Biotechnol.* **15**, 673–677.
- Lovshin, J., Estall, J., Yusta, B., Brown, T. J., and Drucker, D. J. (2001). Glucagon-like peptide-2 action in the murine central nervous system is enhanced by elimination of GLP-1 receptor signaling. *J. Biol. Chem.* **276**, 21489–21499.

- Yusta, B., Boushey, R. P., and Drucker, D. J. (2000). The glucagon-like peptide-2 receptor mediates direct inhibition of cellular apoptosis via a cAMP-dependent protein kinase-independent pathway. *J. Biol. Chem.* **275**, 35345–35352.
- Xiao, Q., Boushey, R. P., Drucker, D. J., and Brubaker, P. L. (1999). Secretion of the intestinotropic hormone glucagon-like peptide 2 is differentially regulated by nutrients in humans. *Gastroenterology* **117**, 99–105.
- Yusta, B., Huang, L., Munroe, D., Wolff, G., Fantaske, R., Sharma, S., Demchyshyn, L., Asa, S. L., and Drucker, D. J. (2000). Enteroendocrine localization of GLP-2 receptor expression. *Gastroenterology* **119**, 744–755.



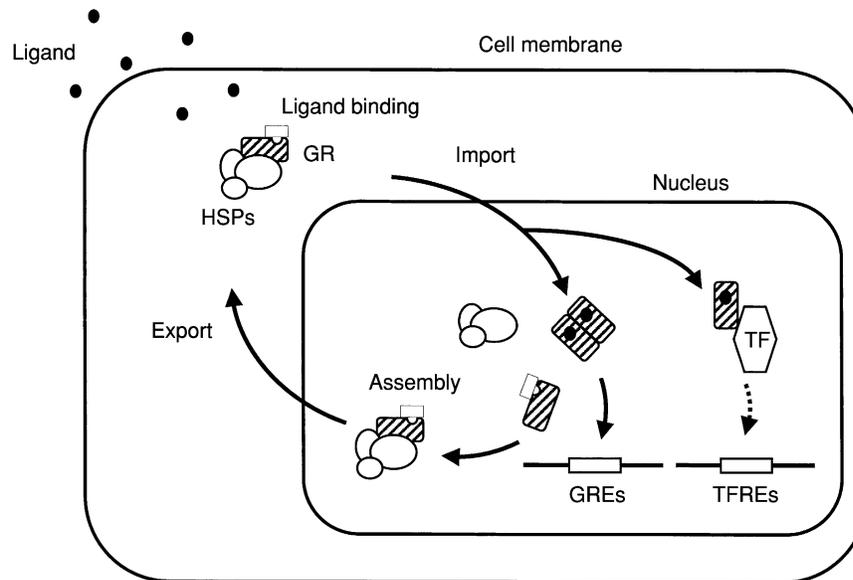
**Figure 1** (A) Genomic, complementary DNA, and protein structures of the human glucocorticoid receptor (hGR). The hGR gene consists of 10 exons. Exon 1 is an untranslated region, exon 2 codes for the amino-terminal domain, exons 3 and 4 code for the DNA-binding domain, and exons 5–9 code for the hinge region and the ligand-binding domain. The glucocorticoid receptor gene contains two terminal exons 9 (exon 9 $\alpha$  and 9 $\beta$ ), alternatively spliced to produce the hGR $\alpha$  and hGR $\beta$  isoforms. (B) Functional domains of the glucocorticoid receptor. The functional domains and subdomains are indicated beneath the linearized protein structures. AF, activation function; DBD, DNA-binding domain; LBD, ligand-binding domain; NLS, nuclear localization signal.

## MOLECULAR MECHANISMS OF GR ACTION

### Nucleocytoplasmic Shuttling of GR

In the absence of ligand, hGR $\alpha$  resides primarily in the cytoplasm of cells as part of a large multiprotein complex, which consists of the receptor polypeptide, two molecules of hsp90, and several other proteins. The hsp90 molecules are thought to sequester hGR $\alpha$  in the cytoplasm of cells by maintaining the receptor in a conformation that masks or inactivates its nuclear localization signals (NLSs). Upon hormone binding, the receptor undergoes an allosteric change, which results in dissociation from hsp90 and other proteins,

unmasking of the NLSs, and exposure of the ligand-binding pocket. In its new conformation, the activated, ligand-bound hGR $\alpha$  translocates into the nucleus, where it binds as homodimer to glucocorticoid-response elements (GREs) located in the promoter region of target genes. hGR $\alpha$  then communicates with the basal transcription machinery and regulates the expression of glucocorticoid-responsive genes positively or negatively, depending on the GRE sequence and promoter context (Fig. 2). The receptor can also modulate gene expression, as a monomer, independently of GRE binding, by physically interacting with other transcription factors, such as activating protein-1 (AP-1) and nuclear factor  $\kappa$ B (NF- $\kappa$ B).



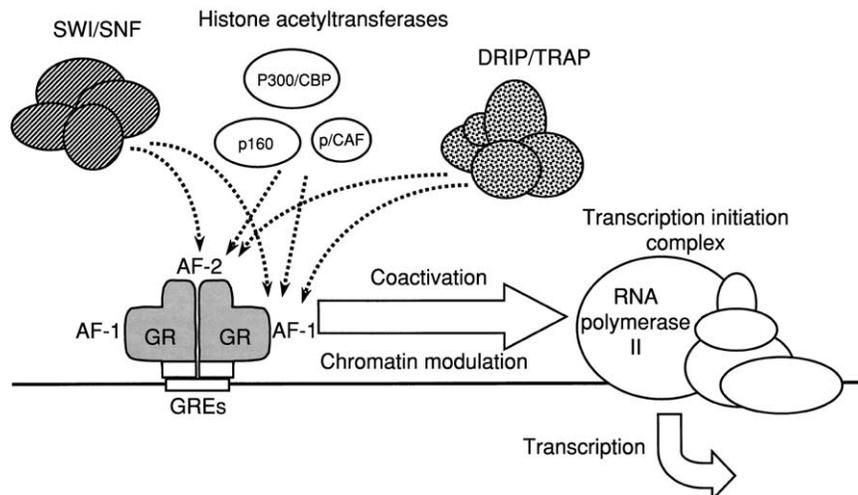
**Figure 2** Nucleocytoplasmic shuttling of the glucocorticoid receptor. GR, glucocorticoid receptor; GREs, glucocorticoid-response elements; HSPs, heat shock proteins; TF, transcription factor; TFREs, transcription factor-responsive elements.

### Mechanisms of Transcriptional Activation by GR

Following binding to GREs, the liganded, activated hGR $\alpha$  enhances the expression of glucocorticoid-responsive genes by regulating the assembly of a transcriptional preinitiation complex at the promoter region of target genes. This is achieved by interaction of the liganded receptor with the basal transcription factors, a group of proteins composed of RNA polymerase II, TATA-binding protein (TBP), and a host of TBP-associated proteins (TAFIIs). The interaction between the activated receptor and the basal transcription factors is mediated by the coactivators, which are nucleoproteins with chromatin-remodeling activity and other enzymatic activities (Fig. 3).

Like other transcriptional activators, hGR $\alpha$  uses its transcriptional activation domains AF-1 and AF-2 as surfaces to recruit chromatin remodeling factors and to interact with the coactivators that link enhancer-bound transcription factors to general transcription factors, thereby initiating transcription. At least two regions of hGR $\alpha$  possess intrinsic transcriptional activation functions: AF-2, which maps to the carboxyl terminus, is glucocorticoid-dependent, with ligand binding promoting the formation of a surface that permits protein-protein contacts between AF-2 and additional regulatory factors. In contrast, AF-1 is located at the amino terminus of the receptor, is glucocorticoid-independent, and can recruit both positive and negative regulatory factors that differentially regulate hGR $\alpha$  transcriptional enhancement.

Several families of nuclear hormone receptor coactivators have been described, including the p160 coactivators, such as the steroid receptor coactivator 1 and the glucocorticoid receptor-interacting protein 1, the p300 and cyclic AMP-responsive element-binding protein (CBP) cointegrators, and the p300/CBP-associated protein (p/CAF). The p160 coactivators are the first to be tethered to the promoter region of steroid target genes, thus playing a pivotal role in hGR $\alpha$ -mediated transactivation. These coactivators interact directly with both the AF-1 of hGR $\alpha$  through their carboxyl-terminal domain and the AF-2 of hGR $\alpha$  through multiple amphipathic LXXLL signature motifs, which are located in their nuclear receptor-binding domain (NRB). They also contain an additional binding site for hGR $\alpha$ , called auxiliary nuclear receptor-interacting domain, which is located between the NRB and the carboxyl-terminal AF-1-binding site. The p160 proteins have intrinsic histone acetyl-transferase activity, which disrupts the DNA nucleosomal interactions at these promoters, allowing the initiation of transcription. Other coactivators that interact with hGR $\alpha$  include the switching/sucrose nonfermenting complex and the newly described chromatin remodeling complex, vitamin D receptor-interacting protein (DRIP)/thyroid hormone-associated protein (TRAP) complex. The DRIP/TRAP complex interacts with both the AF-2 and AF-1 domains of hGR $\alpha$  via its components DRIP205 and DRIP150, respectively. Through coordinated interactions with AF-1 and AF-2, the



**Figure 3** Schematic representation of the interaction of the glucocorticoid receptor with coactivators and other chromatin modulators. AF, activation function; p/CAF, p300/CBP-associated factor; CBP, cAMP-responsive element-binding protein (CREB)-binding protein; DRIP, vitamin D receptor-interacting protein; SNE, sucrose nonfermenting; SWI, mating-type switching; TRAP, thyroid hormone receptor-associated protein.

coactivators enhance the transmission of signals from the DNA-bound hGR $\alpha$  to the transcriptional initiation machinery, loosen chromatin structure, and facilitate access and/or binding of other transcription factors and transcription initiation components to DNA, leading to full transcriptional activity of ligand-bound hGR $\alpha$ .

### Interaction of GR with Other Transcription Factors

In addition to binding to GREs, glucocorticoids may regulate transcription by physically interacting with other transcription factors. Protein-protein interactions between hGR and NF- $\kappa$ B and between AP-1 and signal transducers and activators of transcription result in negative or positive regulation of their responsive genes, mediating many of the anti-inflammatory and immunosuppressive effects of glucocorticoids. hGR may also interact with transcription factor Nur77 to regulate the expression of the proopiomelanocortin gene, p53, which functions as a tumor suppressor gene, as well as others, such as HNF6, Oct-1 and-2, and GATA-1.

## FACTORS THAT MODULATE GR ACTIVITY

### Phosphorylation of GR

Following ligand binding, hGR is phosphorylated at several sites. The mitogen-activated protein kinase

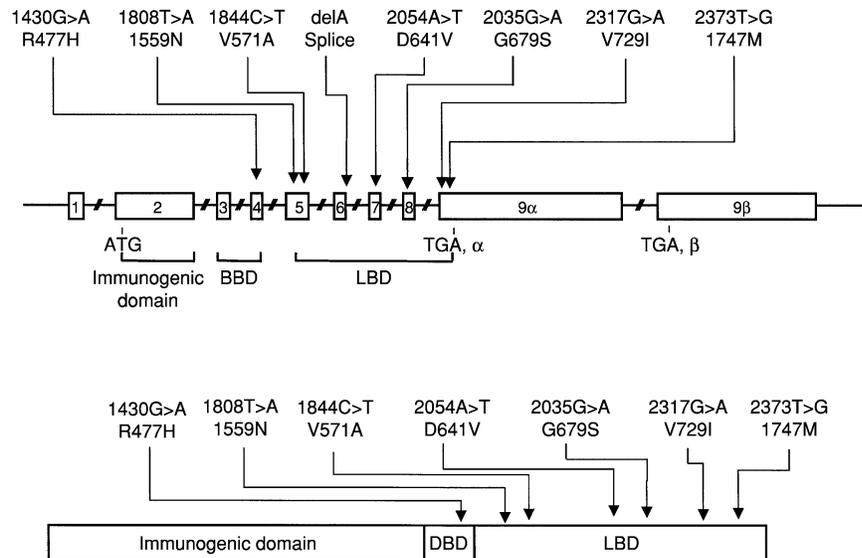
(MAPK) and the cyclin-dependent kinases (CDKs) phosphorylate hGR and modulate its transcriptional activity. MAPK suppresses the activity of hGR in yeast, whereas CDKs stimulate it. JNK, another mitogen-activated kinase, also phosphorylates hGR and suppresses its transcriptional activity. All these kinases phosphorylate hGR at different sites, indicating that the function of hGR may be also regulated by other signal transduction pathways through phosphorylation.

### Chaperones and Cochaperones

hGR forms heterocomplexes with several heat shock proteins, including hsp90, hsp70, hsp56, and possibly hsp23, which may affect its transcriptional activity. hsp90 and receptor-associating protein 46 regulate the transcriptional activity of hGR negatively.

### Chemicals and Compounds

Several chemical compounds may modulate the transcriptional activity of hGR. 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin, a widespread environmental contaminant that produces adverse biological effects such as carcinogenesis, reproductive toxicity, immune dysfunction, hepatotoxicity, and teratogenesis, suppresses the activity of hGR by decreasing binding to glucocorticoids. Geldanamycin, a benzoquinone ansamycin, suppresses hGR function by inhibiting its translocation into the nucleus. On the other hand, thioredoxin, a compound that is accumulated during



**Figure 4** Location of the known mutations of the human glucocorticoid receptor in the genomic structure and its linearized molecule.

oxidative stress, and ursodeoxycholic acid, one of the hydrophilic bile acids, enhance hGR transactivation.

### Natural GR Mutations

Mutations in the hGR $\alpha$  gene may impair one or more of the molecular mechanisms of hGR function, resulting in alterations in tissue sensitivity to glucocorticoids and the clinical phenotype of glucocorticoid resistance. Glucocorticoid resistance is a rare, familial or sporadic condition characterized by generalized, partial end-organ insensitivity to physiologic glucocorticoid concentrations. Patients have compensatory elevations in circulating cortisol and adrenocorticotropic hormone concentrations and resistance of the hypothalamic–pituitary–adrenal axis to dexamethasone suppression, but no clinical evidence of hypo- or hypercortisolism. More than 10 kindreds and sporadic cases with the condition have been reported. Abnormalities of several hGR $\alpha$  characteristics, including cell concentrations, affinity for glucocorticoids, stability, and translocation into the nucleus, have been associated with this condition (Fig. 4).

### Glucocorticoid Receptor- $\beta$ Isoform

hGR $\beta$  functions as a dominant-negative inhibitor of hGR $\alpha$  activity and inhibits hGR $\alpha$ -mediated transactivation of many target genes in a dose-dependent manner. The mechanism(s) underlying this inhibition

has not been fully elucidated, but may involve competition between hGR $\alpha$  and hGR $\beta$  for binding to GREs, formation of hGR $\alpha$ –hGR $\beta$  heterodimers that are transcriptionally inactive, and/or titration or sequestration of coactivators needed by hGR $\alpha$  for full transcriptional activity. The ability of hGR $\beta$  to antagonize the function of hGR $\alpha$  suggests that hGR $\beta$  may play a critical role in regulating target tissue sensitivity to glucocorticoids. Increased expression of hGR $\beta$  has been documented in generalized and tissue-specific glucocorticoid resistance and leads to a reduction in the ability of hGR $\alpha$  to bind to GREs. Therefore, an imbalance in hGR $\alpha$  and hGR $\beta$  expression may underlie the pathogenesis of several clinical conditions associated with glucocorticoid resistance, such as rheumatoid arthritis, systemic lupus erythematosus, and ulcerative colitis.

### See Also the Following Articles

Glucocorticoid Resistance Syndromes and States • Glucocorticoids and Immunity • Glucocorticoids in Aging: Relevance to Cognition • Glucocorticoids, Overview • Growth and Glucocorticoids • Nuclear Factor- $\kappa$ B and Glucocorticoid Receptors

### Further Reading

Auboeuf, D., Honig, A., Berget, S. M., and O'Malley, B. W. (2002). Coordinate regulation of transcription and splicing by steroid receptor coregulators. *Science* 298, 416–419.

- Bamberger, C. M., Schulte, H. M., and Chrousos, G. P. (1996). Molecular determinants of glucocorticoid receptor function and tissue sensitivity to glucocorticoids. *Endocr. Rev.* **17**, 245–261.
- Carson-Jurica, M. A., Schrader, W. T., and O'Malley, B. W. (1990). Steroid receptor family: Structure and functions. *Endocr. Rev.* **11**, 201–220.
- Chrousos, G. P. (1995). The hypothalamic–pituitary–adrenal axis and immune-mediated inflammation. *N. Engl. J. Med.* **332**, 1351–1362.
- Clark, J. K., Schrader, W. T., and O'Malley, B. W. (1992). Mechanism of steroid hormones. In “Williams Textbook of Endocrinology” (J. D. Wilson and D. W. Foster, eds.), pp. 35–90. W. B. Saunders, Philadelphia, PA.
- Dalman, F. C., Scherrer, L. C., Taylor, L. P., Akil, H., and Pratt, W. B. (1991). Localization of the 90-kDa heat shock protein-binding site within the hormone-binding domain of the glucocorticoid receptor by peptide competition. *J. Biol. Chem.* **266**, 3482–3490.
- Hollenberg, S. M., Weinberger, C., Ong, E. S., Cerelli, G., Oro, A., Lebo, R., Thompson, E. B., Rosenfeld, M. G., and Evans, R. M. (1985). Primary structure and expression of a functional human glucocorticoid receptor cDNA. *Nature* **318**, 635–641.
- Kino, T., Vottero, A., Charmandari, E., and Chrousos, G. P. (2002). Familial/sporadic glucocorticoid resistance syndrome and hypertension. *Ann. N. Y. Acad. Sci.* **970**, 101–111.
- McKenna, N. J., and O'Malley, B. W. (2002). Combinatorial control of gene expression by nuclear receptors and coregulators. *Cell* **108**, 465–474.
- McKenna, N. J., Xu, J., Nawaz, Z., Tsai, S. Y., Tsai, M. J., and O'Malley, B. W. (1999). Nuclear receptor coactivators: Multiple enzymes, multiple complexes, multiple functions. *J. Steroid Biochem. Mol. Biol.* **69**, 3–12.
- Miura, T., Ouchida, R., Yoshikawa, N., Okamoto, K., Makino, Y., Nakamura, T., Morimoto, C., Makino, I., and Tanaka, H. (2001). Functional modulation of the glucocorticoid receptor and suppression of NF- $\kappa$ B-dependent transcription by ursodeoxycholic acid. *J. Biol. Chem.* **276**, 47371–47378.
- Rogatsky, I., Logan, S. K., and Garabedian, M. J. (1998). Antagonism of glucocorticoid receptor transcriptional activation by the c-Jun N-terminal kinase. *Proc. Natl. Acad. Sci. USA* **95**, 2050–2055.
- Williams, S. P., and Sigler, P. B. (1998). Atomic structure of progesterone complexed with its receptor. *Nature* **393**, 392–396.



# Glucocorticoid Resistance Syndromes and States

Alessandra Vottero

University of Parma, Italy, and NIH/National Institute of Child Health and Human Development, Bethesda, Maryland, United States

George P. Chrousos

NIH/National Institute of Child Health and Human Development, Bethesda, Maryland, United States

## Glossary

**glucocorticoids** Any of several hormones, including cortisol, corticosterone, and cortisone, released by the cortex of the adrenal gland.

**receptor** Molecule on the cell surface or inside the cell that recognizes a specific antigen or hormone.

**resistance** Degree of nonresponsiveness to a particular hormone or disease in the body. Also called insensitivity.

**G**lucocorticoids (GCs) are major effectors of basal and stress-related homeostasis, influencing cardiovascular function, regulating carbohydrate, protein, and fat metabolism, and modifying the immune/inflammatory response.

## INTRODUCTION

GCs activate the central nervous system and participate in the development and basal functions of several organs and systems. Glucocorticoids bring about these effects by binding to specific intracellular receptor proteins, called glucocorticoid receptors (GRs).

Abnormalities of steroid sensitivity can be divided into two major groups: resistance and hypersensitivity. Resistance syndromes have been described for all steroids; they can be transient or permanent, partial or complete, and compensated or noncompensated. Complete glucocorticoid resistance is incompatible with life because of severe neonatal respiratory distress syndrome, as demonstrated in an *in vivo* study in which GR<sup>-/-</sup> knockout mice died within a few hours after birth. On the other hand, when partial steroid resistance occurs, the hypothalamic–pituitary–adrenal (HPA) axis is reset at higher circulating levels of adrenocorticotrophic hormone (ACTH) and cortisol.

The increased ACTH levels compensate for the insensitivity of tissues to glucocorticoids, but also result in increased secretion of glucocorticoid precursors with mineralocorticoid activity [deoxycorticosterone (DOC) and corticosterone] and adrenal androgens.

Congenital glucocorticoid resistance has been described in over 10 kindreds and in a few sporadic cases and the molecular mechanisms of resistance in some of these patients have been analyzed.

## MOLECULAR ACTIONS OF GLUCOCORTICOIDS

At the cellular level, GCs carry out their actions through an ~94 kDa intracellular receptor protein, the GR. This receptor belongs to the nuclear receptor superfamily (Fig. 1) of steroid/thyroid/retinoic acid receptor proteins, which function as ligand-dependent transcription factors. Since 1985, when the sequence of GR cDNA was first published, two alternative splicing products of the same gene located on chromosome 5 have been described (Fig. 2): the classic, active receptor, GR- $\alpha$ , and the dominant-negative isoform, GR- $\beta$ . The two receptors are highly homologous, differing in just the last 50 and 15 amino acids, respectively. Therefore, the first eight exons of the GR gene containing the 5' noncoding and coding sequences are common to both receptor isoform cDNAs, whereas exons 9 $\alpha$  and 9 $\beta$  containing the coding and 3' noncoding sequences are specific for GR- $\alpha$  and GR- $\beta$ , respectively.

In the absence of the ligand, GR- $\alpha$  resides mostly in the cytoplasm of cells in a multiprotein complex consisting of the receptor polypeptide, two molecules of heat shock protein 90 (Hsp90), and several other proteins. The Hsp90s are thought to allow proper folding and stabilization of the receptor, maintaining the latter in a high-affinity, hormone-binding friendly

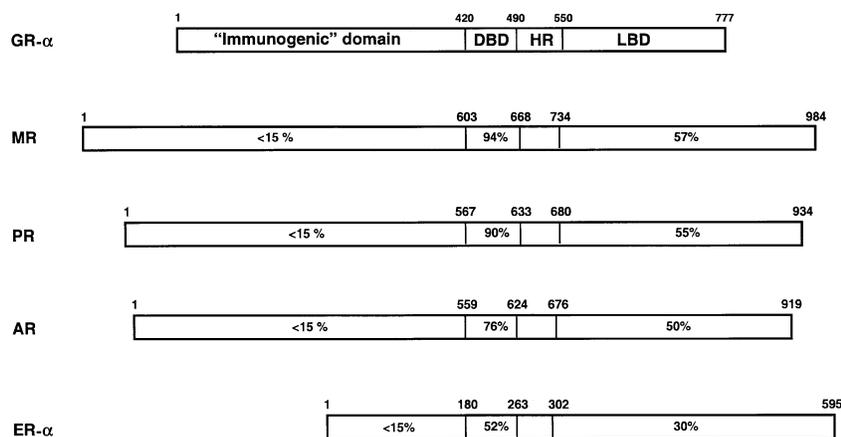


Figure 1 Members of the nuclear receptor superfamily and their major homologies.

state and preventing its interaction with proteins of the importin system, which are involved in the cytoplasm-to-nucleus translocation of many proteins. Experimental evidence supports a model of constant bidirectional shuttling of the complex in the unliganded state between the cytosol and the nucleus. Once the hormone binds, the receptor–ligand complex releases the Hsps and homodimerizes with another activated GR- $\alpha$ . This complex interacts with the importin system and translocates via the nuclear pore into the nucleus, where it positively or negatively regulates gene expression by binding as a homodimer with GC-response elements (GREs) in the presence of coregulators [CBP/p300 and SRC-1 (p160) families of “coactivators”; NcoR/RIP13 and SMRT/TRAC families of “corepressors”] or as a monomer via protein–protein interactions with other transcription factors (Fig. 3A). Through those transcription factors, including activator protein-1 (AP-1), nuclear factor  $\kappa$ B (NF- $\kappa$ B), and Stat-5, the GR modulates the transcription rates of non-GRE-containing genes regulated by these factors (Fig. 3B).

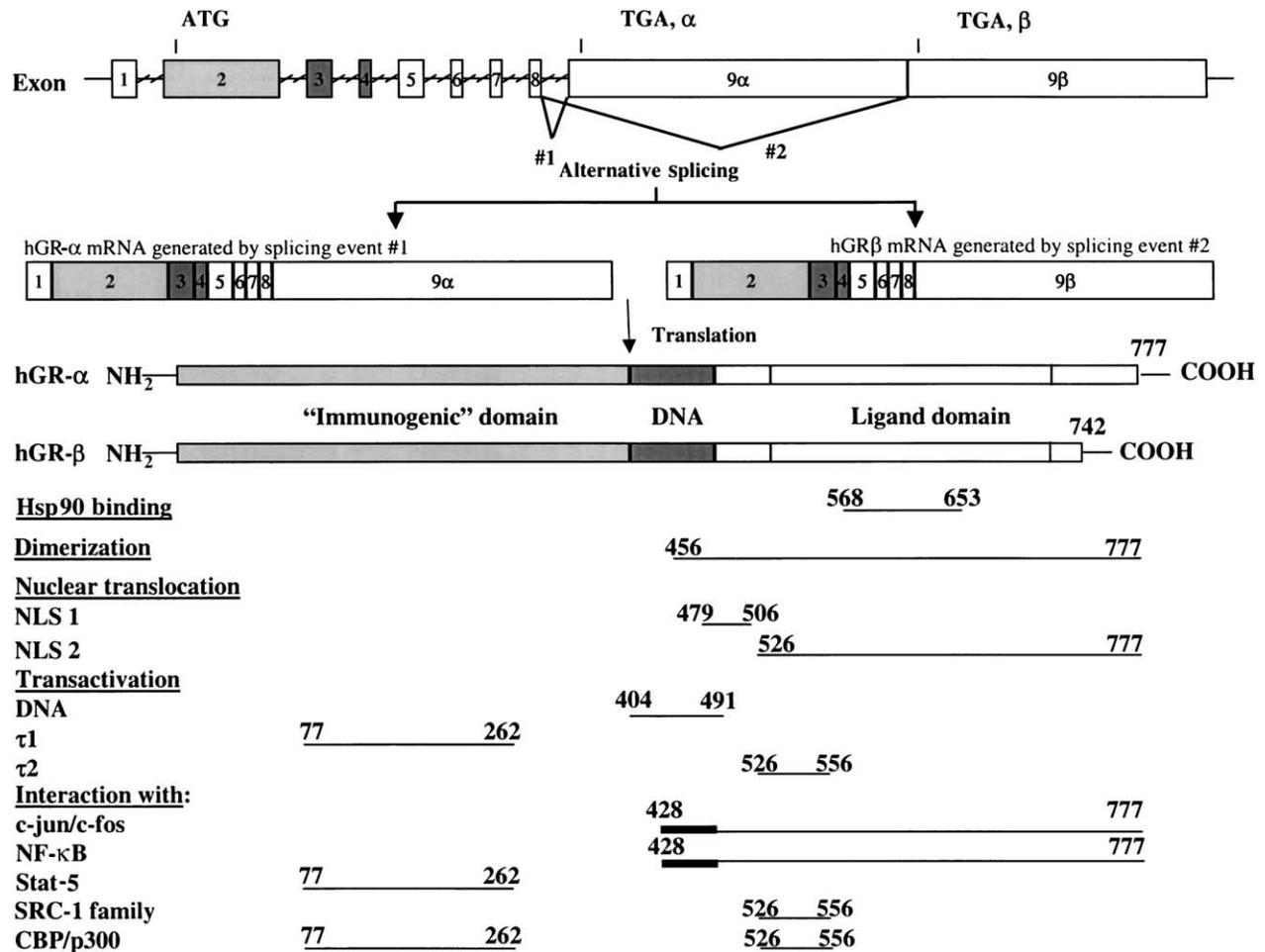
In contrast to GR- $\alpha$ , GR- $\beta$  does not bind to glucocorticoids and functions as weak dominant inhibitor of GR- $\alpha$ . This action is mediated by GRE binding, since no major protein–protein interaction with other transcription factors has been described. The intracellular localization of GR- $\beta$  is uncertain; green fluorescent protein–GR- $\beta$  fusion hybrids have been found in the nucleus of transfected cells, confirming the potential presence of GR- $\beta$  in the nucleus, at a site other than the cytoplasm, under certain conditions even in the absence of ligand. The export of GR- $\beta$  from the nucleus is much slower than that of GR- $\alpha$ , which could explain its accumulation in the nucleus.

## MOLECULAR MECHANISMS OF GLUCOCORTICOID RESISTANCE

The molecular basis of glucocorticoid resistance has been ascribed to mutations in the GR- $\alpha$  gene that impair one or more of the molecular mechanisms of GR function, thus altering tissue sensitivity to GCs. Inactivating mutations within the DNA- and ligand-binding domains, as well as a 4 bp mutation at the 3' boundary of the GR- $\alpha$  gene, have been described in five kindreds and three sporadic cases (Table I).

In 1976, Vingerhoeds *et al.* described the first two patients, a father and a son, with long-term “hypercortisolism” not associated with clinical manifestations of Cushing’s syndrome. In 1982, Chrousos *et al.* demonstrated that these patients had abnormal glucocorticoid receptor properties and suggested that they suffered from glucocorticoid resistance. In 1991, this group described the molecular mechanism underlying the disease in this family: the proband was a homozygote for a single nonconservative point mutation, replacing aspartate at amino acid position 641 with valine, in the hormone-binding domain of the GR. This mutation reduced glucocorticoid receptor-binding affinity for dexamethasone by threefold and caused loss of transactivation activity on the mouse mammary tumor virus (MMTV) promoter.

The second family was described in 1993 by the Chrousos group as well; the proband of this family was a young woman with manifestations of hyperandrogenism. Molecular analysis showed a 4 bp deletion at the 3' boundary of exon 6, removing a donor splice site. This resulted in complete ablation of expression of one of the GR alleles associated with a 50% decrease in GR protein in the affected members of the family.



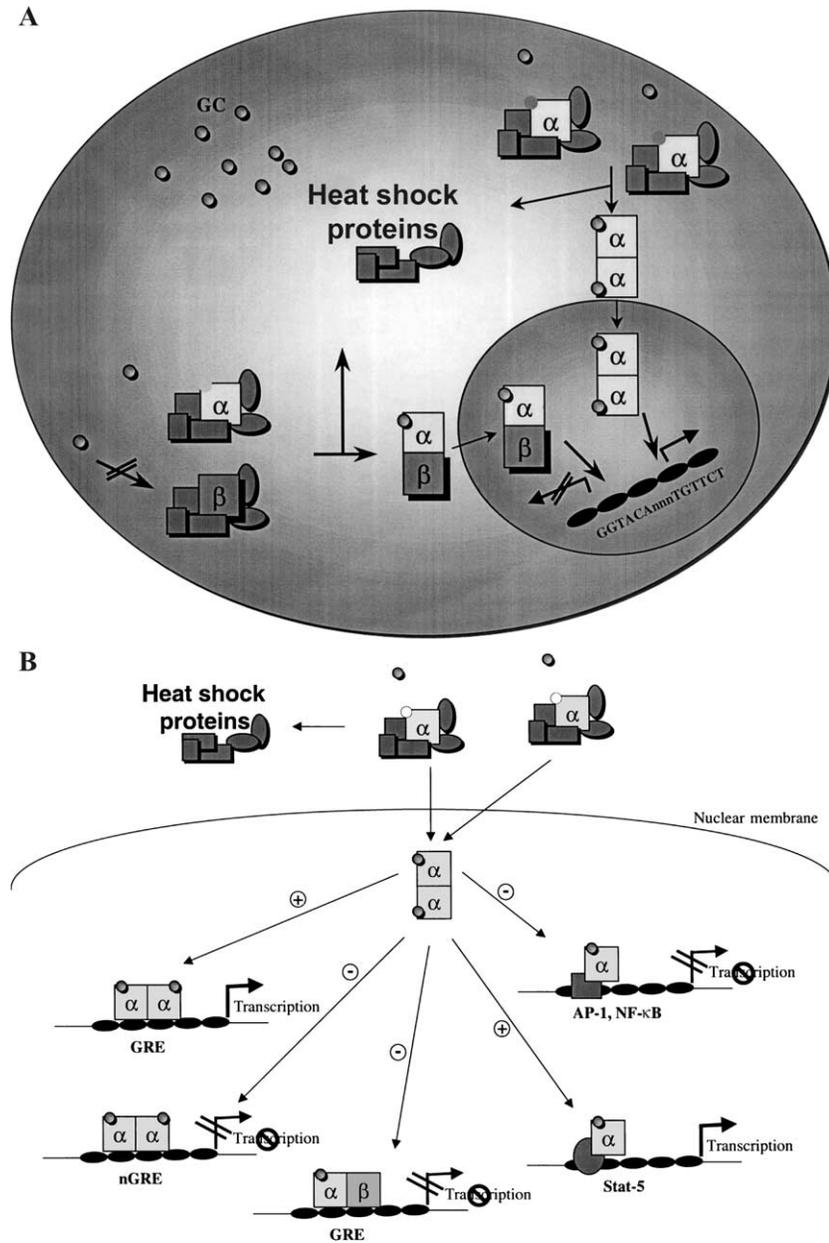
**Figure 2** Structure of the hGR gene and gene products. Alternative splicing events, No. 1 (default splicing pathway) and No. 2 (alternative splicing pathway), generate two different hGR messages, which differ in size due to the use of alternative polyadenylation signals. Translation of the messages produces two isoforms of the glucocorticoid receptor, hGR-α and hGR-β, the structures of which are identical through amino acid 727 but then diverge. hGR-α, 777 amino acids long, has a molecular weight of approximately 98 kDa, whereas hGR-β, 742 amino acids long, has a molecular weight of approximately 94 kDa. Functional domains and their putative sites for these functions are indicated below the linearized GR protein. Boxes and lines represent exons and introns, respectively.

The propositus of the third kindred, a boy with precocious puberty, had a single homozygotic point mutation at amino acid 729 (valine to isoleucine) in the hormone-binding domain, which reduced both the affinity and the transactivation activity of the GR.

In 1996, another interesting, sporadic case of a man with a history of infertility, hypertension, and 5- to 10-fold elevation of urinary free-cortisol levels was described by the Chrousos group. This patient had a *de novo*, germ-line heterozygotic mutation at amino acid 559 (isoleucine to asparagine) in the hormone-binding domain, at the hinge region of the GR. This receptor had a 2- to 3-fold more potent dominant-negative activity than GR-β on the wild-type receptor

and was expressed at a 1:1 ratio with the normal GR-α in the patient’s cells. Interestingly, this mutant receptor prevented translocation of the normal receptor into the nucleus, an effect that would be overcome at very high dexamethasone concentrations. Later this patient developed severe Cushing’s disease due to an ACTH-secreting pituitary adenoma.

A fifth case/kindred with glucocorticoid resistance was studied; the proband was a young woman with signs of hyperandrogenism but no other complaint. Molecular analysis detected a heterozygotic T-to-G substitution at nucleotide 2373 of exon 9α of the GR, in the ligand-binding domain at amino acid position 747, replacing isoleucine with methionine. This



**Figure 3** (A) Putative mechanisms of action of hGR- $\alpha$  and hGR- $\beta$ . In the unliganded state, the classic glucocorticoid receptor- $\alpha$  resides predominantly in the cytoplasm as part of a heteromeric complex including at least five molecules of heat shock proteins. After binding to the hormone, the GR- $\alpha$  molecule changes its conformation and is released by the heat shock proteins. It thus homodimerizes with another hormone-activated receptor molecule and translocates to the cell nucleus, where it can regulate the transcriptional activity of target genes. hGR- $\beta$  is unable to bind glucocorticoids and is transcriptionally inactive, but may have a cell-specific dominant-negative effect on GR- $\alpha$ , primarily on GRE-mediated actions. The GRE consensus sequence (15 bp) is shown. (B) Nuclear actions of GR- $\alpha$  and GR- $\beta$ . After entering the nucleus, ligand-bound GR- $\alpha$  influences the transcription of target genes by different mechanisms: as a homodimer, it may activate transcription by binding to GRE; as a heterodimer with GR- $\beta$ , it may have a diminished ability to transactivate a GRE-containing gene and, hence, may act as a dominant-negative inhibitor. Binding of a GR- $\alpha$  homodimer to a negative GRE (nGRE) may also lead to repression; a GR- $\alpha$ -GR- $\beta$  heterodimer may lose the ability to repress a nGRE. Monomeric GR- $\alpha$  interacts with other transcription factors, such as AP-1 and NF- $\kappa$ B, and prevents them from carrying out their activities, as their target genes contain AP-1 and NF- $\kappa$ B sites, respectively. In a protein-protein interaction similar to that occurring between GR- $\alpha$  and Stat-5, GR- $\beta$  acts synergistically with Stat-5 in the activation of Stat5-responsive genes. There is no evidence that GR- $\beta$  interacts with or influences the activity of AP-1 or NF- $\kappa$ B.

Table I Glucocorticoid Resistance: Molecular Findings

cDNA position	Amino acid position	Mutation	Domain	Reference
2054 (exon 7)	641 <b>D</b> → <b>V</b>	A-to-T mutation (nonconservative) (new site for <i>Hae</i> II)	Hormone binding	Hurley <i>et al.</i> (1991)
2317 (exon 9)	729 <b>V</b> → <b>I</b>	G-to-A mutation	Hormone binding	Malchoff <i>et al.</i> (1993)
1220 (exon 2)	363 <b>A</b> → <b>S</b> (conservative change)	A-to-G mutation (conservative)	Amino-terminal	Karl <i>et al.</i> (1993)
	4 bp deletion (exon 6–intron 6)	Deletion = no splice in 1 allele (segregating)	Hormone binding	Karl <i>et al.</i> (1993)
1808 (exon 5)	559 <b>I</b> → <b>N</b>	T-to-A mutation	Hormone binding	Karl <i>et al.</i> (1996)
1430 (exon 4)	477 <b>R</b> → <b>H</b>	G-to-A mutation	DNA binding	Ruiz <i>et al.</i> (2001)
2035 (exon 8)	679 <b>G</b> → <b>S</b>	G-to-A mutation	Hormone binding	Ruiz <i>et al.</i> (2001)
1844 (exon 5)	571 <b>V</b> → <b>A</b>	C-to-T mutation	Hormone binding	Mendonca <i>et al.</i> (2002)
2373 (exon 9)	747 <b>I</b> → <b>M</b>	T-to-G mutation	Hormone binding	Vottero <i>et al.</i> (2002)

mutation was located close to helix 12 at the C-terminus of the ligand-binding domain, which plays a pivotal role in the formation of activation function-2, a subdomain that interacts with p160 coactivators. The mutant receptor had an approximately 2-fold reduced affinity for dexamethasone and its transcriptional activity on the glucocorticoid-responsive MMTV promoter was compromised by 20- to 30-fold; interestingly, it also had dominant-negative activity on the wild-type receptor, probably secondary to the defective interaction of the mutant receptor with p160 coactivators.

Two novel mutations in the GR gene were found in two unrelated patients with primary cortisol resistance as defined by a pathologic dexamethasone suppression test: R477H in the DNA-binding domain, which is the first reported mutation in that region of the human GR gene, and G679S in the ligand-binding domain. The R477H mutation showed no transactivating capacity, whereas the G679S mutation had reduced transactivation capacity and 50% binding affinity compared to the wild-type GR.

A new phenotype, female pseudo-hermaphroditism and severe hypokalemia, caused by a novel inactivating mutation of the GR gene has been described. A homozygous T-to-C substitution at nucleotide 1844 in exon 5 of the GR gene, which caused a valine to alanine substitution at amino acid 571 in the ligand-binding domain of the receptor, was identified

in the patient. This phenotype indicated that both pre- and postnatal virilization can occur in females with the glucocorticoid resistance syndrome.

The molecular mechanisms of action of some of the natural glucocorticoid receptor mutants causing familial glucocorticoid resistance have been studied. In particular, each of the mutations analyzed imparted different functional defects on the GR signal transduction pathway, which might partly explain the variable clinical phenotype of generalized glucocorticoid resistance. This variability includes patients with no abnormalities in the coding sequences and intron–exon boundaries of their GR genes.

The importance of GR- $\beta$  under physiologic conditions is controversial, but it has been proposed that in pathologic situations, such as glucocorticoid resistance, it might play a pathophysiologic role. Taking into consideration the finding that increasing amounts of GR- $\beta$  produce a dose-dependent decrease in wild-type GR- $\alpha$  transcriptional activity, an imbalance in the expression of these two isoforms might determine an altered sensitivity to glucocorticoids. Supporting a possible role of GR- $\beta$  in glucocorticoid sensitivity, a genetically determined imbalance in the expression of the glucocorticoid receptor isoforms was observed in cultured lymphocytes from a patient with congenital generalized glucocorticoid resistance and chronic leukemia; in this patient, a low GR- $\alpha$  to GR- $\beta$  ratio was found compared to a group of normal controls,

possibly explaining the glucocorticoid resistance since no abnormalities in the sequence of the entire cDNA or in individual exons of this patient's gene were found. In keeping with these findings, a significantly higher number of GR- $\beta$  immunoreactive cells were found in peripheral blood and bronchial lavage cells from glucocorticoid-resistant asthma type 1 patients than from glucocorticoid-sensitive asthmatic patients and normal controls.

Animal models of systemic glucocorticoid resistance, such as New World primates, including squirrel monkeys, marmosets, and owl monkeys, have been described. These animals have total plasma cortisol levels that are 7–20 times higher than in humans or other Old World primates, whereas the concentration, affinity, and predicted amino acid sequence of their GRs are similar to those of the human receptor. Interestingly, these animals exhibit resistance to a variety of other steroid/sterol hormones, including estrogens, progesterone, androgens, aldosterone, and vitamin D. Immunoreactivity of both isoforms of the GR has been found in Epstein-Barr virus-transformed B lymphocytes from marmosets, with the  $\beta$ -isoform being  $\sim$ 10 times overexpressed compared to the corresponding human cells. An altered splicing pattern of the GR pre-mRNA or differential rates of mRNA translation, mRNA degradation, and/or GR protein degradation might contribute to the steroid resistance of these animals. Alternatively, these animals may have decreased coactivator activity and/or increased corepressor activity, leading to their "pansteroid" resistance.

Two sisters with manifestations of glucocorticoid resistance were described by New *et al.* Their evaluation revealed resistance not only to glucocorticoids, but also to mineralocorticoids and androgens; however, they displayed no resistance to vitamin D or thyroid hormones. The diagnosis of these patients was multiple, partial steroid resistance. The New World primate physiologic and biochemical syndrome and the two pathologic human multiple steroid resistance syndrome cases are the first conditions in which a defective steroid receptor coregulator has been suggested to be responsible for an altered clinical and/or biochemical picture.

## CLINICAL CHARACTERISTICS OF GLUCOCORTICOID RESISTANCE

Patients with glucocorticoid resistance syndrome have compensatory elevations in circulating cortisol and ACTH concentrations, which maintain circadian

rhythmicity and appropriate responsiveness to stressors, albeit at higher hormone concentrations, and resistance of the HPA axis to dexamethasone suppression, but no clinical evidence of hypo- or hypercortisolism. The excess of ACTH results in increased production of adrenal steroids with mineralocorticoid and/or androgenic activity.

The clinical spectrum of this disease is quite broad, ranging from completely asymptomatic to mild to severe symptomatic conditions. A large number of subjects may be asymptomatic or display biochemical alterations only. Clinical manifestations due to the excess of mineralocorticoids (including cortisol itself), acting on the intact mineralocorticoid receptor of the patients, are hypertension with or without hypokalemic alkalosis. The increased amounts of androgens, on the other hand, lead to acne, hirsutism, male pattern baldness, menstrual irregularities (oligo-amenorrhea), oligo-anovulation, and infertility in women. In children, early and excessive prepubertal adrenal androgen secretion has been associated with ambiguous genitalia and precocious puberty. In adult men, oligospermia and infertility have been observed, possibly as the result of interference with follicle-stimulating hormone feedback regulation by the excessive adrenal androgens or by the ACTH-induced intratesticular growth of adrenal rests, which may occur as they do in classic and "late-onset" congenital adrenal hyperplasia. Because of the excessive secretion of adrenal androgens, bone mass density is usually high-normal to elevated in patients with glucocorticoid resistance, in contrast to patients with Cushing's syndrome, in whom osteoporosis is observed.

## DIAGNOSIS

The hallmark of the diagnostic evaluation of glucocorticoid resistance is increased serum cortisol and urinary free-cortisol levels without Cushing's syndrome clinical stigmata. Despite cortisol excess, plasma ACTH concentration is normal or high. The circadian rhythm of cortisol and its responsiveness to stress are intact in patients with glucocorticoid resistance who are also resistant to single or multiple doses of dexamethasone. Differential diagnosis includes (1) early or mild forms of Cushing's syndrome; (2) pseudo-Cushing states, such as generalized anxiety disorder, melancholic depression, and/or chronic active alcoholism, conditions associated with obesity and increased cortisol secretion; (3) other causes of mineralocorticoid-induced hypertension; and (4) hirsutism (including idiopathic hirsutism), polycystic

ovary syndrome, and late-onset congenital adrenal hyperplasia.

## THERAPY OF GLUCOCORTICOID RESISTANCE SYNDROME

Asymptomatic, normotensive subjects with primary glucocorticoid resistance do not require any treatment. In contrast, patients with symptomatic generalized glucocorticoid resistance are treated with high, individualized doses of oral dexamethasone (0.5–1.0 mg two or three times daily), a synthetic, potent glucocorticoid with minimal intrinsic mineralocorticoid activity. The goal is to suppress ACTH and, therefore, endogenous cortisol, DOC, corticosterone, and adrenal androgen secretion, correcting the states of mineralocorticoid and androgen excess in these patients.

Hypertensive patients should receive the smallest dose that lowers the serum concentration of cortisol and other mineralocorticoids and corrects electrolyte abnormalities. Hirsute patients should be treated with doses able to reduce the androgen excess.

Untreated patients have no risk of adrenal insufficiency and do not need extra doses of dexamethasone in particularly stressful situations, such as surgery and illness. In contrast, patients undergoing chronic treatment should receive the appropriate glucocorticoid coverage.

### See Also the Following Articles

ACTH (Adrenocorticotropic Hormone) • Glucocorticoid Receptor • Glucocorticoids and Immunity • Glucocorticoids in Aging: Relevance to Cognition • Glucocorticoids, Overview • Growth and Glucocorticoids • Hypercorticism and Cushing's Syndrome

### Further Reading

- Arai, K., and Chrousos, G. P. (1995). Syndromes of glucocorticoid and mineralocorticoid resistance. *Steroids* **60**, 173–179.
- Bronnegard, M., Werner, S., and Gustafsson, J. A. (1986). Primary cortisol resistance associated with a thermolabile glucocorticoid receptor in a patient with fatigue as the only symptom. *J. Clin. Invest.* **78**, 1270–1278.
- de Castro, M., and Chrousos, G. P. (1997). Glucocorticoid resistance. *Curr. Ther. Endocrinol. Metab.* **6**, 188–189.
- Hurley, D. M., Accili, D., Stratakis, C. A., Karl, M., Vamvakopoulos, N., Rorer, E., Constantine, K., Taylor, S. I., and Chrousos, G. P. (1991). Point mutation causing a single amino acid substitution in the hormone binding domain of the glucocorticoid receptor in familial glucocorticoid resistance. *J. Clin. Invest.* **87**, 680–686.
- Iida, S., Gomi, M., Moriwaki, K., Itoh, Y., Hirobe, K., Matsuzawa, Y., Katagiri, S., Yonezawa, T., and Tarui, S. (1985). Primary cortisol resistance accompanied by a reduction in glucocorticoid receptors in two members of the same family. *J. Clin. Endocrinol. Metab.* **60**, 967–971.
- Jenster, G. (1998). Coactivators and corepressors as mediators of nuclear receptor function: An update. *Mol. Cell. Endocrinol.* **143**, 1–7.
- Karl, M., Lamberts, S. W., Detera-Wadleigh, S. D., Encio, I. J., Stratakis, C. A., Hurley, D. M., Accili, D., and Chrousos, G. P. (1993). Familial glucocorticoid resistance caused by a splice site deletion in the human glucocorticoid receptor gene. *J. Clin. Endocrinol. Metab.* **76**, 683–689.
- Karl, M., Lamberts, S. W., Koper, J. W., Katz, D. A., Huizenga, N. E., Kino, T., Haddad, B. R., Hughes, M. R., and Chrousos, G. P. (1996). Cushing's disease preceded by generalized glucocorticoid resistance: Clinical consequences of a novel, dominant-negative glucocorticoid receptor mutation. *Proc. Assoc. Am. Physicians* **108**, 296–307.
- Kino, T., Stauber, R. H., Resau, J. H., Pavlakis, G. N., and Chrousos, G. P. (2001). Pathologic human GR mutant has a transdominant negative effect on the wild-type GR by inhibiting its translocation into the nucleus: Importance of the ligand-binding domain for intracellular GR trafficking. *J. Clin. Endocrinol. Metab.* **86**, 5600–5608.
- Lamberts, S. W., Koper, J. W., Biemond, P., den Holder, F. H., and de Jong, F. H. (1992). Cortisol receptor resistance: The variability of its clinical presentation and response to treatment. *J. Clin. Endocrinol. Metab.* **74**, 313–321.
- Malchoff, C. D., Javier, E. C., Malchoff, D. M., Martin, T., Rogol, A., Brandon, D., Loriaux, D. L., and Reardon, G. E. (1990). Primary cortisol resistance presenting as isosexual precocity. *J. Clin. Endocrinol. Metab.* **70**, 503–507.
- Malchoff, D. M., Brufsky, A., Reardon, G., McDermott, P., Javier, E. C., Bergh, C. H., Rowe, D., and Malchoff, C. D. (1993). A mutation of the glucocorticoid receptor in primary cortisol resistance. *J. Clin. Invest.* **91**, 1918–1925.
- Mendonca, B. B., Leite, M. V., De Castro, M., Kino, T., Elias, L. L. K., Bachega, T. A. S., Arnhold, I. J. P., Chrousos, G. P., Lewicka, S., and Latronico, A. C. (2002). Female pseudohermaphroditism caused by a novel homozygous missense mutation of the GR gene. *J. Clin. Endocrinol. Metab.* **87**, 1805–1809.
- Nawata, H., Sekiya, K., Higuchi, K., Kato, K., and Ibayashi, H. (1987). Decreased deoxyribonucleic acid binding of glucocorticoid-receptor complex in cultured skin fibroblasts from a patient with the glucocorticoid resistance syndrome. *J. Clin. Endocrinol. Metab.* **65**, 219–226.
- Ruiz, M., Lind, U., Gafvels, M., Eggertsen, G., Carlstedt-Duke, J., Nilsson, L., Holtmann, M., Stierna, P., Wikstrom, A. C., and Werner, S. (2001). Characterization of two novel mutations in the glucocorticoid receptor gene in patients with primary cortisol resistance. *Clin. Endocrinol.* **55**, 363–371.

administration. Also, glucocorticoids decrease in a timely manner the expression of L-selectin on bone marrow progenitors and on differentiating neutrophils that, then, reach the bloodstream with low L-selectin levels. Chemokines released from the inflamed tissue are important for activating neutrophils and initiating the program of cell transmigration. The inhibition of the release of chemokines, such as interleukin-8 (IL-8) and other cysteine-X-cysteine (CXC) chemokines, also participates in the inhibition of neutrophil transmigration by glucocorticoids.

Eosinophils, basophils, and T helper 2 (Th2) cells are recruited by chemokines, which are released by immune and nonimmune cells, such as epithelial cells or airway smooth muscle cells. These  $\beta$ -chemokines, or cysteine-cysteine (CC) chemokines, enhance adhesion molecule expression on endothelial cells to allow eosinophils and basophils to roll through the vessel wall. In fact, glucocorticoids prevent tissue invasion of eosinophils by inhibiting the release of CC chemokines, such as eotaxin, eotaxin 2, MCP4 (macrophage inflammatory protein 4), and RANTES (regulated on activated normal T-cell expressed and secreted), released by bronchial epithelial cells. Simultaneously, whereas glucocorticoids decrease the expression of some chemokine receptors, they enhance the expression of others on eosinophils, further complicating the overall picture of eosinophil trafficking.

The flow and movement of monocytes seem tightly regulated by glucocorticoids via similar mechanisms, namely, through the repression of monocytes and endothelial adhesion molecules and the regulation of chemokines and their receptors.

### Innate Immune Response

Glucocorticoids act on the immune system by both suppressing and stimulating a large number of pro-inflammatory or anti-inflammatory mediators. In many ways, glucocorticoids lead to the termination of inflammation by enhancing opsonization and the activity of scavenger systems and by stimulating macrophage phagocytotic ability and antigen uptake. Glucocorticoids stimulate the expression of the mannose receptor or the scavenger receptor CD163, promoting the clearance of microorganisms, dead cell bodies, and antigens. Moreover, they potentiate interferon- $\gamma$  (IFN- $\gamma$ )-induced Fc $\gamma$ RI. At the same time, they prevent inflammation from overshooting, by suppressing the synthesis of many inflammatory mediators, such as several cytokines and chemokines, prostaglandins, leukotrienes, proteolytic enzymes, free oxygen radicals, and nitric oxide.

A large number of cytokines [including IL-1 $\beta$ , tumor necrosis factor  $\alpha$  (TNF $\alpha$ ), IL-6, IL-8, IL-12, and IL-18] are broadly down-regulated by glucocorticoids. Similarly, the secretion of many CC and CXC chemokines is strongly suppressed. Interestingly, anti-inflammatory cytokines such as IL-10 and transforming growth factor- $\beta$  (and to some extent, although it is a matter of some debate, IL-1RA, the IL1 receptor antagonist) are up-regulated by glucocorticoids. Although the negative regulation of pro-inflammatory cytokines by glucocorticoids has been clearly demonstrated both *in vitro* and *in vivo*, glucocorticoids enhance the receptor of some of these pro-inflammatory cytokines and chemokines. It is tempting to speculate that they increase receptor expression to improve the sensitivity of target immune cells to these mediators and help these cells to resolve inflammation. Soluble or decoy receptors inhibiting or further enhancing the inflammatory process are also regulated by glucocorticoids. For example, the decoy receptor IL-1RII, which binds IL-1 without driving its signaling, is enhanced by glucocorticoids. This represents an anti-inflammatory mechanism of action of glucocorticoids. Inversely, the soluble IL-6R is increased by glucocorticoids and further amplifies the pro- or anti-inflammatory response to interleukin-6.

An intriguing inflammatory mediator is the macrophage inhibitory factor (MIF), whose secretion is triggered by glucocorticoids. Synthesized by anterior pituitary cells and macrophages in response to endotoxin challenge and pro-inflammatory cytokines such as TNF $\alpha$ , MIF exerts a major pro-inflammatory effect on macrophages and T cells and overrides the anti-inflammatory and immunosuppressive actions of glucocorticoids. In fact, MIF increases lethality, whereas genetic deletion or therapeutic neutralization confers protection against endotoxemia, acute distress respiratory syndrome, and septic shock. The finding that Toll-like receptor (TLR) 4 expression is increased by MIF underscores the essential role of MIF in the macrophage response to endotoxins and gram-negative bacteria. This indeed may explain why MIF-deficient mice were hyporesponsive to lipopolysaccharide (LPS). Thus, the notion that glucocorticoids enhance the secretion of a major pro-inflammatory cytokine that counteracts their effects represents a yin/yang mechanism of control of the acute-phase response or septic shock.

By inhibiting phospholipase A2, inducible cyclooxygenase 2 (COX2), and inducible prostaglandin (PG) synthase 2 (PGS2), glucocorticoids block the release of arachidonic acid and the synthesis of prostaglandins (PGE2, PGH2, and PGD2) and platelet-activating factor; both constitutive COX1 and PGS1 remain

unaffected by glucocorticoids. Several lines of evidence suggest that glucocorticoids prevent COX2 expression through both posttranscriptional and post-translational mechanisms. Strikingly, glucocorticoids could enhance 5-lipoxygenase and 5-lipoxygenase-activating protein expression in monocytes and eosinophils. This is rather surprising given the overall reduced secretion of leukotrienes after glucocorticoid exposure. It seems that the decreased production of leukotrienes is related to the inhibition of phospholipase A2 expression and activity by glucocorticoids. Glucocorticoids may also suppress the reduced form of nicotinamide adenine dinucleotide phosphate oxidase and superoxide dismutase expression and, hence, the production of free oxygen radicals. Nitric oxide appears to be an important intra- and intercellular signaling molecule in shaping the innate and adaptive immune response with both detrimental and protective effects. Glucocorticoids suppress inducible nitric oxide synthetase expression, which results in a decrease in nitric oxide release by endothelial cells. This inhibition also seems to be mediated by the glucocorticoid second-messenger lipocortin 1 and would prevent early endothelial cell-mediated inflammatory reaction.

## ADAPTIVE IMMUNITY

### Antigen Presentation and Adaptive Immune Response

Dendritic cells (DCs) represent the crucial interplay between innate and adaptive immunity. On encountering microorganisms or antigens, tissue-resident DCs rapidly differentiate and migrate to secondary lymphoid organs. These immature DCs demonstrate increased antigen uptake ability and specialized antigen-processing machinery. Glucocorticoids potentiate this "immature" phenotype. They improve opsonization and the activity of scavenger systems and stimulate macrophage phagocytosis, pinocytotic ability, and antigen uptake. TLRs mediate patterns of microorganism antigen recognition in macrophages, immature DCs, and mature DCs. Interestingly, glucocorticoids modulate the expression of TLR4, the LPS receptor, emphasizing the crucial role of glucocorticoids early in the innate immune response.

During their migration, DCs mature and express major histocompatibility complex (MHC) class II and costimulatory molecules to efficiently present the antigen, as professional APCs (antigen-presenting cells), to naive or memory T cells. Crosstalk between

T cells and dendritic cells through TCR (T-cell receptor)/MHC II-bound antigen, costimulatory molecules, and cytokines allows the development of a T cell immune response and T cell expansion or deletion. When exposed to glucocorticoids during maturation, these DCs have a decreased ability to present antigens and elicit a T cell response, primarily because glucocorticoids prevent MHC class II up-regulation and the expression of the costimulatory molecules, such as B7.2 (CD86) and to some extent B7.1 (CD80), CD40, and the ICAM-1/LFA-1 (intercellular adhesion molecule-1/lymphocyte function associated antigen-1) complex. It is noteworthy that terminally differentiated DCs continue to express these molecules due to their relative resistance to glucocorticoids. The timing of exposure to glucocorticoids thus appears to be essential during dendritic cell maturation.

Once a T cell immune response has flared up, glucocorticoids may modulate and interfere with the type of T cells involved. The differentiation of CD4<sup>+</sup> T cells into T helper 1 (Th1) lymphocytes, which drive cellular immunity, or into Th2 lymphocytes, which drive humoral immunity, depends on the type of antigen encountered and the type of cytokines produced during antigen presentation. Indeed, glucocorticoids block IL-12 secretion by monocytes and DCs. Interleukin-12 is the link between innate and cellular immunity and is crucial for the development of the Th1-directed cellular immune response. On the other hand, glucocorticoids promote Th2 development by enhancing IL-10 secretion by macrophages and immature DCs. In fact, several studies clearly showed, in both mice and humans, that the presence of glucocorticoids during the primary immune response, associated with defective IL-12 production by macrophages or DCs, enhanced Th2 cytokine secretion and decreased Th1 cytokine secretion by CD4<sup>+</sup> lymphocytes on secondary stimulation. This effect seemed IL-12-dependent since it was reversed by the addition of exogenous IL-12 to glucocorticoid-treated APCs during the primary stimulation. Moreover, the poor glucocorticoid regulation of costimulatory molecules in mature DCs ruled out any role for costimulation in this process. Terminally differentiated DCs have been classified into DC1 and DC2, depending on their cytokine secretion profile and their aptitude to force the differentiation of naive T cells into Th1 and Th2 mature lymphocytes. Thus, DC phenotypes seem to be twisted by glucocorticoids into a DC2 program that will ultimately generate a Th2 immune response and the secretion of Th2 cytokines.

## Cellular Immune Response

Interleukin-12 is required for Th1 lymphocyte differentiation and secretion of Th1 cytokines, such as IFN- $\gamma$  and TNF $\alpha$ . Alterations in IL-12 signaling, such as those observed in IL-12<sup>-/-</sup> and IL-12R-deficient mice, are associated with a defective Th1 immune response. In fact, glucocorticoids not only block IL-12 secretion by monocytes/macrophages and DCs, they also suppress IL-12R $\beta$ 1 and IL-12R $\beta$ 2 expression on T cells. They also interfere with IL-12 signaling and prevent IL-12-induced signal transducer and activator of transcription 4 (Stat4) phosphorylation and Stat4-dependent gene expression such as that for interferon regulatory factor-1. Blocking Stat4 activation mimics the situation in Stat4-deficient mice, which are unable to elicit a Th1 immune response. Glucocorticoids also profoundly suppress secretion of the Th1 cytokines IFN- $\gamma$  and TNF $\alpha$ , lessening NK and T cytotoxic effector functions. Such massive inhibition of the Th1 immune response by glucocorticoids led to severe cellular immunodeficiency and impaired defense against intracellular and opportunistic infections.

## Humoral Immune Response

Whereas glucocorticoids profoundly decrease the secretion of Th2 cytokines during primary stimulation, they promote Th2 differentiation during secondary stimulation. Independent of monocytes/macrophages and DCs, glucocorticoids prime naive T cells to Th2 commitment during secondary immune response. Glucocorticoids that are present during primary stimulation promote IL-10 secretion during secondary stimulation in both naive and memory T cells. Sequential contact with glucocorticoids during the primary and/or secondary immune response may thus influence the pattern of Th2 cytokines. This context-dependent action of glucocorticoids may explain some discrepancies reported in the literature on the regulation of Th2 cytokines by glucocorticoids. Very interestingly, by up-regulating IL-10 secretion by macrophage/DCs and Th2 cells, glucocorticoids may participate in the emergence of regulatory T cells, high-IL-10-producing T cells with major *in vivo* immunoregulatory properties as shown in experimental allergic encephalomyelitis.

Following exposure to endogenous or exogenous glucocorticoids, a progressive shift takes place from a cellular Th1 immune response to a humoral Th2 immune response. A major question remains regarding

how glucocorticoids induce Th2 lymphocyte differentiation and the humoral immune response. First, it is well recognized that the lack of commitment in developing a Th1 immune response, namely, the absence of IL-12, is associated with Th2 development. Therefore, the profound suppression of Th1 differentiation by glucocorticoids may undoubtedly participate in Th2 immune response development. Second, glucocorticoids also have differential action on IL-12 and IL-4 signaling. Interleukin-12 and IL-4 activate Stat4 and Stat 6, respectively. Deletion of Stat4 or Stat6 in mice is associated with poor Th1 and Th2 immune responses, respectively. In fact, glucocorticoids block IL-12-mediated Stat4 activation without altering IL-4-induced Stat6 phosphorylation and, hence, help a Th2 immune response to develop. Finally, a growing number of transcription factors have been shown to play a critical role in the hierarchical control of Th1 and Th2 lymphocyte differentiation. Thus, by acting directly on either lymphocytes or DCs, endogenous glucocorticoid hypersecretion or administration of excessive amounts causes a progressive shift from a Th1-directed cellular immune response to a Th2-driven humoral immune response.

It is undeniable that glucocorticoids favor a humoral immune response and antibody production. *In vivo* administration of glucocorticoids raises immunoglobulin E (IgE) serum levels in asthma or atopic patients. Despite the paucity of information, glucocorticoids modulate B cell development. They restrain B cell proliferation and early steps in B cell development but promote the generation of antibody-secreting plasma cells and the secretion of IgE and IgG4. Like DCs and effector T cells, B cells become resistant to the inhibitory actions of glucocorticoids as they proceed through the different stages of differentiation and maturation. Surprisingly, the question of how these hormones influence B cell receptor signaling compared to TCR signaling has never been explored, underscoring the lack of interest in glucocorticoid-mediated actions on B cells. Interleukin-4 is the critical cytokine that induces Th2 differentiation and promotes B cell differentiation and IgE isotype switching. Interestingly, glucocorticoids act in synergy with IL-4 in B cell differentiation and isotype switching, leading to IgE-secreting B cells. Such IgE isotype switching is dependent on the CD40/CD40L complex since it is not observed in X-linked hyper-IgM (CD40L-deficient) patients. In fact, glucocorticoids up-regulate CD40L expression on B cells and this finding may explain their synergistic actions on IgE isotype switching.

## Allergy-Mediated Immune Response and Inflammation

The fact that glucocorticoids are used in the treatment of atopy and asthma and promote humoral Th2 immune response and IgE secretion represents a disturbing and challenging paradox. In truth, glucocorticoids may enhance IgE secretion but they strongly suppress allergic inflammation and chemokine-driven tissue infiltration of eosinophils. Furthermore, it has been suggested that glucocorticoids would inhibit antigen-specific IgE production while raising total IgE levels. This may elucidate why IgE serum levels remained high in asthma patients in clinical remission during steroid treatment. Although they prime DCs and T cells for Th2 development on secondary stimulation, glucocorticoids inhibit Th2 cytokine secretion during primary antigen exposure. Despite an ongoing Th2 immune response, glucocorticoids can still prevent the potentially deleterious IgE-induced allergic immune response by mast cells. Indeed, they interfere with IgE receptor-mediated release of inflammatory mediators and deplete bronchial mucosa from resident mast cells. Suffice it to say, mucosal mast cells are the predominant effector cells orchestrating allergic inflammation. *In vivo* studies suggest that glucocorticoids deplete tissue mast cells by inhibiting essential survival factors such as IL-4 and fibroblast-derived stem cell factor. Cross-linking of their FcεRI with specific IgE results in the degranulation of preformed inflammatory mediators such as histamine, proteases, cytokines, and lipid mediators responsible for the early phase of type I hypersensitivity manifestations. Interestingly, glucocorticoids block IgE-triggered degranulation of human mast cells. They also prevent the late phase of type I hypersensitivity characterized by mast cell activation with *de novo* synthesis of inflammatory mediators and secondary infiltration of Th2 cells, basophils, and eosinophils. Glucocorticoid-induced down-regulation of high-affinity FcεRI and low-affinity FcεRI (CD23) expression may account for the poor response of IgE-triggered mast cell activation only in the late phase. Glucocorticoids also interfere with FcεRI signaling through the disruption of raf-1/heat shock protein 90 and the subsequent mitogen-activated phosphokinase activation and phospholipase A2 responsible for the *de novo* synthesis of arachidonic acid-derived metabolites. Finally, glucocorticoids inhibit cytokine synthesis of pro-inflammatory cytokines and chemokines by mast cells necessary for their own survival and the chemotaxis and expansion of eosinophils and basophils.

## T CELL DEVELOPMENT AND HOMEOSTASIS

Though supra-pharmacologic doses of glucocorticoids induce T cell apoptosis, adrenal- or thymus-derived glucocorticoids or physiologic doses of glucocorticoids can induce T cell survival or apoptosis, depending on the cell type and differentiation stage. Moreover, both the degree of T cell activation and the timing of glucocorticoid exposure (before, during, or after activation) render T cells sensitive or resistant to glucocorticoid-induced apoptosis. Several studies have shown that concomitant TCR signaling and glucocorticoid receptor (GR) signaling promote T cell survival, whereas either TCR signaling alone or GR signaling alone induces T cell apoptosis. Indeed, thymus-derived glucocorticoids appear to play a role in early thymocyte expansion and in central positive selection. Yet, GR<sup>-/-</sup> knockout mice demonstrate a normal thymus, suggesting that glucocorticoid actions on TCR signaling could occur through nongenomic actions that are independent of GR transcriptional activity. Similarly, there is some evidence suggesting that glucocorticoids could influence peripheral T cell development and selection by simultaneously preventing TCR-induced T cell deletion and enhancing T cell survival. Interestingly, glucocorticoids enhance a key cytokine receptor for T cell development, IL-7Rα, whose deletion in mice and humans is associated with a lack of T cells. Moreover, IL-7 potently enhances thymic-independent peripheral expansion and restores immunity in athymic T cell-depleted hosts in mice. The positive regulation of IL-7Rα expression by glucocorticoids suggests their strong influence in the maintenance of peripheral T cell pool homeostasis.

## CONCLUDING REMARKS

The description of the actions of glucocorticoids on the immune response elucidates their positive and negative effects on several components of the innate and adaptive immune responses. Glucocorticoid-induced immunomodulation requires both immunoenhancing and immunosuppressive actions at the same time and these should be integrated in a dynamic, ongoing process. Indeed, pro-inflammatory mediators participate, to the same extent, with anti-inflammatory mediators in the so-called immunosuppressive actions of glucocorticoids. The great advantage of their clinical use in Th1-inflammatory and Th1-autoimmune diseases is obvious because they restrain the inflammatory reaction, prevent tissue

destruction, and block Th1-driven immune responses. Also, the fact that there are glucocorticoids that favor innate immunity and antibody responses may explain why they exhibit clinical benefit in bacterial infections such as in bacterial meningitis. Similarly, the fact that glucocorticoids enhance the phagocytotic ability of neutrophils and macrophages, boosting scavenging and opsonization systems, underscores the clinical benefit and importance of early and continuous administration of glucocorticoids in patients in septic shock. However, these beneficial effects coexist with deleterious adverse effects. Because of their suppressive actions on Th1 immunity, they increase susceptibility to intracellular and opportunistic infections such as tuberculosis or viral infections. Finally, glucocorticoids may worsen disease activity due to their selective immunoenhancing actions. This was hieratically reported in steroid-resistant asthma and ulcerative colitis.

The overall knowledge of the actions of glucocorticoids on adaptive immunity clearly revealed that by decreasing the ability of DCs to elicit a T cell immune response, by promoting Th2 differentiation and helping the development of regulatory T cells, glucocorticoids may induce tolerance. Therefore, glucocorticoids could be extremely helpful in cell therapy through their *ex vivo* action not only on DCs but also potentially on regulatory T cells. This *ex vivo* therapeutic use of glucocorticoids would be very beneficial, thanks to their immune modifier effect, and avoid their long-term deleterious adverse effects. Glucocorticoids could thus represent an adjuvant treatment for cell therapies in autoimmune diseases and organ transplantations.

### See Also the Following Articles

Aging, Immunology and • Corticotropin-Releasing Hormone (CRH) and Inflammation • Cytokine Actions, Cellular Mechanism of • Cytokine Receptors • Glucocorticoid Receptor • Glucocorticoids, Overview • Immune System, Hormonal Effects on

### Further Reading

- Ashwell, J. D., Lu, F. W., and Vacchio, M. S. (2000). Glucocorticoids in T cell development and function. *Annu. Rev. Immunol.* **18**, 309–345.
- Barnes, P. J. (2001). Corticosteroids, IgE, and atopy. *J. Clin. Invest.* **107**, 265–266.
- Blotta, M. H., DeKruyff, R. H., and Umetsu, D. T. (1997). Corticosteroids inhibit IL-12 production in human monocytes and enhance their capacity to induce IL-4 synthesis in CD4<sup>+</sup> lymphocytes. *J. Immunol.* **158**, 5589–5595.
- Calandra, T., Bernhagen, J., Metz, C. N., Spiegel, L. A., Bacher, M., Donnelly, T., Cerami, A., and Bucala, R. (1995). MIF as a glucocorticoid-induced modulator of cytokine production. *Nature* **377**, 68–71.
- Franchimont, D., Galon, J., Vacchio, M. S., Fan, S., Visconti, R., Frucht, D. M., Geenen, V., Chrousos, G. P., Ashwell, J. D., and O’Shea, J. J. (2002). Positive effects of glucocorticoids on T cell function by up-regulation of IL-7 receptor  $\alpha$ . *J. Immunol.* **168**, 2212–2218.
- Hogger, P., Dreier, J., Droste, A., Buck, F., and Sorg, C. (1998). Identification of the integral membrane protein RM3/1 on human monocytes as a glucocorticoid-inducible member of the scavenger receptor cysteine-rich family (CD163). *J. Immunol.* **161**, 1883–1890.
- Jabara, H. H., Brodeur, S. R., and Geha, R. S. (2001). Glucocorticoids upregulate CD40 ligand expression and induce CD40L-dependent immunoglobulin isotype switching. *J. Clin. Invest.* **107**, 371–378.
- Nakagawa, M., Bondy, G. P., Waisman, D., Minshall, D., Hogg, J. C., and van Eeden, S. F. (1999). The effect of glucocorticoids on the expression of L-selectin on polymorphonuclear leukocyte. *Blood* **93**, 2730–2737.
- Piemonti, L., Monti, P., Allavena, P., Sironi, M., Soldini, L., Leone, B. E., Succi, C., and Di Carlo, V. (1999). Glucocorticoids affect human dendritic cell differentiation and maturation. *J. Immunol.* **162**, 6473–6481.
- Strausbaugh, H. J., and Rosen, S. D. (2000). A potential role for annexin 1 as a physiologic mediator of glucocorticoid-induced L-selectin shedding from myeloid cells. *J. Immunol.* **166**, 6294–6300.
- Van Laethem, F., Baus, E., Smyth, L. A., Andris, F., Bex, F., Urbain, J., Kiousis, D., and Leo, O. (2001). Glucocorticoids attenuate T cell receptor signaling. *J. Exp. Med.* **193**, 803–814.
- Wu, C. Y., Sarfati, M., Heusser, C., Fournier, S., Rubio-Trujillo, M., Peleman, R., and Delespesse, G. (1991). Glucocorticoids increase the synthesis of immunoglobulin E by interleukin 4-stimulated human lymphocytes. *J. Clin. Invest.* **87**, 870–877.

## REGULATION OF THE HYPOTHALAMIC–PITUITARY–ADRENAL CORTEX AXIS

Glucocorticoids serve a critical role in adaptation to the changing environment. Their levels are invariably increased by subconscious alterations in homeostasis as well as by what we personally define as “stressful,” be it an external stimulus or an internal perception. On realization of the stressor, higher order neural circuits lift the inhibition imposed by the hippocampus on the hypothalamic–pituitary–adrenal cortex (HPA) axis. As a result, corticotropin-releasing hormone (CRH) is secreted from the paraventricular nucleus of the hypothalamus, which in turn stimulates release of adrenocorticotropin (ACTH) from the anterior pituitary and subsequently secretion of glucocorticoids (CORT: cortisol in humans; corticosterone in rodents) from the adrenal cortex. During its initial phase, this elevation of CORT helps to address the challenge by improving cognition through mineralocorticoid receptors (MRs) in the hippocampus and increasing the energy availability to the brain and skeletal muscles. Once the stressor has been properly addressed, the HPA axis is inhibited by negative feedback mainly through the glucocorticoid receptors (GRs) in the hippocampus, hypothalamus, and pituitary. After the CORT levels return to baseline, MRs in the hippocampus take over the control of regulating basal CORT levels.

## AGE-RELATED ALTERATION IN CORTICOSTERONE PROFILE IN RODENTS

Age-related changes in the human HPA axis are susceptible to a myriad of factors that can be generally divided into genetic differences, nonpsychological environmental factors, complex daily personal interactions, various life histories, and interactions among these factors that define our attitude toward life experiences. In contrast, most of these factors are well controlled in experiments examining aging in laboratory rodents. However, even here, the data have been highly variable. After analyzing the available data on age-related alterations in plasma CORT in several different rat strains, Sapolsky concluded that there was more than a twofold increase in CORT with age. This has been corroborated by other more recent studies. A more descriptive picture of the age-related plasma CORT levels was gained from a longitudinal study in Fischer 344 rats by Sabatino and colleagues. They

found that the free CORT levels increased twofold between middle age and old age. Furthermore, the diurnal pattern showed an expansion of the circadian peak into the trough. This suggests that with progressing age, there is more activation of GRs and for a longer period. Another important result from this study was that when the rats were young, restraint stress-induced CORT levels started to decline soon after termination of the stressor, whereas in old age the stress-induced CORT levels stayed elevated for an additional 30 min prior to the decline. Because these animals were not chronically stressed, these age-related differences arise due to mechanisms that are not well appreciated.

Age also seems to affect the expression levels of the MR and GR. The most consistent observation is the age-related decline in MR mRNA and protein levels within the hippocampus. On the other hand, changes in hippocampal GR levels appear to be strain dependent, with age-related decreases in both mRNA and protein levels in some rat strains but not in others. Interestingly, decreases in hippocampal MRs and GRs are also observed in chronically stressed animals, and the extent of reduction in these receptors may be associated with the severity of stress. Thus, the age- and stress-related modulation of MR and GR expression has a serious impact on regulation of the HPA axis.

In light of the importance of these receptors in the regulation of the HPA axis, it should be mentioned that life histories play an important role in expression of these receptors in the hippocampus. Meaney and colleagues reported that maternal affection (licking) during postnatal development in rodents increases GR expression and that these pups, during their adulthood, show an attenuated CORT response to stress, have a lower age-related increase in CORT, and are less cognitively impaired than their counterparts. In contrast, pups that were deprived of maternal care or infected with endotoxins show a decrease in their hippocampal receptor levels and show an accentuated CORT response to stressors during adulthood. Furthermore, elevated CORT levels during fetal development decrease the expression of these receptors and possibly predispose these animals to age-related increases in CORT. Thus, experiences over a lifetime should be considered as an important variable in age-related CORT elevations.

## CORTICOSTERONE LEVELS, AGE, AND COGNITIVE IMPAIRMENT IN RATS

As with age-related increases in plasma CORT levels, the within-group variability in cognitive performance

among the aged animals is large. In one study by Meaney and colleagues, 20% of the aged rats were cognitively unimpaired (AU) and performed as well as young controls, whereas 28% of the aged rats showed dramatic cognitive impairment (AI). Furthermore, these two aged groups differed significantly for several hippocampus-HPA parameters. The plasma CORT levels across the diurnal cycle in the AU rats were similar to those in the young rats with a single peak of elevated CORT levels at the beginning of the dark cycle, whereas the AI rats showed an expansion of the diurnal peak such that the body was exposed to elevated CORT levels for an extended period of time (12 h); this diurnal pattern is similar to that observed by Sabatino and colleagues mentioned previously. The AI rats also had elevated plasma ACTH levels at the beginning of the acrophase, and this was probably responsible for the extended period of elevated CORT in this group. The termination of stress-induced CORT response was also different in these two aged groups, such that the total CORT exposure (area under the curve) in the AI rats was much greater. Consistent with this profile, the hippocampal MR and GR levels in the AI animals were lower than those in the AU rats.

The data from McEwen's laboratory has clearly established that stress-related increases in CORT are associated with atrophy of the dendritic tree, such that the number of branching points is decreased. This is likely to yield a dramatic reduction in the number of synapses. This is supported by Jucker and colleagues' observation that the number of synaptic boutons in the dentate gyrus region of the hippocampus is associated with performance in the Morris water maze.

An added cellular complexity should also be mentioned. Plasma CORT levels are not necessarily the "true" indication of total cellular CORT exposure. Intracellular CORT levels are regulated by 11 $\beta$ -hydroxysteroid dehydrogenase (11 $\beta$ -HSD). The type 1 isoform, 11 $\beta$ -HSD1, can convert inactive CORT metabolites (11-dehydrocorticosterone, cortisone) back to active CORT. Seckl's group has shown that old mice lacking 11 $\beta$ -HSD1 perform as well as young controls in the Morris water maze, whereas old control mice take twice as long to find the hidden stage. Because the levels of the CORT metabolites are in excess of the free CORT levels, regulation of this activity can prove to be extremely beneficial in decreasing CORT-related cognitive impairment.

## AGE-RELATED ALTERATION IN THE CORTISOL PROFILE IN HUMANS

Despite all of the previously mentioned concerns regarding examination of age-related changes in CORT in humans, it is important to assess whether the human HPA axis is susceptible to dysregulation with age. However, these concerns cannot be undermined and should be evaluated further because doing so may set forth hypotheses that are directly relevant to successful aging as well as to age-related disorders in which etiology remains unclear. For example, it is possible to subdivide healthy volunteers based on their anticipatory stress responses, dexamethasone suppression of CORT levels, and ACTH responsiveness. All of these parameters are quantitative, and if they are to be followed longitudinally, they might allow identification of progressive steps that increase susceptibility to age-related disorders.

The data regarding age-related alteration in basal plasma CORT levels in humans are inconsistent. Because most of these studies have also evaluated cognitive functions, the data are discussed in more detail in the next section. However, an important difference in the experimental design between the rodent studies and the human studies should be emphasized here: in the human studies, unhealthy individuals are either excluded or considered as an independent variable, whereas the same level of discretion is hard to attain, and is often ignored, in rodent studies. Proinflammatory cytokines can activate the HPA axis and may contribute to the age-related cognitive impairment observed in rodents. This may explain some of the inconsistencies observed between the human data and the rodent data.

The HPA axis in humans is not spared by aging. It is now widely accepted that elderly humans also show a delayed termination of the CORT response after a stressful situation. One possible explanation for this increase in latency is the "accumulative effect of stress" over the life span. Allostasis refers to constant alterations in the HPA axis that are necessary to maintain homeostasis in the new environment/situation. As the frequency of HPA activation and/or total exposure to CORT increases, so do the side effects of hypercortisolemia. The HPA axis is designed to minimize this consequence by a negative feedback mechanism, but as the "allostatic load" on the hippocampus increases, the regulation of HPA deteriorates and CORT levels stay elevated for longer than necessary. A subgroup of elderly persons also show elevated

nighttime CORT levels, leading to the suggestion that pharmacological intervention to reduce nighttime CORT may have therapeutic value in these individuals.

### COGNITIVE IMPAIRMENTS AND PLASMA CORTISOL LEVELS IN HUMANS

The human data do not show a direct relationship between age and cognition or between age and cortisol levels. However, elderly persons tend to have a greater risk of both cognitive impairment and elevated cortisol levels. In the MacArthur Foundation Study on Successful Aging, healthy elderly individuals residing in three different communities were followed for  $2\frac{1}{2}$  years. During this period, 20% of the women showed a progressive increase in CORT, whereas 15% showed a decline in urinary CORT levels. In the subgroup with increasing CORT, 75% of the women showed poor performance on the delayed recall of a story task as compared with their performance at the start of the study. In contrast, in the subgroup with decreasing CORT, 70% of the women showed an improvement in their performance on this task. Similarly, Lupien and colleagues found that individuals with increasing CORT over a 5-year period performed poorly on the delayed memory and spatial memory tests as compared with individuals with stable/decreasing CORT levels. Furthermore, those with increasing CORT had a mean 15% decrease in the hippocampal volume. Hippocampal atrophy of a similar extent ( $>10\%$ ) is also seen in AD. More than half of the studies that have examined basal cortisol levels in AD reported an association between AD and increased plasma cortisol. Furthermore, the few studies that have examined the HPA axis overactivity by the dexamethasone suppression test found that a greater number of individuals with AD failed the test as compared with controls. Thus, the age-related risk of elevated CORT levels may also increase the risk of AD.

A decrease in hippocampal volume is also observed in other disorders with hypercortisolemia: depression, posttraumatic stress disorder (PTSD), and Cushing's syndrome. In the cases of PTSD and depression, it is possible that altered brain physiology could have confounding effects on this measure. However, in Cushing's syndrome, hypercortisolemia occurs as a consequence of peripheral tumors that secrete excessive amounts of ACTH or CRH. Moreover, on surgical correction of hypercortisolemia in Cushing's

syndrome, the hippocampal volume increases. These observations support the notion that cortisol leads to cognitive decline by damaging the hippocampus.

### GLUCOCORTICOIDS' EFFECTS ON LONG-TERM POTENTIATION

The associative, specific, and relatively long-lasting characteristics of long-term potentiation (LTP) have led to its acceptance as an excellent model of the molecular mechanism underlying learning and memory. Similar synaptic efficacy can also be elicited by a physiologically patterned, lower threshold primed-burst stimulation. In contrast, low-frequency stimulation of the synapse yields long-term depression (LTD) of the synapse. Both of these mechanisms rely on  $\text{Ca}^{2+}$  influx and the N-methyl-D-aspartate (NMDA glutamate analogue) receptor. A well-known target of NMDA-mediated  $\text{Ca}^{2+}$  influx is the calcium-calmodulin-dependent kinase II (CaMKII), which on activation auto-phosphorylates and becomes autonomously active to maintain the sensitivity of the synapse to future stimulation. Similar mechanisms are also employed during a low-frequency stimulation, but because the  $\text{Ca}^{2+}$  influx is gradual and of low intensity, it fails to induce the  $\text{Ca}^{2+}$ -mediated pathways that are activated during high-frequency stimulation.

CORT shifts the balance between LTP and LTD by differential activation of the MR and GR. For example, treatment of adrenalectomized rats with an MR-specific agonist, aldosterone, enhanced LTP, whereas treatment with a GR-specific agonist, RU-28362, suppressed synaptic efficacy. Under basal conditions, preferential activation of MR in the hippocampus is associated with short-lived  $\text{Ca}^{2+}$  currents that promote LTP. On the other hand, activation of GR increases expression of the NMDA receptor subunit NR2B. This substitution in the NMDA receptor leads to enhanced  $\text{Ca}^{2+}$  influx. Because this alteration is at the receptor level, the effects are observed for a longer time and lead to increased  $\text{Ca}^{2+}$  influx even during low-frequency stimulation. Kim and Yoon proposed that the intracellular  $\text{Ca}^{2+}$  ( $[\text{Ca}^{2+}]_i$ ) has a metaplastic effect on synaptic efficacy. They suggested that if the  $[\text{Ca}^{2+}]_i$  is high, the subsequent stimulus would have to be of much greater intensity to favor synaptic efficacy. If the stimulus is smaller than the previous stimulus, it will fail to induce  $\text{Ca}^{2+}$ -calmodulin pathways but will still elicit other effects of  $\text{Ca}^{2+}$  such as activation of phosphatases and  $\text{Ca}^{2+}$ -dependent  $\text{K}^+$  channels. Activation of  $\text{K}^+$  channels increases the afterhyperpolarization

current (refractory period) of the hippocampal neurons. This would block the cells' ability to respond to high-frequency stimuli, further decreasing the cells' ability to shift back in favor of synaptic efficacy. Consistently, the aged rats have a propensity for LTD and are susceptible to reversal of LTP. In one study, the number of potentiations observed after a primed-burst stimulation in 24-month-old Fischer 344 rats was only 40% that in 6-month-old controls.

## EFFECTS OF CHRONIC DISEASES ON GLUCOCORTICOIDS

One of the earliest systemic responses to immune system activation is the release of interleukin-1 (IL-1), IL-6, and tumor necrosis factor- $\alpha$  (TNF- $\alpha$ ). These cytokines trigger an acute phase response in other organs to alter their function during an infection. These cytokines also activate the HPA axis at several levels; all of these cytokines can also stimulate the HPA axis through the brain; IL-1 and TNF- $\alpha$  can increase CRH, all three increase secretion of ACTH, and IL-6 can also activate the adrenocortical cells directly to secrete CORT. A classic function attributed to CORT is as an anti-inflammatory agent, presumably to reduce "nonspecific" inflammation elsewhere in the body. CORT also increases expression of macrophage migration inhibitory factor (MIF) in various organs in a tissue- and time-dependent manner. MIF exhibits an anti-CORT effect by reversing CORT's effects on both apoptosis and cytokine expression in T lymphocytes and macrophages. This action may serve to overcome the anti-inflammatory effects of CORT within the local area of inflammation. However, in chronic illness, elevated levels of both MIF and CORT may lead to dysregulation by decreasing CORT's effectiveness.

IL-1, IL-6, and TNF- $\alpha$  have been associated with aging and/or age-related disorders and may also be responsible for the observed age-related elevation of CORT in a subgroup of elderly persons. Thus, individuals in preclinical stages of a disease could

have elevated cytokines, elevated CORT, and (as a consequence) cognitive impairment. This is supported by the observation that cognitive impairment is associated with mortality. Furthermore, most of the commonly used rat strains develop multiple lesions with age. Thus, the age-related increase in CORT levels in rodents may be indicative of disease pathology.

In conclusion, age-related increases in CORT might not be a consequence of the aging process but rather might be a consequence of disease, and this elevation of CORT leads to cognitive impairment by damaging the hippocampus.

## See Also the Following Articles

ACTH (Adrenocorticotrophic Hormone) • Aging and Longevity of Human Populations • Alzheimer's Disease and Hormones • Brain, Effects of Steroid Hormones • Functional Genomics of Aging • Glucocorticoids, Overview • Neuroendocrine System and Aging • Stress, Aging, and Central Nervous System Interactions

## Further Reading

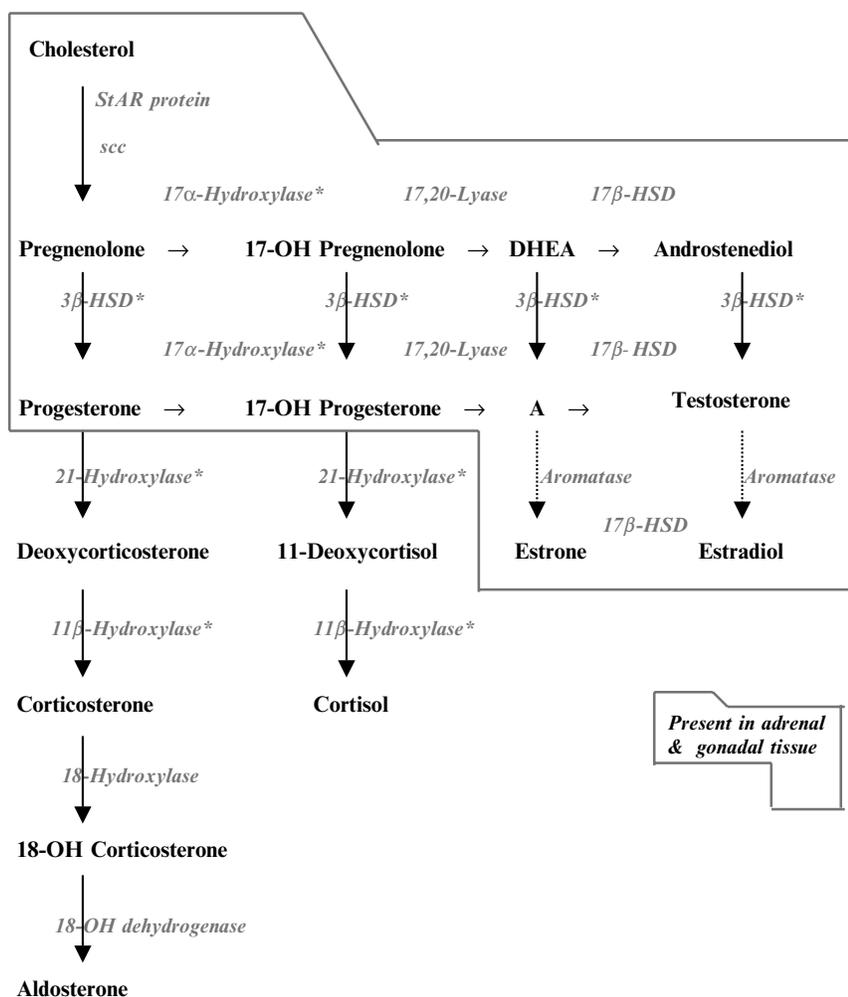
- Harvey, P. D., and Mohs, R. C. (2001). Memory changes with aging and dementia. *In* "Functional Neurobiology of Aging" (P. R. Hof and C. V. Mobs, eds.), pp. 53–63. Academic Press, San Diego.
- Lucassen, P. J., and De Kloet, E. R. (2001). Glucocorticoids and the aging brain: Cause or consequence? *In* "Functional Neurobiology of Aging" (P. R. Hof and C. V. Mobs eds.), pp. 883–905. Academic Press, San Diego.
- Lupien, S. J., Nair, N. P., Briere, S., Maheu, F., Tu, M. T., Lemay, M., McEwen, B. S., and Meaney, M. J. (1999). Increased cortisol levels and impaired cognition in human aging: Implication for depression and dementia in later life. *Rev. Neurosci.* **10**, 117–139.
- Patel, N. V., and Finch, C. E. (2002). The glucocorticoid paradox of caloric restriction in slowing brain aging. *Neurobiol. Aging* **23**, 707–717.
- Sapolsky, R. M. (1999). Glucocorticoids, stress, and their adverse neurological effects: Relevance to aging. *Exp. Gerontol.* **34**, 721–732.
- Turnbull, A. V., and Rivier, C. L. (1999). Regulation of the hypothalamic–pituitary–adrenal axis by cytokines: Actions and mechanisms of action. *Physiol. Rev.* **79**, 1–71.

deoxycorticosterone (DOC). Aldosterone, the most potent 17-deoxysteroid with mineralocorticoid activity, is produced by 11 $\beta$ -hydroxylation of DOC to corticosterone, followed by 18-hydroxylation and 18-oxidation of corticosterone (Fig. 1). The final three steps in aldosterone synthesis are accomplished by a single mitochondrial P450 enzyme, CYP11B2 (P450aldo, aldosterone synthase).

To produce cortisol, CYP17 (P450c17, 17 $\alpha$ -hydroxylase/17,20-lyase) in the endoplasmic reticulum of the zona fasciculata and zona reticularis converts pregnenolone to 17 $\alpha$ -hydroxypregnenolone. 3 $\beta$ -HSD in the zona fasciculata utilizes 17 $\alpha$ -hydroxypregnenolone as a substrate, producing 17 $\alpha$ -hydroxypregesterone. The latter is 21-hydroxylated by CYP21 to form

11-deoxycortisol, which is further converted to cortisol by CYP11B1 (P450c11, 11 $\beta$ -hydroxylase) in the mitochondria.

In the zona reticularis of the adrenal cortex and in the gonads, the 17,20-lyase activity of CYP17 converts 17 $\alpha$ -hydroxypregnenolone to dehydroepiandrosterone (DHEA), a C-19 steroid and sex steroid precursor. DHEA is further converted by 3 $\beta$ -HSD to androstenedione. In the gonads, androstenedione is reduced by 17 $\beta$ -hydroxysteroid dehydrogenase. In pubertal ovaries, aromatase (CYP19, P450c19) can convert androstenedione and testosterone to estrone and estradiol, respectively. Testosterone may be further metabolized to dihydrotestosterone by steroid 5 $\alpha$ -reductase in androgen target tissues.



**Figure 1** Schematic representation of adrenal steroidogenesis. Solid lines indicate major pathways. Dotted lines indicate major pathways in ovaries and minor pathways in adrenals. Asterisks indicate that deficient enzymatic activity results in congenital adrenal hyperplasia (CAH). StAR, steroidogenic acute regulatory protein; scc, cholesterol side-chain cleavage enzyme; 3 $\beta$ -HSD, 3 $\beta$ -hydroxysteroid dehydrogenase; 17 $\beta$ -HSD, 17 $\beta$ -hydroxysteroid dehydrogenase; DHEA, dehydroepiandrosterone; A, androstenedione.

## REGULATION OF CORTISOL SECRETION

Plasma glucocorticoid concentrations are regulated in ways that reflect the varying physiologic needs for the hormones under basal conditions and in response to stress. Cortisol secretion is primarily regulated by ACTH, a 39-amino-acid peptide released by the anterior pituitary. ACTH is synthesized as part of a higher molecular weight precursor peptide, proopiomelanocortin. ACTH is secreted in regular pulses of variable amplitude over 24 h, with peak concentrations attained in the early morning hours (4:00–8:00 AM), thus forming the basis of the circadian pattern of cortisol secretion.

The acute action of ACTH is to increase the flux of cholesterol through the steroidogenic pathway, resulting in the rapid production of steroids. ACTH also influences the remaining steps of steroidogenesis as well as the uptake of cholesterol from plasma lipoproteins, thus ensuring a continuous supply of cholesterol to the mitochondria to meet the demands of activated pregnenolone biosynthesis. It also maintains the size of the adrenal glands, stimulates melanocytes, and results in hyperpigmentation when secreted in excess.

Corticotropin-releasing hormone (CRH) is the principal hypothalamic factor that stimulates the pituitary production of ACTH. It is produced in the paraventricular nuclei of the hypothalamus, but is also found in other parts of the central nervous system as well as in noncentral locations. CRH is secreted in a pulsatile fashion that results in the episodic secretion of ACTH and the circadian variation of cortisol secretion. The magnitude of cortisol response to each ACTH burst remains relatively constant; it is therefore the number of secretory periods, rather than the magnitude of each pulse of CRH or ACTH, that determines the total daily cortisol secretion. In addition to CRH, vasopressin, a peptide product of the posterior pituitary gland, stimulates ACTH release by acting synergistically with CRH. Although CRH increases the amount of ACTH secreted from each responsive corticotroph, vasopressin appears to increase the number of CRH-responsive corticotrophs. In addition to ACTH, other factors may play an important role in the regulation of the adrenal cortex.

Cortisol is the primary negative regulator of basal hypothalamic–pituitary–adrenal (HPA) axis activity through negative feedback on ACTH and CRH secretion. The negative feedback effects of cortisol are exerted at the level of both the hypothalamus and the pituitary and are mediated by type II corticosteroid

receptors. Whether and to what extent direct glucocorticoid feedback on the adrenal cortex itself regulates cortisol synthesis is not clear.

## SECRETION AND METABOLISM

In normal subjects, the secretion of glucocorticoids follows a diurnal pattern, with peak concentrations observed between 6:00 and 8:00 AM and the lowest concentrations observed at approximately 12:00 AM. The cortisol production rate is approximately 12 mg/m<sup>2</sup>/day. More than 90% of circulating cortisol, and to a lesser extent aldosterone, is bound tightly to corticosteroid-binding globulin (CBG) or transcortin. The remaining (10%) of the circulating cortisol is free or loosely bound to albumin. The free and albumin-bound fractions of cortisol represent the biologically active form of the hormone. When plasma cortisol concentrations exceed 20 µg/dl, CBG becomes fully saturated and most of the excess cortisol is biologically active. CBG is synthesized in the liver. Estrogens, thyroid hormones, pregnancy, and oral contraceptives are associated with increased CBG concentrations, whereas hypercortisolism, hepatic disease, or renal disease results in decreased CBG concentrations. In the presence of an intact HPA axis, alterations in CBG concentrations are likely not to affect circulating free cortisol concentrations.

The primary site of cortisol metabolism in humans is the liver, and a number of cytosolic and microsomal enzymes, including cytochrome P450, 5α/5β-reductase, 3α/3β-oxidoreductase, and 11β-hydroxysteroid dehydrogenase, play an important role in the hepatic metabolism of cortisol. The major routes of hepatic metabolism involve A-ring and side-chain reduction followed by conjugation with glucuronic acid and sulfate. The inactive glucuronide and sulfate metabolites are excreted by the kidneys, whereas less than 1% of cortisol is excreted unchanged in the urine. The metabolic clearance of cortisol, therefore, is influenced primarily by factors altering hepatic clearance and to a much lesser degree by factors affecting renal excretion.

## MECHANISMS OF GLUCOCORTICOID ACTION

At the cellular level, the actions of glucocorticoids are mediated by an intracellular receptor protein, the glucocorticoid receptor, which functions as a hormone-activated transcription factor that regulates the expression of glucocorticoid target genes.

## TREATMENT WITH GLUCOCORTICOIDS

Natural and synthetic glucocorticoids can be used for both endocrine and nonendocrine disorders. In clinical practice, glucocorticoids are used to establish the diagnosis and cause of Cushing's syndrome and in the treatment of adrenal insufficiency and congenital adrenal hyperplasia. Glucocorticoids are also given in pharmacologic doses to treat patients with inflammatory, allergic, or immunologic disorders. Chronic therapy has many side effects, ranging from suppression of the HPA axis and Cushing's syndrome to infections and changes in mental status. Factors that influence both the therapeutic and adverse effects of glucocorticoids include the pharmacokinetic properties of the glucocorticoid, daily dosage, timing of doses during the day, individual differences in steroid metabolism, and the duration of treatment.

### Glucocorticoid Replacement Therapy

In deficiency states, physiologic replacement is best achieved with a combination of hydrocortisone and the mineralocorticoid fludrocortisone; hydrocortisone alone does not usually provide sufficient mineralocorticoid activity for complete replacement. In Addison's disease or following adrenalectomy, hydrocortisone at 10–15 mg/m<sup>2</sup> daily by mouth is usually required. This is given in two doses, the larger in the morning and the smaller in the evening, mimicking the normal diurnal rhythm of cortisol secretion. The optimum daily dose is determined on the basis of clinical response. Glucocorticoid therapy is supplemented by fludrocortisone 50 to 300 µg daily. In acute adrenocortical insufficiency, hydrocortisone is given intravenously (preferably as sodium succinate) at doses of 100 mg every 6 to 8 h in 0.9% sodium chloride intravenous infusion. In hypopituitarism, glucocorticoids should be given as in adrenocortical insufficiency, but since the production of aldosterone is regulated by the renin–angiotensin system, a mineralocorticoid is not usually required. Additional replacement therapy with levothyroxine and sex hormones should be given as indicated by the pattern of hormone deficiency.

### Glucocorticoid Therapy

In comparing the relative potencies of corticosteroids in terms of their anti-inflammatory (glucocorticoid) effects, it should be borne in mind that high

glucocorticoid activity in itself is of no advantage unless it is accompanied by relatively low mineralocorticoid activity. The mineralocorticoid activity of fludrocortisone is so high that its anti-inflammatory activity is of no clinical relevance. The equivalent anti-inflammatory doses of corticosteroids are shown in Table I.

The relatively high mineralocorticoid activity of cortisone and hydrocortisone and the resulting fluid retention make them unsuitable for disease suppression on a long-term basis. However, they can be used for adrenal replacement therapy; hydrocortisone is preferred because cortisone requires conversion to hydrocortisone in the liver. Hydrocortisone is used on a short-term basis by intravenous injection for the emergency management of some conditions. The relatively moderate anti-inflammatory potency of hydrocortisone also makes it a useful topical corticosteroid for the management of inflammatory skin conditions because side effects (both topical and systemic) are less marked; cortisone is not active topically.

Prednisolone has predominantly glucocorticoid activity and is the corticosteroid most commonly used by mouth for long-term disease suppression. Betamethasone and dexamethasone have very high glucocorticoid activity but insignificant mineralocorticoid activity. This makes them particularly suitable for high-dose therapy in conditions where fluid retention would be a disadvantage. Betamethasone and dexamethasone also have a long duration of action and this, coupled with their lack of mineralocorticoid action, makes them particularly suitable for conditions that require suppression of ACTH secretion (e.g., congenital adrenal hyperplasia). Some esters of betamethasone and beclomethasone (beclomethasone)

**Table I Equivalent Anti-inflammatory Doses of Corticosteroids**

Prednisolone	5 mg
≡ Betamethasone	750 µg
≡ Cortisone acetate	25 mg
≡ Deflazacort	6 mg
≡ Dexamethasone	750 µg
≡ Hydrocortisone	20 mg
≡ Methylprednisolone	4 mg
≡ Triamcinolone	4 mg

*Source.* Reprinted from the British Medical Association and the Royal Pharmaceutical Society of Great Britain (2002). British National Formulary No. 44, with permission.

*Note.* This table takes no account of mineralocorticoid effects, nor does it take account of variations in duration of action.

exert a considerably more marked topical effect (e.g., on the skin or the lungs) than when given by mouth; use is made of this to obtain topical effects while minimizing systemic side effects (e.g. for skin applications and asthma inhalations). Deflazacort is derived from prednisolone and has high glucocorticoid activity.

## SIDE EFFECTS OF CORTICOSTEROIDS

Overdosage or prolonged use may exaggerate some of the normal physiologic actions of corticosteroids, leading to mineralocorticoid and glucocorticoid side effects. Mineralocorticoid side effects include hypertension, sodium and water retention, and potassium loss. They are most marked with fludrocortisone, but are significant with cortisone, hydrocortisone, corticotropin, and tetracosactide (tetracosactrin). Mineralocorticoid actions are negligible with the high-potency glucocorticoids betamethasone and dexamethasone and occur only slightly with methylprednisolone, prednisolone, and triamcinolone. Glucocorticoid side effects include diabetes and osteoporosis, avascular necrosis of the femoral head, mental disturbances (a serious paranoid state or depression with risk of suicide may be induced, particularly in patients with a history of mental disorder), euphoria, and muscle wasting (proximal myopathy) (Table II). Corticosteroid therapy is also weakly linked with peptic ulceration. High doses of corticosteroids may cause Cushing's syndrome, which is usually reversible on withdrawal of treatment, but this must always be gradually tapered to avoid symptoms of acute adrenal insufficiency. In children, administration of corticosteroids may result in suppression of growth. Other complications include increased susceptibility to infection, poor wound healing, and activation of latent granulomatous infections.

### Adrenal Suppression

During prolonged therapy with corticosteroids, adrenal atrophy develops and may persist for years after stopping. Abrupt withdrawal after a prolonged period may lead to acute adrenal insufficiency, hypotension, or death. Withdrawal may also be associated with fever, myalgia, arthralgia, rhinitis, conjunctivitis, painful itchy skin nodules, and weight loss.

To compensate for a diminished adrenocortical response caused by prolonged corticosteroid treatment, any significant intercurrent illness, trauma, or surgical procedure requires a temporary increase in

**Table II Effects of Chronic Pharmacologic Use of Glucocorticoids**

---

#### Endocrine and metabolic

- Suppression of HPA axis (adrenal suppression)
- Growth failure in children
- Carbohydrate intolerance
  - Hyperinsulinism
  - Insulin resistance
  - Abnormal glucose tolerance test
  - Diabetes mellitus
- Cushingoid features
  - Moon facies, facial plethora
  - Generalized and truncal obesity
  - Supraclavicular fat collection
  - Posterior cervical fat deposition (buffalo hump)
  - Glucocorticoid-induced acne
  - Thin and fragile skin, violaceous striae
- Impotence, menstrual disorders
- Decreased thyroid-stimulating hormone and triiodothyronine
- Hypokalemia, metabolic alkalosis

#### Gastrointestinal

- Gastric irritation, peptic ulcer
- Acute pancreatitis (rare)
- Fatty infiltration of liver (hepatomegaly) (rare)

#### Hematopoietic

- Leukocytosis
- Neutrophilia
  - Increased influx from bone marrow and decreased migration from blood vessels
- Monocytopenia
- Lymphopenia
  - Migration from blood vessels to lymphoid tissue
- Eosinopenia

#### Immunologic

- Suppression of delayed hypersensitivity
  - Inhibition of leukocyte and tissue macrophage migration
  - Inhibition of cytokine secretion/action
- Suppression of the primary antigen response

#### Musculoskeletal

- Osteoporosis, spontaneous fractures
- Aseptic necrosis of femoral and humeral heads and other bones
- Myopathy

#### Ophthalmologic

- Posterior subcapsular cataracts (more common in children)
- Elevated intraocular pressure/glaucoma

#### Neuropsychiatric

- Sleep disturbances, insomnia
- Euphoria, depression, mania, psychosis
- Pseudo-tumor cerebri (benign increase of intracranial pressure)

#### Cardiovascular

- Hypertension
  - Congestive heart failure in predisposed patients
-

corticosteroid dose or, if already stopped, a temporary reintroduction of corticosteroid treatment. Anesthetists must therefore know whether a patient is taking or has been taking a corticosteroid, to avoid a precipitous fall in blood pressure during anesthesia or in the immediate postoperative period. A suitable regimen for corticosteroid replacement, in patients who have taken more than 10 mg prednisolone daily (or equivalent) within 3 months of surgery, is as follows:

- Minor surgery under general anesthesia—usual oral corticosteroid dose on the morning of surgery or hydrocortisone 25–50 mg (usually sodium succinate) intravenously at induction; the usual oral corticosteroid dose is recommenced after surgery.
- Moderate or major surgery—usual oral corticosteroid dose on the morning of surgery and hydrocortisone 25–50 mg intravenously at induction, followed by hydrocortisone 25–50 mg three times a day by intravenous injection for 24 h after moderate surgery or for 48–72 h after major surgery; the usual preoperative oral corticosteroid dose is recommenced on stopping hydrocortisone injections.

Patients on long-term corticosteroid treatment should carry a Steroid Treatment Card, which gives guidance on minimizing risk and provides details of prescriber, drug, dosage and duration of treatment.

### Infections

Prolonged courses of corticosteroids increase susceptibility to infections and severity of infections; clinical presentation of infections may also be atypical. Serious infections, e.g., septicemia and tuberculosis, may reach an advanced stage before being recognized and amebiasis or strongyloidiasis may be activated or exacerbated (they should be excluded before corticosteroid treatment is initiated in those at risk or with suggestive symptoms). Fungal or viral ocular infections may also be exacerbated.

### Chickenpox

Unless they have had chickenpox, patients receiving oral or parenteral corticosteroids for purposes other than replacement should be regarded as being at risk of severe chickenpox. Passive immunization with varicella-zoster immunoglobulin is needed for exposed nonimmune patients receiving systemic corticosteroids or for those who have used them within the previous 3 months; varicella-zoster immunoglobulin

should preferably be given within 3 days of exposure and no later than 10 days. Confirmed chickenpox warrants specialist care and urgent treatment. Corticosteroids should not be stopped and dosage may need to be increased. Topical, inhaled, or rectal corticosteroids are less likely to be associated with an increased risk of severe chickenpox.

### Measles

Patients taking corticosteroids should be advised to take particular care to avoid exposure to measles and to seek immediate medical advice if exposure occurs. Prophylaxis with intramuscular normal immunoglobulin may be needed.

## ADMINISTRATION OF CORTICOSTEROIDS

Whenever possible, local treatment with creams, intra-articular injections, inhalations, eye drops, or enemas should be used in preference to systemic treatment. The suppressive action of a corticosteroid on cortisol secretion is lowest when it is given as a single dose in the morning. In an attempt to reduce pituitary–adrenal suppression further, the total dose for 2 days can sometimes be taken as a single dose on alternate days; alternate-day administration has not been very successful in the management of asthma. Pituitary–adrenal suppression can also be reduced by means of intermittent therapy with short courses. In some conditions, it may be possible to reduce the dose of corticosteroid by adding a small dose of an immunosuppressive drug.

Dosage of corticosteroids varies widely in different diseases and in different patients. If the use of a corticosteroid can save or prolong life, high doses may need to be given, because the complications of therapy are likely to be less serious than the effects of the disease itself. When long-term corticosteroid therapy is used in some chronic diseases, the adverse effects of treatment may become greater than the disabilities caused by the disease. To minimize side effects, the maintenance dose should be kept as low as possible.

## WITHDRAWAL OF CORTICOSTEROIDS

A gradual withdrawal of systemic corticosteroids should be considered in those subjects whose disease is unlikely to relapse and who have: (1) recently received repeated courses (particularly if taken for

longer than 3 weeks); (2) taken a short course within 1 year of stopping long-term therapy; (3) other possible causes of adrenal suppression; (4) received more than 40 mg daily prednisolone (or equivalent); (5) been given repeat doses in the evening; or (6) received treatment for more than 3 weeks.

Systemic corticosteroids may be stopped abruptly in those whose disease is unlikely to relapse and who have received treatment for 3 weeks or less and who are not included in the patient groups described above. During corticosteroid withdrawal, the dose may be reduced rapidly to physiological doses (equivalent to prednisolone at 7.5 mg daily) and then reduced more slowly. Assessment of the disease may be needed during withdrawal to ensure that relapse does not occur.

### See Also the Following Articles

ACTH (Adrenocorticotrophic Hormone) • Adrenal Cortex, Physiology • Adrenal Insufficiency • Adrenal Suppression • Glucocorticoid Receptor • Glucocorticoid Resistance Syndromes and States • Glucocorticoids and Immunity • Glucocorticoids in Aging: Relevance to Cognition • Glucocorticoids, Overview • Growth and Glucocorticoids • Nuclear Factor- $\kappa$ B and Glucocorticoid Receptors

### Further Reading

- British Medical Association and The Royal Pharmaceutical Society of Great Britain. (2002). "British National Formulary No. 44," Chap. 6, pp. 348–351. Pharmaceutical Press, London, UK.
- Clark, J. K., Schrader, W. T., and O'Malley, B. W. (1992). Mechanism of steroid hormones. In "Williams Textbook of Endocrinology" (J. D. Wilson and D. W. Foster, eds.), pp. 35–90. W. B. Saunders, Philadelphia, PA.
- Gower, D. B. (1984). Steroid catabolism and urinary excretion. In "Biochemistry of Steroid Hormones" (H. L. J. Makin, ed.), pp. 349–382. Blackwell Science, Oxford, UK.
- Habib, K. E., Gold, P. W., and Chrousos, G. P. (2001). Neuroendocrinology of stress. *Endocrinol. Metab. Clin. N. Am.* **30**, 695–728.
- Honour, J. W. (1993). The adrenal cortex. In "Clinical Paediatric Endocrinology" (C. G. D. Brook, ed.), pp. 434–452. Blackwell Science, Oxford, UK.
- Keller-Wood, M. E., and Dallman, M. (1984). Corticosteroid inhibition of ACTH secretion. *Endocr. Rev.* **5**, 1–24.
- Simpson, E. R., and Waterman, M. R. (1995). Steroid biosynthesis in the adrenal cortex and its regulation by adrenocorticotropin. In "Endocrinology" (L. J. DeGroot, M. Besser, H. G. Burger, J. L. Jameson, D. L. Loriaux, J. C. Marshall, W. D. Odell, J. T. Potts, Jr., and A. H. Rubenstein, eds.), pp. 1630–1641. W. B. Saunders, Philadelphia, PA.
- Stewart, P. M. (1996). 11 $\beta$ -Hydroxysteroid dehydrogenase: Implications for clinical medicine. *Clin. Endocrinol.* **44**, 493–499.
- Yanase, T., Simpson, E. R., and Waterman, M. R. (1991). 17 $\alpha$ -Hydroxylase/17,20-lyase deficiency: From clinical investigation to molecular definition. *Endocr. Rev.* **12**, 91–108.

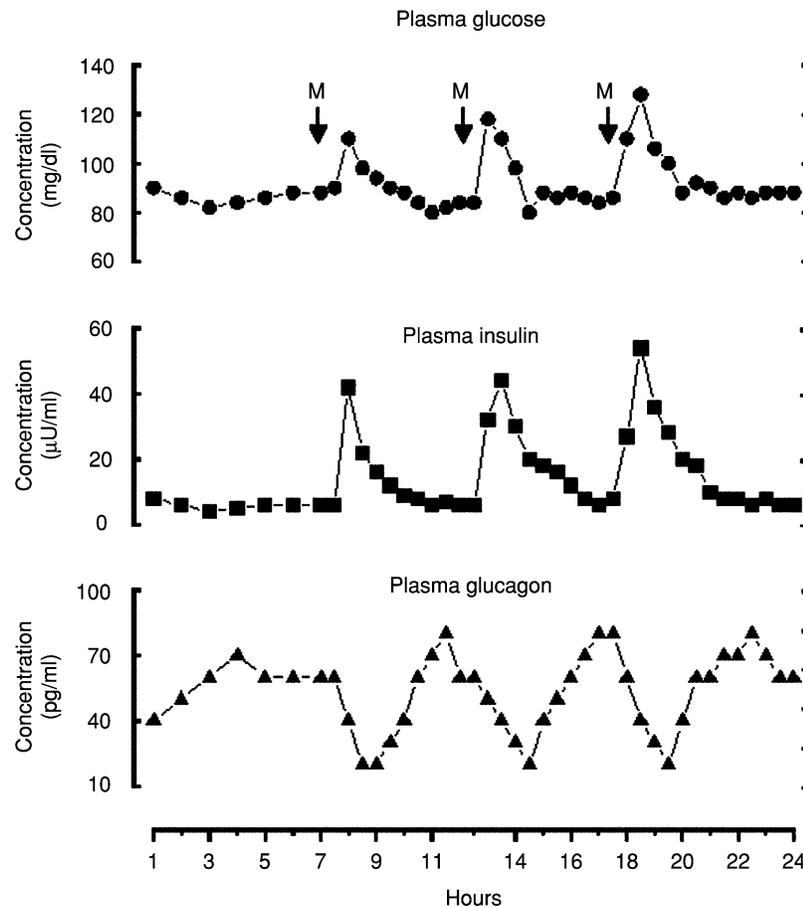


Figure 1 Diurnal changes in plasma glucose, insulin, and glucagon concentrations.

## FACTORS THAT REGULATE GLUCOSE FLUXES

### Insulin

Insulin plays a central role in glucose homeostasis mainly by its action on liver, kidney, adipose tissue, and skeletal muscle. In liver and kidney, it suppresses glucose production by regulating the rate-limiting key enzymes of gluconeogenesis (glucose-6-phosphatase and fructose-1,6-bisphosphatase) and glycogenolysis (glycogensynthase and phosphorylase). In skeletal muscle, its main action is to promote glucose uptake by causing the translocation of Glut-4 transporters to the cell membrane from an intracellular pool. In adipose tissue, insulin will also increase glucose transport and the resultant increase in glycolysis provides  $\alpha$ -glycerophosphate to promote triglyceride formation but its main effect is to inhibit hormone-sensitive lipase. This will lead to a decrease in circulating FFA levels, which will affect glucose production in liver and kidney and glucose uptake in various tissues (Fig. 2).

Binding of insulin to its receptor activates a complex intracellular cascade of autophosphorylation by protein kinases. Intracellular receptor substrates and members of the phosphatidylinositol 3-kinase family seem to be the most important participants. The latter triggers the migration of glucose transporters from an intracellular pool toward the plasma membrane, thereby increasing the number of transporters located on the cellular surface and thus promoting the efficiency of glucose uptake (Fig. 3). This will also result in decreases in intracellular cyclic AMP (cAMP) levels, which mediate the effects of insulin on glucose production and lipolysis. Insulin will also alter the activity of various genes that will affect the amount of certain enzymes. Finally, via its action on amino acid transport, insulin suppresses protein degradation and lipolysis, which will diminish the availability of gluconeogenic precursors and thereby reduce glucose production indirectly.

The main regulator of insulin release is the prevailing plasma glucose concentration. An increase in the

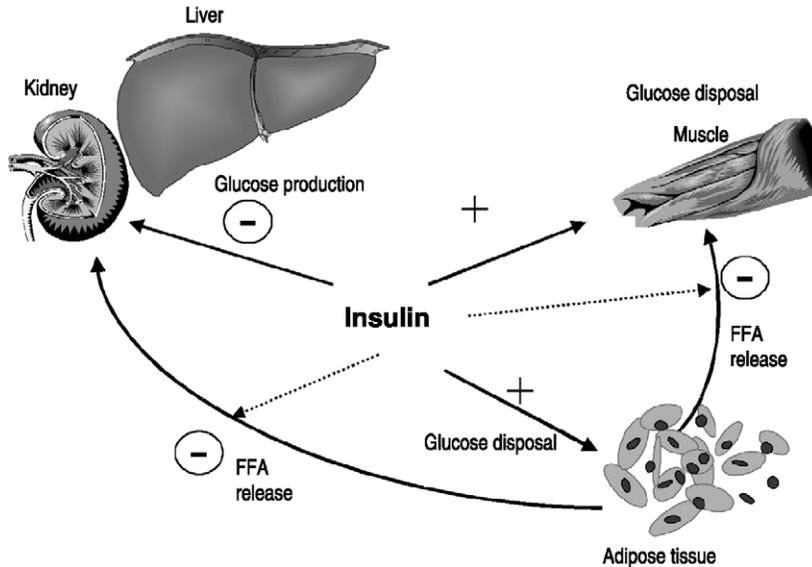


Figure 2 Postprandial insulin secretion; regulation of glucose production and disposal.

plasma glucose concentration stimulates insulin secretion, whereas a decrease in plasma glucose inhibits insulin secretion so that throughout the day, plasma glucose and insulin levels change in parallel (Fig. 1). Amino acids, e.g., arginine, and to a lesser extent FFA can also stimulate insulin secretion. The small intestine produces factors called incretins (e.g., gastrointestinal peptide, glucagon-like peptide-2), which are secreted after meal ingestion and augment postprandial insulin secretion. This explains

why insulin concentrations increase to a greater extent when glucose is given orally than intravenously.

### Glucagon

Glucagon acts solely on the liver, having effects mediated by changes in intracellular cAMP levels that are opposite of those of insulin. Its secretion is also regulated in a reciprocal manner to that of insulin. An increase in plasma glucagon will increase hepatic glucose release within minutes via an increase in glycogen breakdown. Binding of glucagon to its receptor immediately increases intracellular cAMP levels, which increases glycogenolysis and inhibits glycogen synthase by stimulation of phosphorylase and inactivation of glycogen synthase. Prolonged elevation of plasma glucagon can increase gluconeogenesis in the liver, whereas it has no effect on renal gluconeogenesis.

Glucagon secretion is suppressed by increases in plasma glucose and insulin and is increased by hypoglycemia and catecholamines. Amino acids are a potent stimulator of glucagon release. Thus, after protein-rich meals, glucagon release might not be suppressed despite increases in plasma insulin and glucose concentrations.

### Catecholamines

Epinephrine and norepinephrine are released by the adrenal glands and norepinephrine is released from sympathetic nerves during exercise, various stresses (e.g., trauma, infection), and hypoglycemia. They

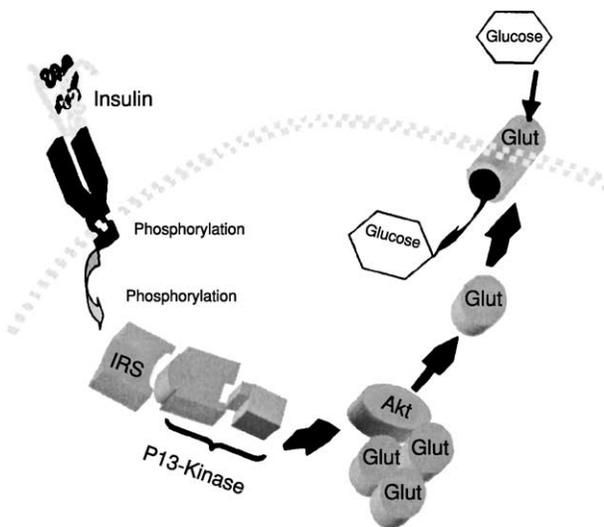


Figure 3 Insulin binding and intracellular cascade. Glut, glucose transporter; IRS, intracellular receptor substrate; PI3-Kinase, phosphatidylinositol 3-kinase.

have complex effects on glucose mediated by both direct and indirect mechanisms. Such actions include stimulation of renal gluconeogenesis, hepatic and muscle glycogenolysis, adipose tissue lipolysis, and glucagon release, which are mediated by  $\beta$ -adrenergic receptors. Catecholamines also inhibit insulin release directly via  $\alpha$ -adrenergic receptors. Indirect effects include suppression of glucose uptake in skeletal muscle due to the elevation of plasma FFA and stimulation of gluconeogenesis in liver and kidney via increases in plasma FFA and gluconeogenic precursors (mainly glycerol from lipolysis and lactate from skeletal muscle glycogenolysis). Along with glucagon, catecholamines are the most important counterregulatory factors protecting against hypoglycemia.

### Growth Hormone, Cortisol, and Thyroid Hormone

Growth hormone, cortisol, and thyroid hormone largely act to regulate the response of target tissues to insulin, glucagon, and catecholamines on a long-term basis, e.g., reducing responses to insulin and increasing responses to glucagon and catecholamines. Under conditions similar to those during which catecholamines are released, growth hormone and cortisol are released and within an hour or two reduce the effectiveness of insulin and enhance the action of glucagon and catecholamines. Prolonged elevation of these hormones, such as is seen in acromegaly and Cushing's syndrome, can cause severe insulin resistance and diabetes mellitus.

### FFA

FFA are a major fuel used by most tissues of the body except the brain, renal medulla, and erythrocytes. Increases in plasma FFA and consequently their uptake into cells have numerous direct and indirect effects that influence glucose homeostasis. These include direct effects on hormone secretion (a moderate stimulating action on insulin secretion and a potent inhibitory action on glucagon and growth hormone) as well as stimulating effects on hepatic and renal gluconeogenesis and an inhibitory effect on muscle glucose uptake. The effects on liver, kidney, and muscle are mediated in part by changes in hormonal environment and competition with glucose as an oxidative fuel (mediated primarily by changes in pyruvate dehydrogenase and interference with insulin signaling pathways, both of these being mediated by coenzyme

A metabolites of FFA). In general, opposite effects are observed when plasma FFA are low.

Circulating FFA levels, like those of glucose, are the net result of changes in FFA entry and exit from plasma. FFA entry into plasma largely depends on the balance between the activation of hormone-sensitive lipase by catecholamines, growth hormone, and cortisol and the inhibition of lipase by insulin. The exit of FFA from plasma is stimulated by insulin.

## THE POSTABSORPTIVE STATE

### General Considerations

In the period after an overnight fast, referred to as the postabsorptive state, plasma glucose ranges between 70 and 110 mg/dl (average 90 mg/dl). This state is considered to represent a steady-state condition since the rate of appearance of glucose approximates its rate of disappearance ( $\sim 10 \mu\text{mol kg}^{-1} \text{min}^{-1}$ ). However, even though removal is often undetectable, the rate of removal is slightly greater than the rate of appearance so that with more prolonged fasting, plasma glucose concentrations decrease. However, even after 72 h of fasting, plasma glucose does normally not decrease below 50 mg/dl (2.9 mM).

### Glucose Utilization

In the postabsorptive state, plasma insulin levels are low and therefore glucose uptake in tissues is largely dependent on tissue needs. The majority of glucose is taken up by the brain ( $\sim 50\%$ ) and is completely oxidized; glucose taken up by muscle ( $\sim 20\%$ ), adipocytes ( $\sim 5\%$ ), erythrocytes ( $\sim 5\%$ ), splanchnic organs ( $\sim 10\%$ ), and kidney ( $\sim 5\%$ ) (Fig. 4) undergoes mostly nonoxidative glycolysis, resulting in the release of 3-carbon precursors (lactate, pyruvate, and alanine), which are used for gluconeogenesis, into the circulation.

### Glucose Production

Glucose production in the postabsorptive state is regulated to match tissue demand, which may increase during exercise or stresses such as infection and trauma. Normally, approximately 50% of the glucose released into the circulation is the result of hepatic glycogenolysis; the remaining 50% is due to gluconeogenesis (30% liver; 20% kidney).

The proportion of glucose produced due to gluconeogenesis increases with the duration of the fast since glycogen stores are rapidly depleted. The liver

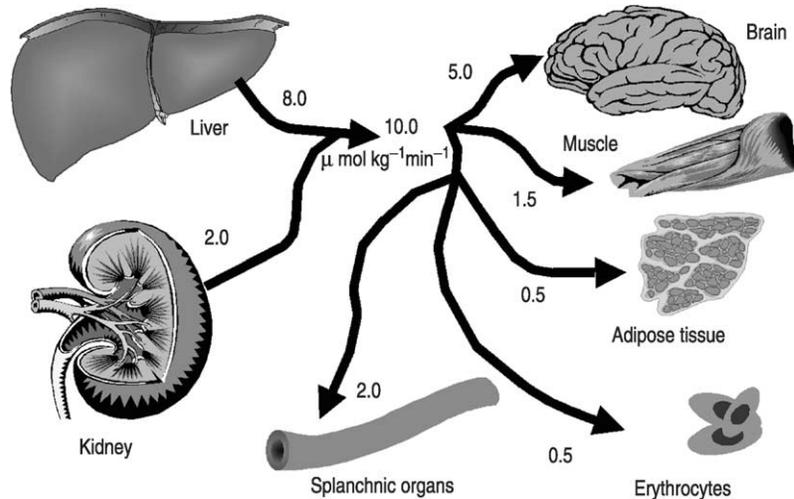


Figure 4 Glucose production and utilization after an overnight fast.

contains a total of 75 g glucose. Assuming that the liver releases glucose from glycogen at a rate of  $5 \mu\text{mol kg}^{-1}\text{min}^{-1}$ , glycogen stores would be depleted within 20 h. Thus, the proportion due to gluconeogenesis must increase so that after 72 h, glucose production by the liver is almost exclusively due to gluconeogenesis. The kidney, in contrast, contains little glycogen stores and the cells that could make glycogen lack glucose-6-phosphatase; consequently, all the glucose released by the kidney is due to gluconeogenesis. (Renal gluconeogenesis increases with fasting to a greater extent than hepatic gluconeogenesis.) Insulin suppresses both hepatic and renal glucose release; however, glucagon promptly increases hepatic glucose release, whereas catecholamines stimulate more renal glucose release.

### Prolonged Fasting

With the duration of fasting, plasma insulin levels decrease, whereas those of glucagon, catecholamines, growth hormone, and cortisol increase. Therefore, the oxidation of glycerol, plasma FFA and FFA products, and the ketone bodies  $\beta$ -hydroxybutyrate and acetoacetate increases. Hepatic glycogen stores become depleted and after 60 h virtually all of glucose released is due to gluconeogenesis. During the first 60–72 h of fasting, the decrease in glucose release is greater than the decrease in glucose uptake, so that plasma glucose levels decrease. At approximately 60 h, with plasma glucose averaging 60 mg/dl, a new pseudo-steady state is achieved (Table I). This stabilization is the basis for the 72 h fast for the diagnosis of patients with hypoglycemia due to insulin-producing tumors of the pancreas (insulinoma). In such patients, insulin

secretion is not appropriately reduced and leads to a further reduction of endogenous glucose production together with increased glucose uptake and consequently to the development of hypoglycemia with plasma glucose levels below 50 mg/dl.

## THE POSTPRANDIAL STATE

### General Considerations

The major function of meal ingestion is to replenish tissue glucose (glycogen) and lipid (triglyceride) stores that have been depleted due to fasting and physical activity. Thus, after meal ingestion, endogenous glucose and FFA release is suppressed, favoring glycogen accumulation. Glucose replaces FFA as the predominant energy fuel as plasma FFA decrease,

Table I Glucose Release and Disposal after Prolonged Fasting (~60 h)

	Glucose release ( $\mu\text{mol kg}^{-1}\text{min}^{-1}$ )		Glucose disposal ( $\mu\text{mol kg}^{-1}\text{min}^{-1}$ )
Overall	6.0	Overall	6.0
Gluconeogenesis	5.5	Oxidation	4.8
Glycogenolysis	0.5	Glycolysis	1.2
Tissues		Tissues	
Liver	2.7	Brain	3.5
Kidney	2.8	Skeletal muscle	1.0
		Splanchnic organs	0.5
		Kidney	0.4
		Adipose tissue	0.2
		Blood cells	0.4

favoring FFA incorporation into triglyceride stores, so that ingested carbohydrate becomes the major fuel used by the body. Various factors, such as the size of the meal, prior physical exercise and duration since the last meal, and composition of the meal, can affect postprandial glucose homeostasis. However, from a practical point of view, the most important factors are changes in insulin and glucagon secretion and their effects on hepatic sequestration of meal carbohydrates, suppression of endogenous glucose production, and finally stimulation of the uptake, storage, glycolysis, and oxidation of glucose in hepatic and posthepatic tissues.

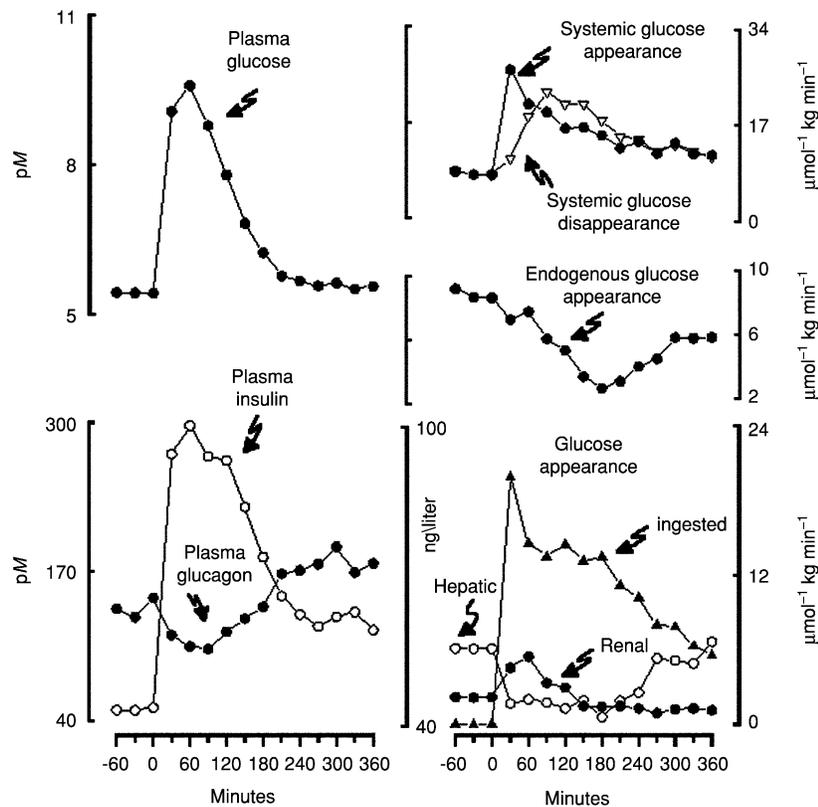
### Postprandial Glucose Concentrations and Fluxes

After a meal, plasma glucose concentrations increase since the rate of appearance of glucose in plasma exceeds the rate of disappearance. Subsequently, plasma glucose decreases when the rate of disappearance exceeds the rate of appearance. The appearance of glucose in plasma represents the sum of glucose from the meal reaching the circulation and the remaining

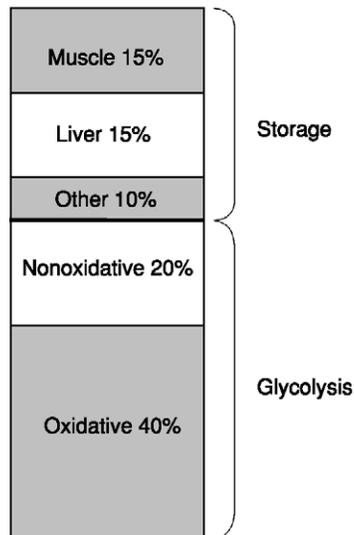
endogenous glucose production. The time courses of these changes are shown in Fig. 5. Endogenous glucose release is suppressed  $\sim 60\%$  after a meal. The appearance of ingested glucose is detected within 15 min after a meal, reaches a maximum at  $\sim 60$  min, and decreases gradually thereafter. Only 75% of the glucose in a meal reaches the systemic circulation; the remaining 25% is sequestered by the splanchnic bed.

Of a theoretical meal containing 100 g of glucose, 20–30% is initially extracted in the splanchnic bed. At least half of this is taken up by the liver and incorporated into hepatic glycogen; the remainder is probably released as lactate due to hepatic glycolysis. Of the glucose in the meal that reaches the systemic circulation, approximately 30–40% is taken up by skeletal muscle to be initially oxidized in favor of FFA and later to be stored as glycogen. Little of the glucose taken up by muscle is released as lactate or other gluconeogenic substrates into the circulation. Of the remaining glucose released into the systemic circulation, 20% is taken up by the brain, 10% by the kidney, and 5% by adipose tissue.

Approximately 40% of the glucose disposed of after a meal is stored predominantly as glycogen in



**Figure 5** Postprandial changes in plasma glucose, insulin, glucagon concentrations, rates of plasma glucose appearance/disappearance, and hepatic and renal glucose production.



**Figure 6** Contribution of storage and glycolysis to the disposal of glucose after a meal.

liver and muscle and, to a lesser extent, as triglycerides in adipose tissue. The remaining 60% is undergoes glycolysis either oxidatively to  $\text{CO}_2$  and  $\text{H}_2\text{O}$  (~40%) or nonoxidatively to lactate (~20%) (Fig. 6). That a substantial amount of glucose undergoes nonoxidative glycolysis is not surprising if one considers that glycolysis is needed to provide 3-carbon fragments for gluconeogenesis and precursors for the indirect pathway of hepatic glycogen formation.

Renal glucose production initially increases after a meal (Fig. 5). The physiological role of this increase still needs to be elucidated. Teleologically, postprandial renal glucose release may facilitate efficient liver glycogen repletion by permitting substantial suppression of hepatic glucose release.

It can be readily appreciated that three main factors regulate postprandial glucose levels: suppression of hepatic glucose release, hepatic sequestration of the ingested glucose, and uptake of glucose from the systemic circulation. Initial suppression of endogenous glucose release and hepatic sequestration depend largely on the reciprocal secretion of glucagon and insulin. Insulin increases the number of glucose transporters in muscle and thereby increases glucose fractional extraction, an index of the efficiency of glucose uptake; glucose fractional extraction by brain and other insulin independent tissues actually decreases.

## SUMMARY

In most circumstances, regulation of glucose production is more important than regulation of glucose utilization in determining plasma glucose

concentrations. Recall that in the fasting state, 80% of glucose utilization is insulin independent. Insulin levels are low and are needed only to suppress excessive endogenous glucose production and lipolysis. Another example of the importance of glucose production is fasting hyperglycemia in type 2 diabetes. Rates of glucose utilization are generally normal, whereas rates of glucose production are increased. Of interest is the finding that renal glucose release initially increases after a meal, whereas hepatic glucose release decreases; these reciprocal changes permit efficient repletion of glycogen stores. This finding and other observations have provided convincing evidence for the concept of hepatorenal reciprocity. According to this concept, when the release of glucose by either liver or kidney is reduced, the other organ will increase its glucose release to maintain euglycemia. Similar reciprocal changes are found during recovery from insulin-induced hypoglycemia in patients with type 2 diabetes. The strongest support for the concept of hepatorenal reciprocity has been provided by the observation that during the anhepatic stage of liver transplantation (when liver glucose release is absent), the kidney increases its release of glucose threefold so that hypoglycemia does not occur.

## See Also the Following Articles

Catecholamines • Diabetes, Type 2 • Glucagon and Glucagon-like Peptides • Glucose, Impaired Tolerance • Glucose Toxicity • Insulin Secretion, Physiology

## Further Reading

- Boden, G. (1997). Role of fatty acids in the pathogenesis of insulin resistance and NIDDM. *Diabetes* **46**, 3–10.
- Cahill, G. (1970). Starvation in man. *N. Engl. J. Med.* **282**, 668–675.
- Dinneen, S., Gerich, J., and Rizza, R. (1992). Carbohydrate metabolism in noninsulin-dependent diabetes mellitus. *N. Engl. J. Med.* **327**, 707–713.
- Ekberg, K., Landau, B., Wajngot, A., *et al.* (1999). Contributions by kidney and liver to glucose production in the postabsorptive state and after 60 h of fasting. *Diabetes* **48**, 292–298.
- Gerich, J. (1993). Control of glycaemia. *Bailliere Clin. Endocrinol. Metab.* **7**, 551–586.
- Havel, R. (1972). Caloric homeostasis and disorders of fuel transport. *N. Engl. J. Med.* **287**, 1186–1192.
- Joseph, S., Heaton, N., Potter, D., *et al.* (2000). Renal glucose production compensates for the liver during the anhepatic phase of liver transplantation. *Diabetes* **49**, 450–456.
- Kelley, D., Mitrakou, A., Marsh, H., *et al.* (1988). Skeletal muscle glycolysis, oxidation, and storage of an oral glucose load. *J. Clin. Invest.* **81**, 1563–1571.
- Kruszynska, Y., Mulford, M., Yu, J., *et al.* (1997). Effects of nonesterified fatty acids on glucose metabolism after glucose ingestion. *Diabetes* **46**, 1586–1593.

- Landau, B., Wahren, J., Chandramouli, V., *et al.* (1996). Contributions of gluconeogenesis to glucose production in the fasted state. *J. Clin. Invest.* **98**, 378–385.
- Magnusson, I., Rothman, D., Gerard, D., *et al.* (1995). Contribution of hepatic glycogenolysis to glucose production in humans in response to a physiological increase in plasma glucagon concentration. *Diabetes* **44**, 185–189.
- Marin, P., Høgh-Kristiansen, I., Jansson, S., *et al.* (1992). Uptake of glucose carbon in muscle glycogen and adipose tissue triglycerides *in vivo* in humans. *Am. J. Physiol.* **263**, E473–E480.
- McGarry, J. (1998). Glucose–fatty acid interactions in health and disease. *Am. J. Clin. Nutr.* **67**, 500S–504S.
- Meyer, C., Dostou, J., Nadkarni, V., and Gerich, J. (1998). Effects of physiological hyperinsulinemia on systemic, renal and hepatic substrate metabolism. *Am. J. Physiol.* **275**, F915–F921.
- Meyer, C., Dostou, J., Welle, S., and Gerich, J. (2002). Role of human liver, kidney and skeletal muscle in postprandial glucose homeostasis. *Am. J. Physiol. Endocrinol. Metab.* **282**, E419–E427.
- Stumvoll, M., Chintalapudi, U., Perriello, G., *et al.* (1995). Uptake and release of glucose by the human kidney: Postabsorptive rates and responses to epinephrine. *J. Clin. Invest.* **96**, 2528–2533.
- Taylor, R., Magnusson, I., and Rothman, D. (1997). Direct assessment of liver glycogen storage by <sup>13</sup>C nuclear magnetic resonance spectroscopy and regulation of glucose homeostasis after a mixed meal in normal subjects. *J. Clin. Invest.* **97**, 126–132.
- Taylor, R., Price, T., Katz, L., *et al.* (1993). Direct measurement of change in muscle glycogen concentration after a mixed meal in normal subjects. *Am. J. Physiol.* **265**, E224–E229.

after as little as a few hours of exposure to high levels of glucose and, in these earlier stages, the damage is reversible. Thus, the impaired insulin secretion observed in early human type 2 diabetes can improve significantly with control of hyperglycemia. Ultimately, however, the damage to the beta cells becomes irreversible (“beta cell failure”).

It is likely that different mechanisms come into play as these processes evolve. The initial events, such as desensitization to glucose, may be related to changes in the levels and/or activities of transcription factors and other key regulatory molecules involved in glucose sensing, whereas the subsequent loss of beta cell function involves down-regulation of insulin mRNA and may proceed to apoptotic cell loss. It is possible, however, that these events represent a continuum that shares a common signaling mechanism. Other workers, for example, Unger, have noted that other nutrients that are found in excess in type 2 diabetes, such as free fatty acids, can also impair beta cell function. These hypotheses—glucose toxicity and “lipotoxicity”—are not mutually exclusive. For example, both free fatty acids and glucose can lead to chronic oxidative stress and both can activate the hexosamine signaling pathway, both being potential mechanisms underlying beta cell failure.

### GLUCOSE-INDUCED INSULIN RESISTANCE IN TYPE 1 AND TYPE 2 DIABETES

Another of the pathophysiologic hallmarks of type 2 diabetes is insulin resistance. Although it is considered by many to be involved in the initial pathogenesis of the disease, there are also numerous studies that illustrate that insulin resistance can be an acquired defect, the result of excess nutrient delivery to tissues. The laboratories of Rossetti, Yki-Jarvinen, and DeFronzo have pioneered work in this area and have also reviewed the topic extensively. As was the case for decreased insulin secretion, the glucose-induced insulin resistance has been convincingly demonstrated in a variety of experimental systems and in humans, nondiabetic as well as those with type 1 or type 2 diabetes. Most often, the experimental paradigm involves decreased insulin-mediated glucose disposal after a period of hyperglycemia, but the definition has been broadened to include any defect in insulin signaling or responsiveness observed after treatment with high concentrations of glucose.

The ultimate mediators of glucose-induced insulin resistance are not known, although much work has

illuminated some of the mechanisms involved. In skeletal muscle, the effect of hyperglycemia is decreased insulin-stimulated glucose uptake caused by a failure of translocation of the specific glucose transporter GLUT4 to the plasma membrane. Why this occurs is less clear. Interest has focused on the early steps in insulin signal transduction, specifically the insulin receptor substrate/phosphatidylinositol 3-kinase/Akt (PKB) pathways, and several defects have been described in these pathways in glucose-induced insulin resistance. However, there is no consistent or universally accepted mechanism that can completely explain the phenomenon.

Glucose-induced insulin resistance shares many features with that induced by excess levels of free fatty acids. This has given rise to the hypothesis that glucose-induced insulin resistance might be more appropriately termed nutrient-induced insulin resistance. A possible common denominator that could explain both glucose- and fatty acid-induced insulin resistance is hexosamine signaling, because high levels of both glucose and fatty acids result in increased hexosamine pathway products. This hypothesis is also attractive in that it would predict that insulin resistance would occur in the presence of excess nutrient *flux* even before nutrient *levels* were significantly increased in plasma. Thus, insulin resistance by this mechanism would be expected to occur before overt hyperglycemia and there could be a unifying mechanism to explain insulin resistance in obesity as well as in clinical diabetes.

Glucose toxicity for insulin action is manifest in the clinical setting in two major ways that deserve mention. First, the typical patient presenting with a prolonged history of weight loss, polyuria, and severe hyperglycemia is likely to improve his or her glucose disposal rate once the hyperglycemia is brought under control; that is, the glucose toxicity will reverse. Thus, it will be harder to achieve glycemic control than to maintain glycemic control. This means that the therapy needed to attain normoglycemia may cause hypoglycemia a few days later as the glucose toxicity component resolves. Conversely, it has also been shown that an initial period of normoglycemia produced by intense insulin therapy will, by breaking the vicious cycle of glucose toxicity, allow better responsiveness to oral agents. Thus, there will be fewer oral agent “failures” after such a period of normoglycemia.

A second major clinical consequence of glucose-induced insulin resistance is that huge doses of insulin may be required to lead to disposal of the nutrient load from a meal and in some cases even the maximal

glucose disposal rate of the individual can be exceeded, meaning that no amount of insulin will lead to normoglycemia. Interestingly, Yki-Jarvinen and others have noted that in such situations, muscle glycogen content and net glucose flux into muscle are essentially normal. Furthermore, there are few if any ill effects of diabetes on skeletal muscle function. This has led to the proposal that nutrient-induced insulin resistance may be an adaptation to excess nutrient flux. That is, skeletal muscle is able to autoregulate glucose uptake by down-regulating glucose transporters in the face of hyperglycemia. If this is the case, it may be more appropriate to view glucose-induced insulin resistance in terms of normal physiology rather than as toxicity.

### MEDIATION OF DIABETIC COMPLICATIONS BY GLUCOSE

There is substantial evidence that most if not all of the classic complications of diabetes are mediated by hyperglycemia and/or excessive nutrient flux. This was most clearly demonstrated in the Diabetes Control and Complications Trial for type 1 diabetes and by the “Kumamoto Study” and the United Kingdom Prospective Diabetes Study for type 2 diabetes. Although not usually referred to as glucose toxicity, some of the pathways leading to these complications—nephropathy, neuropathy, and retinopathy—may be shared and it may be informative to group them together with beta cell failure and insulin resistance as “adverse consequences of excess nutrient flux.”

### PROPOSED MECHANISMS FOR GLUCOSE TOXICITY

The precise mechanisms and mediators of glucose toxicity are not known. Given the complex regulation of glucose metabolism, the many pathways of known regulation by glucose, and the many fates of intra- and extracellular glucose, it is likely that there will be multiple pathways through which glucose exerts its adverse effects. Some of the leading proposals can be very briefly summarized as follows:

- Nonenzymatic glycation of proteins. Covalent attachment of free glucose to amino groups in proteins or other macromolecules can occur via formation of a Schiff's base. These can subsequently rearrange, leading to the elaboration of advanced glycation end-products that can have toxic effects on

cells. Nonenzymatic glycation is contrasted to the enzymatic linkage of carbohydrates (usually uridine diphosphate amino sugars) to specific residues of proteins. Given that the effects of high levels of glucose on insulin secretion and insulin action can be seen in hours to a few days and that these effects are likely to be mediated intracellularly, it is unlikely that this process plays a large role in beta cell failure or insulin resistance.

- Aldose reductase pathway. Sorbitol, generated by the enzyme aldose reductase, has been postulated to act normally as an intracellular osmolyte to buffer changes in extracellular osmolality. In the presence of chronic hyperglycemia, however, untoward effects of sorbitol accumulation, such as depletion of myoinositol or increased oxidative stress, may occur. This pathway has not been specifically linked to insulin resistance or beta cell failure.

- Protein kinase C. Hyperglycemia-induced activation of diacylglycerol/protein kinase C-dependent pathways has been postulated to lead to changes in gene regulation that could play a role in diabetic complications, although no studies provide a specific link to the classic glucose toxicity effects.

- Hexosamines. Since the demonstration by Marshall that glucosamine can induce insulin resistance in adipocytes, and his discovery of the hexosamine signaling pathway, much evidence that this pathway plays an important role in nutrient sensing and adaptation to excess nutrient flux has accumulated. In particular, it has been shown in several animal and cell culture models that high concentrations of glucose lead to high levels of products of the hexosamine biosynthesis pathway that in turn can mimic very well the diabetic phenotypes of beta cell failure and glucose-induced insulin resistance. The hypothesis is further attractive in that the pathway may provide a common mechanism for the similar effects of free fatty acids, which also result in increased hexosamine flux. The signaling function of the pathway is thought to occur by the enzymatic O-linked glycosylation of cytosolic signaling proteins. This dynamic process occurs on residues otherwise used for regulatory phosphorylation, so that high levels of hexosamine flux can generate a dominant signal downstream to abrogate hormone signaling in situations of excess nutrients. Thus, insulin would no longer be able to stimulate glucose uptake or glycogen synthesis if intracellular fuel stores were already replete, protecting the cell from excess nutrient accumulation and allowing the excess calories to be shunted to adipocytes.

- Oxidative stress. This has been shown to be a factor in a large variety of tissues that are harmed by the diabetic milieu. Brownlee has proposed that many features of glucose toxicity and diabetic complications may be functionally linked to this mechanism, including activation of the hexosamine signaling pathway.

### See Also the Following Articles

Diabetes, Type 1 • Diabetes, Type 2 • Glucose, Impaired Tolerance • Glucose Physiology, Normal • Insulin-Resistant States, Role of Free Fatty Acids (FFA) • Obesity and Diabetes, Regulation of Food Intake

### Further Reading

- Du, X.-L., Edelstein, D., Rossetti, R., Fantus, I. G., Goldberg, H., Ziyadeh, F., Wu, J., and Brownlee, M. (2000). Hyperglycemia-induced mitochondrial superoxide overproduction activates the hexosamine pathway and induces plasminogen activator inhibitor-1 expression by increasing Sp1 glycosylation. *Proc. Natl. Acad. Sci. USA* **97**, 12222–12226.
- The Diabetes Control and Complications Trial Research Group. (1993). The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N. Engl. J. Med.* **329**, 977–986.
- Leahy, J. L., Bonner-Weir, S., and Weir, G. C. (1992). Beta-cell dysfunction induced by chronic hyperglycemia: Current ideas on mechanism of impaired glucose-induced insulin secretion. *Diabetes Care* **15**, 442–455.
- Marshall, S., Bacote, V., and Traxinger, R. R. (1991). Discovery of a metabolic pathway mediating glucose-induced desensitization of the glucose transport system. *J. Biol. Chem.* **266**, 4706–4712.
- McClain, D. A. (2001). Hexosamines as mediators of nutrient sensing: Relevance to obesity, insulin resistance, and diabetes. *Curr. Opin. Endocrinol. Diabetes* **8**, 186–191.
- Robertson, R. P., Olson, L. K., and Zhang, H. J. (1994). Differentiating glucose toxicity from desensitization: A new message from the insulin gene. *Diabetes* **43**, 1085–1089.
- Rossetti, L., Giaccari, A., and DeFronzo, R. A. (1990). Glucose toxicity. *Diabetes Care* **6**, 610–630.
- Rossetti, L. (2000). Perspective: Hexosamines and nutrient sensing. *Endocrinology* **141**, 1922–1925.
- UK Prospective Diabetes Study (UKPDS) Group. (1998). Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet* **352**, 837–853.
- Yki-Jarvinen, H. (1992). Glucose toxicity. *Endocr. Rev.* **13**, 415–431.

Table 1 Diagnostic Criteria for Impaired Glucose Levels

Organization	Impaired fasting glucose <sup>a</sup>	Impaired glucose tolerance <sup>b</sup>
World Health Organization	Not defined	140–199 mg/dl 7.8–11.1 mmol/liter
American Diabetes Association	110–125 mg/dl 6.1–6.9 mmol/liter	140–199 mg/dl 7.8–11.1 mmol/liter

<sup>a</sup>After an 8 to 12 h overnight fast.

<sup>b</sup>2 h value on 75 g OGTT.

Prospective studies in people with IGT have revealed rates of progression to diabetes (usually type 2) in the range of 5–10% per year. The rates translate to a three- to sevenfold increase in the risk of diabetes relative to people with normal glucose levels. Factors that may alter the risk of diabetes among people with IGT vary from study to study, but include glucose levels, insulin resistance, B cell defect, obesity, lipid levels, and blood pressure (greater values for each factor indicate increased risk). Clinical cardiovascular events have been reported to be more frequent in people with IGT than in people with normal glucose tolerance in some, but not all studies. In general, the excess risks were accounted for by the presence of cardiovascular risk factors other than mild hyperglycemia.

## PATHOPHYSIOLOGY

In general, glucose levels rise above normal when insulin secretion from the pancreas is insufficient to meet the insulin needs of tissues that make glucose (the liver) or take it up in response to insulin (skeletal muscle and adipose tissue). Diabetes reflects a large imbalance between these two factors. IGT represents a relatively milder imbalance, albeit not a trivial one. Cross-sectional data from the Insulin Resistance and Atherosclerosis Study indicate that people with IGT have ~60% less insulin secretion for their degree of insulin resistance than people with normal glucose tolerance. Similar findings have been reported for women with gestational diabetes, a separate form of impaired glucose tolerance. Causes of insufficient insulin secretion vary among people with IGT. Some people have circulating autoimmune markers directed at the pancreatic islets or insulin-secreting B cells. Those individuals appear to have IGT as part of evolving type 1 diabetes. Other individuals with IGT (the majority) are obese or have other reasons to have increased insulin requirements of their tissues (“insulin resistance”). They may make considerable insulin, but less than equally insulin-resistant people

with normal glucose tolerance. Evidence from the Troglitazone in Prevention of Diabetes (TRIPOD) study indicates that, at least in Hispanic Americans, the increased insulin secretory demands that are imposed on pancreatic B cells by chronic insulin resistance cause loss of B cell function. Treatment of insulin resistance can delay or prevent progressive B cell failure and diabetes.

The insulin resistance from which IGT frequently evolves is at the center of a cluster of clinical conditions that are collectively known as the insulin resistance syndrome. Like IGT, atherosclerosis is a component of this syndrome. Many mechanisms have been proposed as links between insulin resistance and atherosclerosis, including hyperinsulinemia, dyslipidemia (especially elevated triglycerides and low levels of high-density lipoprotein cholesterol), hypertension, impaired fibrinolysis, and chronic inflammation. In people with IGT, mild hyperglycemia could promote atherosclerosis as well, although solid evidence for this in humans (i.e., lowering of mild hyperglycemia to mitigate atherosclerosis) is lacking.

## TREATMENT

There is no standard treatment for IGT. Four different randomized trials to evaluate the effects of different interventions on the risk of diabetes in people with IGT have been completed. In two of them [the Finnish Diabetes Prevention Study (DPS) and U.S. Diabetes Prevention Program (DPP)], intensive lifestyle interventions were designed to achieve modest weight reduction (e.g., 7% of body weight in the DPP) and a modest increase in physical activity (e.g., brisk walking or the equivalent for 150 min per week in the DPP) in adult men and women with IGT. The risk of diabetes was reduced 58% compared to the control group in each study. The U.S. DPP also included an arm in which subjects were randomized to metformin 1500 mg/day instead of intensive lifestyle intervention. The risk of diabetes in the metformin arm was reduced by 31% compared to the control group. In the STOP-NIDDM trial, acarbose (100 mg three times daily) was given to adult men and women with IGT. The incidence of diabetes was reduced by 25% compared to placebo-treated subjects. Survival curves in the Finnish DPS, U.S. DPP, and STOP-NIDDM trials revealed a slowing of the onset of diabetes in treated arms rather than true diabetes prevention. In the TRIPOD study cited above, a medium dose of an insulin-sensitizing thiazolidinedione drug (troglitazone, which is no longer available for clinical use) reduced the incidence of diabetes by

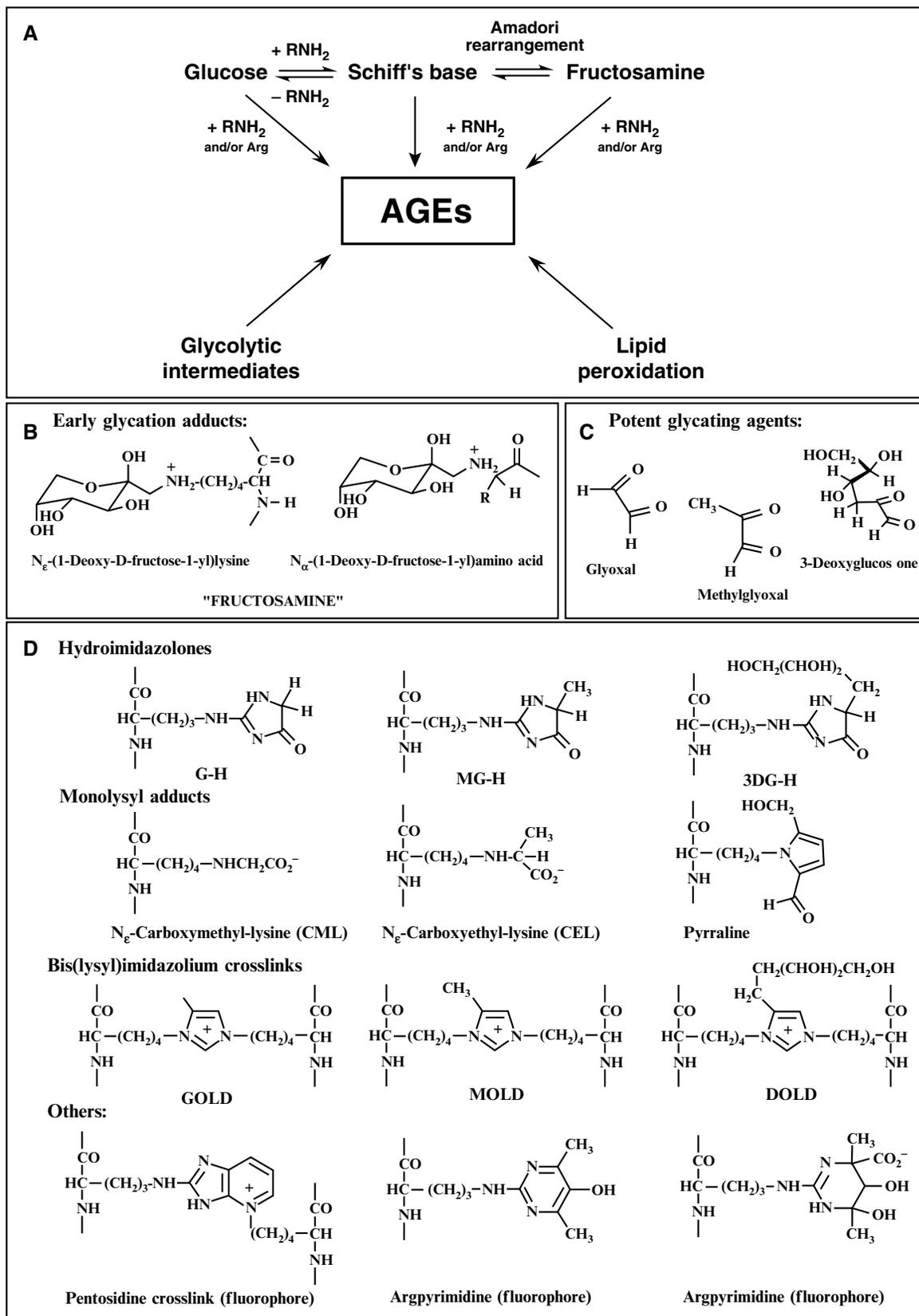
55% in Hispanic American women with prior gestational diabetes. Women who responded to the drug with reduced endogenous insulin requirements had complete stabilization of pancreatic B cell function and glucose levels for 4.5 years, indicating true diabetes prevention during that period of time. These four studies reveal that it is possible to delay or prevent the onset of diabetes in people with IGT using behavioral and/or pharmacological interventions. Efficient strategies for screening to find people with IGT and optimal approaches to clinical management remain to be defined. Likewise, whether treatment of mild hyperglycemia per se in IGT can reduce cardiovascular events independent of changes in other cardiovascular risk factors remains to be determined. People with IGT should be evaluated and treated for standard cardiovascular risk factors.

### See Also the Following Articles

Atherosclerosis • Cardiovascular Disease in Diabetes • Diabetes, Type 1 • Diabetes, Type 2 • Glucose Physiology, Normal • Glucose Toxicity • Insulin Secretion: Functional and Biochemical Aspects

### Further Reading

- American Diabetes Association. (1997). Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care* **20**, 1183–1197.
- Buchanan, T. A., Xiang, A. H., Peters, R. K., Kjos, S. L., Marroquin, A., Goico, J., Ochoa, C., Tan, S., Berkowitz, K., Hodis, H. N., and Azen, S. P. (2002). The TRIPOD Study: Preservation of pancreatic B-cell function and prevention of type 2 diabetes by pharmacological treatment of insulin resistance in high-risk Hispanic women. *Diabetes*, **51**, 2796–2803.
- Chiasson, J. L., Josse, R. G., Gomis, R., Hanefeld, M., Karasik, A., and Laakso, M. (2002). Acarbose for prevention of type 2 diabetes: The STOP-NIDDM randomised trial. *Lancet* **359**, 2072–2077.
- Diabetes Prevention Program Research Group. (2002). Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N. Engl. J. Med.* **346**, 393–403.
- Harris, M. I., Flegal, K. M., Cowie, C. C., Eberhardt, M. S., Goldstein, D. E., Little, R. R., Wiedmeyer, H.-M., and Byrd-Holt, D. D. (1998). Prevalence of diabetes, impaired fasting glucose, and impaired glucose tolerance in U.S. adults. *Diabetes Care* **21**, 518–524.
- Haffner, S. M., Agostino, R. D., Jr., Saad, M. F., O'Leary, D. H., Savage, P. J., Rewers, M., Selby, J., Bergman, R. N., and Mykkanen, L. (2000). Carotid artery atherosclerosis in type-2 diabetic and nondiabetic subjects with and without symptomatic coronary artery disease (The Insulin Resistance Atherosclerosis Study). *Am. J. Cardiol.* **85**, 1395–1400.
- Tuomilehto, J., Lindstrom, J., Eriksson, J. G., Valle, T. T., Hamalainen, H., Ilanne-Parikka, P., Keinanen-Kiukaanniemi, S., Laakso, M., Louheranta, A., Rastas, M., Salminen, V., and Uusitupa, M. (2001). Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *N. Engl. J. Med.* **344**, 1343–1350.
- World Health Organization. (1998). Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus: Provisional report of a WHO consultation. *Diabetic Med.* **15**, 539–545.



efficient enzymatic detoxification of these  $\alpha$ -oxoaldehydes. Glyoxal and methylglyoxal are detoxified to glycolate and D-lactate, respectively, by the glutathione-dependent glyoxalase system. 3-Deoxyglucosone is detoxified by NADPH-dependent 3-deoxyglucosone reductase to 3-deoxyfructose. In addition, there is a high cysteinyl thiol pool that binds  $\alpha$ -oxoaldehydes reversibly and thereby suppresses their irreversible reactions to form AGEs. Fructosamine residues are removed from proteins by phosphorylation to fructoseamine-3-phosphate, catalyzed by fructosamine 3-phosphokinase, and degraded to 3-deoxyglucosone. Once formed *in vivo*, AGEs must be repaired or replaced and degraded. The reactions of saccharide derivatives with proteins, nucleotides, and phospholipids to form AGEs occurs intra- and extracellularly. Inside cells, the impact of glycation is countered by high turnover (short half-life) of many cellular proteins, phospholipids, and RNA as well as by mechanisms of DNA repair. Degradation of extracellular glycationed proteins occurs by specific recognition by receptors, internalization, and proteolytic processing of the glycationed protein ligand.

### PROTEIN GLYCATION IN MONITORING GLYCEMIC CONTROL

Fructosamine concentrations change in response to persistent hyperglycemia but are unresponsive to short postprandial hyperglycemia. The assay of fructosamine residues is a diagnostic marker of medium-term glycemic control in diabetes mellitus. Approximately 6 to 15% of the human serum albumin (HSA) is glycationed with a fructosamine *in vivo*, primarily at Lys-525. Glycationed HSA reflects glycemic control during the 2 to 3 weeks preceding analysis. Glycationed hemoglobin HbA<sub>1</sub> accounts for approximately 7.5% of total hemoglobin in normal human individuals. The various forms are designated HbA<sub>1a1</sub>, HbA<sub>1a2</sub>, HbA<sub>1b</sub>, and HbA<sub>1c</sub>. HbA<sub>1c</sub> is the most abundant of the minor components in normal human red blood cells *in vivo*, accounting for approximately 5% of total hemoglobin. It too is a mixture of mostly the fructosamine adducts of the  $\beta$ -val-1 (60%) and  $\alpha$ -Lys-61 (40%). HbA<sub>1a1</sub> and HbA<sub>1a2</sub> are the  $\beta$ -chain N-terminal adducts of

fructose-1,6-bisphosphate and glucose-6-phosphate, respectively. HbA<sub>1b</sub> is thought to result from a deamidation in the  $\beta$ -chain of HbA<sub>1c</sub>. The measurement of HbA<sub>1c</sub> reflects glycemic control over the 6 to 8 weeks preceding analysis. In diabetes mellitus, the concentration of fructosamine of serum proteins is typically increased twofold relative to that of normal human individuals (approximately 5 vs 2 nmol/mg protein). Glycationed hemoglobin HbA<sub>1c</sub> is typically increased from 5% of total hemoglobin in normal controls to  $\leq 7\%$  (good glycemic control), 7 to 10% (moderate control), and  $> 10\%$  (poor control). There is no significant increase of fructosamine and glycationed hemoglobin in prediabetic impaired glucose tolerance.

### PROTEIN GLYCATION A RISK MARKER OF DIABETIC COMPLICATIONS

Glycationed hemoglobin is a risk factor for the development of chronic clinical complications, probably as a surrogate indicator of hyperglycemia. Increased concentrations of AGE have been associated with diabetic complications in cross-sectional studies. AGE content of skin collagen (CML, pentosidine, and others) was shown to be a risk marker for microvascular complications of diabetes. Serum protein AGE content at baseline was shown to be a risk predictor for development of early morphological kidney damage leading to diabetic nephropathy.

### FUNCTIONAL CONSEQUENCES OF PROTEIN GLYCATION

Glycation occurs inside and outside of cells. Glycation of nucleotides in DNA normally induces DNA repair but may induce apoptosis if glycation is excessive. Glycation of basic phospholipids may induce changes in membrane structure and lipid peroxidation. Glycation of cellular proteins produces changes in structure and loss of enzymatic activity. These effects are countered by protein degradation and renewal. Glycation of the extracellular matrix produces changes in macromolecular structure affecting matrix-matrix and matrix cell interactions associated with decreased

---

**Figure 1** Formation of early glycation adducts and advanced glycation end products: (A) glycation adducts, (B) early glycation adducts (fructosamines), (C) reactive  $\alpha$ -oxoaldehydes in physiological glycation process, and (D) molecular structures of AGEs. 3DG-H, N<sub>8</sub>-[5-(2,3,4-trihydroxybutyl)-5-hydro-4-imidazolone-2-yl]ornithine; DOLD, 3-deoxyglucosone-derived lysine dimer, 1,3-di(N<sup>ε</sup>-lysino)-4-(2,3,4-trihydroxybutyl)-imidazolium salt; G-H, N<sub>8</sub>-(5-hydro-4-imidazolone-2-yl)ornithine; GOLD, glyoxal-derived lysine dimer, 1,3-di(N<sup>ε</sup>-lysino)imidazolium salt; MG-H, N<sub>8</sub>-(5-hydro-5-methyl-4-imidazolone-2-yl)-ornithine; MOLD, methylglyoxal-derived lysine dimer, 1,3-di(N<sup>ε</sup>-lysino)-4-methyl-imidazolium salt.

elasticity and increased fluid filtration across arterial walls and endothelial cell adhesion. When the concentration of AGEs increases above a critical level, cell surface AGE receptors become activated. This is associated with increased expression of extracellular matrix proteins, vascular adhesion molecules, cytokines, and growth factors. Depending on the cell type and concurrent signaling, this is associated with chemotaxis, angiogenesis, oxidative stress, and cell proliferation or apoptosis. These processes are thought to contribute to disease mechanisms associated with the development of diabetic complications. Strategies for therapeutic interventions to counter the effects of glycation by preventing increased  $\alpha$ -oxoaldehyde formation in hyperglycemia, scavenging of  $\alpha$ -oxoaldehydes, and blocking of AGE receptor activation are in experimental development, and all hold promise for the eventual prevention of diabetic complications clinically.

## POLYOL PATHWAY AND DIABETIC COMPLICATIONS

Aldose reductase (ALR2) is the first enzyme in the polyol pathway (Fig. 2). It catalyzes the NADPH-dependent reduction of glucose and many other carbonyl compounds to sorbitol and the corresponding alcohol derivatives. ALR2 has a low affinity (high  $K_M$ ) for glucose. Hence, there is an increased flux of glucose metabolism via the polyol pathway in hyperglycemia; in human red blood cells, glucose metabolism increases from 3% of total glucose metabolism in normoglycemia to 11% in hyperglycemia. Sorbitol produced in the first step of the polyol pathway is oxidized by  $NAD^+$ -dependent sorbitol dehydrogenase (SDH) to fructose in the second step. The complete traverse of the polyol pathway involves oxidation of NADPH to  $NADP^+$  and reduction of  $NAD^+$  to NADH, that is, a net hydride transfer from NADPH in the pentosephosphate pathway to  $NAD^+$  in the Embden–Meyerhof pathway. Therefore, there is a net reduction of  $NAD^+$  to NADH in the Embden–Meyerhof pathway and an increase in the cytosolic NADH/ $NAD^+$  ratio (Fig. 2). The effects of activation of the polyol pathway in diabetes have been attributed to osmotic stress arising from intracellular accumulation of sorbitol, decreased  $Na^+K^+$  ATPase activity, *in situ* inhibition of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) by change in the NADH/ $NAD^+$  ratio, and *in situ* inhibition of glutathione reductase by depletion of NADPH-exacerbating intracellular oxidative stress. Diabetes induced in homozygous knock-out mice deficient in ALR2 neither decreased the

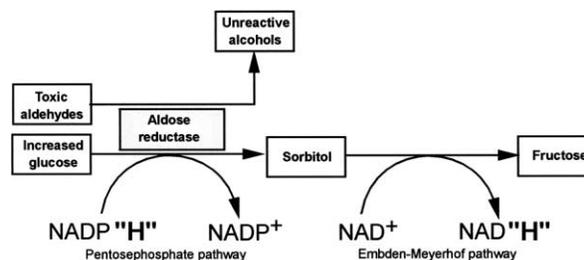


Figure 2 The polyol pathway.

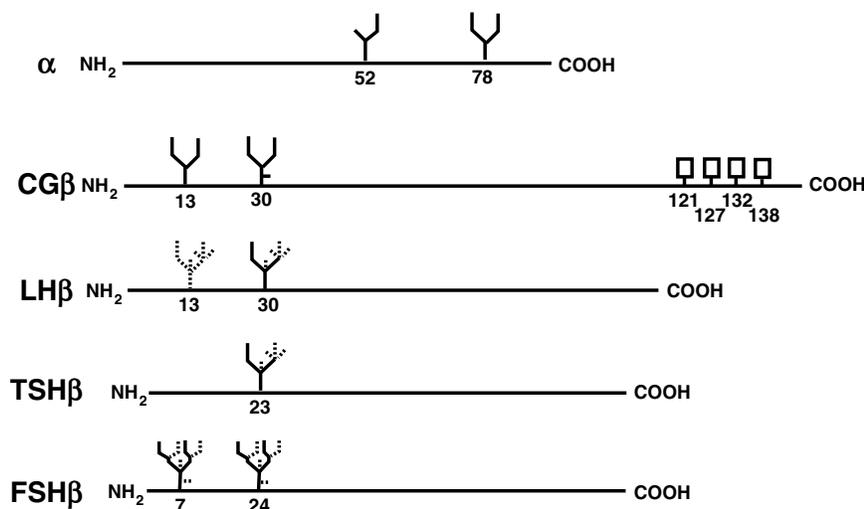
glutathione content of sciatic nerve nor had decreased motor nerve conduction velocity—in contrast to wild-type mice. Studies of the inhibition of the polyol pathway *in vivo* with specific inhibitors have given inconsistent results. An ALR2 inhibitor prevented the development of diabetic neuropathy in dogs but failed to prevent retinopathy or thickening of the capillary basement membrane in the retina, kidney, and muscle. Several clinical trials have been performed with negative outcomes. The prevention of clinical diabetic neuropathy with a potent aldose reductase inhibitor was found in a multiple-dose, placebo-controlled trial. Effect on the redox balance is the most likely mechanism by which increased flux through the polyol pathway has deleterious consequences.

## See Also the Following Articles

Diabetic Nerve Disease, Neuropathy • Diabetes, Type 1 • Glucose Physiology, Normal • Glucose Toxicity • Glycoproteins

## Further Reading

- Brownlee, M. (2001). Biochemistry and molecular cell biology of diabetic complications. *Nature* **414**, 813–820.
- Carington, A. L., and Litchfield, J. C. (1999). The aldose reductase pathway and nonenzymatic glycation in the pathogenesis of diabetic neuropathy: A critical review for the end of the 20th century. *Diab. Rev.* **7**, 275–299.
- Greene, D. A., Arezzo, J. C., and Brown, M. B. (1999). Effect of aldose reductase inhibition on nerve conduction and morphometry in diabetic neuropathy. *Neurology* **53**, 580–591.
- McCance, D. R., Dyer, D. G., Dunn, J. A., Baiue, K. E., Thorpe, S. R., Baynes, J. W., and Lyons, T. J. (1993). Maillard reaction products and their relation to complications in insulin-dependent diabetes mellitus. *J. Clin. Invest.* **91**, 2470–2478.
- Thornalley, P. J. (1998). Cell activation by glycated proteins: AGE receptors, receptor recognition factors, and functional classification of AGEs. *Cell. Molec. Biol.* **44**, 1013–1023.
- Thornalley, P. J. (1999). Clinical significance of glycation. *Clin. Lab.* **45**, 263–273.
- Williamson, J. R., Chang, K., Frangos, M., Hasan, K. S., Ido, T., Kawamura, T., Nyengaard, J. R., van den Enden, M., Kilo, C., and Tilton, R. G. (1993). Hyperglycaemic pseudohypoxia and diabetic complications. *Diabetes* **42**, 801–813.



**Figure 1** Structures of the subunits of human glycoprotein hormones. N-linked sugar chains and O-linked sugar chains are shown by “trees” and “boxes,” respectively, together with their positions. The N-linked sugar chains of  $\alpha$ -subunit are those of hCG, and the  $\alpha$ -subunit of other glycohormones have different sugar chains. Dotted portions of the trees in the  $\beta$ -subunits of LH, TSH, and FSH indicate heterogeneity in branching structures. LH $\beta$  in the case of human hormone is glycosylated at the Asn 30. In contrast, a sugar chain is linked only at the Asn 13 of LH $\beta$  of most other animal species. Reproduced, with slight modifications, from Bielinska and Boime (1995).

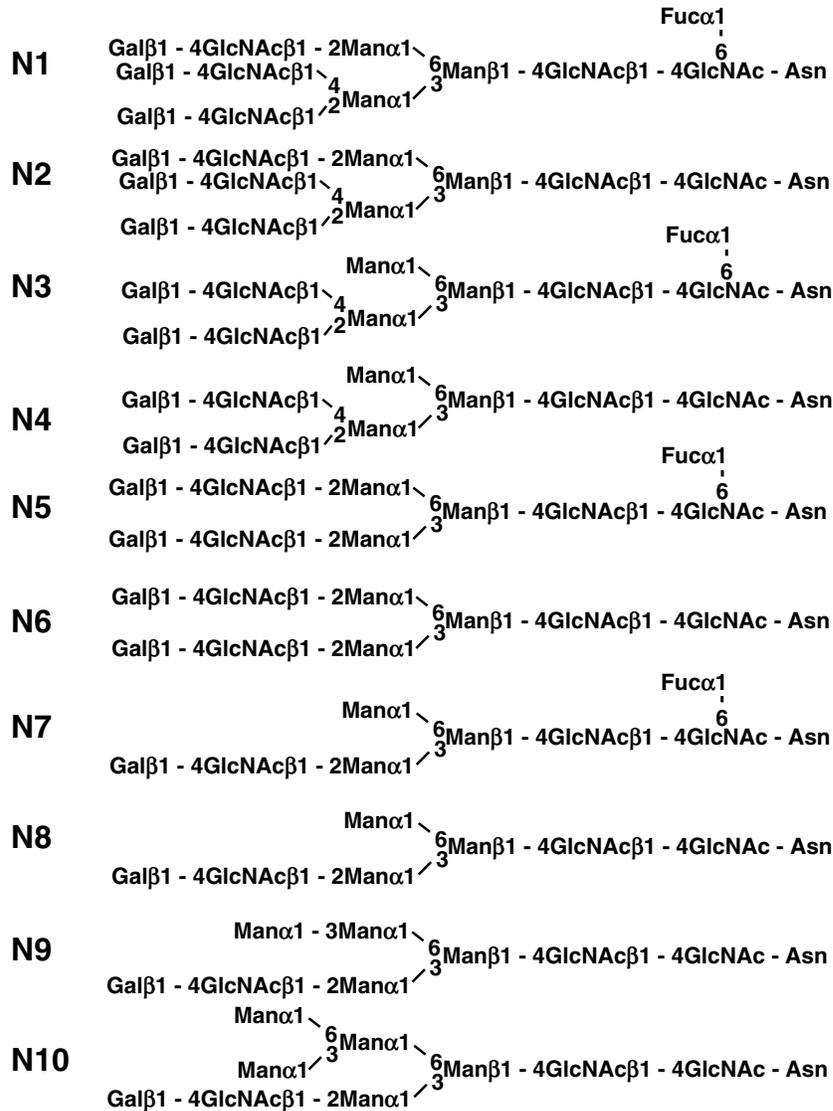
sugar chains of glycoproteins. However, comparative study of the N-linked sugar chains of hCG $\alpha$  and hCG $\beta$  revealed that sialylated N5 is distributed mainly in hCG $\beta$  and that sialylated N8 is distributed mainly in hCG $\alpha$ , whereas sialylated N6 is detected as a major sugar chain of both subunits. The specific distribution of different N-linked sugar chains at the four N-glycosylation sites of hCG molecule cannot be explained by our current knowledge of the biosynthetic mechanism of the N-linked sugar chains. An unknown control mechanism involving the steric effects of the polypeptide moiety may play a role in the formation of N-linked sugar chains of hCG. This assumption was supported by the study of the N-linked sugar chains of the free  $\alpha$ -subunit.

A small amount of  $\alpha$ -subunit occurs in free form in the urine of pregnant women. Interestingly, this free  $\alpha$ -subunit cannot bind to hCG $\beta$ , in contrast to hCG $\alpha$  dissociated from hCG heterodimer. Because the free  $\alpha$ -subunit has the same amino acid sequence as does hCG $\alpha$ , it was assumed that the free  $\alpha$ -subunit contains different sugar chains from hCG $\alpha$ . Structural studies of the sugar chains of the free  $\alpha$ -subunit revealed that it contains sialylated N5 as its major sugar chains. This evidence indicated that bulky sialylated N5 on the free  $\alpha$ -subunit may sterically inhibit its association with hCG $\beta$ . Maturation of the N-linked sugar chains of the free  $\alpha$ -subunit to larger complex-type sugar chains might be induced because

the subunit did not bind to  $\beta$ -subunit. Therefore, uneven distribution of the N-linked sugar chains at the four N-glycosylation sites of hCG may be produced only when the two subunits are associated before the N-linked sugar chains start maturation at the Golgi apparatus.

Later on, the structures of the N-linked sugar chains of LH and TSH of various mammals, including human (hLH and hTSH), were elucidated (as shown in Fig. 4), as were those of human FSH (hFSH) (as shown in Fig. 5). In contrast to hCG, the three pituitary glycoprotein hormones contain triantennary and tetraantennary complex-type sugar chains, and their sialic acids are linked at the C-3 and C-6 positions of galactose residues. Occurrence of bisected sugar chains was also found. The most interesting evidence is that a part of the N-linked sugar chains of hLH and hTSH contain the SO<sub>4</sub>-4GalNAc $\beta$ 1-4GlcNAc group as their outer chains.

An *N*-acetylgalactosaminyltransferase, which forms the GalNAc $\beta$ 1-4GlcNAc group, was found to occur in the pituitary but not in the placenta. This enzyme requires presence of the -Pro-X-Arg- motif in the polypeptide portion of the substrate glycoprotein. Such sequence is present in hLH and hTSH but not in hFSH, explaining the absence of *N*-acetylgalactosamine residue in hFSH. The enzyme responsible for sulfation of the sugar chains of pituitary

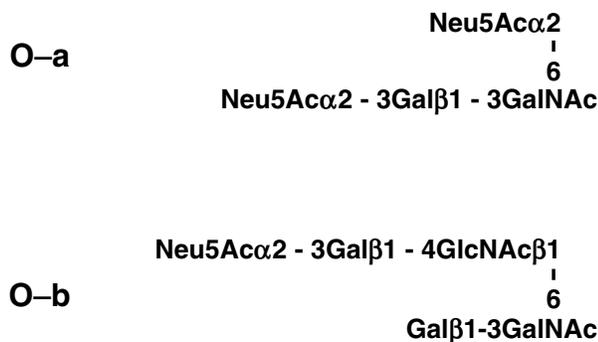


**Figure 2** Structures of desialylated N-linked sugar chains isolated from various hCG preparations. All sialic acid residues are exclusively linked at the C-3 position of the galactose residues. Gal, galactose; GlcNAc, *N*-acetylglucosamine; Man, mannose; Fuc, L-fucose.

glycoprotein hormones was detected in the Golgi membrane preparation of pituitary gland by the transfer of sulfate residues from 3'-phosphoadenosine 5'-phosphosulfate. This 4-*O*-*N*-acetylgalactosamine sulfotransferase does not require any specific peptide motif and can transfer sulfate even to the trisaccharide: GalNAcβ1-4GlcNAcβ1-2Man. Therefore, addition of a β-*N*-acetylgalactosamine residue to the sugar chains is a prerequisite for the sulfation of the sugar chains.

Both FSH and LH are synthesized by gonadotrophs in the anterior pituitary. Although they share the same α-subunit polypeptide, their sugar chains

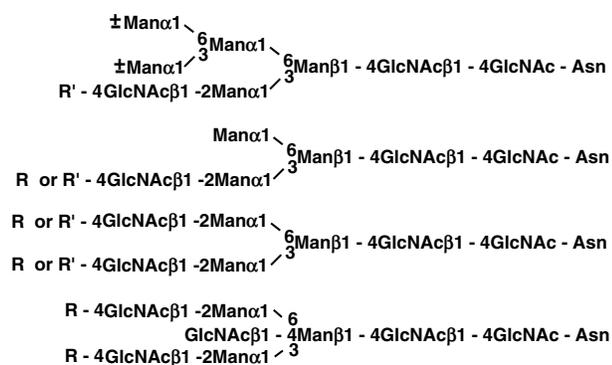
differ in branching and terminal modification. Therefore, assembly of the two subunits should control the maturation of the sugar chains, as discussed previously for the site-directed maturation of the sugar chains of hCG. Actually, assembly of the two subunits of glycoprotein hormones was found to occur when the N-linked sugar chains are still in the state of high mannose types, which are sensitive to endo-β-*N*-acetylglucosaminidase H digestion. In further processing and maturation steps to lead to complex-type sugar chains, association to different β-subunits will result in different sugar chain formation.



**Figure 3** Structures of O-linked sugar chains found in hCG. Neu5Ac, *N*-acetylneuraminic acid; GalNAc, *N*-acetylgalactosamine. Other abbreviations are the same as in Fig. 2.

### hCGs OF TROPHOBLASTIC DISEASES

A large amount of hCG is detected in the sera and urine of patients with various trophoblastic diseases such as hydatidiform mole and choriocarcinoma. Some of the hydatidiform moles show malignant characteristics, such as invasion into the surrounding tissues and metastasis, and are discriminated by the term “invasive mole.” Structural studies of the sugar chains of hCG samples purified from the urine of these patients revealed many interesting structural alterations that can be used for the diagnostic purposes. The N-linked sugar chain patterns of the urinary hCG samples, obtained from patients with hydatidiform mole, were identical to those of the samples from healthy pregnant women. In contrast, quite different patterns were obtained from the hCG samples, purified from the urine of patients with choriocarcinoma. In all choriocarcinoma hCGs, eight



**Figure 4** Structures of N-linked sugar chains found in LH and TSH. R represents the Neu5Ac $\alpha$ 2–3 or 6Gal $\beta$ 1– group, and R' represents the SO<sub>4</sub>–4GalNAc $\beta$ 1– group. Abbreviations of the sugar units are the same as in Figs. 2 and 3.

oligosaccharides, N1 to N8 as shown in Fig. 2, were detected in either their sialylated or nonsialylated forms. Interestingly, the hCG samples purified from the urine of patients with invasive mole gave another alteration in their glycosylation pattern. These samples contained sialylated N1, N2, N5, N6, N7, and N8 but not sialylated N3 and N4.

Although the structural alteration found in the sugar chains of choriocarcinoma hCGs is quite complicated, it can be explained by the changes induced in the two glycosyltransferases. An increase in the molar ratio of total fucosylated oligosaccharides (N1, N3, N5, and N7 in Fig. 2) indicated that expression of the fucosyltransferase responsible for formation of the Fuca1–6GlcNAc group is enhanced in choriocarcinoma cells. Appearance of oligosaccharide N7 may indicate that the enhanced fucosyltransferase has wider specificity than that in normal syncytiotrophoblasts. Structurally, oligosaccharides N1, N2, N3, and N4 can be formed by addition of the Gal $\beta$ 1–4GlcNAc $\beta$ 1–4 outer chain to N5, N6, N7, and N8, respectively. Therefore, *N*-acetylglucosaminyltransferase IV (GnT-IV), which forms the GlcNAc $\beta$ 1–4Man group and is suppressed in syncytiotrophoblasts, must be strongly expressed in choriocarcinoma. Detection of N1 and N2 in invasive mole indicated that GnT-IV is expressed in these cells as well. However, absence of N3 and N4 indicates that the newly expressed GnT-IV can transfer an *N*-acetylglucosamine residue to biantennary sugar chains but not to monoantennary sugar chains.

Cloning of the structural gene of GnT-IV from a human liver cDNA library revealed that two active GnT-IV genes are present. The translation products of these two genes were named GnT-IVa and GnT-IVb. When activities of the five glycosyltransferases (GnT-I to GnT-V), which are related to the formation of the antennary portion of the complex-type sugar chains, were studied comparatively in normal placenta and several choriocarcinoma cell lines, only GnT-IV activity was strikingly increased in the cancer cells. Northern blot analysis revealed that the GnT-IVa gene was strongly overexpressed in the cancer cells, whereas the GnT-IVb gene was expressed at the same level as in the normal placenta. So far, no difference in the substrate specificities of GnT-IVa and GnT-IVb has been found. The data have indicated that overexpression of the GnT-IVa gene and the resulting increase in GnT-IV activity are the enzymatic basis of formation of the abnormal sugar chains in choriocarcinoma cells. An enzymatic basis to form the different sugar patterns of choriocarcinoma and invasive mole hCGs remains to be elucidated.

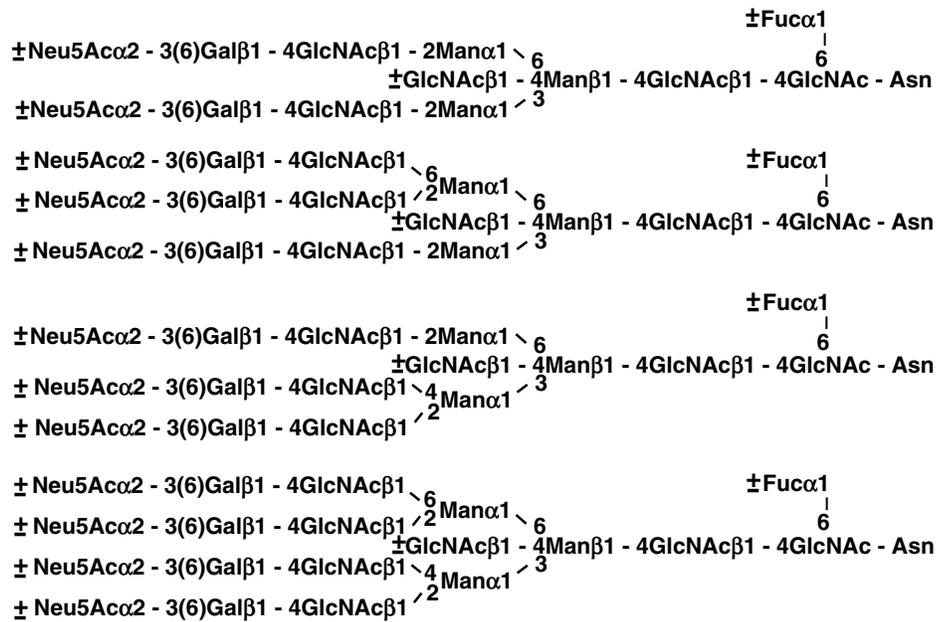


Figure 5 Structures of the N-linked sugar chains found in hFSH. Abbreviations of the sugar units are the same as in Figs. 2 and 3.

Because oligosaccharides N1 and N2 were detected in invasive mole and choriocarcinoma hCGs but not in normal pregnant hCGs and hydatidiform mole hCGs, this difference can be used for the diagnosis of malignant trophoblastic diseases. Actually, a *Datura stramonium* agglutinin-Sepharose column, which has an affinity to the Gal $\beta$ 1-4GlcNAc $\beta$ 1-4(Gal $\beta$ 1-4GlcNAc $\beta$ 1-2)Man group, can discriminate hCGs in the urine of patients with malignant diseases from those in the urine of pregnant women and patients with hydatidiform mole.

Alteration of the O-linked sugar chains of hCG by malignant transformation was also found. hCG samples obtained from patients with hydatidiform mole, invasive mole, and choriocarcinoma all contain both O-a and O-b in Fig. 3. However, the proportion of O-b increased prominently in choriocarcinoma. Although hydatidiform mole hCGs contain a similar proportion of the sialylated tetrasaccharide to normal hCG, the proportion increased moderately but significantly in invasive mole hCGs.

### FUNCTIONAL ROLE OF THE SUGAR CHAINS OF GLYCOPROTEIN HORMONES

The specific distribution of different sugar chains in the two subunits of hCG indicated that even the

smallest N-linked sugar chain, such as sialylated N8, may work as an important signal in expression of the hormonal action of hCG. That modification of the N-linked sugar chains of hCG alters its hormonal activity was proposed by many studies. Complete removal of sialic acid residues from hCG reduces its hormonal activity (as measured by cyclic AMP (cAMP) production and steroidogenesis in the target cells) to 50%, although its binding to target cells is enhanced. Removal of the whole N-linked sugar chains from hCG further increases the binding of hCG to its target cells but completely eliminates its hormonal activity. These results indicate that absence of the sugar chains dissociates receptor binding of hCG from its signal transduction. The deglycosylated hCG behaves as an antagonist to native hCG. It was also reported that the glycopeptides mixture, obtained from hCG by exhaustive pronase digestion, blocked hCG signal transduction. These results suggest that binding of hCG to a lectin-like membrane component in addition to an hCG receptor is necessary to induce the signal transduction.

Comparative study of the biological activities of hCGs, in which only one of the N-glycosylation sites was eliminated by recombinant DNA technology, revealed that N-glycosylation at the Asn-52 of hCG $\alpha$  is essential for the expression of hormonal activity, whereas removal of the N-linked sugar chains at either Asn-13 or Asn-30 of hCG $\beta$  or Asn-78 of hCG $\alpha$  had no effect. Other important evidence

shown by this line of study is that removal of Asn-78 glycosylation of hCG $\alpha$  markedly reduces its assembly with hCG $\beta$ . Glycosylation of the two N-glycosylation sites of hCG $\beta$  is not essential for its assembly with hCG $\alpha$ , but elimination of Asn-30 glycosylation inhibits the secretion of uncombined hCG $\beta$ . These results indicated that the N-linked sugar chains of hCG are important for constructing the correct conformation of each subunit. That the presence of at least one *N*-acetylglucosamine residue at each of the four N-glycosylation sites of hCG is enough to keep the correct folding of the two subunits was shown two decades ago by investigating the effects of digestion with exo- and endoglycosidases on the folding and assembly of the two subunits.

Although exact structures of the sugar chains essential for proper expression of the functional role of hCG were not presented, the role of sialic acid residues was further investigated. As described previously, removal of all sialic residues of the N-linked sugar chains of hCG, which exclusively occur as the Neu5Ac $\alpha$ 2-3Gal group, reduces the hormonal activity of hCG to 50%. When the desialylated hCG was resialylated by incubation with CMP-Neu5Ac and Gal $\beta$ 1-4GlcNAc: $\alpha$ 2-6 sialyltransferase, the isomeric hCG containing only the Neu5Ac $\alpha$ 2-6Gal group gave almost the same dose-response curve as did the natural hCG. This recovery of hormonal activity is not obtained by the addition of the Gal $\alpha$ 1-3 residue to the nonreducing terminal galactose residues. Interestingly, extensive sialylation of desialylated hCG reduced the hormonal activity of the isomeric hCG. Further investigation revealed that sialylation of the outer chain on the Man $\alpha$ 1-3 arm, rather than the Man $\alpha$ 1-6 arm, of the N-linked sugar chains of hCG is favorable for the signal transduction.

Addition of 2 mM *N*-acetylneuraminic acid hexamer, obtained by incomplete sialidase digestion of colominic acid, to the mixture of hCG and MA-10 cells, a mouse Leydig tumor cell line established by Ascoli, revealed that the oligosaccharide did not inhibit the binding of hCG to the surface of the target cells but that cAMP production was reduced to 50%. This result indicated that the hexasaccharide can inhibit the interaction of the sialic acid residues of hCG with the specific binding site on the cell surface but does not influence the binding of the peptide portion of hCG to the hCG receptor. The possibility that the action of the sialic acid hexamer may be due to a nonspecific anionic polymer effect was defuted because the addition of fucoidin did not show any inhibition of the [ $^3$ H]hCG binding to the cell surface receptor or cAMP production by hCG.

Presence of the sialic acid binding site on the surface of MA-10 cells was confirmed by using 3'-sialyllactose-conjugated bovine serum albumin as a probe. Based on the data indicating that sialic acid residues bind directly to the cell surface, a model of the hCG-receptor complex was proposed. A dual interaction of the peptide portion and the sialylated sugar chain of hCG with respective binding sites is essential for the signal transduction. Interestingly, a region homologous to the soybean lectin was found in the hCG receptor.

Compared to hCG, the functional roles of the sugar chains of pituitary hormones were not investigated aggressively. Like other serum glycoproteins, sialylation of the N-linked sugar chains of glycoprotein hormones will protect them from clearance by hepatic galactose-binding receptor. In contrast, sulfation of the N-linked sugar chains of hLH and hTSH leads to more rapid clearance of them. It was reported that sulfated LH is cleared from circulation by binding to a receptor, which specifically recognizes the SO $_4$ -4GalNAc $\beta$ 1-4GlcNAc group, on the surface of hepatic endothelial cells. Fully sialylated hTSH, made by recombinant technique, showed a longer half-life in circulation than did natural TSH. Because the sulfation versus sialylation pattern of TSH is regulated by thyrotropin-releasing hormone, the rapid clearance may be important in controlling the serum level of this glyco-hormone.

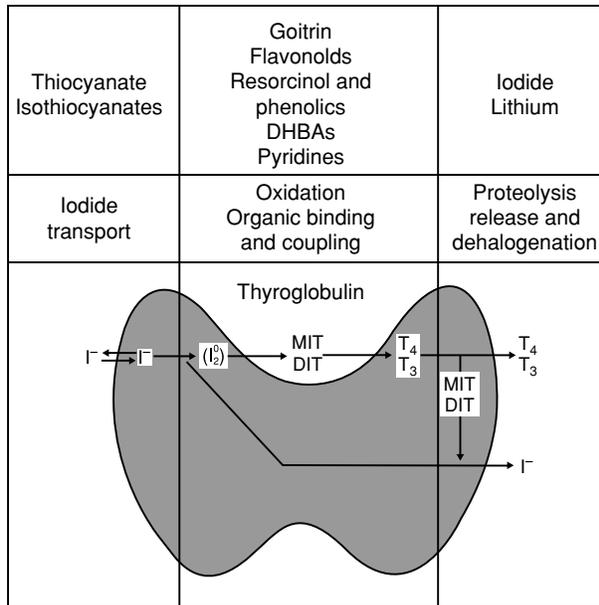
### See Also the Following Articles

FSH (Follicle-Stimulating Hormone) • Glycation- and/or Polyol Pathway-Inducing Complications • LH (Luteinizing Hormone) • TSH Function and Secretion

### Further Reading

- Baenziger, J. U., Kumar, S., Brodbeck, R. M., Smith, P. L., and Beranek, M. C. (1992). Circulatory half-life but not interaction with the lutropin/chorionic gonadotropin receptor is modulated by sulfation of bovine lutropin oligosaccharides. *Proc. Natl. Acad. Sci. USA* **89**, 334-338.
- Bielinska, M., and Boime, I. (1995). The glycoprotein hormone family: Structure and function of the carbohydrate chains. In "Glycoproteins" (J. Montreuil, H. Schachter, and J. F. G. Vliegthart, eds.), pp. 565-587. Elsevier, Amsterdam.
- Fiete, D., Srivastava, V., Hindsgaul, O., and Baenziger, J. U. (1991). A hepatic reticuloendothelial cell receptor specific for SO $_4$ -4GalNAc beta 1,4GlcNAc beta 1,2Man alpha that mediates rapid clearance of lutropin. *Cell* **67**, 1103-1110.
- Kobata, A., and Amano, J. (1995). Structures and function of the N-linked sugar chains of glyco-hormones. In "Biopolymers and Bioproducts: Structure, Function, and Application" (J. Svasti, ed.), pp. 210-217. Samakkhisan Publishing, Bangkok, Thailand.

- Kobata, A., and Takeuchi, M. (1999). Structure, pathology, and function of the N-linked sugar chains of human chorionic gonadotropin. *Biochim. Biophys. Acta* **1455**, 315–326.
- Nemansky, M., DeLeeuw, R., Wijands, R. A., and Van den Eijnden, D. H. (1995). Enzymic remodelling of the N- and O-linked carbohydrate chains of human chorionic gonadotropin: Effects on biological activity and receptor binding. *Eur. J. Biochem.* **227**, 880–888.
- Skelton, T. P., Hooper, L. U., Srivastava, V. V., Hindsgaul, O., and Baenziger, J. U. (1991). Characterization of a sulfotransferase responsible for the 4-O-sulfation of terminal beta-N-acetyl-D-galactosamine on asparagine-linked oligosaccharides of glycoprotein hormones. *J. Biol. Chem.* **266**, 17142–17150.
- Takamatsu, S., Oguri, S., Yoshida, A., Minowa, T., Nakamura, K., Takeuchi, M., and Kobata, A. (1999). Unusually high expression of N-acetylglucosaminyltransferase-IVa in human chorionic carcinoma cell lines: A possible enzymatic basis of the formation of abnormal biantennary sugar chain. *Cancer Res.* **59**, 3949–3953.
- Varki, A., Cummings, R., Esko, J., Freeze, H., Hart, G., and Marth, J. (eds.) (1999). “Essentials of Glycobiology,” Cold Spring Harbor Laboratory Press, New York.
- Weishaar, G., Hiyama, J., and Renwick, A. G. C. (1991). Site-specific N-glycosylation of human chorionic gonadotropin: Structural analysis of glycopeptides by one- and two-dimensional <sup>1</sup>H NMR spectroscopy. *Glycobiology* **1**, 393–404.



**Figure 1** Environmental antithyroid–goitrogenic compounds and their site of action in the thyroid gland. Goitrin, L,5-vinyl-2-thioxazolidone; DHBAs, dihydroxy-benzoic acids;  $I^-$ , iodide; MIT, monoiodotyrosine; DIT, diiodotyrosine;  $T_4$ , thyroxine;  $T_3$ , triiodothyronine. Reproduced with permission from the *Annual Review of Nutrition*, volume 10, (©)1990 by Annual Reviews.

affected by goiter. Cassava, a staple food in these areas, has antithyroid effects in humans and experimental animals. Thus, daily consumption of cassava, in the presence of severe iodine deficiency, is thought to be the cause of endemic goiter and cretinism in these areas of Zaire. The goitrogenic action of cassava is due to endogenous release of thiocyanate from linamarin, a cyanogenic glucoside present in cassava, particularly in the tuberous roots. Thiocyanate is also present in Pearl millet, the staple food of people living in iodine-deficient endemic goiter areas of western Sudan. Pearl millet is rich in C-glycosylflavones, which in combination with thiocyanate exert additive and complementary antithyroid and goitrogenic effects. Thiocyanate is also found in high concentrations (1 g/L) in wastewater effluents of coal conversion processes and in body fluids as a metabolite of hydrogen cyanide gas consumed while smoking. Studies in Sweden indicate that cigarette smoking may produce goiter. Similarly, goiter and hypothyroidism were documented in patients receiving long-term thiocyanate treatment for hypertension. This goitrogenic effect of thiocyanate is more evident in the presence of iodine deficiency. Several observations suggest that thiocyanate crosses the human placenta and may cause both goiter and neonatal hypothyroidism.

### Thiocyanate

Thiocyanate and thiocyanate-like compounds primarily inhibit the iodine-concentrating mechanism of the thyroid, and their goitrogenic activity can be overcome by iodine administration (Fig. 1). Thiocyanate at low concentrations inhibits iodide transport by increasing the velocity constant of iodide efflux from the thyroid gland. At high concentrations, the iodide efflux is greatly accelerated, whereas the unidirectional iodide clearance into the gland is inhibited. Thiocyanate at these high concentrations also inhibits the incorporation of iodide into thyroglobulin by competing with iodide at the thyroid peroxidase (TPO) level. Thiocyanate is rapidly converted to sulfate in the thyroid gland. Administration of thyroid-stimulating hormone (TSH) increases the intrathyroidal catabolism of thiocyanate and is capable of reversing the block of iodide uptake produced by this ion.

### Isothiocyanates

The isothiocyanates and cyanogenic glycosides act on the thyroid mainly by their rapid conversion to thiocyanate. However, isothiocyanates also react spontaneously with amino groups to form thiourea derivatives, which produce a thiourea-like antithyroid effect. Isothiocyanates also possess intrinsic antithyroid activity, as demonstrated by *in vitro* inhibition of iodide uptake in the case of methyl- and allyl-isothiocyanates and of both iodide uptake and organification in the case of butyl-isothiocyanate.

### Thio-Oxazolidone (Goitrin)

The thionamide or thiourea-like goitrogens interfere in the thyroid gland with the organification of iodide and formation of the active thyroid hormones, and their action usually cannot be antagonized by iodine. Naturally occurring goitrin is representative of this category (Fig. 1). Long-term administration of goitrin to rats results in increased thyroid weight and decreased radioactive iodide uptake and hormone synthesis by the thyroid gland. Actually, goitrin possesses 133% of the potency of propylthiouracil in humans. Goitrin is unique in that it is not degraded like thioglycosides. Additive antithyroidal effects of thiocyanate, isothiocyanate, and goitrin also occur with combinations of these naturally occurring goitrogens.

### Disulfides

The small aliphatic disulfides ( $R-S-S-R$ ;  $R$  = methyl-, ethyl-, *n*-propyl, phenyl-), the major components of

onion and garlic, exert marked thiourea-like antithyroid activity. None of these disulfides inhibits *in vitro* the TPO enzyme, but fractions with sulfur-bearing organic compounds, possibly aliphatic disulfides from the goitrogenic well supplying a Colombian district with endemic goiter, inhibited *in vitro*  $^{125}\text{I}$ -organification. Disulfides are also present in high concentrations (0.3–0.5 g/L) in aqueous effluents from coal conversion processes, and they have also been identified as water contaminants in the United States, where the most frequently isolated compounds are dimethyl, diethyl, and diphenyl disulfides.

## FLAVONOIDS

Flavonoids are important stable organic constituents of a wide variety of plants. Flavonoids are universally present in vascular plants and in a large number of food plants. Because of their widespread occurrence in edible plants such as fruits, vegetables, and grains, flavonoids are an integral part of the human diet. They are present in high concentrations in polymeric (tannins) and oligomeric (pigments) forms in various staple foods in the Third World such as millet, sorghum, beans, and ground nuts.

Flavonoids are polyhydroxyphenolic compounds with a  $\text{C}_6\text{--C}_3\text{--C}_6$  structure. Mammalian organisms are unable to synthesize the flavone nucleus. Flavonoids are strictly exogenous food components of exclusively vegetable origin. They have high chemical reactivity with multiple important biological implications. Flavonoids are quickly metabolized in higher organisms, and that is the reason why they are not found in normal tissue constituents. Most flavonoids are present as  $\beta$ -glucosides that cannot be absorbed in tissues. No mammalian enzymes have been found that deglycosylate these compounds to their bioactive aglycone species. Following ingestion by mammals, flavonoid glycosides are hydrolyzed by intestinal microbial glycosidases to flavonoid aglycones. These may be absorbed and undergo metabolism by mammalian tissues, or they may be further metabolized by intestinal micro-organisms to undergo B-ring hydroxylation and middle-ring fission, with production of various metabolic monomeric compounds, including phenolic acids, phloroglucinol, resorcinol, and gallic acid. Each metabolic step is characterized by a marked increase in antithyroid effects. Flavonoid aglycones, such as apigenin and luteolin present in Fonio millet (*Digitaria*

*exilis*), and a variety of flavonoid metabolites (e.g., phloroglucinol, resorcinol, phenolic acids) are several times more potent than the parent glycosides glucosylvitexin, glucosylorientin, and vitexin present in Pearl millet (*Pennisetum* [L.] *leek*, also known as typhoides or americanum), as inhibitors of TPO, the enzyme-catalyzing iodide oxidation and hormone synthesis in the thyroid gland. This greater inhibitory effect is further enhanced by the additive effects exerted by mixtures of flavonoid aglycones and flavonoid metabolites that are formed after ingestion of mixtures of flavonoid glycosides present in many plant foodstuffs. In addition, these metabolic products may produce adverse effects on other parameters of thyroid function not observed with the glycosides. As a result, the antithyroid effects of flavonoid glycosides in foodstuffs may be greatly enhanced by metabolic alterations after ingestion by mammals, as in the case of the flavonoids present in the Pearl millet grain, the staple food of people living in iodine-deficient endemic goiter areas of western Sudan, which make a major contribution to and are primarily responsible for its antithyroid and goitrogenic effects. Furthermore, antithyroid effects *in vivo* of vitexin, one of the three major flavonoids in Pearl millet, has been demonstrated to provide evidence that C-glycosylflavones are the goitrogens in this cereal grain. It is of interest that a significant portion of the flavonoids isolated from Fonio millet, the staple food of people living in the severely affected endemic goiter area of Guinea in Western Africa, are already present as the aglycones apigenin and luteolin, with more potent antithyroid activity than their parent glycosides. Flavonoids not only inhibit TPO but, acting on iodothyronine deiodinase enzymes, also inhibit the peripheral metabolism of thyroid hormones. Flavonoids also affect serum thyroid hormone binding and thyrotropin (TSH) regulation. Thus, this class of compounds alters thyroid hormone economy in a complex manner.

At this point, there is substantial evidence indicating, first, that various millet species used as staple food by the populations in the semi-arid tropics are rich in flavonoids; second, that flavonoids have potent and diverse antithyroid properties; and third, that under the appropriate environmental dietary conditions of low iodine and protein-calorie intakes, which are prevalent in most countries of the Third World, flavonoids become an important etiological determinant of endemic goiter and hypothyroidism.

## POLYHYDROXYPHENOLS AND PHENOL DERIVATIVES

Coal is a source of a large variety of antithyroid and goitrogenic compounds such as, phenol, dihydroxyphenols (resorcinol), substituted dihydroxybenzenes, thiocyanate, disulfides, phthalic acids, pyridines, and halogenated and polycyclic aromatic hydrocarbons (PAH) (Table I). Most of these compounds have been identified in drinking water from the iodine-sufficient goitrous areas of Kentucky (United States) and Colombia (South America). Phenolics are the major organic pollutants in wastewater effluents from various types of coal treatment processes. Resorcinol, substituted resorcinols, and other antithyroid phenolic pollutants are present at levels of as high as 5 g/L in coal-derived effluents. Up to 8% of shale bitumen is also composed of phenols. Phenol, dihydroxyphenols, trihydroxyphenols, and halogenated phenols are readily absorbed from the gastrointestinal tract. Phenol, resorcinol, and catechol, in suitable preparations, are readily absorbed through human skin. Essentially all phenols, polyhydroxyphenols, and halogenated phenols are readily absorbed after injection. A major route of metabolism of polyhydroxyphenols, polyhydroxyphenolic acids, and halogenated phenols is by conjugation to glucuronic or sulfuric acids. The major route of excretion of these compounds is the urinary tract, and various amounts of the free parent compound and its monoglucuronide and monosulfate conjugates are excreted in the urine.

Resorcinol, the prototype of this group of compounds, is antithyroid and goitrogenic both in man and in experimental animals. During the early 1950s, the goitrogenic effect of resorcinol was demonstrated when patients applying resorcinol ointments for the treatment of varicose ulcers developed goiter and hypothyroidism. Several observations also suggest that resorcinol crosses the human placenta and may cause both goiter and neonatal hypothyroidism. Resorcinol has been shown both *in vivo* and *in vitro* to inhibit thyroidal organification of iodide. A comparison of the antiperoxidase activity of resorcinol (1,3-dihydroxybenzene), catechol (1,2-dihydroxybenzene), and hydroquinone (1,4-dihydroxybenzene) (Table I) indicates the importance of hydroxyl groups in the meta position for maximal activity. Furthermore, the net antiperoxidase effects of mixtures of dihydroxyphenols, as well as dihydroxyphenols and thiocyanate (also a coal-derived pollutant), are equivalent to or greater than the sum of the effects produced by individual compounds, indicating that the true goitrogenic potential of the major water-soluble compounds present

in coal and shales are due to the combined effects of the individual constituents rather than to any single compound. Demonstration *in vivo* and *in vitro* of antithyroid and goitrogenic activities of coal-water extracts from iodine-sufficient goiter areas indicate that shale- and coal-derived organic pollutants appear to be a major factor contributing to the high goiter prevalence and associated disorders observed in certain areas with aquifers and watersheds rich in these organic rocks. Studies of the physical state of organic goitrogens in water indicate that the active compounds form dissociable complexes and that they are part of larger organic molecules, possible humic substances (HS). Furthermore, resorcinol and other parent antithyroid phenolic and phenolic-carboxylic compounds are degradation monomeric by-products of reduction, oxidation, and microbial degradation of HSs. HSs, high-molecular-weight complex polymeric compounds, are the principal organic compounds of soils and waters. More than 90% of total organic matter in water consists of HSs, which are also present in coals and shales. Decaying organic matter becomes the substrate of lignin and flavonoid types of HS during the process of fossilization (or coalification). Actually, cyanidin, a naturally occurring flavonol used as a model subunit of flavonoid-type HS, yields the following antithyroid compounds by reductive degradation: resorcinol, phloroglucinol, orcinol, and 3,4-dihydroxybenzoic acid (Table I). Demonstration *in vivo* and *in vitro* of antithyroid effects of vitexin, a major C-glucosylflavone in Pearl millet, provides evidence that flavonoid structures are the link for phenolic goitrogens in foodstuffs (e.g., millet) and those present in coals, shales, soils, and water, all of which are an obligatory step and integral part of the biogeochemical cycle of organic-phenolic goitrogens in nature (Fig. 2).

In addition to thiocyanate, cigarette smoke contains various goitrogenic resorcinol derivatives, flavonoids, and hydroxypyridines. As mentioned previously, cigarette smoking may produce goiter, and smoking increases the severity and metabolic effects of hypothyroidism, probably by alteration of both thyroid function and hormone action.

The presence of halogenated organic compounds with known or potential harmful effects has prompted public health and environmental concerns. These compounds are produced by the chlorination of water supplies, sewage, and power plant cooling waters. Present at micrograms-per-liter concentrations (parts per billion) in treated domestic sewage and cooling waters, 4-chlororesorcinol and 3-chloro-4-hydroxybenzoic acid possess antithyroid activities as inhibitors of TPO and thyroidal iodide organification.

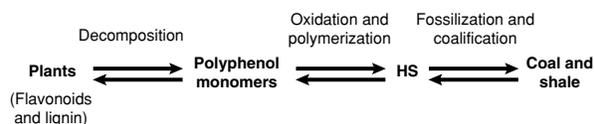
Table I Environmental Agents Producing Goitrogenic and/or Antithyroid Effects

Compound	Goitrogenic/Antithyroid effects		
	<i>In vivo</i>		<i>In vitro</i>
	Humans	Animals	
<b>Sulfurated organics</b>			
Thiocyanate (SCN) <sup>a</sup>	+	+	+
Isothiocyanates	NT	+	+
L-5-vinyl-2-thiooxazolidone (goitrin)	+	+	+
Disulfides (R-S-S-R)	NT	+	0,+ (?) <sup>b</sup>
<b>Flavonoids (polyphenols)</b>			
Glycosides	NT	+	+
Aglycones	NT	+	+
C-ring fission metabolites (e.g., phloroglucinol, phenolic acids)	NT	+	+
<b>Polyhydroxyphenols and phenol derivatives</b>			
Phenol	NT	NT	+
Catechol (1,3-dihydroxybenzene)	NT	NT	+
Resorcinol (1,3-dihydroxybenzene) <sup>a</sup>	+	+	+
Hydroquinone (1,4-dihydroxybenzene)	NT	NT	+
m-Dihydroxyacetophenones	NT	NT	+
2-Methylresorcinol	NT	+	+
5-Methylresorcinol (orcinol)	NT	+	+
4-Methylcatechol	NT	NT	+
Pyrogallol (1,2,3-trihydroxybenzene)	NT	+	+
Phloroglucinol (1,3,5-trihydroxybenzene)	NT	+	+
4-Chlororesorcinol	NT	+	+
3-Chloro-4-hydroxybenzoic acid	NT	NT	+
2,3-Dinitrophenol	+	+	0
<b>Pyridines</b>			
3-Hydroxypyridine	NT	NT	+
Dihydroxypyridines	NT	+	+
<b>Phthalate esters and metabolites</b>			
Diisobutyl phthalate	NT	NT	0
Dioctyl phthalate	NT	+	0
o-Phthalic acid	NT	NT	0
m-Phthalic acid	NT	NT	0
3,4-Dihydroxybenzoic acid (DHBA)	NT	NT	+
3,5-Dihydroxybenzoic acid	NT	NT	+
<b>Polychlorinated (PCB) and polybrominated (PBB) biphenyls</b>			
PCBs (Aroclor)	NT	+	NT
PBBs and PBB oxides	+	+	NT
<b>Other organochlorines</b>			
Dichlorodiphenyltrichloroethane (p,p <sup>1</sup> -DDT)	NT	+	NT
Dichlorodiphenyldichloroethane (p,p <sup>1</sup> -DDE) and dieldrin	NT	+	NT
2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD)	NT	+	NT
<b>Polycyclic aromatic hydrocarbons (PAHs)</b>			
3,4-Benzopyrene (BaP)	NT	+(?)	NT
3-Methylcolanthrene (MCA)	NT	+	NT
7,12-Dimethylbenzanthracene (DMBA)	NT	+	NT
9-Methylanthracene (MA)	NT	+	0
<b>Inorganics</b>			
Excess iodine <sup>a</sup>	+	+	+
Lithium <sup>a</sup>	+	+	+

Note. +, active; 0, inactive; NT, nontested.

<sup>a</sup>Agents also used as medications.

<sup>b</sup>Inactive in TPO assay; active (?) in thyroid slices assay.



**Figure 2** Simplified scheme of the biogeochemical cycle of phenolic goitrogens and interrelationships among plant flavonoids (e.g., millet), polyhydroxyphenols, humic substances (HS), coals, and shales. Reproduced with permission from the *Annual Review of Nutrition*, volume 10, (©)1990 by Annual Reviews.

Whether these pollutants exert additive or synergistic antithyroid effects or act as “triggers” of autoimmune thyroiditis, or both, requires investigation, particularly because more than 60 soluble chloro-organics have been identified in the primary and secondary effluents of typical domestic sewage treatment plants.

Derivatives of 2,4-dinitrophenol (DNP) are widely used in agriculture and industry. An insecticide, a herbicide, and a fungicide, DNP is also used in the manufacture of dyes, to preserve timber, and as an indicator as well as being a by-product of ozonization of parathion. DNP is readily absorbed through intact skin and respiratory tract. DNP causes toxicity by the uncoupling of oxidative phosphorylation in the mitochondria of cells throughout the body. Administration of 2,4-DNP to human volunteers resulted in rapid and pronounced decline of circulating thyroid hormones. A decrease in TSH secretion results in decreased synthesis and release of thyroxine ( $T_4$ ) and triiodothyronine ( $T_3$ ) and possibly involution of the thyroid gland. The antithyroid effect of 2,4-DNP is due in part to an inhibition of the pituitary TSH mechanism. Once  $T_4$  and  $T_3$  are released into the circulation, they are instantaneously bound to serum carrier proteins. DNP also interferes with  $T_4$  binding, further decreasing serum  $T_4$  concentration. In addition to inhibiting the TSH mechanism and interfering with  $T_4$  binding, DNP accelerates the disappearance of  $T_4$  from the circulation; thus, the serum concentration is lowered even more. The public health impact of this pollutant on the thyroid is still unknown.

## PYRIDINES

Hydroxypyridines also occur in aqueous effluents from coal conversion processes as well as in cigarette smoke. Dihydroxypyridines and 3-hydroxypyridine are potent inhibitors of TPO, producing effects comparable to or greater than those of propylthiouracil. After ingestion, mimosine, a naturally occurring amino acid in the seeds and foliage of the tropical

legume *Leucaena leucocephala*, is metabolized to 3,4-dihydroxypyridine (3,4-DHP), a potent antithyroid agent that produces goiter in mammals. 3,4-DHP crosses the placenta barrier, producing goitrous offspring. The phenolic properties of the 3-hydroxy group in various hydroxypyridines are reflected in the metabolism of these compounds *in vivo*. 3-hydroxypyridine fed to rabbits is converted to ethereal glucuronide and sulfate conjugates. 3,4-DHP glucuronide and sulfate conjugates account for the majority of 3,4-DHP in the blood of cattle grazing on leucaena. The ring structure of dihydroxypyridines does not appear to be broken down in the body and also appears to be relatively stable to bacterial degradation.

## PHTHALATE ESTERS AND METABOLITES

Phthalic acid esters, or phthalates, are ubiquitous in their distribution and have been frequently identified as water pollutants. Dibutyl phthalates (DBPs) and dioctyl phthalates (DOPs) have been isolated from water-supplying areas of endemic goiter in western Colombia and eastern Kentucky. Although phthalate esters are most commonly the result of industrial pollution, they also appear naturally in shale, crude oil, petroleum, plants, and fungal metabolites and as emission pollutants from coal liquefaction plants.

Phthalate esters are well absorbed from the gastrointestinal tract. Prior to intestinal absorption, there is hydrolysis to the corresponding monoester metabolite. This is particularly true of the longer chain derivatives such as DOPs. Phthalates are widely distributed in the body, with the liver being the major initial repository organ. Clearance from the body is rapid. Short-chain phthalates can be excreted unchanged or following complete hydrolysis to phthalic acid. Prior to excretion, most longer chain compounds are converted, by oxidative metabolism, to polar derivatives of the monoesters. The major route of phthalate esters elimination from the body is urinary excretion.

Phthalate esters are commonly used as plasticizers to impart flexibility to plastics, particularly polyvinylchloride polymers (PVCs), which have a wide variety of biomedical and other uses such as building and construction, home furnishings, cars, clothing, and food wrappings. A small fraction of phthalate esters are used as nonplasticizers for pesticide carriers, oils, and insect repellents. Phthalates may be present in concentrations of up to 40% of the weight of the plastic.

Phthalate esters are known to leach out from finished PVC products into blood and physiological solutions. The entry of these plasticizers into a patient's bloodstream during blood transfusion, intravenous fluid administration, or hemodialysis has become a matter of concern among public health officials and the medical community. A high incidence of goiter in patients receiving maintenance hemodialysis has been reported. It remains to be determined whether phthalate ester metabolites, contaminants in the water entering the patient's bloodstream, or middle molecules (e.g., hydroxybenzoic and vanillic acids) that accumulate in uremic serum and are poorly removed by hemodialysis are responsible for this condition.

Although phthalate esters and phthalic acids do not possess intrinsic antithyroid activity (Table I), they undergo degradation by gram-negative bacteria to form dihydroxybenzoic acid (DHBA). DHBAs are known to possess antithyroid properties (Table I). The 3,4- and 3,5-DHBAs also inhibit *in vitro* TPO and the incorporation of iodide into thyroid hormones. The proven effective role of gram-negative bacteria in phthalate biodegradation may explain, in part, the relationship established between frequency of goiter and bacterial contamination of water supplies. Furthermore, marked ultrastructural changes of the thyroid gland, similar to those seen after administration of TSH, and decreased serum T<sub>4</sub> concentration have been observed in rats treated with phthalic acid esters. Thus, phthalates may become goitrogenic under appropriate conditions and are also actively concentrated and metabolized by several species of fish. Whether these widely distributed pollutants exert deleterious effects on the thyroids of humans has not been investigated.

## POLYCHLORINATED AND POLYBROMINATED BIPHENYLS

These are aromatic compounds containing two benzene nuclei with two or more substituent chlorine or bromine atoms. They have a wide variety of industrial applications, including electric transformers, capacitors, and heat transformers. There is growing evidence that atmospheric transport is the primary mode of global distribution of polychlorinated biphenyls (PCBs) from sites of use and disposal. Plant foliage accumulates the vapor of PCBs from the atmosphere. In addition to their occurrence in surface water (e.g., rivers, lakes), PCBs have been detected in drinking water. Perhaps the most significant human exposures are limited to individuals consuming freshwater fish

from contaminated streams and lakes and to occupational exposure of industrial workers. PCBs can also be found in the milk of nursing mothers who have eaten large amounts of sport fish or who have been occupationally exposed. Currently, PCBs are targeted by bioremediation strategies, and some strains of *Pseudomonas* spp. (*Pseudomonas cepacia*) can degrade these stable aromatic pollutants. PCBs and polybrominated biphenyls (PBBs) have high-lipid solubility and resistance to physical degradation. They are slowly metabolized, and their excretion is limited. Long-term low-level exposure to the organohalides results in their gradual accumulation in fat, including the fat of breast milk. PCBs have been found in the adipose tissue of 30 to 45% of the general population. The biological and toxicological properties of PCB mixtures may vary depending on their isomeric composition. Oral administration of PCBs to various mammals results in rapid and nearly complete (90%) intestinal absorption. The degradation and elimination of PCBs depend on the hepatic microsomal enzyme system. The excretion of PCBs is related to the extent of their metabolism. Those with greater chlorine content have a correspondingly longer biological half-life in mammals. This resistance to metabolism is reflected in their deposition in adipose tissue. The PBBs, however, have very low acute toxicity in all animal species tested, and PBBs have biological properties similar to PCBs.

Despite the lack of evidence that dietary PCBs and PBBs have any deleterious effects on health, there is a growing concern and uncertainty about the long-range effects of bioaccumulation and contamination of our ecosystem with these chemicals. The uncertainty extends to the potential harmful effects of these pollutants on the thyroid. For instance, an increased prevalence of primary hypothyroidism (11%) was documented among workers from a plant that manufactured PBBs and PBB oxidases. These individuals had elevated titers of TPO-microsomal antibodies, indicating that hypothyroidism was probably a manifestation of lymphocytic autoimmune thyroiditis, perhaps a PBB-induced pathogenic autoimmune response or exacerbation of underlying subclinical disease.

PCBs are potent hepatic microsomal enzyme inducers. Rats exposed to PCBs exhibit a greatly enhanced biliary excretion of circulating T<sub>4</sub>. The T<sub>4</sub> is excreted as a glucuronide that is then lost in the feces. This response is probably secondary to induction of hepatic microsomal T<sub>4</sub>-uridine diphosphate-glucuronyl transferase. The enhanced peripheral metabolism and reduced binding of T<sub>4</sub> to serum proteins

in PCB-treated animals result in markedly decreased serum  $T_4$  concentrations. These low levels stimulate the pituitary–thyrotropin–thyroid axis, and this eventually results in goiter formation. Although PCB-treated animals exhibit decreased serum  $T_4$ , their  $T_3$  levels are unchanged. The relative iodine deficiency brought about by the accelerated metabolism of  $T_4$  may induce increased thyroidal  $T_3$  secretion as well as increased peripheral deiodination of  $T_4$  to  $T_3$ . PBBs appear to act similarly to PCBs. There is, however, some indication that they may also interfere directly with the process of hormonal synthesis in the thyroid gland.

## OTHER ORGANOCHLORINES

DDT (2,2-bis(p-chlorophenyl)-1,1,1-trichloroethane) is polychlorinated and nondegradable. The substance is practically insoluble in water and is resistant to destruction by light and oxidation. Its stability has created difficulties in residue removal from water, soil, and foodstuffs. The dominant degradative reaction of DDT is dehydrochlorination to form DDE (2,2-bis(p-chlorophenyl)-1,1-dichloroethylene), which has the same low solubility in water and high lipid–water partitioning as did its precursor. This substance is almost undegradable, both biologically and environmentally. Dieldrin is one of the cyclodiene insecticides. It is almost insoluble in water and, like DDT and DDE, is very stable, both environmentally and biologically. DDT has been used extensively, both in malaria control and in agriculture, all over the world. Because of biomagnification and persistence, DDT and its breakdown products, DDE and DDD (dichlorodiphenyldichloroethane), are ubiquitous contaminants of water and of virtually every food product. Most of the fish from Lake Michigan in North America contain DDT residues. The substance is also present in milk; humans are at the top of the food pyramid, so human milk is especially contaminated. The situation is basically similar for dieldrin, which is found in surface waters virtually everywhere. Dieldrin is heavily bioconcentrated in the lipids of terrestrial and aquatic wildlife, humans, and foods, especially animal fats and milk. Global distribution of high concentrations of organochlorines, including DDT, DDE, DDD, and dieldrin, were recently found not only in developing countries but also in industrialized countries, which continue to be highly contaminated even though the use of many of these compounds is restricted.

DDT is reductively dechlorinated in biological systems to form DDE and DDD. DDE, the predominant

residue stored in tissues (reaching about 70% in humans), is much less toxic than DDT. DDE is eliminated from the body slowly, and little is known about its degradation pathway. DDT is also eliminated from the human body slowly through reduction to DDD and other more water-soluble derivatives.

DDT is known to cause marked alterations in thyroid gland structure such as thyroid enlargement, follicular epithelial cell hyperplasia, and progressive loss of colloid in birds, and DDD is known to cause goiter and increased hepatobiliary excretion of thyroid hormones in rats. All of these compounds (DDT, DDE, DDD, and dieldrin) induce microsomal enzyme activity that may affect thyroid hormone metabolism in a similar way to that of the polyhalogenated biphenyls and PAH. The impact of these pollutants on the human thyroid is unknown.

Dioxin (or tetrachlorodibenzodioxin, TCDD), one of the most toxic small organic molecules, is a contaminant in the manufacturing process of several pesticides and herbicides, including Agent Orange. Also a potent inducer of hepatic microsomal enzymes, TCDD markedly enhances the metabolism and biliary excretion of  $T_4$ -glucuronide. Rats treated with TCDD concomitantly develop hypothyroxinemia, increased serum TSH concentrations, and goiter, probably as a result of  $T_4$  loss in the bile. The impact on the thyroid of humans exposed to this agent is unknown, and studies of thyroid gland function and thyroid hormone metabolism in those affected are needed.

## POLYCYCLIC AROMATIC HYDROCARBONS

PAHs have been found repeatedly in food and domestic water supplies as well as in industrial and municipal waste effluents. They also occur naturally in coal, soils, ground water, and surface water as well as in their sediments and biota. One of the most potent of the carcinogenic PAH compounds, 3,4-benzpyrene (BaP), is widely distributed and, as in the case of other PAHs, is not efficiently removed by conventional water treatment processes.

The PAH carcinogens, BaP and 3-methylcholanthrene (MCA) by enhancement of hepatic UDP-glucuronyltransferase and glucuronidation, accelerate  $T_4$  metabolism and excretion of  $T_4$ -glucuronide, resulting in decreased serum  $T_4$  concentrations, activation of the pituitary–thyrotropin–thyroid axis, and (eventually) goiter formation. There is also indication that MCA interferes directly with the process of hormonal synthesis in the thyroid gland. Furthermore, MCA, as well as 7,12-dimethylbenzanthracene,

induces goitrous thyroiditis in the BUF rat. Thus, MCA exerts its deleterious effects on the thyroid gland by at least three different mechanisms. Finally, the coal-derived PAH methylanthracene (MA), which has been identified in drinking water from the goitrous coal-rich district of eastern Kentucky, was found to produce goiter in the BUF rat without alteration of hormone synthesis or lymphocytic infiltration of the thyroid gland.

### See Also the Following Articles

Graves' Disease • Hypothyroidism, Causes of • Iodine Deficiency • Nontoxic Goiter • Smoking and the Thyroid • Thyroid Autoimmunity • Thyroid Carcinoma • Thyroid Disease, Epidemiology of • Thyroid Gland, Anatomy and Physiology • Toxic Multinodular Goiter • TSH Function and Secretion

### Further Reading

- Cody, V., Middleton, E., Jr., and Harborne, J. B., eds. (1986). "Plant Flavonoids in Biology and Medicine: Biochemical, Pharmacological, and Structure-Activity Relationships." Liss, New York.
- Cody, V., Middleton, E., Jr., Harborne, J. B., and Beretz, A., eds. (1988). "Plant Flavonoids in Biology and Medicine II: Biochemical, Cellular, and Medicinal Properties." Liss, New York.
- Dumont, J. E., Lamy, F., Roger, P., and Maenhaut, C. (1992). Physiological and pathological regulation of thyroid cell proliferation and differentiation by thyrotropin and other factors. *Physiol. Rev.* **72**, 667-697.
- Gaitan, E., ed. (1989). "Environmental Goitrogenesis." CRC, Boca Raton, FL.
- Gaitan, E. (1990). Goitrogens in food and water. *Annu. Rev. Nutr.* **10**, 21-39.
- Gaitan, E. (1996). Flavonoids and the thyroid. *Nutrition* **12**, 127-129.
- Gaitan, E. (1997). Environmental goitrogens. In "Diseases of the Thyroid" (L. E. Braverman, ed.), pp. 331-348. Humana Press, Totowa, NJ.
- Gaitan, E., and Dunn, J. T. (1992). Epidemiology of iodine deficiency. *Trends Endocrinol. Metab.* **3**, 170-175.
- Gaitan, E., Nelson, N. C., and Poole, G. V. (1991). Endemic goiter and endemic thyroid disorders. *World J. Surg.* **15**, 205-215.
- Lindsay, R. H., Hill, J. B., Gaitan, E., Cooksey, R. C., and Jolley, R. L. (1992). Antithyroid effects of coal-derived pollutants. *J. Toxicol. Envir. Health* **37**, 467-481.
- Weetman, A. P., and McGregor, A. M. (1994). Autoimmune thyroid disease: Further development in our understanding. *Endocrine Rev.* **15**, 788-830.

Table I Comparison of Various Ovulation Induction Agents

Agent	How given	Mechanism of action	Indications	Multiple pregnancy rate (percentage)
Clomiphene citrate	Oral	Blockade of estrogen receptors	Ovulatory dysfunction Superovulation	8
Dopamine agonists Bromocriptine Cabergoline Pergolide	Oral	Dopamine antagonist	Hyperprolactinemia	1
GnRH	Subcutaneous or intramuscular pump	Restoration of GnRH pulses	Hypothalamic amenorrhea	1
Gonadotropins Human menopausal gonadotropins Follitropin	Subcutaneous or intramuscular injections	Direct stimulation of ovarian follicle growth	Ovulatory dysfunction Superovulation	20–25

Note. GnRH, gonadotropin-releasing hormone.

is not discussed here because it is not used in humans. The categories of gonadotropins in clinical use and their trade names are summarized in Table II.

hMG products were the first, and for many years the only, gonadotropins available for human use.

Table II Gonadotropin Preparations

Drug	Trade names
Human menopausal gonadotropins (hMG)	Humegon
	Merional
	Lepori
	Pergonal
Urofollitropin (urinary derived)	Menogon
	Pertisol
	Menopur
	Repronex
	Bravelle
	Fertinex
Follitropin (recombinant)	Fertinorm
	Fostimon
	Metrodin
	Follistim
	Gonal-F
Lutropin (LH) [not available–United States]	Puregon
	Luveris
Human chorionic gonadotropin (hCG) (urinary derived)	A.P.L.
	Choragon
	Novarel
	Pregnyl
	Profasi
Human chorionic gonadotropin (hCG) (recombinant)	Ovidrel
	Ovitrelle

They contain an equal mixture of FSH and LH and are extracted from the urine of postmenopausal donors. The medications must be reconstituted before each use and are given intramuscularly or subcutaneously. In the United States, they are sold under the trade names Pergonal and Repronex.

The earliest preparations of purified FSH were extracted from urinary sources much like hMG (e.g., urofollitropin, Metrodin) and were indicated for intramuscular use only. Techniques were then developed for purification of urinary products to enable subcutaneous administration (e.g., Fertinex, Bravelle). Today, most FSH used comes from recombinant DNA technology and mammalian cell culture. Recombinant products provide a much higher degree of purity and batch-to-batch consistency and include follitropin- $\alpha$  (Gonal-F) and follitropin- $\beta$  (Follistim). They are designed for subcutaneous use but may be injected intramuscularly as well. Until recently, all preparations required fresh daily reconstitution from powder, although one multidose formulation is currently available in the United States.

The final stages of ovulation induction require a surge in circulating luteinizing hormone (LH) levels. In gonadotropin cycles, LH is substituted with human chorionic gonadotropin (hCG). Because LH and hCG work via the same receptor and hCG has a far longer circulating half-life, hCG is a logical agent to induce ovulation. hCG has been available for many years as a urinary-derived product for intramuscular injection. More recently, these older products, as well as newer recombinant products, have been given by subcutaneous injection.

## INDICATIONS FOR THERAPY

There are essentially two categories of indications for treatment with gonadotropins. The first subset is the use of these medications in anovulatory women. This group may be further broken down into two groups. The first group consists of women with hypogonadotropic hypogonadism, also known as hypothalamic amenorrhea, or WHO class I. These women do not have regular menstrual cycles due to very low levels of endogenous gonadotropins, are hypoestrogenic, and do not respond to clomiphene. They are prime candidates for hMG injections and have excellent pregnancy rates. The other subgroup consists of anovulatory women with normal estrogen levels and includes women with polycystic ovarian syndrome and luteal phase defects. Often, these patients have failed clomiphene- and/or insulin-sensitizing agents. The goal in all anovulatory women on gonadotropins is the production and release of one or two mature oocytes.

The second category for gonadotropin use is for controlled ovarian hyperstimulation (COH), also known as superovulation. With this technique, higher doses of gonadotropins are used to recruit multiple follicles, usually as an adjunct to intrauterine insemination (IUI) or *in vitro* fertilization (IVF). Indications for COH with IUI or IVF include unexplained, tubal, cervical, male, and endometriosis-related infertility. The objective in non-IVF gonadotropin cycles is to recruit no more than three or four mature eggs, but with IVF even more oocytes may be desirable.

## THE TREATMENT CYCLE

The gonadotropin treatment cycle begins on the third day of the menstrual cycle, commonly referred to as "day 3." At this time, a transvaginal ultrasound is performed, and blood is obtained for estradiol and FSH levels. On that day, gonadotropin injections are begun. Doses range from 75 to 600 IU of FSH or hMG daily or some combination thereof. They may be daily or twice a day and may be intramuscular or subcutaneous.

Patients then return to the clinic on a frequent basis over the next 7 to 10 days for serial ultrasounds to monitor follicular growth. At the same visits, serum estradiol levels are obtained as an adjunct in follicular monitoring, and serum LH levels are checked to ensure that premature luteinization does not occur. One should note that this monitoring protocol may differ significantly from clinic to clinic.

When the dominant follicle(s) reaches a mature size as seen on ultrasound, usually somewhere between 16 and 20 weeks, the decision is made to trigger ovulation. Human chorionic gonadotropin at 5,000 to 10,000 IU or 250  $\mu$ g is given by intramuscular or subcutaneous injection. Approximately 36 h later, ovulation will occur, and at this time IUI is performed or timed intercourse is encouraged. If an IVF procedure is planned, oocyte retrieval will be performed at this time.

During the subsequent week, many clinicians will choose to provide support of the luteal phase. This may be accomplished by supplemental lower dose hCG injections or progesterone by various routes. Ovulation may be confirmed by a mid-luteal serum progesterone level.

Gonadotropin cycles may be performed back to back without skipping a month in between provided that no large cysts are noted on the interval baseline ultrasound. Doses of gonadotropins may be adjusted in between or during cycles. The course of therapy is continued for three or four cycles. If it is unsuccessful, IVF or other treatments may be contemplated.

## SIDE EFFECTS

Gonadotropins are potent medications, and although they are generally safe in experienced hands, they clearly have the potential to cause serious problems when not used with caution. Common side effects include headache, abdominal discomfort, pain at the injection site, and mood swings. One may also see skin reactions when hMG is used subcutaneously. More troublesome is the potential for multiple gestation and ovarian hyperstimulation.

The chance for multiple pregnancy among women who achieve pregnancy on gonadotropins is 20 to 25%. Although twins and triplets predominate, quadruplets and other higher order multiple gestations have become more commonplace. This phenomenon is due almost exclusively to gonadotropin use. The risks for many complications of pregnancy, including preterm delivery, diabetes, and preeclampsia, increase exponentially with increasing numbers of feti. Therefore, it behooves the clinician to limit the number of mature follicles permitted and, if necessary, to recommend cycle cancellation to the patient. Multifetal reduction may be used to reduce triplet or higher gestations down to twins. Although fraught with ethical and emotional concerns, this procedure can significantly reduce maternal and fetal morbidity

in such cases. However, as stated earlier, the best treatment is prevention.

The other serious concern is the development of the ovarian hyperstimulation syndrome (OHSS). This disorder, which occurs only with ovulation induction, may in some cases involve massive ovarian enlargement, ascites, abdominal discomfort, hydrothorax, and/or a hypercoagulable state. The syndrome has been classified into mild, moderate, and severe, based on symptomatology, ovarian size, and the degree of hemoconcentration. Treatment is generally supportive, with maintenance of urine output, bed rest, and monitoring for increase of severity. The use of albumin infusions for prophylaxis remains controversial. More severe cases may require hospitalization, paracentesis, and/or (occasionally) prophylaxis for deep venous thrombosis, intensive care unit (ICU) admission, and dopamine drips. More extensive reviews can be found elsewhere. As with multiple gestation, the best treatment is prevention.

The long-term concern with gonadotropin therapy revolves around the issue as to whether these medications increase the risk for ovarian cancer. Whittemore and colleagues demonstrated an elevated risk for ovarian tumors (malignant and low-malignant potential) in patients on various ovulation-inducing drugs, a risk that disappeared when treatment was successful. Rossing and colleagues showed an increased risk of ovarian cancer in women on long-term clomiphene but not short-term clomiphene or gonadotropins. Several studies from large IVF programs that use gonadotropins suggested no change in the risk of ovarian carcinoma. In contrast, Shushan and colleagues found a threefold increase in ovarian cancer in women on gonadotropin preparations. In summary, one can conclude that the bulk of the literature suggests no effect of these drugs on cancer risk, but the question is far from answered.

## FUTURE DIRECTIONS

The future of ovulation induction likely will involve a refinement in the agents available to the clinician as well as the development of protocols for optimal stimulation. Recombinant DNA technology has enabled pharmaceutical companies to develop recombinant FSH products during recent years. These new products have enabled easier delivery of the drug via the subcutaneous route and have been a great leap forward in quality control. In the future, it might even be possible to develop an orally active

gonadotropin via emerging technologies, further enhancing ease of use.

Extension of molecular engineering techniques may also enable the development of “designer gonadotropins” in which the potency and half-life may be altered. One might even envision a time where very specific molecules might be engineered to meet the needs of individual patients. By optimizing the stimulation of each patient, we may better succeed in the concurrent goals of improving pregnancy rates while minimizing multiple gestation and ovarian hyperstimulation.

## CONCLUSION

Gonadotropin-induced ovulation involves the use of FSH with or without LH to stimulate the growth of ovarian follicles, followed by injection of hCG to induce release of the oocytes. This treatment may be used for women who are anovulatory but refractory to clomiphene or for controlled ovarian hyperstimulation in preparation for insemination or IVF. These drugs are potentially dangerous and require monitoring with serial ultrasounds and estradiol levels. Women on injectable gonadotropins are at risk for multiple gestation and the ovarian hyperstimulation syndrome. For these reasons, gonadotropin therapy is best prescribed by clinicians familiar with the use of the drugs and with the management of the associated complications.

## See Also the Following Articles

FSH (Follicle-Stimulating Hormone) • Gonadotropin-Releasing Hormone (GnRH) Actions • In Vitro Fertilization (IVF) • Infertility, Overview • LH (Luteinizing Hormone) • Ovarian Hyperstimulation Syndrome • Ovarian-Follicular Apparatus • Superovulation and Intrauterine Insemination

## Further Reading

- Adashi, E. Y., and McClamrock, H. C., eds. (1994). “Managing the Anovulatory State: Medical Induction of Ovulation” [ACOG technical bulletin]. American College of Obstetricians and Gynecologists, Washington, DC.
- Baird, D. T. (2001). Is there a place for different isoforms of FSH in clinical medicine? IV. The clinician’s point of view. *Hum. Reprod.* **16**, 1316–1318.
- Daya, S., and Gunby, J. (2000). Recombinant versus urinary follicle stimulating hormone for ovarian stimulation in assisted reproduction cycles. *Cochrane Database Syst. Rev.* **4**, CD002810.
- Dodson, W. C. (1998). “Induction of Ovarian Follicle Development and Ovulation with Exogenous Gonadotropins” [report]. American Society for Reproductive Medicine, Birmingham, AL.
- Whelan, J. G., and Vlahos, N. F. (2000). The ovarian hyperstimulation syndrome. *Fertil. Steril.* **73**, 883–896.

Table I Summary of Gene Mutations Known to Cause Hypogonadotropic Hypogonadism in Humans

Gene	Principle effects	Inheritance
KAL1	Hypothalamic (disrupts neuronal migration)	X-linked recessive
AHC	Hypothalamus, pituitary, adrenal	X-linked recessive
LEP	Hypothalamus—leptin deficiency	Autosomal recessive
LEPR	Hypothalamus—leptin resistance	Autosomal recessive
GNRHR	Pituitary—GnRH resistance	Autosomal recessive
PROP1	Pituitary—combined pituitary deficiency	Autosomal recessive
HESX1	Pituitary—septo-optic dysplasia	Autosomal recessive
FSH $\beta$	Pituitary—isolated FSH deficiency	Autosomal recessive
LH $\beta$	Pituitary—isolated LH deficiency	Autosomal recessive

neurons fail to synapse normally and Kallmann syndrome ensues.

Up to about half of obvious X-linked Kallmann syndrome families possess KAL1 mutations, and of these, half also have unilateral renal agenesis. The frequency of KAL1 mutations in unselected IHH males is considerably less (5–10%), indicating that this X-linked gene is not the most common cause of human IHH, as was originally proposed. Although the sample size is small, males with KAL1 mutations uniformly tend to demonstrate the most severe gonadotropin defect—an apulsatile pattern of LH secretion. No KAL1 gene mutations have been identified in females with IHH and anosmia, suggesting that other autosomal genes may be involved.

### AHC Gene

Males with adrenal hypoplasia congenita (AHC) display adrenal failure during infancy or childhood because of a failure to form the permanent zone of the adrenal gland. They are deficient in both glucocorticoid and mineralocorticoid hormones. If they are properly treated, they fail to undergo puberty due to IHH. Mutations in the AHC gene have been

demonstrated to cause both adrenal failure (AHC) and IHH. The AHC gene encodes a protein termed DAX1. DAX1 is an orphan receptor (with no known ligand) that belongs to the steroid hormone superfamily. DAX1 is a transcription factor that is important in the development of the adrenal cortex and pituitary gonadotropes and that appears to regulate gonadotropin secretion at both the hypothalamus and the pituitary.

A large number of AHC gene mutations have been described in patients with AHC/IHH, most of which are nonsense and frameshift inactivating mutations throughout the gene. Missense mutations occur almost exclusively in the C terminus. Although AHC is an X-linked recessive disease affecting males, one mutation was identified in a female with normal adrenal function and IHH only. Mutations in the AHC gene probably do not cause IHH without adrenal disease given that no mutations were detected by DNA sequencing of DNAs from 100 IHH males. Although the gene was originally hypothesized to be an ovarian determinant gene because of its localization within the DSS region of the X chromosome, a region of Xp that, when duplicated, is associated with sex reversal (feminization) of 46,XY males, conditional knockout studies do not support its role in ovarian determination but do indicate a role in spermatogenesis.

Table II KAL1 Expression in Chick and Human Correlates with Phenotypic Findings of KAL1 Mutations (1–10)

Olfactory epithelium	Anosmia
Cerebellum	Nystagmus, abnormal balance
Corticospinal tracts	Synkinesia (mirror movements)
Oculomotor nucleus	Visual abnormalities
Retina	Visual defects
Facial mesenchyme	Midline facial clefting
Mesonephros and metanephros	Unilateral renal agenesis
Limb buds	Pes cavus

### GNRHR Gene

The GnRH receptor (GNRHR), a G protein-coupled receptor, constituted the first autosomal-recessive gene identified to possess mutations in human IHH. Although the hypogonadal mouse had a naturally occurring GnRH gene deletion, no GnRH gene mutations had been identified in humans. Several investigators characterized mutations in the larger receptor for GnRH—the GNRHR gene. One group

hypothesized that partial loss-of-function GNRHR mutations could occur in patients with incomplete IHH, that is, IHH patients with evidence of puberty. Others suggested that the GNRHR gene was a likely candidate gene for mutation in IHH patients because of the absence of mutations in the ligand and the variability in response to exogenous GnRH that some IHH patients have.

More than a dozen GNRHR mutations have been identified to date, but they are almost exclusively compound heterozygous missense mutations affecting binding and/or signal transduction. The phenotypic spectrum of GNRHR mutations producing GnRH resistance ranges from complete IHH (no evidence of puberty or fertility) to incomplete IHH (partial pubertal defects). Even in patients who have no evidence of puberty, GnRH administration may improve pituitary gonadotropin responses, suggesting marginal GNRHR function. The true prevalence of GNRHR mutations among all IHH patients is difficult to assess but in one study was 2.2% (1 of 46) of total IHH patients and 7.1% (1 of 14) of patients where an affected female was present. A more recent study corroborated these findings.

### LEP and LEPR Genes

Leptin (LEP) possesses important roles in metabolism and puberty. LEP-deficient (*ob/ob*) and LEP-resistant (*db/db*) mice manifest obesity and hypogonadotropic hypogonadism. LEP mutations have been described in only a few families with obesity, low serum LEP concentrations, and absent pubertal development. Similar to LEP-deficient *ob/ob* mice, patients with LEP mutations had extreme obesity, hyperinsulinemia, and hypogonadism, but unlike the mice, they did not have stunted height, hyperglycemia, or hypercortisolemia. An LEP receptor (LEPR) mutation resulted in a similar phenotype of obesity and hypogonadotropic hypogonadism except that serum LEP levels were elevated, indicating LEP resistance. These rare human autosomal-recessive diseases involving LEP and its receptor strongly implicate a role for LEP action in normal pubertal development.

### PROPI Gene

Mutations in *Prop1*, the mouse orthologue of human *PROPI*, produce the Ames dwarf mouse. Human *PROPI* mutations cause combined pituitary hormone deficiency (CPHD), with varying deficiencies of growth hormone, thyroid-stimulating hormone

(TSH), prolactin, FSH, LH, and (occasionally) adrenocorticotrophic hormone (ACTH). Affected individuals display short stature, absent puberty, and hypothyroidism. A variety of *PROPI* mutations have been identified, including missense mutations and deletions, but a particular two-base pair deletion is common, being a hot spot in the *PROPI* gene for mutations. In a large cohort of patients with IHH only, no *PROPI* mutations were identified, indicating that *PROPI* mutations generally cause a more universal pituitary failure.

### HESX1 Gene

Mutations in the *HESX1* gene have been identified in some families with septo-optic dysplasia, a disorder characterized by panhypopituitarism, optic nerve atrophy, and other midline central nervous system (CNS) abnormalities such as agenesis of the septum pellucidum and corpus callosum. Both autosomal-recessive and autosomal-dominant *HESX1* mutations have been described in this disorder. *Hesx1*, the mouse orthologue of *HESX1*, is expressed in the early forebrain and later becomes restricted to Rathke's pouch, which ultimately becomes the anterior pituitary gland. Hypogonadotropic hypogonadism, due to pituitary gonadotropin deficiency, is a common component of the phenotype in patients with septo-optic dysplasia.

### FSH $\beta$ Gene

FSH, LH, TSH, and human choriogonadotropin (hCG) comprise the pituitary glycoprotein hormones. Each dimeric protein consists of a common  $\alpha$ -subunit encoded by a single gene and a specific  $\beta$ -subunit. No human  $\alpha$ -subunit mutations have been described, but similar to the  $\alpha$ -subunit knockout mouse, the phenotype would be expected to include hypogonadotropic hypogonadism and hypothyroidism. It is possible that  $\alpha$ -subunit gene mutations could be lethal in humans given that hCG would be predicted to be deficient (mice do not have hCG).

Female homozygous *FSH $\beta$*  knockout mice had low serum FSH levels and were sterile because of arrested ovarian follicular development. Surprisingly, serum estradiol was normal in these mice despite unmeasurable serum FSH levels. Male homozygous *FSH $\beta$*  knockout mice had normal levels of serum testosterone, small testes, and oligospermia but were fertile. The phenotype of humans with *FSH $\beta$*  mutations is similar except that estradiol levels are low. Affected

females presented with absent or incomplete breast development, low FSH and estradiol, high LH, and sterility. Despite the elevated LH level similar to that of women with polycystic ovary syndrome (PCOS), hirsutism is not present, suggesting that FSH might be necessary for normal LH-induced androgen production by the theca cells.

Males with FSH $\beta$  mutations present with azoospermia, but puberty may be normal or absent. Contrary to the FSH $\beta$  knockout mice, humans with mutations suggested that FSH was essential for spermatogenesis in human males. In addition, more severe mutations affect androgen production, manifested clinically by delayed puberty and low testosterone.

### LH $\beta$ -hCG $\beta$ Genes

The LH $\beta$ -hCG $\beta$  gene complex, consisting of a single LH $\beta$  gene and six hCG $\beta$  genes, is polymorphic, but there is only one known human mutation in the LH $\beta$  gene. The proband presented with pubertal delay, bilaterally small descended testes, low testosterone, and elevated gonadotropins. He responded to hCG administration, suggesting that the LH ligand might be defective. He was homozygous for a missense mutation of the LH $\beta$  gene, which was detectable by immunoassay but undetectable by radioreceptor assay.

Although LH $\beta$  polymorphisms have been identified, it is more problematic that they are definitive causes of LH dysfunction; however, they may modify LH effects. No LH $\beta$  mutations have been identified in females, but the phenotype would be expected to be normal pubertal development and anovulation without hirsutism.

### CONCLUSION

The genetic basis of IHH remains unknown for most patients; however, mutations in X-linked and autosomal genes may account for approximately 20% of cases. It is highly likely that additional autosomal genes that cause IHH will be identified. However, it is also possible that a large percentage of patients will have complex disease caused by a large number of modifying gene mutations and environmental factors.

### Acknowledgments

Portions of this work were supported by a grant from the U.S. Public Health Service, National Institute of Child Health and Human Development (HD33004).

### See Also the Following Articles

Congenital Adrenal Hypoplasia Syndromes • Gonadotropin-Releasing Hormone (GnRH) Actions • Gonadotropin-Releasing Hormone Receptor Gene, Mutation of • Leptin

### Further Reading

- Achermann, J. C., and Jameson, J. L. (1999). Fertility and infertility: Genetic contributions from the hypothalamic-pituitary-gonadal axis. *Mol. Endocrinol.* **13**, 812–818.
- Georgopoulos, N. A., Pralong, F. P., Seidman, C. E., Seidman, J. G., Crowley, W. F., Jr., and Vallejo, M. (1997). Genetic heterogeneity evidenced by low incidence of KAL-1 gene mutations in sporadic cases of gonadotropin-releasing hormone deficiency. *J. Clin. Endocrinol. Metab.* **82**, 213–217.
- Hardelin, J. P. (2001). Kallmann syndrome: Towards molecular pathogenesis. *Mol. Cell Endocrinol.* **179**, 75–81.
- Layman, L. C. (1999a). The genetics of human hypogonadotropic hypogonadism. *Am. J. Med. Genet.* **89**, 240–248.
- Layman, L. C. (1999b). The molecular basis of human hypogonadotropic hypogonadism. *Mol. Genet. Metab.* **68**, 191–199.
- Layman, L. C., Cohen, D. P., Jin, M., Xie, J., Li, Z., Reindollar, R. H., Bolbolan, S., Bick, D. P., Sherins, R. J., Duck, L. W., Musgrove, L. C., Sellers, J. C., and Neill, J. D. (1998). Mutations in the gonadotropin-releasing hormone receptor gene cause hypogonadotropic hypogonadism. *Nat. Genet.* **18**, 14–15.
- Layman, L. C., and McDonough, P. G. (2000). Mutations of the follicle stimulating hormone-beta and its receptor in human and mouse: Phenotype/Genotype. *Mol. Cell Endocrinol.* **161**, 9–17.
- Oliveira, L. M., Seminara, S. B., Beranova, M., Hayes, F. J., Valkenburgh, S. B., Schipani, E., Costa, E. M., Latronico, A. C., Crowley, W. F., Jr., and Vallejo, M. (2001). The importance of autosomal genes in Kallmann syndrome: Genotype-Phenotype correlations and neuroendocrine characteristics. *J. Clin. Endocrinol. Metab.* **86**, 1532–1538.
- Rugarli, E., and Ballabio, A. (1993). Kallmann syndrome from genetics to neurobiology. *JAMA* **270**, 2713–2716.
- Zhang, Y-H., Guo, W., Wagner, R. L., Huang, B-L., McCabe, L., Vilain, E., Burris, T. P., Anyane-Yeboah, K., Burghes, A. H. M., Chitayat, D., Chudley, A. E., Genel, M., Gertner, J. M., Klingensmith, G. J., Levine, S. N., Nakamoto, J., New, M. I., Pagon, R. A., Pappas, J. G., Quigley, C. A., Rosenthal, I. M., Baxter, J. D., Fletterick, R. J., and McCabe, E. R. B. (1998). DAX1 mutations provide insight into structure-function relationships in steroidogenic tissue development. *Am. J. Hum. Genet.* **62**, 855–864.

receptors demonstrated a 421-fold preference for GnRH-II versus GnRH-I. Unlike the GnRH-II receptor and most other GPCRs, the GnRH-I receptor contains no large cytoplasmic C-terminal tail. The C termini of both types of receptors are phosphorylated in response to GnRH, leading to receptor desensitization.

GnRH activation of its receptor results in stimulation of several signaling pathways. Phospholipase C transmits its signal to diacylglycerol (DAG) and inositol 1,4,5-trisphosphate (IP3). DAG activates the intracellular protein kinase C pathway, and IP3 stimulates release of intracellular calcium. Protein kinase C activation in response to GnRH also leads to increases in the mitogen-activated protein kinase (MAPK) pathway in pituitary cells. These pathways modulate gonadotropin synthesis and release from pituitary cells.

## DEVELOPMENTAL REGULATION OF GnRH

In contrast to most cells of the CNS and all other hypothalamic-releasing factor neurons, GnRH neurons do not initially develop within the brain. GnRH neurons originate in the olfactory placode, migrate across the cribriform plate, and ultimately reside in the diencephalic region of the basal hypothalamus in a dispersed neuronal network. Although born in the olfactory placode, GnRH neurons are not derived from olfactory neurons. In rodents, GnRH neuronal migration begins on embryonic day 10.5 to 11 and final migration occurs by day 15 or 16. During this migratory phase, GnRH neurons express low levels of GnRH mRNA and protein. Identification of the factors that modulate both migration and expression of GnRH in migratory neurons is currently an area of active research. Cell adhesion molecules, including neural cell adhesion molecule (NCAM), are thought to be important. In addition, an adhesion related kinase (Ark) has been shown, in cell culture models, to modulate both neuronal migration and expression of GnRH. Soluble factors such as  $\gamma$ -amino butyric acid (GABA) and netrins have also been implicated in GnRH neuronal migration.

The human hypothalamus only contains a few thousand GnRH neurons (in contrast to approximately 800 neurons in the rodent) that are also arranged in a neuronal network in the basal hypothalamus after migration. Release of GnRH from these neurons occurs in a pulsatile fashion. Immortalized GnRH neuronal cell culture models, as well as isolated

GnRH neurons, demonstrate that the GnRH “pulse generator” is intrinsic to the neuronal cells. During human development, the GnRH pulse generator and subsequent gonadotropin release is active in the early neonate but decreases by 6 months of age. Until puberty, the GnRH pulse generator is repressed. The exact mechanisms of repression are not known and may involve central GABAergic neuronal activity. Similarly, the factors that reactivate the GnRH pulse generator at puberty are not completely understood. Neuronal decreases in GABAergic activity and increased glutaminergic, norepinephrine, and neuropeptide Y activities have been implicated in reactivation of the GnRH pulse generator at the time of sexual maturation. Studies have also suggested a role for transforming growth factor- $\alpha$ /erbB-1 and neuregulin/erbB-4 secreted from glial cells, and nitric oxide from central endothelial cells, as factors that may influence the reactivation of GnRH pulsatile release at the time of puberty.

## DISEASE STATES ASSOCIATED WITH GnRH

Precocious puberty is defined as puberty that occurs prior to 8 years of age in girls and prior to 9 years of age in boys. Central precocious puberty, which is more frequent in girls, is caused by premature reactivation of the GnRH pulse generator and is usually idiopathic. Hypothalamic lesions or tumors, which are more common in boys, have been shown to induce central precocious puberty. The exact mechanism is unknown but may involve premature activation of the glial production of transforming growth factor- $\alpha$  to cross-talk with GnRH neurons.

In contrast to precocious puberty, patients with hypogonadotropic hypogonadism (HH) display delayed or absent puberty, hypogonadism, and infertility. HH arises due to deficiency in either GnRH production or GnRH signaling to the pituitary. In Kallmann's syndrome, GnRH neurons fail to migrate to the hypothalamus from the olfactory placode; therefore, GnRH production is compromised. Anosmia, or lack of smell, is also associated with Kallmann's syndrome. This finding suggests that a common mechanism between GnRH neuronal migration and olfactory neuronal maturation is defective. A genetic defect in the KAL gene, whose product is anosmin, leads to incomplete migration of GnRH and olfactory neurons in an X-linked inheritance. The molecular bases of other forms of GnRH deficiency are unknown. GnRH deficiency affects 1

in 7500 males and 1 in 70,000 females. Other causes of HH involve defective signaling pathways downstream of GnRH. Mutations in pituitary GnRH receptors affect the ability of GnRH to activate gonadotropin synthesis and secretion. In addition, patients have been reported with mutations of the LH $\beta$  or FSH $\beta$  genes resulting in defective sexual maturation.

Clinically, GnRH is used in a pump delivery system to result in an episodic delivery to restore defects in patients with HH. Paradoxically, GnRH agonists such as nafarelin, leuprolide, histrelin, and goserelin block GnRH action. The pituitary is sensitive to pulsatile stimulation by GnRH to synthesize and release the gonadotropins. Continuous stimulation of pituitary GnRH receptors by exogenously administered GnRH agonists, rather than by pulsatile stimulation, down-regulates and desensitizes GnRH receptors. The ultimate effect of chronic stimulation of the pituitary GnRH receptors is to decrease LH and FSH production, with subsequent decreases in circulating estrogen or testosterone. Nafarelin, leuprolide, and histrelin are indicated for central precocious puberty. Nafarelin, leuprolide, and goserelin are indicated for treatment of endometriosis. Leuprolide and goserelin are also indicated for palliative treatment of advanced breast and prostate tumors due to the ability of GnRH agonists to decrease sex steroid hormone levels. More recently, GnRH antagonists have been used to directly block GnRH receptors on the pituitary and to inhibit LH and FSH production. They are indicated for the treatment of female infertility as adjunct therapy during ovarian hyperstimulation for *in vitro* fertilization.

## CONCLUSION

GnRH is the hypothalamic hormone that stimulates the pituitary gland to produce FSH and LH. The reactivation of GnRH expression in the hypothalamus is responsible for the initiation of the onset of puberty

and sexual maturation. Aberrant reactivation of GnRH expression is associated with either premature, delayed, or absent puberty. Therefore, the GnRH pathway is a target for several drug therapies that target either inappropriately early or delayed puberty. In addition to their role in normal reproductive development, the sex steroids may have a pathological role in hormone-dependent tumors such as breast and prostate. As the first step in the hypothalamic–pituitary–gonadal axis, the GnRH pathway is also a therapeutic target to inhibit the production of sex steroids from the gonads in the treatment of hormone-dependent malignancies.

## See Also the Following Articles

Gonadotropin-Induced Ovulation • Gonadotropin-Releasing Hormone Deficiency, Congenital Isolated • Gonadotropin-Releasing Hormone Receptor Gene, Mutation of • Gonadotropins and Testicular Function in Aging • Gonadotropin-Secreting Tumors • Precocious Puberty, Central (Female) • Precocious Puberty, Central (Male)

## Further Reading

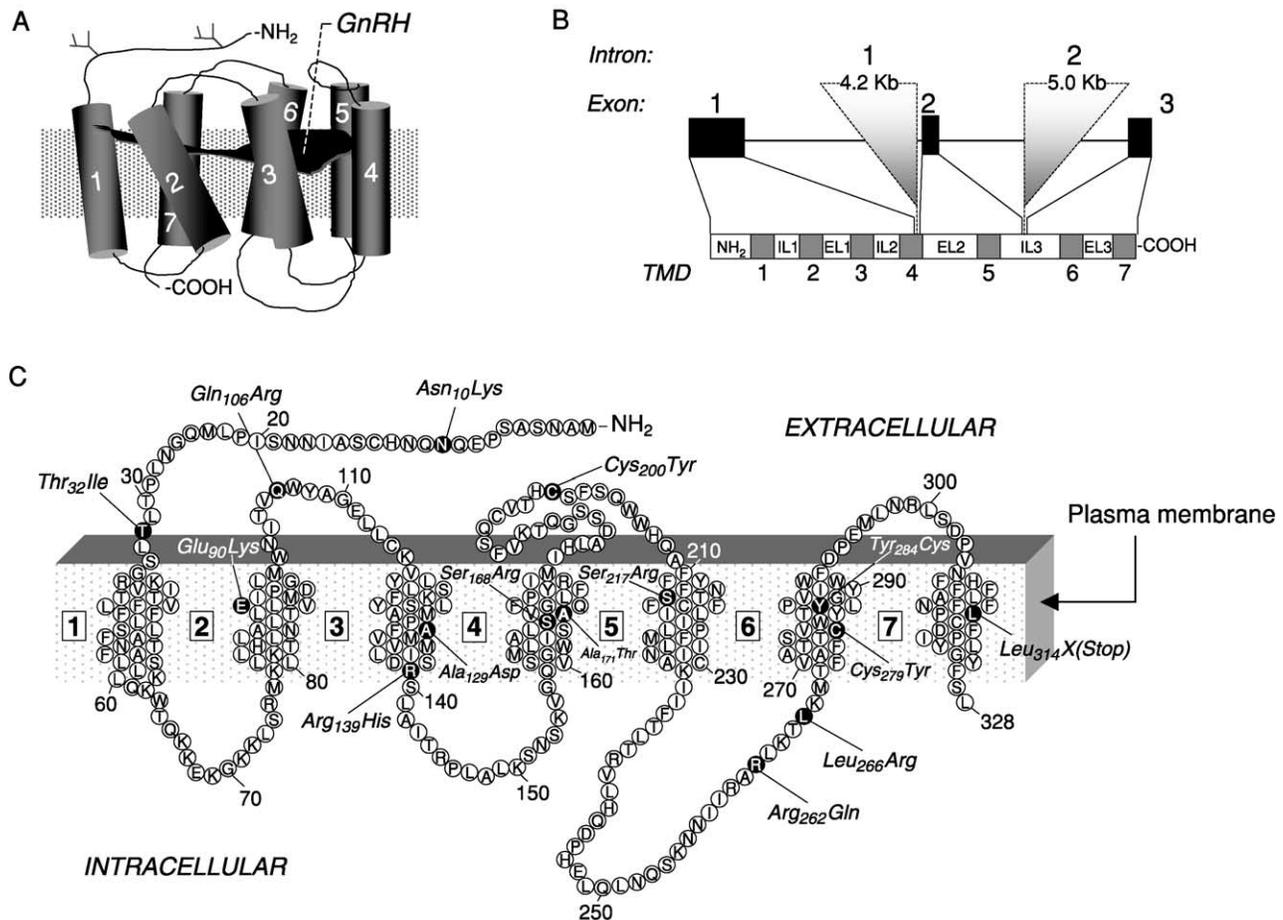
- MacColl, G., Quinton, R., and Bouloux, P. M. (2002). GnRH neuronal development: Insights into hypogonadotrophic hypogonadism. *Trends Endocrinol. Metab.* **13**, 112–118.
- Neill, J. D. (2002). Minireview: GnRH and GnRH receptor genes in the human genome. *Endocrinology* **143**, 737–743.
- Parhar, I. S. (ed.) (2002). “Gonadotropin-Releasing Hormone: Molecules and Receptors,” *Progress in Brain Research*, vol. 141. Elsevier, Amsterdam.
- Prevot, V. (2002). Glial–neuronal–endothelial interactions are involved in the control of GnRH secretion. *J. Neuroendocrinol.* **14**, 247–255.
- Tobet, S. A., Bless, E. P., and Schwarting, G. A. (2001). Developmental aspect of the gonadotropin-releasing hormone system. *Mol. Cell. Endocrinol.* **185**, 173–184.
- Wierman, M. E. (1996). Gonadotropin-releasing hormone. In “Reproductive Endocrinology, Surgery, and Technology” (E. Y. Adashi, J. A. Rock, and Z. Rosenwaks, eds.), pp. 666–681. Lippincott–Raven, Philadelphia.

detected to date and briefly discusses their impact on the phenotypic expression of HH in individuals bearing such receptor defects.

## THE GnRH RECEPTOR

The GnRHR belongs to the superfamily of G protein-coupled receptors (GPCRs), specifically the family related to the rhodopsin/ $\beta$ -adrenergic receptors, which is the best-known member in terms of its structural and functional characteristics. Seven transmembrane hydrophobic domains (TMDs) oriented roughly perpendicular to the plasma membrane plane, with an extracellular  $\text{NH}_2$  terminal, an intracellular  $\text{COOH}$  terminal, and three alternating

intra- and extracellular hydrophilic loops connecting the TMDs, characterize the structure of these receptors (Figs. 1A and 1C). The mammalian GnRHR exhibits more than 85% amino acid identity among the several species that have been cloned. Unlike other members of the rhodopsin/ $\beta$ -adrenergic subfamily of GPCRs, the GnRHR exhibits several unique features, including the reciprocal exchange of the conserved Asp and Asn residues in TMDs 2 and 7, the replacement of Tyr with Ser in the Asp-Arg-Tyr motif located in the junction of the TMD 3 and the intercellular loop (IL) 2, and the lack of the  $\text{COOH}$  terminal domain. This latter feature, which is not exhibited by GnRH receptors from nonmammalian vertebrate species such as fish and avian GnRH



**Figure 1** The human GnRH receptor. (A) Counter-clockwise orientation of a prototypical G protein-coupled receptor from TMDs 1 to 7. The closed-loop structure is representative of receptors for peptide ligands such as GnRH. In this arrangement, the core consists mainly of TMDs 2, 3, 5, and 6, whereas domains 1 and 7 are peripherally sequestered. (B) Schematic of the human GnRHR gene. The open reading frame is distributed among three exons (black rectangles), spanning 18.9 Kb, that encode amino acids 1 to 174, 175 to 248, and 249 to 328, respectively. Intron 1 is located between amino acids 174 and 175 in the putative transmembrane domain (TMD) 4, and intron 2 is located between amino acids 248 and 249 in the intracellular loop (IL) 3 (shaded inverse triangles). (C) Sequence of the human GnRHR and location of the inactivating mutations identified to date (black circles).

receptors, is unique among the thousands of members of mammalian GPCRs and is apparently associated with differential physiological regulation (internalization, desensitization, and cell surface expression) of the receptor in mammals and nonmammals. In humans, the GnRHR is located in 4q13.2-3 and consists of three exons and two introns that encode for a 328-amino acid protein (Fig. 1B).

The GnRHR is preferentially coupled to the trimeric  $G_{q/11}$  protein, localized in the cytoplasm, and associated with the intracellular domains of the receptor. Activation of the GnRHR by its ligand is associated with conformational changes in the receptor molecule that lead to activation of the  $G_{\alpha_{q/11}}$  subunit followed by its dissociation from the  $G\beta\gamma$  complex. The activation of the GnRHR- $G_{\alpha_{q/11}}$  protein complex stimulates the effector enzyme phospholipase- $C\beta$ , which in turn induces phosphatidylinositol 4,5-bisphosphate hydrolysis, leading to formation of the second messengers inositol 1,4,5-triphosphate (IP3) and diacylglycerol. The former diffuses through the cytoplasm, promoting the release of intracellular calcium and the secretion of gonadotropins, whereas the latter activates the enzyme protein kinase C, triggering a cascade of protein-protein phosphorylation and interactions that eventually lead to the expression of biological effects.

## MUTATIONS IN THE HUMAN GnRHR GENE

Structural alterations in key residues of the receptor molecule or the G proteins may potentially lead to altered function of the receptor-G protein system. Mutations in sites involved in ligand binding usually result in altered receptors unable to recognize the ligand and to become activated (loss-of-function mutations), whereas mutations in sites involved in receptor activation or G protein coupling may lead either to loss of function or to constitutive activation (activation in the absence of ligand or gain-of-function mutations) of the receptor molecule. Spontaneous mutations of this latter type have not yet been detected in the GnRHR.

Resistance to GnRH by inactivating (loss-of-function) mutations in the human GnRHR gene leads to distinct forms of autosomal-recessive HH (Table I). GnRHR mutations may also occur in individuals with sporadic HH, but with a lower frequency. The detected mutations in the human GnRHR gene are distributed along the entire coding sequence of the

receptor, including the  $NH_2$  terminus, TMDs 2 to 7, extracellular loops 1 and 2, and the IL3 (Fig. 1C). However, two "hot spots" have been identified: the  $Gln_{106}Arg$  and the  $Arg_{262}Gln$  mutations. Expression of 15 mutated GnRH receptors in heterologous cell systems has shown that these mutations may influence ligand binding, receptor expression at the cell surface, and/or intracellular signal transduction. Although the expression of a given functional defect is apparently related to the structural modification introduced on specific microdomains involved in particular receptor functions, further studies are still necessary for this viewpoint to be accepted with certainty. In fact, in a number of these mutant receptors the function may be partially or completely rescuable *in vitro* by genetic or pharmacologic means, which indicates that the structural alteration may additionally lead to altered intracellular trafficking and reduced cell surface expression. Studies employing these GnRHR mutants have found that the  $Arg_{262}Gln$ ,  $Thr_{32}Ile$ ,  $Cys_{200}Tyr$ ,  $Leu_{266}Arg$ , and  $Cys_{279}Tyr$  mutations predominantly affect IP3 production, whereas in the  $Gln_{106}Arg$ ,  $Ala_{129}Asp$ ,  $Ser_{168}Arg$ ,  $Asn_{10}Lys$ , and  $Arg_{139}His$  mutations, agonist binding at the cell surface is decreased or severely impaired; all of these mutations apparently do not affect the membrane expression of the altered GnRHR in a significant manner. On the other hand, the  $Tyr_{284}Cys$  and  $Glu_{90}Lys$  mutations profoundly affect membrane receptor density as a consequence of either disruption of trafficking to the membrane or instability and subsequent degradation of the mutant receptor. Interestingly, deletion of  $K_{191}$  (which results in increased membrane expression and reduced rate of internalization of the wild-type human GnRHR) in the latter GnRHR mutant efficiently restores membrane expression and agonist-induced, receptor-mediated intracellular signaling. In the  $Ser_{217}Arg$  mutation as well as in the  $Leu_{314}X(Stop)$  mutation, GnRH binding is practically abolished. In this latter mutation, which leads to partial deletion of the TMD 7, the mRNA levels of the receptor were reduced considerably, suggesting that the truncated protein might not be adequately translated and expressed *in vivo*. A truncated nonfunctional GnRHR has also been reported for the homozygous splice junction mutation (G to A replacement) at the intron 1-exon 2 boundary, which results in a transcript showing splicing of exon 1 to exon 3 (i.e., complete deletion exon 2). Finally, the  $Ala_{171}Thr$  mutation in the TMD 4 of the GnRHR presumably disrupts receptor function by impeding conformational mobility of the TMD 3 and 4,

Table 1 Mutations in the Human GnRH Receptor Gene, Genotypes, and Impact on Receptor Function

Mutation	Genotype	Impact on GnRHR function
Gln <sub>106</sub> Arg Arg <sub>262</sub> Gln	Compound heterozygous	Decreased ligand binding—NME? Decreased IP3 production
Arg <sub>262</sub> Gln Tyr <sub>284</sub> Cys	Compound heterozygous	Decreased IP3 production and Decreased membrane expression
Ala <sub>129</sub> Asp Arg <sub>262</sub> Gln	Compound heterozygous	Undetectable ligand binding—NME? Decreased IP3 production
Gln <sub>106</sub> Arg <sup>a</sup> Ser <sub>217</sub> Arg <sup>a</sup> Arg <sub>262</sub> Gln	Compound heterozygous	Decreased ligand binding Undetectable ligand binding Decreased IP3 production
Ser <sub>168</sub> Arg	Homozygous	Undetectable ligand binding—NME
Gln <sub>106</sub> Arg Leu <sub>314</sub> X(Stop)	Compound heterozygous	Decreased ligand binding Undetectable ligand binding
Thr <sub>32</sub> Ile Cys <sub>200</sub> Tyr	Compound heterozygous	Absent IP3 production—NME Absent IP3 production—NME
Leu <sub>266</sub> Arg Gln <sub>106</sub> Arg	Compound heterozygous	Absent IP3 production—NME Decreased ligand binding
Cys <sub>279</sub> Tyr	Homozygous	Absent IP3 production—NME
Asn <sub>10</sub> Lys Gln <sub>106</sub> Arg	Compound heterozygous	Decreased ligand binding—NME Decreased ligand binding
Arg <sub>139</sub> His	Homozygous	Undetectable ligand binding—NME
Glu <sub>90</sub> Lys	Homozygous	Absent membrane expression
Exon2, Complete deletion	Homozygous	Unfunctional truncated receptor
Gln <sub>106</sub> Arg Ala <sub>171</sub> Thr	Compound heterozygous	Undetectable ligand binding—NME

Note. NME, normal membrane expression.

<sup>a</sup>Mutations in the same allele.

resulting in stabilization of the receptor in its inactive conformation.

Individuals with HH due to GnRHR mutations exhibit a strikingly wide spectrum of clinical and biochemical phenotypes, including variable alterations in pubertal development, plasma gonadotropin and sex steroid levels, response to exogenous GnRH administration, and pulsatile pattern of gonadotropin release. These alterations usually occur in the absence of anatomical or functional abnormalities of the hypothalamic–gonadotrope axis. Thus, the hypogonadism due to GnRHR mutations can be complete or partial. The differences between phenotypes in HH due to inactivating GnRHR mutations could be related to the particular allelic combination of the coexisting mutations, with the functional activity of a given mutant being more or less severely affected than that exhibited by the other. Nevertheless, the fact that different phenotypes may be present within affected kindred bearing the same molecular alteration suggests that other factors or mechanisms may be implicated in the phenotypic expression of the GnRH-resistant HH.

## CONCLUSION

GnRH resistance due to inactivating mutations in the GnRHR gene leads to impaired synthesis and secretion of the pituitary gonadotropins and to a distinct form of hypogonadotropic hypogonadism. Several naturally occurring mutations in the GnRHR have been described. These mutations may occur along the entire coding sequence of the receptor and usually result in distinct alterations in receptor function. Naturally occurring mutations in the GnRHR represent unique models for the analysis of the structure–activity relationships of this particular receptor.

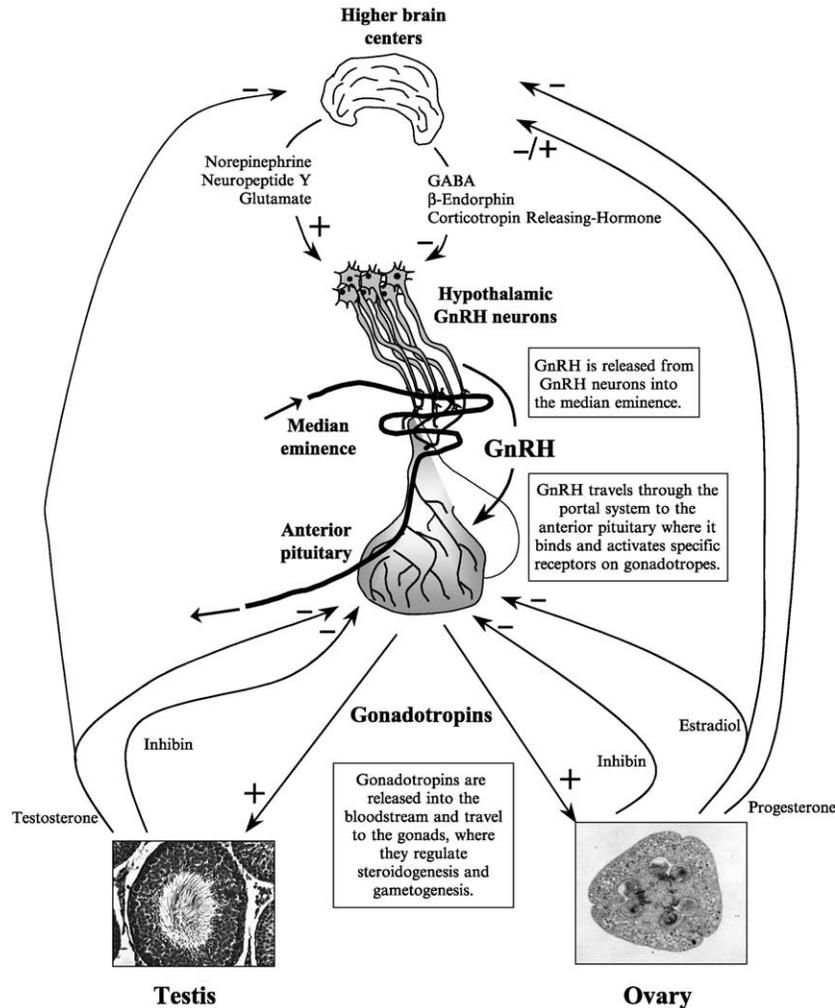
## See Also the Following Articles

Gonadotropin-Releasing Hormone (GnRH) Actions • Gonadotropin-Releasing Hormone Deficiency, Congenital Isolated

## Further Reading

Beranova, M., Oliveira, L. M. B., Bédécarrats, G. Y., Schipani, E., Vallejo, M., Ammini, A. C., Quintos, J. B., Hall, J. E., Martin, K.

- A., Hayes, F. J., Pitteloud, N., Kaiser, U. B., Crowley, W. F., Jr., and Seminara, S. B. (2001). Prevalence, phenotypic spectrum, and modes of inheritance of gonadotropin-releasing hormone receptor mutations in idiopathic hypogonadotropic hypogonadism. *J. Clin. Endocrinol. Metab.* **86**, 1580–1588.
- Caron, P., Chauvin, S., Christin-Maitre, S., Bennet, A., Lahlou, N., Counis, R., Bouchard, P., and Kottler, M. L. (1999). Resistance of hypogonadic patients with mutated GnRH receptor genes to pulsatile GnRH administration. *J. Clin. Endocrinol. Metab.* **84**, 990–996.
- Costa, E. M. F., Bedecarrats, G. Y., Mendonca, B. B., Arnhold, I. J. P., Kaiser, U. B., and Latronico, A. C. (2001). Two novel mutations in the gonadotropin-releasing hormone receptor gene in Brazilian patients with hypogonadotropic hypogonadism and normal olfaction. *J. Clin. Endocrinol. Metab.* **86**, 2680–2686.
- de Roux, N., Young, J., Brailly-Tabard, S., Misrahi, M., Milgrom, E., and Schaison, G. (1999). The same molecular defects of the gonadotropin-releasing hormone receptor determine a variable degree of hypogonadism in affected kindred. *J. Clin. Endocrinol. Metab.* **84**, 567–572.
- de Roux, N., Young, J., Misrahi, M., Genet, R., Chanson, P., Schaison, G., and Milgrom, E. (1997). A family with hypogonadotropic hypogonadism and mutations in the gonadotropin-releasing hormone receptor. *N. Engl. J. Med.* **337**, 1597–1602.
- Karges, B., Karges, W., Mine, M., Ludwig, L., Kühne, R., Milgrom, M., and De Roux, N. (2003). Mutation Ala<sub>171</sub>Thr stabilizes the gonadotropin-releasing hormone receptor in its inactive conformation, causing familial hypogonadotropic hypogonadism. *J. Clin. Endocrinol. Metab.* **88**, 1873–1879.
- Kottler, M. L., Chauvin, S., Lahlou, N., Harris, C. E., Johnston, C. J., Lagarde, J. P., Bouchard, P., Farid, N. R., and Counis, R. (2000). A new compound heterozygous mutation of the gonadotropin-releasing hormone receptor (L314X, Q106R) in a woman with complete hypogonadotropic hypogonadism: Chronic estrogen administration amplifies the gonadotropin defect. *J. Clin. Endocrinol. Metab.* **85**, 3002–3008.
- Layman, L.C., Cohen, D.P., Jin, M., Xie, J., Li, Z., Reindollar, R.H., Bolbolan, S., Bick, D.P., Sherins, R. R., Duck, L. W., Musgrove, L. C., Sellers, J. C., and Neill, J. D. (1998). Mutations in gonadotropin-releasing hormone receptor gene cause hypogonadotropic hypogonadism. *Nat. Genet.* **18**, 14–15.
- Lu, Z.-L., Saldanha, J. W., and Hulme, E. C. (2002). Seven-transmembrane receptors: Crystals clarify. *Trends Pharmacol. Sci.* **23**, 140–146.
- Maya-Núñez, G., Janovick, J., and Conn, P. M. (2000). Combined modification of intracellular and extracellular loci on human gonadotropin-releasing hormone receptor provides a mechanism for enhanced expression. *Endocrine* **13**, 401–409.
- Maya-Núñez, G., Ulloa-Aguirre, A., Söderlund, D., Conn, P. M., and Méndez, J. P. (2000). Molecular basis of hypogonadotropic hypogonadism: Restoration of mutant (E<sup>90</sup>K) GnRH receptor function by a deletion at a distant site. *J. Clin. Endocrinol. Metab.* **87**, 2144–2149.
- Pralong, F. P., Gómez, F., Castillo, E., Cotecchia, S., Abuin, L., Aubert, M. L., Portmann, L., and Gaillard, R. C. (1999). Complete hypogonadotropic hypogonadism associated with a novel inactivating mutation of the gonadotropin-releasing hormone receptor. *J. Clin. Endocrinol. Metab.* **84**, 3811–3816.
- Silveira, L. F. G., Stewart, P. M., Thomas, M., Clark, D. A., Bouloux, P. M. G., and MacColl, G. S. (2002). Novel homozygous splice acceptor site GnRH receptor (GnRHR) mutation: human GnRHR “knockout.” *J. Clin. Endocrinol. Metab.* **87**, 2973–2977.
- Söderlund, D., Canto, P., de la Chesnaye, E., Ulloa-Aguirre, A., and Méndez, J. P. (2001). A novel homozygous mutation in the second transmembrane domain of the gonadotropin releasing hormone receptor gene. *Clin. Endocrinol. (Oxford)* **54**, 493–498.
- Ulloa-Aguirre, A., and Conn, P. M. (1998). G protein-coupled receptors and the G protein family. In “Handbook of Physiology, Section 7: The Endocrine System, Volume 1: Cellular Endocrinology” (P. M. Conn and H. M. Goodman, eds.), pp. 87–124. Oxford University Press, New York.
- Ulloa-Aguirre, A., Janovick, J. A., Leños-Miranda, A., and Conn, P. M. (2003). Misrouted cell surface receptors as a novel disease aetiology and potential therapeutic target: the case of hypogonadotropic hypogonadism due to gonadotropin-releasing hormone resistance. *Expert Opin. Ther. Targets* **7**, 175–185.



**Figure 1** Hypothalamo-Hypophyseal (pituitary)-Gonadal axis. Through its control of the hypothalamo-pituitary-gonadal axis, GnRH regulates reproductive development.

GnRH-II (chicken GnRH-II) is the most ubiquitous form, present in the brain in all vertebrate classes except Agnathans. The ratfish and dogfish, both fish species of ancient origin in the class Chondrichthyes, possess a GnRH-II- and GnRH-I-type isoform, respectively. The ancestral GnRH molecule likely arose prior to the protochordates. In all GnRH peptides in vertebrates and protochordates, certain regions of the molecule have been highly conserved, including the  $\text{NH}_2$  terminus (pGlu<sup>1</sup>) and Ser<sup>4</sup>, and the COOH terminus. These regions and the length of the molecule have remained unchanged during the 500 million years of evolution of the chordates. The conservation of the  $\text{NH}_2$  terminus (pGlu<sup>1</sup>) and Ser<sup>4</sup>, and the COOH terminus suggests that these regions are significant for active conformation of the peptide, effective receptor binding, resistance to enzymatic

degradation, and receptor-mediated events required for gonadotropin release. This significance has been supported by numerous activity studies of GnRH analogues in mammalian and other vertebrate systems.

Further comparative studies on the GnRH family will help to provide clues on the evolution of reproductive mechanisms and insights into our understanding of gene duplication during the early development of vertebrates, structure-activity relations, and the molecular evolution and functional diversity of GnRH.

### GnRH PRECURSOR

The prepro-GnRH molecule is tripartite, encoding a signal peptide that is followed directly by the GnRH

Table I GnRH Family of Peptides

GnRH	1	2	3	4	5	6	7	8	9	10	Percentage identity to mGnRH
Vertebrates											
Mammal	pGlu	His	Trp	Ser	Tyr	Gly	Leu	Arg	Pro	GlyNH <sub>2</sub>	X
Guinea pig	pGlu	<u>Tyr</u>	Trp	Ser	Tyr	Gly	Val	Arg	Pro	GlyNH <sub>2</sub>	80
Chicken-I	pGlu	His	Trp	Ser	Tyr	Gly	Leu	<u>Gln</u>	Pro	GlyNH <sub>2</sub>	90
Rana	pGlu	His	Trp	Ser	Tyr	Gly	Leu	<u>Trp</u>	Pro	GlyNH <sub>2</sub>	90
Seabream	pGlu	His	Trp	Ser	Tyr	Gly	Leu	<u>Ser</u>	Pro	GlyNH <sub>2</sub>	90
Salmon	pGlu	His	Trp	Ser	Tyr	Gly	<u>Trp</u>	<u>Leu</u>	Pro	GlyNH <sub>2</sub>	80
White fish	pGlu	His	Trp	Ser	Tyr	Gly	<u>Met</u>	<u>Asn</u>	Pro	GlyNH <sub>3</sub>	80
Medaka	pGlu	His	Trp	Ser	<u>Phe</u>	Gly	Leu	<u>Ser</u>	Pro	GlyNH <sub>2</sub>	80
Catfish	pGlu	His	Trp	Ser	<u>His</u>	Gly	Leu	<u>Asn</u>	Pro	GlyNH <sub>2</sub>	80
Herring	pGlu	His	Trp	Ser	<u>His</u>	Gly	Leu	<u>Ser</u>	Pro	GlyNH <sub>2</sub>	80
Chicken-II	pGlu	His	Trp	Ser	<u>His</u>	Gly	<u>Trp</u>	<u>Tyr</u>	Pro	GlyNH <sub>2</sub>	70
Dogfish	pGlu	His	Trp	Ser	<u>His</u>	Gly	<u>Trp</u>	<u>Leu</u>	Pro	GlyNH <sub>2</sub>	70
Lamprey-III	pGlu	His	Trp	Ser	<u>His</u>	<u>Asp</u>	<u>Trp</u>	<u>Lys</u>	Pro	GlyNH <sub>2</sub>	60
Lamprey-I	pGlu	His	<u>Tyr</u>	Ser	<u>Leu</u>	<u>Glu</u>	<u>Trp</u>	<u>Lys</u>	Pro	GlyNH <sub>2</sub>	50
Invertebrates											
Tunicate-I	pGlu	His	Trp	Ser	<u>Asp</u>	<u>Tyr</u>	<u>Phe</u>	<u>Lys</u>	Pro	GlyNH <sub>2</sub>	60
Tunicate-II	pGlu	His	Trp	Ser	<u>Leu</u>	<u>Cys</u>	<u>His</u>	<u>Ala</u>	Pro	GlyNH <sub>2</sub>	60

Note. There are currently 16 known forms of GnRH: 14 in vertebrates and 2 in invertebrates.

decapeptide, a dibasic cleavage site, and the GnRH-associated peptide (GAP), as diagrammed in Fig. 2.

Similar to other neuropeptides, GnRH is first synthesized as a larger precursor protein, called prepro GnRH, which is then processed to its final decapeptide form. Neuropeptide precursors consist of at least 60 amino acids and include a signal sequence at the amino terminus and a spacer region at the carboxy terminus. The spacer region is separated from the neuropeptide by a dibasic amino acid cleavage site and may or may not have a biological function. The amino acid sequence of prepro GnRH has been indirectly identified by the isolation of its cDNA sequence.

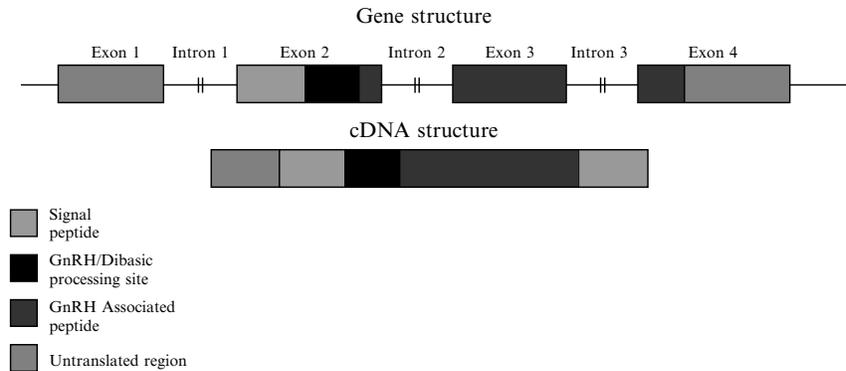
Table III shows a complete list of the GnRH cDNAs/genes identified to date.

Of the 16 forms of GnRH identified, the cDNA or gene sequences of 11 forms have been determined. Due to the low identity of the prepro-GnRH cDNA sequence among forms, investigators are restricted to the use of degenerate oligonucleotide primers or probes based on the amino acid sequence of the specific GnRH for identification of GnRH. Classical molecular biological techniques were used for isolating the first GnRH cDNA. This entailed screening a genomic library with degenerate oligonucleotide probes. The full cDNA sequence was then cloned

Table II The Lineages of GnRH

Type GnRH	Brain distribution/origin	Primary GnRH structures identified in vertebrates
GnRH-I	Hypothalamus, diencephalon/ Olfactory origin	Mammal GnRH in mouse, primate, human, sheep, pig, eel, newt, frog; chicken GnRH-I in chicken, lizard; salmon GnRH in goldfish, salmon; catfish GnRH in catfish; dogfish GnRH in dogfish
GnRH-II	Midbrain/Ventricular ependyma origin	Chicken GnRH-II in mouse, primate, human, chicken, lizard, frog, newt, eel, goldfish, catfish, salmon, medaka, red seabream, tilapia, ratfish
GnRH-III	Telencephalon/Olfactory origin	Salmon GnRH in medaka, red seabream, tilapia
GnRH-IV	Hypothalamus, diencephalon/ ventricular origin	Lamprey GnRH-I and lamprey GnRH-III in lamprey

Note. GnRH is divided into four types based on a combination of function, location of expression, and molecular phylogenetic analysis.



**Figure 2** Prepro-GnRH tripartite structure. The GnRH precursor is made up of three regions encoding the signal peptide, the GnRH decapeptide, and the GnRH-associated peptide.

from a brain cDNA library. Salmon GnRH (type GnRH-III) was also isolated by screening a brain cDNA library with degenerate oligonucleotides. Later in the 1990s, the polymerase chain reaction (PCR) was used to amplify specific gene sequences from a messenger ribonucleic acid (mRNA) population or from cDNA or genomic libraries. This method has now been used to isolate the nine additional GnRH cDNA sequences. The first cDNA sequences for chicken GnRH-I, chicken GnRH-II, catfish GnRH, and seabream GnRH all were determined using reverse transcription-PCR (RT-PCR) and/or rapid amplification of cDNA ends (RACE). Both of these methods are based on amplification via PCR.

In the African cichlid, the salmon GnRH gene was first isolated in 1991 by Bond and colleagues, and later the prepro-chicken GnRH-II gene was isolated from the same species. The presence of two cDNA sequences confirmed that multiple GnRH forms present in an organism were not produced from differential processing of one GnRH gene but rather were, in fact, transcribed from separate genes. Since that time, multiple GnRH genes have been isolated from many species.

The signal peptide contains a core of hydrophobic amino acids that is followed by alternating polar and nonpolar amino acids, except in the case of seabream GnRH, which contains only polar residues at the carboxy terminus of the signal peptide. The signal peptide is considered to aid in the prohormone's transport across the endoplasmic reticulum membrane for posttranslational processing.

The characteristic hydrophobic signal peptide targets the prohormone to the proper cellular organelle. The mechanism for this is currently being elucidated. It is thought that the hydrophobic residues may be interacting with the signal recognition particle

(SRP), which targets proteins to the rough endoplasmic reticulum (RER). The SRP is a ribonucleoprotein complex containing a 7S RNA molecule and six polypeptides. One of the associated polypeptides, SRP54, is a 54-kDa protein that binds to the signal portion of a prepro peptide as it is translated at the ribosome. SRP54 contains a putative guanosine 5'-triphosphate (GTP)-binding domain as well as a methionine-rich region. GTP is required for releasing SRP from the signal peptide and ribosome. The region that is rich in methionine residues is proposed to be the region that binds to the signal peptide. The methionine-rich region is predicted to form four amphipathic helices that position the hydrophobic residues entirely on one face, with polar residues exclusively on the opposite face. This secondary structure would suggest that the hydrophobic signal peptide could bind in the hydrophobic pocket lined by amphipathic helices. The methionine residues of SRP54 are proposed to form methionine bristles to accommodate a wide variety of hydrophobic signal sequences, and this could explain the lack of apparent consensus sequence of signal peptides apart from their generally high hydrophobicity.

After the signal peptide is cleaved from the prohormone, the glutamine residue at GnRH position 1 is cyclized to pyro-glutamate. Although the mechanism of this cyclization is unknown, the physiological spontaneous rate of formation occurs very slowly, and enzymatic activity catalyzing this formation has been described in brain, pituitary, and lymphocytes.

Following the GnRH decapeptide in the precursor hormone structure is a glycine residue that donates its amide group to the terminal glycine residue of GnRH. This amidation is catalyzed by peptidyl glycine  $\alpha$ -amidating mono-oxygenase only after GAP is removed. This enzyme requires that the amide donor

Table III GnRH cDNAs and Genes Identified to Date

Isoform	Organism	Year	Reference	
Mammalian	Human	1984	Seeburg <i>et al.</i> , <i>Nature</i>	
	Norway rat	1986	Adelman <i>et al.</i> , <i>Proc. Natl. Acad. Sci. USA</i>	
	Human	1986	Adelman <i>et al.</i> , <i>Proc. Natl. Acad. Sci. USA</i>	
	<b>Mouse</b>	1986	Mason <i>et al.</i> , <i>Science</i>	
	<b>Norway rat</b>	1989	Bond <i>et al.</i> , <i>Mol. Endocrinol.</i>	
	African clawed frog	1994	Hayes <i>et al.</i> , <i>Endocrinology</i>	
	Rhesus monkey	1996	Dong <i>et al.</i> , <i>Mol. Cell Endocrinol.</i>	
	Tree shrew	1996	Kasten <i>et al.</i> , <i>Gen. Comp. Endocrinol.</i>	
	Japanese eel	1999	Okubo <i>et al.</i> , <i>Zool. Sci.</i>	
	Bullfrog	2001	Wang <i>et al.</i> , <i>J. Exp. Zool.</i>	
	Guinea pig	Guinea pig	1997	Jimenez-Linan <i>et al.</i> , <i>Endocrinology</i>
Chicken-I	Chicken	1993	Dunn <i>et al.</i> , <i>J. Mol. Endocrinol.</i>	
Rana	Frog	2000	Yoo <i>et al.</i> , <i>Mol. Cell Endocrinol.</i>	
Medaka	Medaka	2000	Okubo <i>et al.</i> , <i>Biochem. Biophys. Res. Commun.</i>	
Catfish	African catfish	1994	Bogerd <i>et al.</i> , <i>Eur. J. Biochem.</i>	
Seabream	Gilthead seabream	1995	Gothilf <i>et al.</i> , <i>Mol. Mar. Biol. Biotechnol.</i>	
	<i>Haplochromis burtoni</i>	1995	White <i>et al.</i> , <i>Proc. Natl. Acad. Sci. USA</i>	
	Red sea bream	1996	Okuzawa, unpublished	
	<b>Striped sea bass</b>	1998	Chow <i>et al.</i> , <i>J. Mol. Endocrinol.</i>	
	<b><i>Haplochromis burtoni</i></b>	1998	White <i>et al.</i> , <i>Gen. Comp. Endocrinol.</i>	
	European sea bass	2002	Zmora <i>et al.</i> , <i>J. Endocrinol.</i>	
	Barfin flounder	2002	Amano <i>et al.</i> , <i>Gen. Comp. Endocrinol.</i>	
	<b>Nile tilapia</b>	2002	Farahmand <i>et al.</i> , unpublished	
	Chicken-II	<i>Haplochromis burtoni</i>	1994	White <i>et al.</i> , <i>Proc. Natl. Acad. Sci. USA</i>
		African catfish	1994	Bogerd <i>et al.</i> , <i>Eur. J. Biochem.</i>
Tree shrew		1996	Kasten <i>et al.</i> , <i>Gen. Comp. Endocrinol.</i>	
Goldfish		1996	Lin <i>et al.</i> , <i>Gen. Comp. Endocrinol.</i>	
Human		1998	White <i>et al.</i> , <i>Proc. Natl. Acad. Sci. USA</i>	
<b>Striped sea bass</b>		1998	Chow <i>et al.</i> , <i>J. Mol. Endocrinol.</i>	
<b><i>Haplochromis burtoni</i></b>		1998	White <i>et al.</i> , <i>Gen. Comp. Endocrinol.</i>	
<b>Human</b>		1998	White <i>et al.</i> , <i>Proc. Natl. Acad. Sci. USA</i>	
House shrew		1998	White <i>et al.</i> , unpublished	
Silver-Gray brushtail possum		1999	Lawrence <i>et al.</i> , unpublished	
Rio cauca caecilian		1999	Ebersole <i>et al.</i> , unpublished	
Rhesus monkey		1999	Urbanski <i>et al.</i> , <i>Endocrinology</i>	
Japanese eel		1999	Okubo <i>et al.</i> , <i>Zool. Sci.</i>	
Medaka		2000	Okubo <i>et al.</i> , <i>Biochem. Biophys. Res. Commun.</i>	
Australian bonytongue		2001	Okubo and Aida, <i>Gen. Comp. Endocrinol.</i>	
Bullfrog		2001	Wang <i>et al.</i> , <i>J. Exp. Zool.</i>	
European sea bass		2002	Zmora <i>et al.</i> , <i>J. Endocrinol.</i>	
Zebra fish		2002	Whitlock <i>et al.</i> , unpublished	
Zebra fish		2002	Stevens <i>et al.</i> , unpublished	
Barfin flounder		2002	Amano <i>et al.</i> , <i>Gen. Comp. Endocrinol.</i>	
Common carp	2002	Li <i>et al.</i> , unpublished		
Salmon	<i>Haplochromis burtoni</i>	1991	Bond <i>et al.</i> , <i>Mol. Endocrinol.</i>	
	Atlantic salmon	1992	Klungland <i>et al.</i> , <i>Mol. Cell Endocrinol.</i>	
	<b>Atlantic salmon</b>	1992	Klungland <i>et al.</i> , <i>Mol. Cell. Endocrinol.</i>	
	Rainbow trout	1992	Alestrom <i>et al.</i> , <i>Mol. Mar. Biol. Biotechnol.</i>	
	Cherry salmon	1992	Suzuki <i>et al.</i> , <i>J. Mol. Endocrinol.</i>	

(continues)

Table III (continued)

Isoform	Organism	Year	Reference
	<b>Brook trout</b>	1992	Klungland <i>et al.</i> , <i>Mol. Cell Endocrinol.</i>
	<b>Chinook salmon</b>	1992	Klungland <i>et al.</i> , <i>Mol. Cell Endocrinol.</i>
	<b>Rainbow trout</b>	1992	Klungland <i>et al.</i> , <i>Mol. Cell Endocrinol.</i>
	<b>Brown trout</b>	1992	Klungland <i>et al.</i> , <i>Mol. Cell Endocrinol.</i>
	Plainfin midshipman	1995	Grober <i>et al.</i> , <i>Gen. Comp. Endocrinol.</i>
	Sockeye salmon	1995	Ashihara <i>et al.</i> , <i>J. Mol. Endocrinol.</i>
	<b>Sockeye salmon</b>	1995	Coe <i>et al.</i> , <i>Mol. Cell Endocrinol.</i>
	Goldfish	1996	Lin <i>et al.</i> , <i>Gen. Comp. Endocrinol.</i>
	<b><i>Haplochromis burtoni</i></b>	1998	White <i>et al.</i> , <i>Gen. Comp. Endocrinol.</i>
	Medaka	2000	Okubo <i>et al.</i> , <i>Biochem. Biophys. Res. Commun.</i>
	Zebrafish	2000	Torgersen <i>et al.</i> , unpublished
	Australian bonytongue	2001	Okubo and Aida, <i>Gen. Comp. Endocrinol.</i>
	European sea bass	2002	Zmora <i>et al.</i> , <i>J. Endocrinol.</i>
	<b>Zebrafish</b>	2002	Torgersen <i>et al.</i> , unpublished
	Barfin flounder	2002	Amano <i>et al.</i> , <i>Gen. Comp. Endocrinol.</i>
	<b>Mozambique tilapia</b>	2002	Molina <i>et al.</i> , unpublished
Lamprey-III	Sea lamprey	2002	Silver <i>et al.</i> , <i>Am. Zool.</i>
	Pacific sea lamprey	2002	Silver <i>et al.</i> , <i>Am. Zool.</i>
	Australian lamprey	2002	Silver <i>et al.</i> , <i>Am. Zool.</i>
	Pouched lamprey	2002	Silver <i>et al.</i> , <i>Am. Zool.</i>
	American brook lamprey	2003	Silver <i>et al.</i> , <i>Am. Zool.</i>
	Northern brook lamprey	2003	Silver <i>et al.</i> , <i>Am. Zool.</i>
	Western brook lamprey	2003	Silver <i>et al.</i> , <i>Am. Zool.</i>
	Silver lamprey	2003	Silver <i>et al.</i> , <i>Am. Zool.</i>
Lamprey-I	Sea lamprey	2001	Suzuki <i>et al.</i> , <i>J. Mol. Endocrinol.</i>

Note. There are currently 57 GnRH precursor cDNAs and 17 genes (in boldface) identified and sequenced.

be at the C terminus of the peptide. It has been proposed that GAP-releasing enzyme initiates this posttranslational modification by an endoproteolytic event. The enzyme is a serine protease that recognizes the lysine<sup>12</sup>-arginine<sup>13</sup> dibasic cleavage site sequence that typically acts in brain prohormones for proteolytic cleavage of the active peptide from the precursor molecule and cleaves the amide bond directly following the residues. The next enzyme proposed to be involved is hypothalamic carboxypeptidase E, which sequentially removes the arginine<sup>13</sup> and the lysine<sup>12</sup> residues. This leaves the glycine<sup>11</sup> at the C terminus of the peptide as a substrate for peptidyl glycine  $\alpha$ -amidating mono-oxygenase.

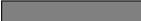
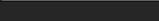
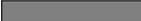
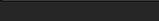
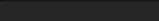
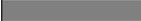
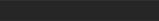
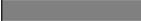
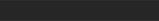
The initial cleaving event between residues 13 and 14 is thought to occur during vesicle formation at the Golgi apparatus, whereas further processing occurs within the secretory granules where GAP and bioactive GnRH are stored until needed for release. The function of GAP remains undetermined. GAP, in conjunction with the signal peptide, most likely functions to ensure that the hormone is long enough for

insertion into the endoplasmic reticulum for proper processing to occur. Researchers have questioned a possible biological role for GAP because there is *in vitro* evidence that human GAP is responsible for both inhibiting prolactin release and stimulating gonadotropin hormone (GTH) release in rat pituitary as well as inhibiting prolactin release in the teleost fish tilapia. A bioactive role for GAP is also supported by rat *in vivo* studies, in which the first 13 residues of the GAP molecule were shown to stimulate the release of GTH.

Although the molecular organization of GnRH precursors appears to be similar throughout various species, the amino acid composition is highly divergent. The sequence encoding the GnRH peptide and the following cleavage sites are highly conserved among all of the vertebrates, whereas the signal peptide and GAP remain quite divergent (Fig. 3).

In contrast to other known vertebrate GnRH precursors, which typically have one transcript (and in two cases have two transcripts), three distinct transcripts have been isolated and sequenced in lampreys.

## Mammalian GnRH precursor identities

	Signal peptide	GnRH	GAP
<i>Homo sapiens</i> mGnRH			
	67%	100%	70%
<i>Mus musculus</i> mGnRH			
	61%	100%	70%
<i>Rattus norvegicus</i> mGnRH			
	61%	100%	84%
<i>Tupaia glis</i> mGnRH			
	22%	100%	34%
<i>Xenopus laevis</i> mGnRH			
	22%	100%	26%
<i>Rana catesbeiana</i> mGnRH			
	23%	100%	34%
<i>Anguilla japonica</i> mGnRH			

**Figure 3** Prepro-GnRH sequence comparison. The GnRH precursor is highly conserved in structure and in the GnRH decapeptide sequence.

The lamprey GnRH-I (type GnRH-IV) transcripts, termed GAP49, GAP50, and GAP58, differed in the length of the GAP coding sequence and were demonstrated to be the products of a single gene. Analysis of the lamprey GnRH-I gene intron-2 splice junction demonstrated that alternate splicing produces the different lamprey GnRH-I transcripts. Lamprey GnRH-I is the first GnRH gene demonstrated to use splice sequence variants to produce multiple transcripts that may reflect an ancestral gene regulatory mechanism.

The GnRH genes that have been identified have maintained a highly conserved pattern of introns and exons (Fig. 2). The coding region is dispersed over four exons, where the first exon is untranslated; the second encodes the signal peptide, the GnRH decapeptide, the cleavage site, and the N-terminal portion of GAP; the third consists entirely of the middle portion of the GAP molecule; and the fourth contains the C-terminal of GAP as well as the 3' untranslated region. The only noticeable difference between the salmon and mammalian GnRH gene structures is their relative intron sizes. Because of the similar architecture, the emergence of multiple forms of GnRH has been attributed to nucleotide base changes and not to differential splicing of message or variable processing of the precursor protein.

## GNRH RECEPTOR

In light of the crucial role GnRH plays in human physiology and disease, its receptor has been a subject of intense research for many years. Numerous studies

on the binding characteristics of the GnRH receptor were performed throughout the 1970s and 1980s. In 1992, the first successful cloning of a GnRH receptor from the mouse using a homology-based PCR amplification scheme was reported. Later that year, the human GnRH receptor cDNA was reported as well. Since these landmark studies, the GnRH receptor primary structure has been reported in several other organisms (Table IV). Multiple receptor forms have been reported in brains and pituitaries of primates, teleost fish (e.g., goldfish, medaka), and amphibians (e.g., bullfrog, Korean frog [*Rana dybowskii*]). The GnRH receptors have been classified into two major groupings, type I and type II, based primarily on sequence identity and major structural characteristics. Further subdivisions of these two groups have been proposed, but more studies on expression, activity, and pharmacological characteristics are required to establish a solid subdivision of these two major groups.

Analysis of the sequences of the first identified GnRH receptors defined them as belonging to the G protein-coupled receptor (GPCR) superfamily of receptors. The members of this superfamily share a common general structure composed of seven hydrophobic  $\alpha$ -helical transmembrane domains connected by hydrophilic intracellular and extracellular loops, an extracellular N-terminal tail, and an intracellular C-terminal tail. The GnRH receptor belongs to the class A rhodopsin-like GPCR subfamily. This subfamily is characterized by conserved amino acids, or motifs, in certain positions of the receptor. The GnRH receptors exhibit an interesting pattern of evolutionary change from the otherwise highly conserved motifs of the class A GPCRs (Fig. 4). Of the structural variations seen in the GnRH receptors, most notable is the absence of any C-terminal tail in the type I mammalian GnRH receptors. These receptors are the only GPCRs without a C-terminal intracellular tail. The implications of this drastic modification of structure are still not well understood.

The identification of the GnRH receptors in the goldfish and the bullfrog marked the first time that multiple receptors had been cloned from one species. Recently, several studies have been published reporting the identification of multiple GnRH receptors in brains and pituitaries of other organisms. Two distinct receptors have been cloned in another teleost fish, the medaka, and the three receptor types identified in the bullfrog have been supported by isolation of three very similar receptors in another frog, the Korean frog *R. dybowskii*. A second GnRH receptor has been identified in the marmoset, and similar novel

Table IV GnRH Receptors Identified to Date

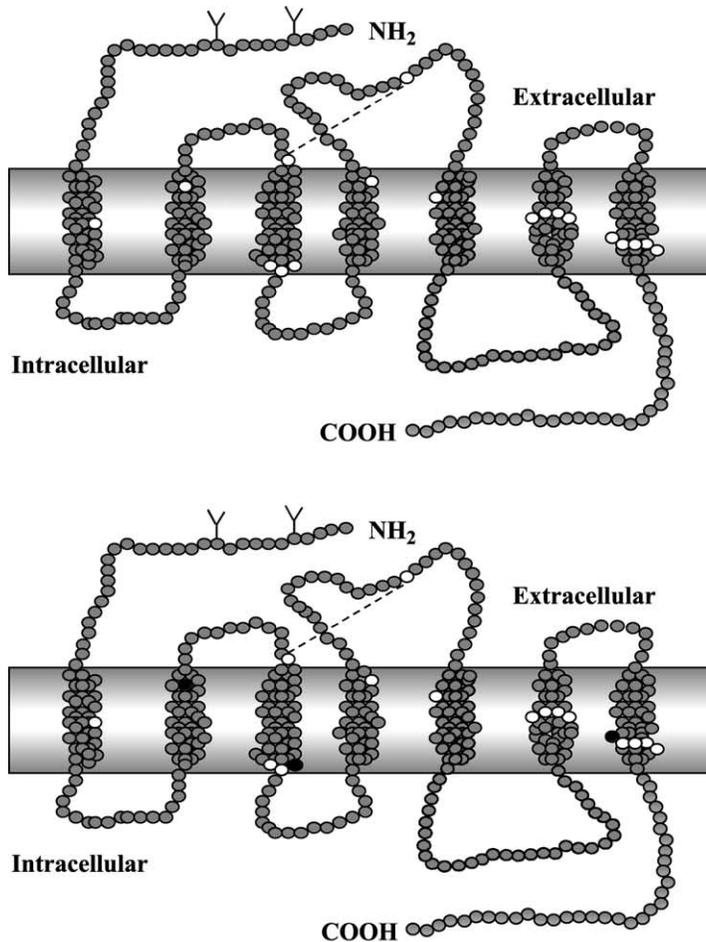
Organism (receptor number)	Year	Reference
<b>Human</b>	1992	Kakar <i>et al.</i> , <i>Biochem. Biophys. Res. Commun.</i>
	1993	Chi <i>et al.</i> , <i>Mol. Cell Endocrinol.</i>
<b>Mouse</b>	1992	Tsutsumi <i>et al.</i> , <i>Mol. Endocrinol.</i>
	1992	Perrin <i>et al.</i> , <i>Biochem. Biophys. Res. Commun.</i>
	1993	Reinhart <i>et al.</i> , <i>J. Biol. Chem.</i>
<b>Rat</b>	1992	Eidne <i>et al.</i> , <i>Mol. Cell Endocrinol.</i>
	1992	Kaiser <i>et al.</i> , <i>Biochem. Biophys. Res. Commun.</i>
	1993	Kudo <i>et al.</i> , <i>Zool. Sci.</i>
<b>Cow</b>	1993	Kakar <i>et al.</i> , <i>Domest. Anim. Endocrinol.</i>
<b>Pig</b>	1994	Weesner and Matteri, <i>J. Anim. Sci.</i>
<b>Sheep</b>	1996	Campion <i>et al.</i> , <i>Gene</i>
<b>Marmoset</b>	1999	Byrne <i>et al.</i> , <i>J. Endocrinol.</i>
<b>Australian brushtail possum</b>	2000	King <i>et al.</i> , <i>Gen. Comp. Endocrinol.</i>
<b>Bonnet monkey</b>	2000	Santra <i>et al.</i> , <i>Mol. Hum. Reprod.</i>
<b>Dog</b>	2000	Cui <i>et al.</i> , <i>Mol. Endocrinol.</i>
<b>Tamar wallaby</b>	2002	Cheung <i>et al.</i> , <i>Reprod. Fertil. Dev.</i>
African catfish	1997	Tensen <i>et al.</i> , <i>Eur. J. Biochem.</i>
Goldfish A (1)	1999	Illing <i>et al.</i> , <i>Proc. Natl. Acad. Sci. USA</i>
Goldfish B (2)	1999	Illing <i>et al.</i> , <i>Proc. Natl. Acad. Sci. USA</i>
African clawed frog	2000	Troskie <i>et al.</i> , <i>Endocrinology</i>
Japanese eel	2000	Okubo <i>et al.</i> , <i>Gen. Comp. Endocrinol.</i>
Rainbow trout	2000	Madigou <i>et al.</i> , <i>Biol. Reprod.</i>
Striped sea bass	2000	Alok <i>et al.</i> , <i>Mol. Cell Endocrinol.</i>
African green monkey	2001	Neill <i>et al.</i> , <i>Biochem. Biophys. Res. Commun.</i>
Bullfrog (1)	2001	Wang <i>et al.</i> , <i>Proc. Natl. Acad. Sci. USA</i>
Bullfrog (2)	2001	Wang <i>et al.</i> , <i>Proc. Natl. Acad. Sci. USA</i>
Bullfrog (3)	2001	Wang <i>et al.</i> , <i>Proc. Natl. Acad. Sci. USA</i>
Chicken	2001	Sun <i>et al.</i> , <i>J. Biol. Chem.</i>
Marmoset (2)	2001	Millar <i>et al.</i> , <i>Proc. Natl. Acad. Sci. USA</i>
Medaka (1)	2001	Okubo <i>et al.</i> , <i>Endocrinology</i>
Medaka (2)	2001	Okubo <i>et al.</i> , <i>Endocrinology</i>
Korean frog (1)	2003	Seong <i>et al.</i> , <i>Endocrinology</i>
Korean frog (2)	2003	Seong <i>et al.</i> , <i>Endocrinology</i>
Korean frog (3)	2003	Seong <i>et al.</i> , <i>Endocrinology</i>
Lamprey	2001	Nucci <i>et al.</i> , <i>Amer. Zool.</i>

*Note.* The amino acid sequences of 33 GnRH receptors have been deduced. Receptors that lack the evolutionarily conserved C-terminal tail are in boldface, whereas receptors that retain this domain are in regular type.

receptors have been described from the African green monkey and the rhesus monkey *Macaca mulatta*. These receptors have been shown to contain intracellular C-terminal tails, unlike all previously identified mammalian GnRH receptors. The discovery and description of these receptors opens up exciting new possibilities in the study of the importance of the C-terminal tail in receptor signaling, cycling, expression, and desensitization.

The gene structure of the GnRH receptor has also been described in multiple species: human, sheep,

mouse, rat, dog, Japanese eel, African clawed frog, and medaka. With the exception of the medaka receptor 2 and the clawed frog receptor, all of these genes have shown a conserved pattern of introns and exons. This pattern is composed of three exons and two introns. Exon 1 extends from the 5' end of the transcript through the middle of the region encoding transmembrane helix 4, exon 2 encodes from the middle of transmembrane helix 4 to the middle of intracellular loop 3, and exon 3 includes the rest of the coding sequence and the 3' untranslated region.



### Class A GPCRs

Motif	Sequence
TM 1	N
TM 2	D
TM 3	DRY
TM 4	W
TM 5	P
TM 6	CWXP
TM 7	NPXXY
Thiol bond	Present
C-terminal tail	Present

### GnRH Receptors

Motif	Type 1 (most mammalian)	Type 2 (most other vertebrate forms)
TM 1	N	N
TM 2	N	D
TM 3	DRS	DRH/DRQ
TM 4	W	W
TM 5	P	P
TM 6	CWXP	CWXP
TM 7	DPXXY	DPXXY
Thiol bond	Present	Present
C-terminal tail	Absent (in mammals)	Present

**Figure 4** Class A GPCR and GnRH-R motifs. The GnRH receptors exhibit an intriguing pattern of motif change over the course of vertebrate evolution. Most notably, mammalian type I GnRH receptors are the only GPCRs that lack an intracellular tail.

Investigations into the expression level and localization of the GnRH receptor have also indicated the presence of great complexity in the function and regulation of GnRH at the level of the receptor. These studies have shown wide distribution of the GnRH receptor in the brain, particularly in the hypothalamus and midbrain, and also in the pituitary (as expected). In addition, GnRH receptors have been found in detectable levels in other organs. In humans, GnRH receptor mRNA was shown to be present in the pituitary, ovary, testis, breast, and prostate. In fish and amphibians, GnRH receptor mRNAs have been found in the brain, pituitary, ovary, testis, and retina.

Interestingly, in the studies that have characterized multiple GnRH receptors in the same species, differential tissue distribution of receptor isoforms has been demonstrated. In the goldfish, differential tissue-specific expression was reported, whereas in the bullfrog, differential expression was also shown but only

within regions of the brain and pituitary rather than among many tissues. The variability of expression among the three bullfrog GnRH receptors was also shown to correspond with stages of the reproductive cycle. This is the first solid evidence of regulation of reproduction by expression variance of multiple GnRH receptors.

The extensive cloning of GnRH receptors from various species and the characterization and conservation of gene structure have enabled researchers to explore the structure–function aspects of the receptors and how their differences and similarities give insight into the complex integration and function of the GnRH system. Many studies have investigated the importance of the C-terminal tail in receptor signaling, internalization, desensitization, and expression. The unusual GnRH receptor variations in the conserved motifs typically found in the class A GPCRs have been investigated as well, as have specific

structural components of the GnRH receptor. This body of data, combined with modeling studies of the GnRH peptide and the hundreds of activity studies on GnRH analogues, has produced a putative picture of how GnRH binds to its receptor. In the absence of a crystallographic structure of the GnRH-bound receptor, these studies provide the best information on the mechanics of GnRH binding.

## GnRH IN MAMMALS

GnRH was first isolated from porcine and ovine hypothalamic extracts, giving rise to the popularly held view that a single form of GnRH (mammalian GnRH) regulates the hypothalamo-pituitary-gonadal axis in all mammals. The question that has arisen over the years is how one GnRH differentially regulates the release of two gonadotropins: LH and FSH. Although there have been a number of different models proposed, questions remain and have fueled the quest for a separate FSH-releasing factor or novel GnRH isoform.

Recent studies have demonstrated that at least two different GnRH isoforms are expressed within the brain of a single vertebrate species. Current thought suggests that one GnRH (type GnRH-I) functions as a neurohormone regulating the pituitary in mediating the release of gonadotropin. The other form (type GnRH-II) may have a neurotransmitter or neuromodulatory function and is generally localized in areas outside the hypothalamus, particularly in midbrain regions. In mammals, the second GnRH isoform that has been identified to date is chicken GnRH-II. The only direct evidence of the existence of GnRH forms other than mammalian GnRH in mammals has come from the identification of GnRH complementary DNAs in three species: the tree shrew, the guinea pig, and the human. Two prepro-GnRH mRNAs identified in the tree shrew encoded mammalian and chicken GnRH-II. In the guinea pig, a prepro GnRH encodes for a unique form of GnRH, guinea pig GnRH. Most recently in the human, two genes for mammal GnRH and chicken GnRH-II were demonstrated.

Using indirect methods in mammals, a limited number of species, monotremes, marsupials, rats, and primitive eutherians have been examined for variant forms of GnRH through immunocytochemistry (ICC), high-performance liquid chromatography (HPLC), and radioimmunoassay (RIA) using specific antibodies to various GnRH forms. In most of the species examined, immunoreactive chicken GnRH-II

was shown to be present generally in the midbrain. In recent studies in rats using a double-labeled immunocytochemical technique, lamprey GnRH-III neurons not only were observed in regions that control FSH release but also were colocalized with mammalian GnRH neurons in areas primarily controlling LH release. These studies used indirect techniques with various heterologous antibodies. Immunoreactive (ir) GnRH will need confirmation by identification of these ir-variant GnRH forms through determination of the primary structure by protein purification or molecular cloning followed by extensive biological activity studies.

The first study to demonstrate that a lamprey-like GnRH was present in the human hypothalamus and median eminence used a combination of ICC, HPLC, and RIA. In these same studies, the hypothalamic distribution of immunopositive lamprey-like GnRH neurons was similar to that observed for those containing the mammalian GnRH. More recently, also using similar indirect methods, it was demonstrated that a chicken GnRH-II-like form was found in stump-tail and rhesus monkeys but that only a few of the ir-chicken GnRH-II cells were in the posterior basal hypothalamus; most of the immunopositive neurons to chicken GnRH-II antiserum were shown to be in the midbrain. These studies did not screen for other forms of GnRH such as the lamprey GnRH forms, nor did their data suggest that chicken GnRH-II is a neurohormone. Thus, there is incomplete and contrasting data on the nature of GnRHs in primates. Confirmation of the exact nature of a second or possibly third form of GnRH in primates and other mammals will require isolation, sequence analysis, determination of localization, and biological function studies.

## STATUS OF AN FSH-RELEASING FACTOR

Although research has shown convincingly that GnRH regulates LH release, there are not any definitive data supporting GnRH control of FSH release. However, it has been suggested that regulation of FSH secretion involves a complex balance among stimulation by GnRH from the hypothalamus, inhibitory feedback by sex steroids (testosterone and estradiol) and inhibin B from the gonads, and autocrine/paracrine modulation by activin and follistatin within the pituitary. As a result, for many years there has been a quest for the identification of a separate FSH-releasing factor in mammals. It has been shown

that lamprey GnRH-III (type GnRH-IV) is highly selective in stimulating FSH release in rats and cows. Combined with the earlier immunocytochemical data of a lamprey GnRH-like form in various species of mammals, this provides convincing evidence for a lamprey GnRH-like molecule in the mammalian hypothalamus. In light of this more recent evidence that there might be a novel GnRH in mammals and other vertebrates, the possibility exists that a lamprey-like peptide might be an FSH-releasing factor. Thus, the possibility exists that there is a second GnRH-like molecule that is also a hypothalamic factor.

## GNRH ANALOGUES

Since 1971, when the primary structure of mammalian GnRH was determined, more than 10,000 analogues have been made to GnRH and been tested in hundreds of studies in mammals. As a result of these studies, several mammalian GnRH analogues have been shown to be highly successful and are currently being used for sterilization, conception, and other therapeutic and clinical applications. For example, Lupron Depot, a GnRH analogue, is now one of the leading chemical treatments for advanced prostate cancer and endometriosis in humans. Continuous treatment with Lupron Depot results in decreased levels of LH and FSH. In males, testosterone is reduced to castrate levels; in premenopausal females, estrogens are reduced to postmenopausal levels. GnRH antagonists are also used clinically, and the most effective GnRH antagonists have substitutions of one amino acid in position 6 as well as substitutions in positions 1, 2, and 3. However, because of the lack of systematic examination in the characterization of these various analogues, there is still critical information that is needed in our understanding of their biological activity. Information on the comparative activities of the GnRH family in vertebrates may provide important hypotheses about the role of specific amino acids in GnRH binding and receptor activation and may lead to the development of more potent GnRH analogues.

## CONCLUSION

Since the discovery of GnRH in 1971, research has exploded relating to the structure and function

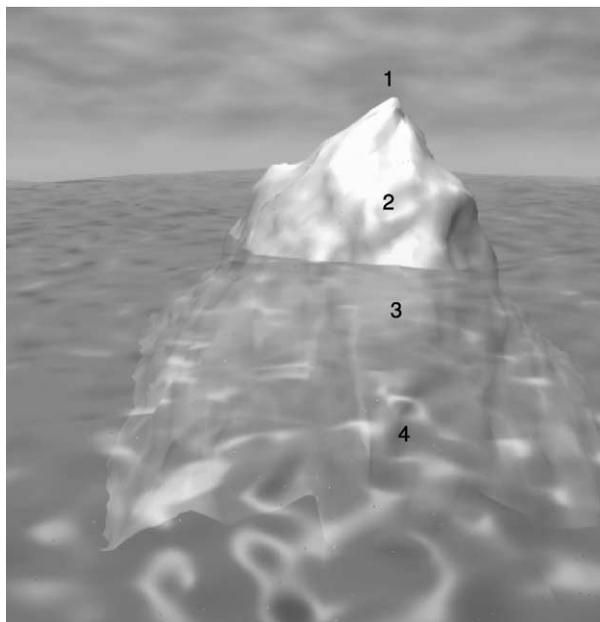
of GnRH and its receptor. Our understanding of GnRH, as the central figure in the control of reproduction, is critical at the molecular, biochemical, and physiological levels. We have come to realize that the GnRH system has proven to be extremely complex and that there is still much to be learned. Importantly, GnRH has become an important subject of not only physiological research but also medical treatments.

## See Also the Following Articles

Corticotropin-Releasing Hormone, Family of

## Further Reading

- Dubois, E. A., Zandbergen, M. A., Peute, J., and Goos, H. J. (2002). Evolutionary development of three gonadotropin-releasing hormone (GnRH) systems in vertebrates. *Brain Res. Bull.* **57**, 413–418.
- Fernald, R. D., and White, R. B. (1999). Gonadotropin-releasing hormone genes: Phylogeny, structure, and functions. *Front Neuroendocrinol.* **20**, 224–240.
- Jennes, L., Eyigor, O., Janovick, J. A., and Conn, P. M. (1997). Brain gonadotropin releasing hormone receptors: Localization and regulation. *Recent Prog. Horm. Res.* **52**, 475–490.
- McCann, S. M., Karanth, S., Mastronardi, C. A., et al. (2001). Control of gonadotropin secretion by follicle-stimulating hormone-releasing factor, luteinizing hormone-releasing hormone, and leptin. *Arch. Med. Res.* **32**, 476–485.
- Norwitz, E. R., Jeong, K. H., and Chin, W. W. (1999). Molecular mechanisms of gonadotropin-releasing hormone receptor gene regulation. *J. Soc. Gynecol. Invest.* **6**, 169–178.
- Parhar, I. S., and Sakuma, Y. (eds.) (1997). "GnRH Neurons: Gene to Behavior." Brain Shuppan, Tokyo.
- Sealfon, S. C., Weinstein, H., and Millar, R. P. (1997). Molecular mechanisms of ligand interaction with the gonadotropin-releasing hormone receptor. *Endocrinol. Rev.* **18**, 180–205.
- Shacham, S., Harris, D., Ben-Shlomo, H., et al. (2001). Mechanism of GnRH receptor signaling on gonadotropin release and gene expression in pituitary gonadotrophs. *Vitamin Horm.* **63**, 63–90.
- Silver, M. R., Lee, K. J., Takahashi, A., Kawauchi, H., Joss, J., Nozaki, M., and Sower, S. A. (2001). Molecular phylogenetic analysis within the petromyzontiforme lineage using the cDNA for lamprey gonadotropin releasing hormone-III. *Am. Zool.* **41**, 1587.
- Sower, S. A., and Kawauchi, H. (2001). Update: Brain and pituitary hormones of lampreys. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **129**, 291–302.
- Suzuki, K., Gamble, R. L., and Sower, S. A. (2000). Multiple transcripts encoding lamprey gonadotropin-releasing hormone-I precursors. *J. Mol. Endocrinol.* **24**, 365–376.



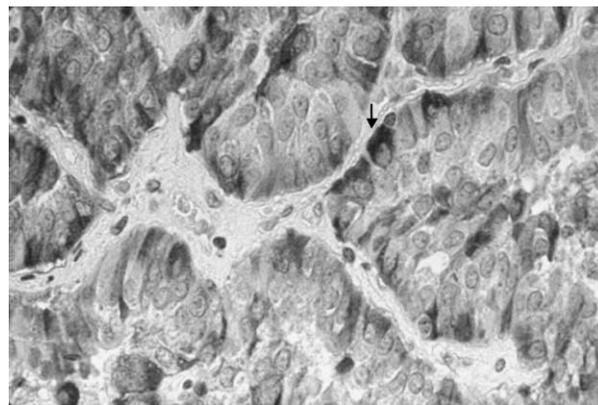
**Figure 1** Gonadotropinoma “iceberg.” (1) Gonadotropinomas in which there is hormonal secretion with clinical effects; (2) gonadotropinomas in which hormonal secretion occurs without clinical effects; (3) gonadotropinomas in which there is no hormonal secretion (silent adenoma) but there is positive immunohistochemistry; and (4) silent adenomas with gonadotropin gene expression on Northern blot analysis. Reprinted from Beckers (2003), with permission.

are possible (inhibin  $A_{\alpha\beta A}$  and inhibin  $B_{\alpha\beta B}$ ). Activin, a member of transforming growth factor- $\beta$  growth hormone superfamily, is a homodimer of two disulfide-linked  $\beta$ -subunits of inhibin. Thus, three possible forms of activin exist:  $A_{\beta A\beta A}$ ,  $B_{\beta B\beta B}$ , and  $AB_{\beta A\beta B}$ . Activin is produced in many tissues, including the gonads and pituitary, where it has autocrine and paracrine effects. It stimulates FSH secretion from normal rat gonadotroph cells and human gonadotropinoma cells. Follistatin is produced in the pituitary by folliculostellate and gonadotroph cells and is a negative modulator of the effects of activin. It is believed that a decrease in follistatin could be responsible for enhanced activin activity, which could account for the increased cell division and FSH secretion characteristic of gonadotroph adenomas. Other studies found a positive correlation between the concentrations of activin, follistatin, and FSH secreted by gonadotroph adenomas *in vitro* and concluded that the production of activin A might explain the relatively high levels of FSH and follistatin, but the amount of follistatin was apparently insufficient to antagonize the stimulatory effects of activin A. Finally, an inhibitory effect of activin on DNA synthesis has been reported in a

subgroup of clinically nonfunctioning pituitary adenomas (NFPA) *in vitro*, suggesting an antiproliferative effect of this locally secreted peptide. This antiproliferative effect involves an up-regulation of p21, a cyclin-dependent kinase inhibitor. It is concluded that defects in activin-mediated signal transduction pathways and dysregulation of modulating peptides may be involved in abnormal activin-mediated growth control in pituitary tumors. It is likely that an alteration in these control mechanisms may play a role in the tumorigenesis of gonadotropinomas but further studies are required.

## PATHOLOGY

As is the case with other pituitary adenomas, gonadotroph adenoma cells are usually arranged in cords and may vary in size from one adenoma to another. Before the advent of immunohistochemical techniques, NFPA were recognized by their chromophobe, agranulocytic appearance after Herlant tetrachromic staining. Electronic microscopy first revealed that these supposedly nonsecretory tumors contained cytoplasmic secretory granules. Later, immunohistochemical studies using monoclonal antibodies directed against the  $\alpha$ -subunit or against the intact LH, FSH, hCG, and their respective  $\beta$ -subunits clearly identified the secretory nature of gonadotropinomas (Fig. 2). Prognostic proliferation markers, such as proliferating cell nuclear antigen (PCNA), p53, and Ki67, have been investigated in gonadotroph adenomas but with inconclusive results. The expression of the polysialylated neural cell adhesion



**Figure 2** Gonadotropinoma immunohistochemistry. Staining was performed using anti-FSH antibodies. Positivity is observed at the periphery of the cell (vascular pole). Reprinted from Beckers (2003), with permission; also reprinted with kind permission from Dr. J. Trouillas, Lyon, France.

molecule, however, seems to be strongly associated with tumor invasion and an aggressive profile.

## CLINICAL ASPECTS

The clinical recognition of gonadotroph adenoma is relatively difficult, as these tumors do not secrete efficiently and their secretory products, intact gonadotropins and/or their subunits, generally do not cause a recognizable clinical syndrome. Consequently, gonadotropinomas can grow undetected until they become sufficiently large to cause symptoms due to tumor expansion, such as visual impairment, headaches, pituitary insufficiency, and, more rarely, seizures, meningitis, and cerebrospinal fluid rhinorrhea.

Oligomenorrhea or amenorrhea can present in premenopausal female patients and is correlated with increased secretion of intact FSH, hyperprolactinemia caused by pituitary stalk compression, and/or secondary hypogonadism attributable to the destruction of normal residual pituitary gland. However, FSH secretion can also cause in premenopausal women a sustained ovarian stimulation, inducing multiple large ovarian cysts, elevated serum estradiol, and infertility.

There have been few reported cases of males with gonadotroph adenomas associated with a characteristic clinical syndrome due to hypersecretion of LH causing testis hyperstimulation (testicular enlargement, elevated serum testosterone). Most gonadotropinomas in men, however, cause secondary hypogonadism with decreased libido, normal to low serum testosterone levels, and normal testicular volume.

## DIAGNOSIS

The overwhelming majority of NFPA are gonadotropinomas but, as noted above, they seldom cause a clinical hormonal hypersecretory syndrome.

Gonadotroph adenomas can be recognized by a supranormal basal serum concentration of intact gonadotropins, often FSH but less commonly intact LH and/or their subunits ( $\alpha$ -and/or  $\beta$ -subunits). In relatively few cases, high levels of hCG have been found. When intact LH is elevated, male patients may also have supranormal serum testosterone concentrations.

When a supranormal serum  $\alpha$ -subunit level exists as the sole abnormality in the presence of a pituitary mass, the diagnosis may be either a thyroid-stimulating hormone-secreting adenoma or a gonadotropinoma. In these cases, thyrotropin-releasing hormone (TRH)-induced stimulation of intact FSH and/or LH or their subunits would confirm a gonadotroph origin.

Basal concentrations of gonadotropins are not helpful in postmenopausal women because it is not possible to differentiate a sellar mass from normal postmenopausal gonadotroph cells as the source of gonadotropins. Only a marked discrepancy among FSH, LH, and/or their subunits would make such measurements useful in the diagnosis of a gonadotropinoma in a postmenopausal female patient.

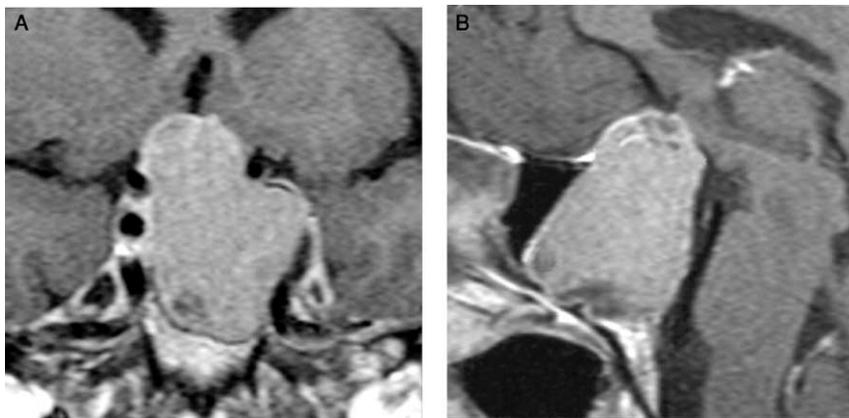
Administration of TRH typically produces an increase in the secretion of FSH, LH, and especially the LH  $\beta$ -subunit and this finding could be of assistance in making a diagnosis in cases where gonadotroph adenomas are associated with basal levels of gonadotropins and/or their subunits in the normal range. However, other studies have suggested that the TRH test is unable to detect silent gonadotropinomas, but is useful in postmenopausal women and also in the confirmation of a diagnosis in patients with secreting gonadotroph adenomas. Usually, gonadotropinomas also remain sensitive to GnRH stimulation. It is difficult to interpret the response to the GnRH in NFPA, as the hormonal response may originate from the gonadotropinoma or from the remaining normal gonadotroph cells of the pituitary.

Pituitary magnetic resonance imaging is by far the best radiological tool to assess the dimensions and the invasiveness of gonadotropinomas (Fig. 3). Almost all reported gonadotropinomas are macroadenomas and nearly one-third of them are considered to be invasive. The incidence of microadenoma is largely unknown but should be suspected when a pituitary microincidentaloma is found.

## TREATMENT

The initial treatment of gonadotroph adenoma is by transsphenoidal surgery, especially if visual disturbances or hemorrhage is present. However, tumor excision is usually incomplete in large macroadenomas and recurrence rates average approximately 30% in NFPA after a mean follow-up of 6 years. Incomplete tumor removal determines a low cure rate and additional treatment may be required. Some centers advocate routine radiotherapy after surgery to reduce the risk of tumor regrowth. When conventional irradiation is administered following surgery, it is usually effective in preventing regrowth of the adenoma. The efficacy of stereotactic radiation remains to be determined.

When the surgery is not curative, several drugs have been utilized in the treatment of gonadotroph adenoma, but thus far none have been found to reduce tumor size consistently and substantially.



**Figure 3** Gonadotroph macroadenoma extending far above the sella, elevating the optic chiasm, and invading the cavernous sinus as observed with magnetic resonance imaging in coronal (A) and sagittal (B) views. Reprinted from Beckers (2003), with permission.

Dopamine receptors have been identified on gonadotroph adenomas and dopamine agonists decrease the levels of the gonadotropins and the  $\alpha$ -subunit *in vitro* and *in vivo*. However, the use of dopamine agonists (e.g., bromocriptine, cabergoline, and quina-golide) has not been found to be useful in the majority of gonadotroph adenomas, although 20% of treated patients did experience a response. In any case, a significant positive correlation has been reported between the uptake score measured using pituitary scintigraphy with  $^{123}\text{I}$ -labeled methoxybenzamide and the percentage inhibition of  $\alpha$ -subunit levels and tumor shrinkage during dopamine agonist treatment.

Somatostatin analogues have been used to treat gonadotroph adenomas, as somatostatin receptors have been identified on gonadotropinomas and somatostatin can decrease the secretion of gonadotropins and/or their subunits by gonadotropinoma cells *in vitro*. Several clinical trials have studied the use of somatostatin analogues in patients with gonadotroph adenomas. Although dramatic decreases in the size of gonadotroph adenomas and some associated improvements in visual field defects have been occasionally reported following somatostatin analogue treatment, the majority of patients experience little if any improvement in adenoma size or visual fields.

GnRH agonists and antagonists have been considered as potential therapeutic agents for gonadotroph adenomas. However, their administration generally does not produce an effect on adenoma size.

Troglitazone, a PPAR (peroxisome proliferator activated receptor)- $\gamma$  ligand, has been studied in mice as a novel medical therapy for NFPA with encouraging results.

## CONCLUSION

In conclusion, the approach to gonadotroph adenomas is guided by the tumor size, the presence of effects of the tumor mass, and hypopituitarism. Surgery remains the first curative choice for gonadotroph macroadenoma. The gonadotroph microadenoma may be observed and treated only if significant growth is documented over the course of a year or more.

Radiotherapy may be useful in patients who are poor surgical candidates, in patients whose tumors are surgically inaccessible, or after surgery to reduce the risk of tumor regrowth.

Several pharmacologic treatments have been tried but only a minority of patients experience significant benefits. Further studies are necessary and novel drugs that can reduce the size of the tumor and/or prevent its regrowth are greatly needed.

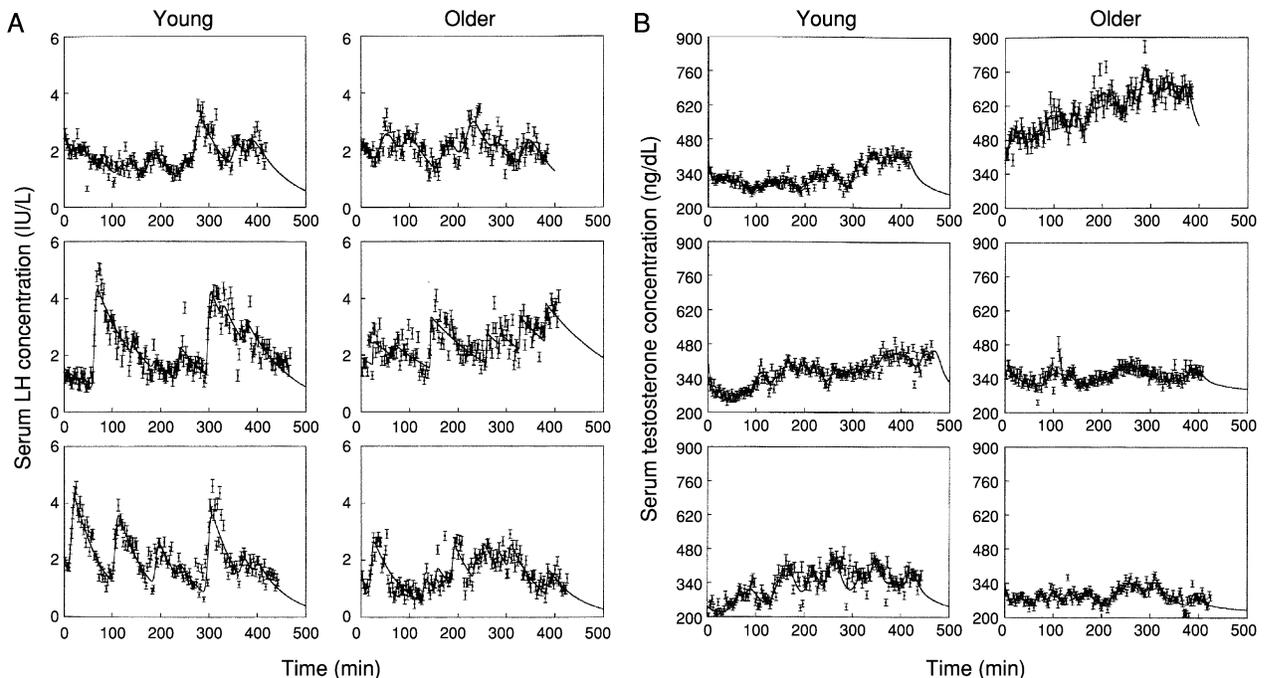
## See Also the Following Articles

Adrenal Tumors, Molecular Pathogenesis • Gonadotropin-Releasing Hormone (GnRH) Actions • Pituitary Adenomas, TSH-Secreting • Pituitary Region, Non-Functioning Tumors of • Pituitary Tumors, ACTH-Secreting • Pituitary Tumors, Clonality • Pituitary Tumors, Molecular Pathogenesis • Pituitary Tumors, Surgery • Testicular Tumors

## Further Reading

Asa, S. L., Bamberger, A. M., Cao, B., Wong, M., Parker, K. L., and Ezzat, S. (1996). The transcription activator steroidogenic factor-1 is preferentially expressed in the human pituitary gonadotroph. *J. Clin. Endocrinol. Metab.* **81**, 2165–2170.

- Beckers, A. (2003). "Pituitary Adenomas," CD-ROM, 3rd ed. Graphmed Ltd., Liège, Belgium.
- Beckers, A., Stevenaert, A., Mashiter, K., and Hennen, G. (1985). Follicle-stimulating hormone-secreting pituitary adenomas. *J. Clin. Endocrinol. Metab.* **61**, 525–528.
- Boggild, M. D., Jenkinson, S., Pistorello, M., Boscaro, M., Scanarini, M., McTernan, P., Perrett, C. W., Thakker, R. V., and Clayton, R. N. (1994). Molecular genetic studies of sporadic pituitary tumors. *J. Clin. Endocrinol. Metab.* **78**, 387–392.
- Chanson, P., De Roux, N., Young, J., Bidart, J. M., Jacquet, P., Misrahi, M., Milgrom, E., and Schaison, G. (1998). Absence of activating mutations in the GnRH receptor gene in human pituitary gonadotroph adenomas. *Eur. J. Endocrinol.* **139**, 157–160.
- Colao, A., Ferone, D., Lastoria, S., Cerbone, G., Di Sarno, A., Di Somma, C., Lucci, R., and Lombardi, G. (2000). Hormone levels and tumour size response to quinagolide and cabergoline in patients with prolactin-secreting and clinically non-functioning pituitary adenomas: Predictive value of pituitary scintigraphy with <sup>125</sup>I-methoxybenzamide. *Clin. Endocrinol.* **52**, 437–445.
- Danila, D. C., Inder, W. J., Zhang, X., Alexander, J. M., Swearingen, B., Hedley-Whyte, E. T., and Klibanski, A. (2000). Activin effects on neoplastic proliferation of human pituitary tumors. *J. Clin. Endocrinol. Metab.* **85**, 1009–1015.
- Heaney, A. P., Fernando, M., and Melmed, S. (2003). PPAR- $\gamma$  receptor ligands: Novel therapy for pituitary adenomas. *J. Clin. Invest.* **111**, 1381–1388.
- Penabad, J. L., Bashey, H. M., Asa, S. L., Haddad, G., Davis, K. D., Herbst, A. B., Gennarelli, T. A., Kaiser, U. B., Chin, W. W., and Snyder, P. J. (1996). Decreased follistatin gene expression in gonadotroph adenomas. *J. Clin. Endocrinol. Metab.* **81**, 3397–3403.
- Pichard, C., Gerber, S., Laloi, M., Kujas, M., Clemenceau, S., Ponvert, D., Bruckert, E., and Turpin, G. (2002). Pituitary carcinoma: Report of an exceptional case and review of the literature. *J. Endocrinol. Invest.* **25**, 65–72.
- Poncin, J., Stevenaert, A., and Beckers, A. (1999). Somatic MEN 1 gene mutation does not contribute significantly to sporadic pituitary tumorigenesis. *Eur. J. Endocrinol.* **140**, 573–576.
- Risbridger, G. P., Schmitt, J. F., and Robertson, D. M. (2001). Activins and inhibins in endocrine and other tumors. *Endocr. Rev.* **22**, 836–858.
- Roncaroli, F., Nosé, V., Scheithauer, B. W., Kovacs, K., Horvath, E., Young, W. F., Lloyd, L. V., Bishop, M. C., Hsi, B., and Fletcher, J. A. (2003). Gonadotrophic pituitary carcinoma: HER/neu expression and gene amplification. *J. Neurosurg.* **99**, 402–408.
- Shimon, I., Rubinek, T., Bar-Hava, I., Nass, D., Hadani, M., Amsterdam, A., and Harel, G. (2001). Ovarian hyperstimulation without elevated serum estradiol associated with pure follicle-stimulating hormone-secreting pituitary adenoma. *J. Clin. Endocrinol. Metab.* **86**, 3635–3640.
- Snyder, P. J. (2002). Gonadotroph adenomas. In "The Pituitary" (S. Melmed, ed.), 2nd ed., pp. 575–591. Blackwell Science, Cambridge, UK.
- Trouillas, J., Daniel, L., Guigard, M. P., Tong, S., Gouvernet, J., Jouanneau, E., Jan, M., Perrin, G., Fischer, G., Tabarin, A., Rougon, G., and Figarella-Branger, D. (2003). Polysialylated neural cell adhesion molecules expressed in human pituitary tumors and related to extrasellar invasion. *J. Neurosurg.* **98**, 1084–1093.
- Valdes-Socin, H., Jaffrain-Rea, M. L., Tamburrano, G., Cavagnini, F., Ciccarelli, E., Colao, A., Delemer, B., Brue, T., Rohmer, V., Wemeau, J. L., Levasseur, S., Teh, B. T., Stevenaert, A., and Beckers, A. (2002). Familial isolated pituitary tumors: Clinical and molecular studies in 80 patients. In "The Endocrine Society's 84th Annual Meeting, San Francisco, California, 19–22 June 2002." Abstract 647.
- Verges, B., Boureille, F., Goudet, P., Murat, A., Beckers, A., Sassolas, G., Cougard, P., Chambe, B., Montvernay, C., and Calender, A. (2002). Pituitary disease in MEN type 1 (MEN1): Data from the France–Belgium MEN1 multicenter study. *J. Clin. Endocrinol. Metab.* **87**, 457–465.
- Wessels, H. T., Hofland, L. J., van der Wal, R., van Gastel, L., van Koetsveld, P. M., de Herder, W. W., and de Jong, F. H. (2001). *In vitro* secretion of FSH by cultured clinically nonfunctioning and gonadotroph pituitary adenomas is directly correlated with locally produced levels of activin A. *Clin. Endocrinol.* **54**, 485–492.



**Figure 1** Serum LH (A) and testosterone (B) concentration profiles in three young and three older men sampled at 2.5-min intervals during sleeping hours. The continuous lines are predicted by deconvolution analysis. Deconvolution revealed loss of high-amplitude LH and testosterone secretory pulses in older individuals with preserved basal secretion rates. Entropy analyses disclosed more disorderly (less regular) release of LH and testosterone in older men. Reproduced with permission from the American Society of Andrology. Veldhuis, J. D. (1999). Recent insights into neuroendocrine mechanisms of aging of the human male hypothalamic-pituitary-gonadal axis. *J. Androl.* **20**, 1–17.

frequency with loss of high-amplitude LH secretory pulses, even though basal secretions are preserved. The dysfunctional GnRH-LH pulsing mechanism results in decreased LH mass per pulse values (Fig. 1). The release of LH showed more disorderliness in older men than in younger men. Moreover, sophisticated mathematical modeling techniques show that there is more disruption of synchrony between LH and FSH, as well as between LH and testosterone secretion, in older men than in their younger counterparts. This multifold synchrony disruption of neuroendocrine control of LH and subsequent testosterone secretion results in uncoupling of testosterone-related nocturnal penile tumescence activity that may be of physiological importance.

The abnormal neuroendocrine regulation of LH pulse secretion is also evident in the aging rat model, where decreased LH pulse amplitude is observed. Hypothalamic GnRH pulse generator dysfunction is suggested by the decreased GnRH response to excitatory amino acid stimulation.

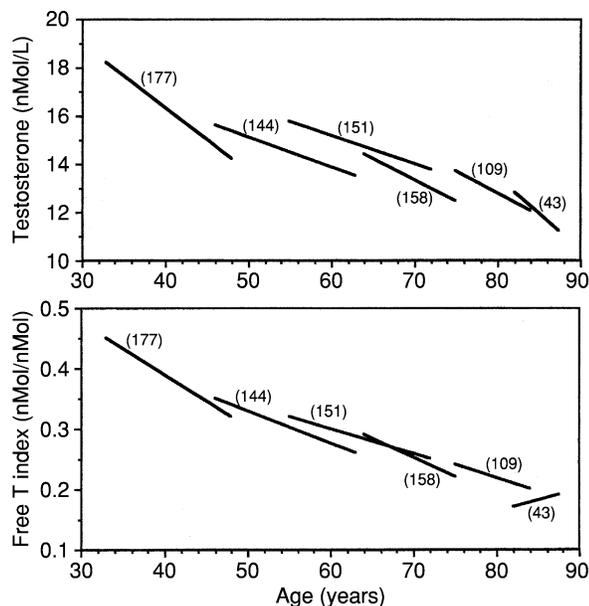
There is evidence to show that inducible nitric oxide synthase (NOS) activity is increased in the aging GnRH secretory neurons in the hypothalamus.

The increase in inducible NOS activity might be a result of increased release of reactive oxygen species, cytokines, or other substances associated with aging. Increased NOS activity results in nitrosylation of cells and increased apoptosis of GnRH-secreting cells in the hypothalamus of aging rats.

From population-based studies, cross-sectional analyses fail to show significant trends of serum prolactin with aging, but the longitudinal data show a sharp increase in morning serum prolactin levels at 5.3% per year. In detailed studies of nocturnal prolactin secretory patterns in elderly men, serum basal concentration and secretory mass per burst of prolactin are decreased at night and serum prolactin is correlated with serum testosterone concentrations. The significance of the changes in prolactin with aging is not known but has been postulated to be reflective of neuroendocrine effects of aging.

## TESTICULAR FUNCTION IN AGING

There is now ample evidence not only from cross-sectional studies but also from longitudinal



**Figure 2** Longitudinal effects of aging on date-adjusted testosterone and free testosterone index. Linear segment plots for total testosterone and free testosterone index versus age are shown for men with testosterone and sex hormone-binding globulin values on at least two visits. Each linear segment has a slope equal to the mean of the individual longitudinal slopes during each decade and is centered on the median age for each cohort of men from the second to the ninth decades. Numbers in parentheses represent the numbers of men in the various cohorts. With the exception of free testosterone index during the ninth decade, segments show significant downward progression at every age, with no significant change in slopes for testosterone or free testosterone index over the entire age range. Reproduced with permission from The Endocrine Society. Harman, S. M., Melter, E. J., Tobin, J. D., Pearson, J., and Blackman, M. R. (2001). *J. Clin. Endocrinol. Metab.* **86**, 724–731.

epidemiological studies to show that aging in men is associated with a slow but progressive decline in serum total testosterone and bioavailable or free testosterone levels (Fig. 2). Serum total testosterone concentrations decline by 1.6% and bioavailable testosterone declines by 2.3% per year, and these decreases are accompanied by significant increases (1.3% per year) in serum sex hormone-binding globulin concentrations. There is also a concomitant decline in serum estrogens and adrenal androgens, androstenedione and dehydroepiandrosterone (DHEA) and DHEA sulfate. In contrast, serum 5 $\alpha$  dihydrotestosterone (DHT) concentrations tend to increase in older men (3.4% per year). Such increases in serum DHT are associated with stable intraprostatic DHT levels. The rise in DHT in the presence of declining testosterone may be due to increased 5 $\alpha$  reductase activity in the liver, skin, or prostate. The decline in serum testosterone is

associated with ill-defined symptoms of andropause, including asthenia, loss of motivation, failure of concentration, decreased libido and sexual activity, loss of muscle and bone mass, loss of strength, increased body fat, depressed mood, and decreased quality-of-life measures. The decrease in serum testosterone concentration is the result of defective Leydig cell steroidogenic capacity coupled with decreased responsiveness of the Leydig cells to endogenous LH. Studies in rodent models of the aging confirmed the decreased steroidogenic capacity of Leydig cells both *in vitro* and *in vivo*. Our laboratory has shown that although Leydig cell numbers are not reduced in aged animals, the Leydig cell volume is decreased. There is also evidence showing that some Leydig cells undergo apoptosis without evidence of renewal.

Despite the decline in steroidogenic capacity, the ability of the seminiferous tubules to produce normal spermatozoa appears to persist in elderly healthy men. There is some evidence that older men might have slightly lower sperm concentration and motility, with little impact on their fertility. Studies in aging men showed that circulating inhibin B (a marker of Sertoli cell function) is maintained. Serum inhibin B is related to serum FSH but not to age or serum testosterone levels in older men. The maintenance of normal serum inhibin B concentrations in older men indicates normal Sertoli cell function and spermatogenic activity.

Our laboratory and others have shown in the brown Norway rat model that aging is associated with decreased testis volume due to decreased seminiferous tubule volume and sperm concentration. The decrease in germ cells is nonuniform and affects some but not all tubules. We have also showed that this decline in germ cells is related to accelerated apoptosis affecting germ cells. Because the loss of germ cells appears to be patchy, it is likely that, in addition to decreased testosterone secretory capacity of Leydig cells, other factors such as increased cytokines, reactive oxygen species, and other paracrine factors induced by decreased blood flow or neuronal input may play important roles.

## IMPLICATIONS IN MANAGEMENT OF AGING MEN

There is general agreement that aging in men is associated with declining serum testosterone and bioavailable/free testosterone concentrations. Although serum testosterone may be decreased, seminiferous tubule

function appears to remain relatively normal, with normal sperm output and reproductive capability even to the seventh decade. Low serum testosterone levels are implicated as a possible cause of the clinical features associated with andropause. Lack of energy, frailty, decreased general well-being, sexual dysfunction, loss of muscle, and loss of bone mass may be ameliorated by testosterone supplementation. Because of the probable abnormal functioning GnRH pulse generator in aging men, stimulation of the axis by clomiphene citrate might not result in favorable increases in LH pulse amplitude and testosterone secretion. Furthermore, because Leydig cell dysfunction appears to be the predominant cause of androgen deficiency in aging males, treatment with recombinant LH or human chorionic gonadotropin might not elicit a satisfactory response, although gonadotropin administration has not been tested in a large number of individuals. Based on our understanding of the gonadotropins and testicular dysfunction in aging men, replacement with androgens at the current time and replacement with selective androgen receptor modulators in the future appear to be the best choices for treatment of symptoms or signs of andropause. Studies are ongoing to determine the benefit versus risk ratios of androgen replacement therapy in aging men.

### See Also the Following Articles

Androgen Biosynthesis and Gene Defects • Gonadotropin-Releasing Hormone (GnRH) Actions • Hypergonadotropic Hypogonadism • Impotence and Aging • Spermatogenesis, Endocrine Control of

### Further Reading

- Feldman, H. A., Longcope, C., Derby, C., Johannes, C. B., Araujo, A. B., Coviello, A., Breckner, W. J., and McKinley, J. B. (2002). Age trends in the level of serum testosterone and other hormones in middle-aged men: Longitudinal results from the Massachusetts Male Aging Study. *J. Clin. Endocrinol. Metab.* **87**, 589–598.
- Harman, M., Metter, E. J., Tobin, J. D., Pearson, J., and Blackman, M. R. (2001). Longitudinal effects of aging on serum total and free testosterone levels in healthy men. *J. Clin. Endocrinol. Metab.* **86**, 724–731.
- Veldhuis, J. D. (1999). Recent insights into neuroendocrine mechanisms of aging of the human male hypothalamic–pituitary–gonadal axis. *J. Androl.* **20**, 1–17.
- Vermuelen, A. (2001). Androgen replacement therapy in the aging male: A critical evaluation. *J. Clin. Endocrinol. Metab.* **80**, 2380–2389.
- Wang, C., Sinha Hikim, A., Ferrini, M., Bonavera, J. J., Vernet, V., Leung, A., Lue, Y. H., Gonzalez-Cadavid, N., and Swerdloff, R. S. (2002). Male reproductive aging using the brown Norway rat as a model for men. In “Endocrine Facets of Aging” (Novartis Foundation Symposium 242, D. J. Chadwick and J. A. Goode, eds.), pp. 82–97. Wiley, Chichester, UK.

deposited in the dermal tissues. These deposits lead to disruption of the collagen bundles and marked edema. Lymphocytic infiltration is not always apparent.

## ETIOLOGY

The reason why localized myxedema occurs in Graves' disease is unknown. The strong association with Graves' hyperthyroidism, and especially ophthalmopathy, suggests that it should have an autoimmune pathogenesis. Although nearly all patients afflicted with myxedema have high titers of thyroid-stimulating hormone (TSH) receptor autoantibodies, there is no clear relation between these titers and the severity of the skin lesions. Nevertheless, these antibodies may very well play a role given that skin fibroblasts express the TSH receptor. However, studies in which serum immunoglobulins (IgGs) from patients with dermopathy were added to cultured skin fibroblasts did not show any effect of the immunoglobulins on the production of glycosaminoglycans by the fibroblasts. Other studies showed that serum from patients with Graves' disease contains a factor—other than IgG—that is able to stimulate cultured skin fibroblasts to produce glycosaminoglycans. This factor may be a cytokine produced by lymphocytes.

The preponderance for the lower legs in Graves' dermopathy is unclear. One reason may be the influence of gravity and stasis, but there is also evidence that fibroblasts from different sites have a different regulation of glycosaminoglycan synthesis.

## TREATMENT

Treatment is often not indicated because the lesions are usually asymptomatic. Only when there is local discomfort or the lesions become unsightly should

treatment be considered. There are no controlled studies on any treatment for Graves' dermopathy, but uncontrolled studies suggest that topical application of corticosteroids is beneficial. The corticosteroids containing cream should be applied directly to the lesions and then covered with a plastic occlusive dressing. After 3 to 10 weeks, the dose and frequency of the application may be gradually tapered. Compressive stockings during the daytime will diminish the fluid accumulation.

## See Also the Following Articles

Graves' Disease • Graves' Disease, Hyperthyroidism in • Graves' Ophthalmopathy • Thyroid Disease, Epidemiology of

## Further Reading

- Daumerie, C., Ludgate, M., Costagliola, S., and Many, M. C. (2002). Evidence for thyrotropin receptor immunoreactivity in pretibial connective tissue from patients with thyroid-associated dermopathy. *Eur. J. Endocrinol.* **146**, 35–38.
- Fatourechi, V. (2000). Localized myxedema and thyroid acropachy. In "Werner and Ingbar's the Thyroid: A Fundamental and Clinical Text" (L. A. Braverman and R. D. Utiger, eds.), 8th ed., pp. 548–555. Lippincott, Williams & Wilkins, Philadelphia.
- Peacey, S. R., Flemming, L., Messenger, A., and Weetman, A. P. (1996). Is Graves' dermopathy a generalized disorder? *Thyroid* **6**, 41–45.
- Schwartz, K. M., Fatourechi, V., Ahmed, D. D. F., and Pond, G. (2002). Dermopathy of Graves' disease (pretibial myxedema): Long-term outcome. *J. Clin. Endocrinol. Metab.* **87**, 438–446
- Smith, T. J., Bahn, R. S., and Gorman, C. A. (1989). Connective tissue, glycosaminoglycans, and diseases of the thyroid. *Endocrinol. Rev.* **1**, 366–391.

trigger in developing Graves' thyroid disease. Various infectious diseases have been considered risk factors. Infections with *Yersinia enterocolitica* have been implicated because this pathogen has thyroid-stimulating hormone (TSH) binding sites that presumably are homologous to the TSH receptor, the prime autoantigen in Graves' hyperthyroidism. The association between the occurrence of Graves' disease and the therapeutic use of interferon- $\alpha$  (IFN- $\alpha$ ) in patients infected with the hepatitis C virus suggests that other conditions resulting in high Interferon alpha levels, such as certain viral infections, may also increase the risk of Graves' disease. Finally, the clear female preponderance suggests that estrogen levels may play an important role. Graves' disease often occurs during the postpartum period.

## PATHOGENESIS

Although the cause of Graves' disease is unknown, autoimmunity directed against the TSH receptor is the hallmark of Graves' thyroid disease. Autoantibodies against this receptor mimic the action of its natural ligand, TSH, inducing hyperthyroidism and goiter. Whether these antibodies also are responsible for the other, extra-thyroidal manifestations of Graves' disease is uncertain, although the TSH receptor has been found in fibroblasts residing in the retrobulbar tissues and in the pretibial skin. The pathogenetic importance of TSH receptor autoantibodies as the cause of Graves' hyperthyroidism is, however, without doubt. Transfer of human immunoglobulin (IgG) samples containing these antibodies to rodents causes a prolonged release of previously radiolabeled thyroid hormone from mouse or guinea pig thyroids. This bioassay was used to make the diagnosis of Graves' disease in the past, and the transferred antibodies were named "long-acting thyroid stimulator" (LATS). Nowadays, different tests are applied. Patient IgGs can be incubated with cells transfected with the human TSH receptor, and the response of these cells in terms of the release of cyclic AMP (cAMP) can be measured in a TSH receptor-stimulating immunoglobulin (TSI) bioassay. Another widely used and less cumbersome assay measures the inhibition of binding of  $^{125}\text{I}$ -labeled TSH to TSH receptor containing preparations by patient IgGs in the so-called TSH-binding inhibitory immunoglobulins (TBII) assay.

In the majority of patients, other antithyroidal autoantibodies are also present. In approximately 70% of patients, antibodies against thyroid peroxidase (TPO) can be detected. TPO antibodies are the

hallmark of Hashimoto's hypothyroidism, and their titers are related to the degree of lymphocytic infiltration of the thyroid gland. The importance of these antibodies in Graves' hyperthyroidism is unclear, but they might be related to the fact that some patients with Graves' hyperthyroidism become hypothyroid in the long run.

## VARIOUS MANIFESTATIONS OF GRAVES' DISEASE

There are three manifestations that may occur in patients with Graves' disease: hyperthyroidism, ophthalmopathy, and pretibial myxedema. Hyperthyroidism is the most common manifestation, and some ophthalmopathy occurs in 25 to 50% of these patients, although severe eye disease is rather rare and develops in only approximately 5% of patients. Pretibial myxedema is even less common and is seen in fewer than 1% of patients. The reason why such diverse organs—thyroid, orbit, skin—can become affected in one patient is unclear, and it has even been reasoned that these three manifestations are, in fact, the expression of distinct disorders frequently occurring more or less simultaneously. This assumption cannot be disproved entirely, but it seems unlikely for a number of reasons.

First, there is a close temporal relationship between the onset of Graves' thyroid disease and the onset of Graves' ophthalmopathy; 80% of patients develop one manifestation within 18 months of the onset of the other. Second, when sensitive methods are used to assess eye involvement in patients with Graves' hyperthyroidism without clinically apparent ophthalmopathy, evidence for orbital tissue swelling is found in virtually all of them. Third, it has been observed that some patients with multinodular goiter develop Graves' hyperthyroidism and ophthalmopathy several months after treatment with radioactive iodine. Multinodular goiter is not an autoimmune thyroid disease and, as such, is not associated with ophthalmopathy. Radioactive iodine leads to a slow destruction of the thyroid, and thyroidal proteins (including the TSH receptor) will leak into the circulation and be presented to the immune system. A minority of patients subsequently mount an autoimmune response to the TSH receptor, leading to hyperthyroidism, and a small number also develop ophthalmopathy. This chain of events suggests that the two disorders are linked to each other. In fact, it supports another theory explaining the multiple-organ character of Graves' disease: autoantibodies against a thyroidal

antigen cross-react with a similar antigen in the skin and the retrobulbar tissues. The nature of this antigen is unknown, but the TSH receptor is a good candidate because TSH receptor autoantibodies are the cause of the hyperthyroidism and the TSH receptor is expressed on retrobulbar fibroblasts and possibly also on fibroblasts derived from skin affected by pretibial myxedema.

### See Also the Following Articles

Goitrogens, Environmental • Graves' Dermopathy • Graves' Disease, Hyperthyroidism in • Graves' Ophthalmopathy •

Hashimoto's Disease • Iodine • Smoking and the Thyroid • Thyroid Autoimmunity • Thyroid Gland, Anatomy and Physiology • TSH Function and Secretion

### Further Reading

- Davies, T. F. (2000). Autoimmune thyroid disease. *Endocrinol. Metab. Clin. North America* **29**, 239–442.
- McKenna, T. J. (2001). Graves' disease. *Lancet* **357**, 1793–1796.
- Prummel, M. F., and Laurberg, P. (2003). Interferon- $\alpha$  and autoimmune thyroid disease. *Thyroid* **13**, 547–551.
- Weetman, A. P., and McGregor, A. M. (1994). Autoimmune thyroid disease: Further developments in our understanding. *Endocrinol. Rev.* **15**, 788–830.

**Table I Common Clinical Manifestations of Thyrotoxicosis**

Symptoms	Nervousness, hyperactivity	
	Fatigue, weakness	
	Increased perspiration, heat intolerance	
	Palpitations	
	Tremor	
	Weight loss despite increased appetite	
	Menstrual disturbances	
	Signs	Hyperactivity
		Tachycardia, atrial fibrillation
		Systolic hypertension
Warm moist skin		
Only in Graves' disease	Tremor	
	Hyperreflexia	
	Stare and eyelid retraction	
	Diffusely enlarged thyroid (goiter)	
	Ophthalmopathy	
	Dermopathy	

concentrations remain unaffected. Hypercalcemia and hypercalciuria are common but usually only moderate, serum magnesium levels tend to be decreased, and phosphate levels tend to be higher. There is no clear evidence for direct toxic effects of elevated thyroid hormones on the liver, although mild elevations of alkaline phosphatase and bilirubin are sometimes seen. An increased frequency of bowel movements is not uncommon and is associated with an increased motility of the intestines. Peristalsis of the esophagus is also increased, and patients occasionally may complain of swallowing difficulties. The red blood cells may be microcytic, but anemia is usually not seen. Mild thrombocytopenia due to shortened platelet survival may occur. Some muscular weakness and wasting occurs in nearly all patients, and this may lead to fatigue and weakness of the larger proximal limb muscles. Nervousness, anxiety, irritability, attention deficits, and some memory loss are common manifestations, but true psychiatric syndromes occur in only a minority of patients. In women, oligomenorrhea is common, but fertility is not severely impaired. In men, there is no evidence for an effect on spermatogenesis, but the biological activity of estrogens is increased and mild gynecomastia can often be found. Hyperthyroidism is a known risk factor for osteoporosis, and the bone mass is frequently decreased in hyperthyroidism. Most of these systemic effects of hyperthyroidism are restored to normal on successful reversal to the euthyroid state.

## DIAGNOSIS

The biochemical diagnosis of thyrotoxicosis is usually straightforward. Laboratory tests will show elevated levels of  $T_4$  and  $T_3$  in the presence of low or undetectable TSH values. In Graves' disease,  $T_3$  levels are relatively higher than  $T_4$  levels, and in some patients only  $T_3$  levels may be increased ( $T_3$  toxicosis). To discriminate Graves' disease from other forms of hyperthyroidism (e.g., toxic multinodular goiter, toxic adenoma) or thyrotoxicosis (e.g., silent or subacute thyroiditis, exogenous thyroid hormone use), additional diagnostic procedures may be useful. A thyroid scintigram will show homogenous and often increased uptake. Determination of TSH receptor autoantibodies may sometimes be useful. An assay measuring the capability of patient immunoglobulins (IgGs) to displace binding of labeled TSH to TSH receptors (TSH binding inhibitory immunoglobulin assay, TBII) has a sensitivity of 96% for the diagnosis of Graves' hyperthyroidism. TBIIs are sometimes present in patients with multinodular goiter and destructive thyroiditis; hence, the specificity is only 84%.

## MANAGEMENT

Initial treatment with a beta blocking agent will alleviate most of the signs and symptoms of hyperthyroidism because many are mediated through the sympathetic nerve system. This symptomatic treatment is usually followed by a therapy aimed at diminishing the synthesis and secretion of thyroid hormones. This can be achieved effectively by anti-thyroid drugs such as methimazole, carbimazole, and propylthiouracil (PTU); by ablating large parts of the thyroid gland with radioactive iodine; or by partly removing the gland surgically. Each of these options has its advantages and disadvantages.

Therapy with 30 mg of methimazole taken once daily, or with 300 to 400 mg of propylthiouracil in three or four divided doses, will render most patients euthyroid in 4 to 6 weeks. When euthyroidism has been achieved, one can choose between diminishing the dose ("titration method") or maintaining the same dose and adding L-thyroxine ("block and replace-ment"). This antithyroid treatment is then continued for 12 to 24 months, after which time approximately 50% of patients will enter into a complete remission. In general, chances to enter into a remission increase if there is a decrease in goiter size, TSH levels become normal during treatment, TSH receptor antibodies disappear, and the patient is a nonsmoker.

Disadvantages of antithyroid drugs include allergic reactions consisting of an often transient pruritic rash, and agranulocytosis, a very rare but life-threatening event. Prompt discontinuation of the antithyroid drugs will restore the leukocyte count to normal within a week. Other rare side effects include arthralgias, hepatitis, and drug fever.

The advantage of subtotal thyroidectomy is the rapid and instantaneous reversal of the thyrotoxic state. It is preferably performed in patients who have been rendered euthyroid by antithyroid drugs, but in emergency situations this can also be achieved by the administration of large doses of iodine (five drops of Lugol's solution daily for a maximum of 10 days) or by treatment with lithium carbonate or amiodarone. Although mortality rates are negligible, a subtotal thyroidectomy carries a small but unavoidable risk for damage to the recurrent laryngeal nerve leading to vocal cord paralysis and for hypoparathyroidism. Postoperative hypothyroidism occurs in a substantial minority of patients.

Radioactive iodine is a simple and economical way in which to control hyperthyroidism. It can be given to patients who have been rendered euthyroid by antithyroid drugs or can be used as a first line of therapy. Its main disadvantage is the induction of hypothyroidism, which occurs in 40 to 70% of patients within 10 years after treatment. Although hypothyroidism is easily treated with L-thyroxine substitution, not all patients like the idea of lifelong substitution treatment. Radioactive iodine may aggravate concomitant ophthalmopathy, especially in smokers, in patients with high pretreatment T<sub>3</sub> levels, and in patients whose eye disease is active. This can be prevented by concomitant administration of corticosteroids. Radiation thyroiditis occurs in a minority of patients during the first weeks after treatment and very rarely may result in a thyrotoxic storm. This is seen mainly in elderly patients with severe

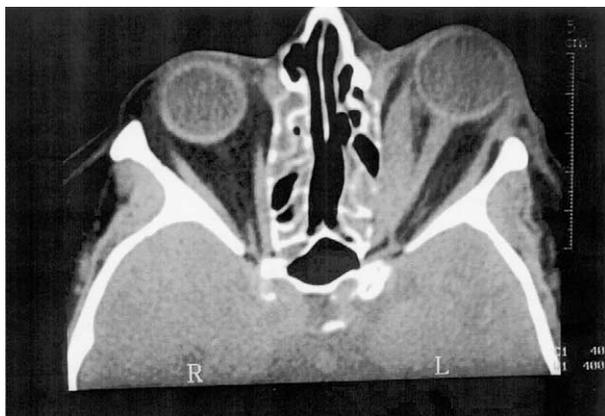
hyperthyroidism; hence, the common practice is to treat these patients with antithyroid drugs first. There is no evidence of a carcinogenic or leukemogenic effect of radioactive iodine; in fact, a recent large follow-up study actually found a 10% decrease in risk for cancer mortality. Radioactive iodine and a 1-year course of antithyroid drugs are the most commonly used therapies in patients with newly diagnosed Graves' hyperthyroidism. Subtotal thyroidectomy is reserved for young patients, especially when a course of antithyroid drugs has failed. In the United States, a preference is seen for the use of high doses of radioactive iodine aimed at induction of hypothyroidism. In Europe, patients are more often treated with a course of antithyroid drugs in light of the 50% remission rate.

### See Also the Following Articles

Graves' Dermopathy • Graves' Disease • Graves' Ophthalmopathy • Hyperthyroidism, Subclinical • Iodine, Radioactive • Thyrotoxicosis: Diagnosis • Thyrotoxicosis, Overview of Causes

### Further Reading

- Franklyn, J. A., Maisonneuve, P., Sheppard, M., Betteridge, J., and Boyle, P. (1999). Cancer incidence and mortality after radioiodine treatment for hyperthyroidism: A population-based cohort study. *Lancet* **353**, 2111–2115.
- Lazarus, J. H. (1997). Hyperthyroidism. *Lancet* **349**, 339–342.
- McDougall, I. R. (1991). Graves' disease: Current concepts. *Med. Clin. North America* **75**, 79–95.
- Pedersen, I. B., Knudsen, N., Perrild, H., Ovesen, L., and Laurberg, P. (2001). TSH receptor antibody measurement for differentiation of hyperthyroidism into Graves' disease and multinodular toxic goitre: A comparison of two competitive binding assays. *Clin. Endocrinol.* **55**, 381–390.
- Solomon, B., Glinioer, D., Lagasse, R., and Wartofsky, L. (1990). Current trends in the management of Graves' disease. *J. Clin. Endocrinol. Metab.* **70**, 1518–1524.



**Figure 1** CT scanning of the orbits of a patient with unilateral Graves' ophthalmopathy. The eye muscles in the right orbit are small and ribbon-like, whereas one eye muscle (the medial rectus) on the left is obviously swollen.

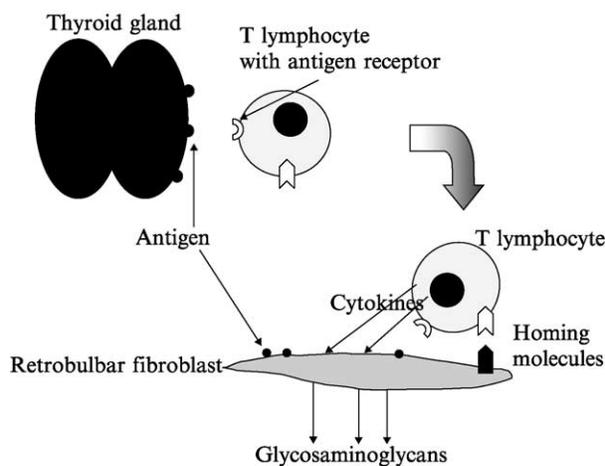
## ETIOLOGY

Like Graves' hyperthyroidism, Graves' ophthalmopathy is an autoimmune disorder with a multifactorial background. The majority of patients with Graves' ophthalmopathy have family members afflicted by Graves' hyperthyroidism, but a family history of ophthalmopathy itself is uncommon. This may be partly due to the relative rareness of this complication. Another plausible explanation is that certain environmental factors are necessary to bring about the eye disease in patients with (a genetic predisposition for) Graves' hyperthyroidism. No genetic loci have been found to be associated with an increased risk for ophthalmopathy, with the possible exception of a CTLA-4 polymorphism. Interestingly, one protective locus has been identified (HLA-DPB1\*201).

On the other hand, an environmental factor, smoking, has been identified beyond any doubt as a strong risk factor for the development of this eye disease. Smoking increases the risk of ophthalmopathy 7.7-fold and is especially associated with more severe eye disease. The pathogenetic importance of smoking is further underscored by the fact that the results of various therapeutic interventions are not as good in smokers as in nonsmokers. The reason behind the influence of smoking on the eye disease is unknown. Another environmental risk factor for ophthalmopathy is the treatment of the underlying hyperthyroidism with radioactive iodine. This treatment carries a small but unequivocally higher risk than does antithyroid drug treatment for the development of (usually mild and transient) ophthalmopathy.

## PATHOGENESIS

On histology, the enlarged retrobulbar tissues show a marked lymphocytic infiltration consisting mainly of T lymphocytes and only a small number of B cells. The autoantigen responsible for attracting these immunocompetent cells is unknown, unlike the situation in Graves' hyperthyroidism where the thyroid-stimulating hormone (TSH) receptor is the autoantigen. The current hypothesis suggests that autoreactive T lymphocytes recognize an antigen that is present in both the thyroid and the retrobulbar fibroblasts (Fig. 2). These T cells then secrete various cytokines, facilitating an amplification of the immune response by inducing adhesion molecules such as ICAM-1 on retrobulbar fibroblasts, which direct other immune cells to the target tissues. T-cell-derived cytokines may also activate B lymphocytes to differentiate into plasma cells capable of producing autoantibodies against unknown autoantigens. In addition, several cytokines possess the ability to stimulate glycosaminoglycan (GAG) production by fibroblasts and fibroblast proliferation. This chain of events leads to swelling and active inflammation of the retrobulbar tissues. As for the nature of the responsible autoantigen, one might suggest the TSH receptor given that this protein is also expressed on retrobulbar fibroblasts at certain stages of their differentiation.



**Figure 2** Schematic representation of the pathogenesis of Graves' ophthalmopathy. T lymphocytes recognize and process an antigen on the thyroid gland and mount an antithyroidal autoimmune attack, resulting in hyperthyroidism. These T cells circulate throughout the body and may become attracted to the orbit, where retrobulbar fibroblasts express homing molecules. There, the T cells recognize an antigen similar (or even identical) to the thyroidal antigen and start an autoimmune attack against the fibroblasts. The cells will produce several cytokines, which stimulate the production of glycosaminoglycans by the retrobulbar fibroblasts.

## RELATIONSHIP WITH GRAVES' HYPERTHYROIDISM

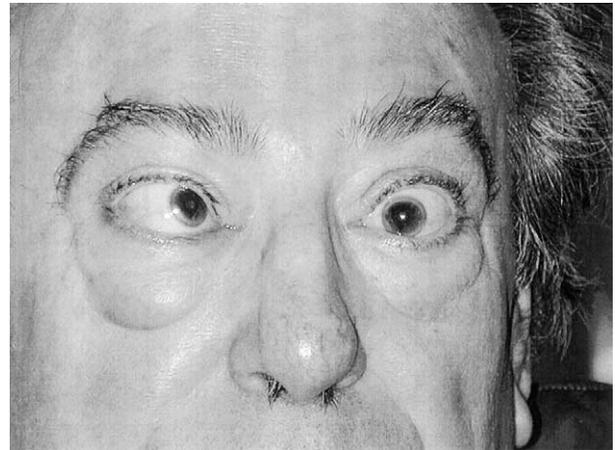
Graves' ophthalmopathy is usually seen in patients in whom a diagnosis of Graves' hyperthyroidism is well established. In the majority of patients, the eye disease develops concomitantly with or some months after the onset of the thyroid disease. However, in 25% of patients, the eye disease precedes a diagnosis of hyperthyroidism by a number of months. Sometimes, there is considerable delay between the onset of the two manifestations that in rare cases may even be several years. In these circumstances, it can be difficult to make the correct diagnosis. In a small minority (5%), patients with ophthalmopathy present with primary hypothyroidism instead of hyperthyroidism. The pathogenesis of this thyroid failure is unclear but may be due to so-called TSH-blocking autoantibodies against the TSH receptor. These antibodies render the thyroid gland devoid of the stimulating effect of TSH, resulting in atrophy and hypothyroidism.

## CLINICAL MANIFESTATIONS

Graves' ophthalmopathy causes a plethora of signs and symptoms that are summarized in [Table I](#), the so-called NO SPECS classification. This classification system was first developed in 1969, was subsequently revised on a number of occasions, and now serves as a good memory aid. The various signs and symptoms might seem complex ([Fig. 3](#)), but all can be explained from a mechanical point of view.

### Class 1: Only Signs, No Symptoms

This refers to the upper eyelid retraction frequently observed in patients with Graves' hyperthyroidism.



**Figure 3** A patient with severe Graves' ophthalmopathy. Note the abnormal eyeball position, resulting in double vision, swelling of the eyelids, increased lid aperture, and periorbital swelling.

The lid retraction causes stare and lid lag on downward gaze (Von Graefe's sign) and can be due to swelling of the superior levator muscle. However, thyrotoxicosis per se can also induce this sign by increasing the sympathetic tone, so Von Graefe's sign is sometimes also present in hyperthyroidism not caused by Graves' disease. Sympathetic overactivity is not the only cause of lid retraction given that upper eyelid retraction frequently remains present when ophthalmopathy patients are rendered euthyroid. If severe, lid retraction may lead to lagophthalmos, that is, incomplete closure of the eyelids at night.

### Class 2: Soft Tissue Involvement

This entails chemosis (edema of the conjunctiva), conjunctival injection and redness, swelling of the

**Table I** NO SPECS Classification of Eye Changes in Graves' Ophthalmopathy

Class	Description	Signs and symptoms	Cause
0	No signs or symptoms	—	—
1	Only signs, no symptoms	Lid retraction, stare, lid lag	Increased sympathetic tone
2	Soft tissue involvement	Swelling of eyelids, chemosis, photophobia, grittiness	Impaired venous drainage, herniated orbital fat
3	Proptosis	Exophthalmos	Increased retrobulbar pressure pushing globe forward
4	Extraocular muscle involvement	Restricted eyeball motility (often with diplopia)	Swollen eye muscles
5	Corneal involvement	Keratitis, corneal ulcer	Overexposure of cornea
6	Sight loss	Decreased visual acuity due to optic nerve involvement, impaired color vision, visual field defects	Pressure on optic nerve, apical crowding

caruncle, and swelling of the upper and lower eyelids (periorbital swelling). These findings are partly explained by impaired venous drainage as a result of the increase in volume of the retrobulbar tissues. Periorbital swelling is also due to herniation of retrobulbar fatty tissues through openings in the orbital septum covering the retrobulbar cavity.

### Class 3: Proptosis

Because of the confining bony surroundings, the swollen retrobulbar tissues have no other outlet than pushing the globe forward. Hence, exophthalmos may be seen as “nature’s own decompression.”

### Class 4: Extraocular Muscle Involvement

One can easily imagine that swelling of the normally very thin extraocular eye muscles leads to impaired mobility. If the impairment is asymmetrical, the patient will have double vision. However, if the impairment is symmetrical, no diplopia will occur. Sometimes, the patient will keep the neck in a certain position (usually bent backward) to correct for impaired motility. This so-called ocular torticollis may lead to painful neck muscles.

### Class 5: Corneal Involvement

Exophthalmos, lid retraction, lagophthalmos, and less frequent blinking all contribute to an excessive exposure of the cornea to air that can lead to inflammation of the cornea (keratitis). Early signs are photophobia, a gritty sensation, intolerance to contact lenses, and blurred vision. This phenomenon is different from diplopia in that the abnormal images disappear after blinking.

### Class 6: Sight Loss

Sight loss can occur if the enlarged eye muscles compress the optic nerve. This can occur in the apex of the orbital cavity where the optic nerve leaves the orbit. On CT scanning, no room is seen between the optic nerve and the swollen muscles; this is called “apical crowding.” Early signs of optic nerve involvement are impaired color vision and visual field defects. This severe complication is more often seen in males and in patients without significant proptosis. In those patients, a tight orbital septum precludes forward displacement of the globe, causing a rise in retrobulbar pressure that is damaging to the optic nerve.

## Other Signs and Symptoms

Apart from the manifestations described in the NO SPECS classification, patients in the active stages of the eye disease often complain of a dull pain or pressure on or behind the eyeball. Pain can also be felt during attempted up, side, or down gaze. Perhaps the most important complaint is the change in appearance. Bulging of the eyes, swelling of the eyelids or even only lid retraction, and an abnormal position of the globes all contribute to a sometimes very marked change in appearance. This will be better appreciated if the patient shows photographs of himself or herself taken before onset of the disease.

## QUALITY OF LIFE

In view of these many and different clinical manifestations, it is not surprising that patients suffer from a diminished quality of life. The changes in appearance as a consequence of proptosis and periorbital swelling can be profound. Diplopia will hamper many activities in daily life such as reading and driving. In fact, research has established that even patients with mild to moderately severe eye disease already have a markedly decreased sense of well-being. They rate their degree of social and role functioning lower than do patients with other chronic diseases such as diabetes mellitus. Use of a disease-specific quality of life questionnaire developed with the aid of patient self-support groups has shown unequivocally that Graves' ophthalmopathy is a seriously disabling disorder. The disease leads to feelings of social isolation in as many as 40% of these patients. Half of the patients notice unpleasant reactions from others, and many do not want to appear in photographs. As a consequence, as many as 70% of patients with mild to moderately severe ophthalmopathy report a marked decrease in self-confidence.

## DIAGNOSIS

In patients with clear evidence of Graves' hyperthyroidism who present with bilateral proptosis, a diagnosis of Graves' ophthalmopathy is easily made. However, this diagnosis may be less obvious in patients with unilateral disease or in patients without a thyroid disorder. Nevertheless, even in these patients, the most likely diagnosis will be Graves' eye disease. It should be noted that there is no diagnostic procedure establishing a diagnosis of thyroid-associated eye disease unequivocally, but several imaging procedures may be helpful. CT scanning of the orbits typically

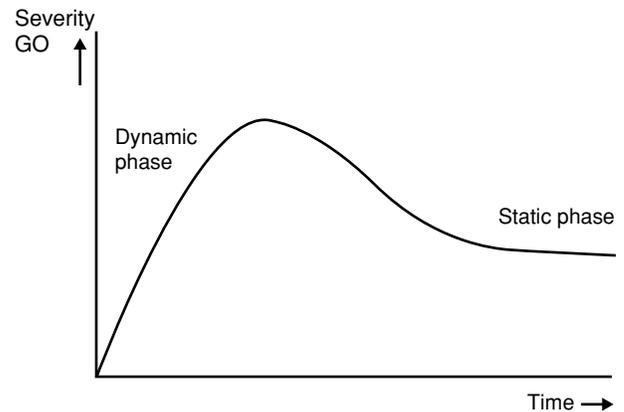
will reveal swelling of the extraocular eye muscles, of the retrobulbar tissues, or of both (Fig. 1). Magnetic resonance imaging (MRI) will show the same but may also demonstrate a prolonged T2 relaxation time, suggestive of edematous swelling. Other diagnostic imaging procedures include octreotide scintigraphy, where labeled octreotide binds to somatostatin receptors on activated orbital lymphocytes, and ultrasound. However, enlargement of the eye muscles or connective tissues is not definitive proof of the existence of ophthalmopathy. Other diagnoses that should be entertained are lymphomas or metastases of carcinomas to the orbit and the rare disease entity of an orbital pseudotumor. It is the combination of various signs and symptoms of ophthalmopathy, together with evidence of an autoimmune thyroid disease, that will trigger the physician to make the correct diagnosis.

## GENERAL THERAPEUTIC PRINCIPLES

The first step in the management of a patient with ophthalmopathy is adequate treatment of the underlying thyroid disorder. Restoration of the euthyroid state frequently leads to some amelioration of soft tissue involvement and sometimes even of diplopia in 2 to 3 months. Generous application of lubricants (eye drops should be applied at least six times daily) prevents corneal damage. Lagophthalmos is frequent, and application of a protective gel at night protects the exposed cornea during sleep. Further therapeutic measures depend on the severity and stage of the eye disease.

## DISEASE ACTIVITY

From observations by the Australian physician, F. F. Rundle, it has become clear that the disease has a tendency toward spontaneous improvement (Fig. 4). After reaching a peak in severity, the signs and symptoms gradually ameliorate over a highly variable period of time, from several months to a number of years. However, in most patients, a complete restoration to the premorbid state is seldom reached. On histology, the active state is characterized by lymphocytic infiltration, whereas during the chronic phase, fibrotic scar tissue is found. It is conceivable that immunosuppressive therapies will be effective during the active stage only and that rehabilitative surgery is the treatment of choice when the disease is in its inactive phase. Unfortunately, reliably assessing the stage of the disease in individual patients has been difficult. Obviously, when the eye condition is rapidly



**Figure 4** Rundle's curve showing the natural history of Graves' ophthalmopathy and its tendency toward spontaneous improvement over a variable period of time. Note that the  $x$  axis (time) is not defined.

deteriorating, it is in its active stage. On the other hand, when the ophthalmopathy has been stable for 6 months, it is likely to have become inactive. However, in many patients with moderately severe disease, such changes are less obvious and one prefers to assess the disease stage without having to observe symptomatic patients during several months without therapy. Therefore, various methods to assess disease activity have been developed, and the ones used most frequently are mentioned in Table II. However, none of these measures is entirely satisfactory in predicting a successful outcome of immunosuppressive therapy. If one or more of the measures from Table II indicates activity, it is reasonable to assume that the disease is still amenable to immunosuppressive therapy (if the severity of the disease warrants treatment).

## MANAGEMENT DURING THE ACTIVE STAGE

One option during the initial phase of the disease is to observe the patient over the course of several months. When the disease is not severe, this is certainly a good option in view of the natural tendency toward spontaneous improvement and because of the side effects associated with the various treatment options. Adequate antithyroid treatment and lubrication of the cornea will alleviate some of the symptoms until the disease has become inactive and surgical rehabilitation is possible. In more severe cases, nonsurgical (immunosuppressive) treatment can be indicated. Both corticosteroids and retrobulbar irradiation have been used for this indication for many decades. The

**Table II Commonly Used Parameters for Disease Activity and Their Cutoff Values Indicating Active Disease**

Parameter	Cutoff value
Duration of signs or symptoms	<18 months
Clinical activity score (for each sign present, 1 point is given):	
<i>Pain or oppressed feeling on globe</i>	≥4
<i>Pain on attempted up, side, and/or down gaze</i>	
<i>Redness of eyelids</i>	
<i>Redness of conjunctiva</i>	
<i>Swelling of eyelids</i>	
<i>Swelling of caruncle</i>	
<i>Chemosis</i>	

rationale for the use of corticosteroids is their anti-inflammatory and immunosuppressive effects. Radiotherapy probably exerts its effect through the killing of infiltrating lymphocytes and activated (GAG-secreting) fibroblasts. Both therapies have an overall response rate of approximately 65% and are effective at reducing periorbital swelling and the degree of diplopia. They have very little, and clinically insignificant, effects on the degree of proptosis. A typical oral prednisone course starts with 60 mg for 2 weeks, followed by 40 mg for 2 weeks, 30 mg for 4 weeks, and 20 mg for another 4 weeks, after which the dose is tapered to zero in 1 to 2 months. Intravenous administration of large doses (500–1000 mg) of methylprednisolone for 1 to 3 days, repeated several times, is also effective and may be better tolerated. Radiotherapy is administered in 10 daily fractions of 200 cGy. Improvement of eye signs occurs earlier when corticosteroids are used (usually some effect is seen within 2–4 weeks), whereas a beneficial effect of irradiation may take 3 to 6 months. The usual prednisone dose is almost always associated with side effects, whereas radiotherapy causes no significant untoward effects and is safe in the long term. The only contraindication is diabetes mellitus because irradiation can aggravate diabetic retinopathy.

In patients with optic nerve involvement, characterized by apical crowding on CT scanning, visual field defects, and/or a decrease in visual acuity, a more aggressive approach is warranted. In those patients, methylprednisolone pulses can be administered (1000 mg intravenously on 3 consecutive days, repeated once 1 week later, followed by oral prednisone for 3–5 months), and this may even be combined

with orbital irradiation. In case of an unsuccessful outcome, acute surgical decompression of the orbits is the only option left.

Intravenously administered immunoglobulins have also been found to be beneficial, with a response rate similar to that of oral corticosteroids. Because of its high cost, the treatment is seldom used. Octreotide and other (long-acting) somatostatin analogues may cause some regression of eye signs. They are well tolerated but seem to be less effective than prednisone. Other immunosuppressive treatments used in autoimmune diseases, such as ciamexon and azathioprine, are not effective in Graves' ophthalmopathy. Cyclosporine given as monotherapy is ineffective but can be combined with oral prednisone, in which case a lower dose of prednisone can be used to achieve a similar response rate.

## SURGICAL TREATMENT

Once the disease has become inactive, most patients still experience disabling and often disfiguring symptoms, even after otherwise successful immunosuppressive therapy. These remaining manifestations can be treated successfully with various targeted surgical procedures by an experienced orbital surgeon. The surgical procedures should be done in a strict order. First, an orbital decompression is performed if there is significant proptosis. During this procedure, several of the orbital walls are partly removed to increase the retrobulbar space. The degree of proptosis reduction varies with the number of walls removed. A three-wall decompression (coronal, transantral, and swinging eyelid approaches) typically results in a 6- to 8-mm reduction in proptosis. Complications include numbness of parts of the skin and worsening or even induction of diplopia. This latter complication occurs in 10 to 20% of patients and is the reason why a decompression should always be done before squint (strabismus) surgery.

Squint surgery consists of a recession of restricted extraocular eye muscles. The tendon of the muscle is severed from the eyeball and reinserted in a new position determined preoperatively during extensive orthoptic evaluations. Binocular single vision in all directions of gaze is seldom achieved, but single vision in the primary and reading positions is reached in 50 to 80% of patients after one operation. Thus, frequently two or even more procedures are required to obtain an acceptable situation.

The final step in the rehabilitation of these patients consists of eyelid surgery. Upper eyelid retraction can

be treated by recessing the levator muscle with or without a sclera interpositioning. Upper eyelid surgery is difficult, and in most patients various procedures (under local anesthesia) are needed. Lower eyelid surgery is often successful after just one attempt. Finally, redundant connective and fatty tissues can be removed by a blepharoplasty of the upper and/or lower eyelids.

### See Also the Following Articles

Eye Disease in Diabetes • Graves' Dermopathy • Graves' Disease • Graves' Disease, Hyperthyroidism in

### Further Reading

- Bahn, R. S. (2000) Understanding the immunology of Graves' ophthalmopathy: Is it an autoimmune disease? *Endocrinol. Metab. Clinics North America* **29**, 287–296.
- Bahn, R. S. (2001). "Thyroid Eye Disease." Kluwer Academic, Norwell, MA.
- Prummel, M. F. (2000). "Recent Developments in Graves' Ophthalmopathy." Kluwer Academic, Norwell, MA.
- Wiersinga, W. M., and Prummel, M. F. (2000). An evidence-based approach to the treatment of Graves' ophthalmopathy. *Endocrinol. Metab. Clinics North America* **29**, 297–319.
- Wiersinga, W. M., and Prummel, M. F. (2002). Graves' ophthalmopathy: A rational approach to treatment. *Trends Endocrinol. Metab.* **13**, 280–287.

prepubertal period, but there is a marked decrease in growth velocity during the years preceding puberty, and puberty is also delayed. For other conditions, such as Crohn's disease, the disease occurs during the preadolescent period, the pubertal changes stop, and the pubertal growth spurt is delayed. Besides the delayed onset of puberty, other abnormalities of puberty have been reported such as a slow clinical progression of puberty and delayed menarche in females as well as a decrease in the duration and intensity of the pubertal growth spurt, resulting in an irreversible loss of height during puberty.

Therefore, abnormal growth during puberty is the major mechanism of short stature in chronically diseased patients.

### Chronic Renal Failure

When renal disease and its consequences occur early during infancy, a marked height deficit may develop due mainly to inadequate nutrition. Additional factors are water and electrolyte losses, metabolic acidosis, renal osteodystrophy, and catabolic response to infection.

During the mid-childhood period, the growth pattern of a child with congenital chronic renal failure (CRF) follows the percentile achieved at the end of infancy. In children who develop CRF after 2 years of life, growth usually follows the percentile achieved after stabilization of the disease. The degree of renal dysfunction is the principal determinant of growth during this period. Relative height tends to decrease in patients with a glomerular filtration rate (GFR) of less than 25 ml/min/1.73 m<sup>2</sup>; growth is usually stable when the GFR is above this threshold.

The onset of puberty is usually delayed in adolescents with CRF, with an average delay of approximately 2 years for the appearance of clinical signs of puberty. The degree of delay of the pubertal growth spurt (mean delay 2.5 years) correlates with the duration of CRF. The mean pubertal height gain is only 50% of that of normal late-maturing children. Hence, it appears that height potential in children with CRF is lost mainly during two developmental stages of rapid growth: infancy and puberty. These abnormal growth patterns have severe consequences for final height.

According to the Registry of the European Dialysis and Transplant Association, 50% of patients who started dialysis before 15 years of age have a final height below the 3rd percentile of the normal population. Children who had continued dialysis until

adulthood reached a slightly lower mean final height than did those with a renal transplant. Less than 25% of patients with preterminal CRF reached a final height above the 50th percentile, indicating that even moderate CRF alters the genetic growth potential. Despite general improvements in renal replacement therapy, the mean final height in adolescents with CRF did not increase during the past decade.

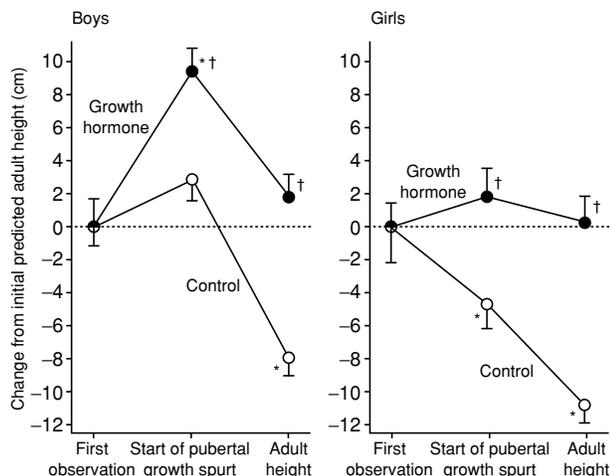
Modern treatment of children with CRF and short stature nowadays includes growth hormone (GH) administration. This has changed the height prognosis of these children, who may reach a normal adult height provided that the GH treatment is administered early in the course of their disease (Fig. 1).

### Juvenile Idiopathic Arthritis

Growth retardation is a prominent feature of children with idiopathic arthritis (JIA). It is present in patients given glucocorticoids, but short stature was described long before the era of modern treatment. In 1887, Sir George Still wrote the following about JIA:

*A remarkable feature in these cases is the general arrest of development that occurs when the disease begins before the second dentition. A child of 12 and a half years would easily have been mistaken for 6 or 7 years, which another of 4 years looked more like 2 and a half or 3 years.*

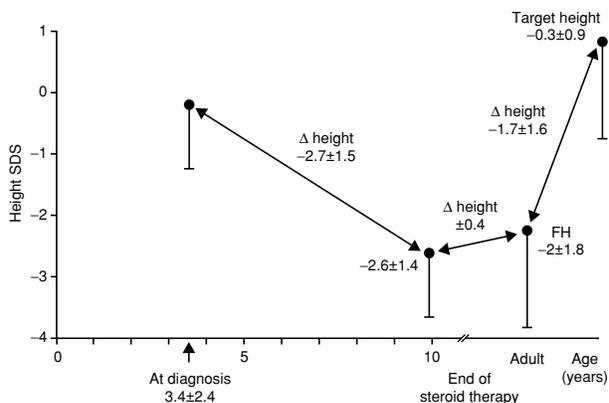
Although growth retardation was described a long time ago, few precise studies are available on cohorts of patients treated with corticoids. We have studied linear growth and final height in a large group of patients who were followed until final height (Fig. 2). Mean age at onset of the disease was  $3.4 \pm 2.4$  years, and prednisone therapy was initiated at disease onset for a mean duration of  $13.6 \pm 5.0$  years. A significant loss of height of  $-2.7 \pm 1.5$  SDS (standard deviation score) occurred during the first 4 years of the disease and was correlated with prednisone therapy duration. Consequently, mean height for chronological age (approximately 12 years) at prednisone discontinuation was significantly lower than mean target height ( $-2.6 \pm 1.4$  vs  $-0.3 \pm 0.9$  SDS). After prednisone discontinuation, 17 patients (70%) had catch-up growth, but 7 patients (30%) had a persistent loss of height. Consequently, mean final height was significantly different between these two groups ( $-3.6 \pm 1.2$  SDS in the group without catch-up growth vs  $-1.5 \pm 1.6$  SDS in the group with catch-up growth). Overall, mean final height was  $-2.0 \pm 1.8$  SDS and was strongly correlated with mean height at prednisone discontinuation.



**Figure 1** Change from initially predicted adult height at baseline in 38 children (32 boys and 6 girls) with chronic renal failure who received growth hormone treatment, as compared with 50 control children with chronic renal failure who did not receive growth hormone, according to gender. Values are means  $\pm$  standard deviations. Asterisks indicate significant differences from the previous period ( $P < 0.001$ ); daggers indicate significant differences from the children who were not treated with growth hormone ( $P < 0.001$ ). Reproduced from Haffner, D., Schaefer, F., Nissel, R., Wühl, E., Tönshoff, B., and Mehls, O., for the German Study Group for Growth Hormone Treatment in Chronic Renal Failure. (2000). Effect of growth hormone treatment on the adult height of children with chronic renal failure. *N. Engl. J. Med.* **343**, 923–930, © 2000 Massachusetts Medical Society. All rights reserved.

Altogether, two periods had a critical influence on final height in these patients: the active phase of the disease characterized by a significant loss of height and the period following prednisone discontinuation, during which catch-up growth occurred in most, but not all, of the patients. Earlier studies reported that linear growth retardation during the active phase of JIA was dependent mainly on disease severity and duration but was worsened by high-dose glucocorticoid therapy. In fact, it has been established that glucocorticoid therapy slows growth when the dose is at least 0.25 mg/kg/day prednisone equivalent. Moreover, many patients with JIA, particularly younger ones, experience partial catch-up growth when their disease becomes inactive. Catch-up growth is less common in patients with longer disease duration.

Few studies have reported final height in patients with a history of JIA. In a study of 65 patients with various forms of JIA, Zak and co-workers found that final height SDS was less than  $-2$  SDS in 11% of patients, all of whom had had polyarticular disease treated with glucocorticoid therapy, and that polyarticular or systemic disease was associated with a final height shorter than the target height. In our study,



**Figure 2** Systemic form of juvenile chronic arthritis: growth and final height. Reproduced from Simon, D., Lucidarme, N., Prieur, A. M., Ruiz, J-C., and Czernichow, P. (2002). Treatment of growth failure in juvenile chronic arthritis. *Horm. Res.* **58**(Suppl. 1), 28–32, by permission of Karger, Basel.

41% of patients suffering from systemic forms of JIA had a final height less than  $-2$  SDS and that 87% of patients had a final height below their target height. Altogether, these data suggest that polyarticular or systemic JIA treated with glucocorticoids is associated with an increased risk of reduced final height. The final height of these patients may vary from one patient to another, depending mainly on severity of the disease and on their linear growth during and after prednisone therapy.

## Crohn's Disease

Approximately 30 to 40% of patients under 18 years of age with Crohn's disease (CD) suffer from impaired linear growth, lack of weight gain, retarded bone development, and delayed onset of sexual maturation. It is not uncommon to find that growth failure precedes any clinical evidence of bowel disease, often by years. Decreased growth rate has been recorded in 40% of patients before symptoms were noted and in 60% of patients at the time of diagnosis. At least one-third of these patients had decreased height velocity before any weight loss. Thus, both abnormal linear growth and abnormal ponderal growth are significant presenting features of inflammatory bowel disease (IBD) in children and may be the earliest indicators of the disease.

A prospective 3-year study, carried out during the early 1990s on pediatric CD patients, found a linear growth delay in 40% of the patients; their actual height was less than 95% of the expected median for their age, and 10% of them had severe growth

impairment below the 5th percentile ( $Z$  score  $< -1.64$ ). A similar percentage (37%) had bone age that was more than 2 SDS below normal and retarded sexual maturation. During the follow-up, 24% had retarded linear growth velocity.

A more recent multicenter prospective 2-year study (1996–1999) of 82 pediatric CD patients indicated that severe growth failure (height  $\leq -2$  SDS) was initially present in 12 (15%) of the cases, with significant retardation of bone age and sexual maturation. Of these, 11 remained less than  $-2$  SDS after 2 years of follow-up. In addition, 6 (8%) of the patients who were initially within the normal range showed impaired growth velocity during the follow-up, leading to a total of 17 (21%) of the patients less than  $-2$  SDS at 2 years follow-up. No differences in the sex, age at diagnosis, age at inclusion, and parental height for this group were found.

During the 1970s, a retrospective study analyzed the long-term growth of patients from adolescence into adulthood and found that a population suffering from IBD during infancy was significantly shorter than the normal population at diagnosis (12.2 years) and after reaching adulthood (follow-up at 8 years). Similar data were reported during the 1990s, with persistent stunting into adulthood of 15 and 31%, respectively.

Modern treatments, including nutrition, have improved the height prognosis of children with CD. More studies are needed to identify those children who remain at risk for short stature.

### Celiac Disease

Childhood celiac disease (CCD) is caused by an intolerance to gluten, leading to damage of the small intestinal mucosa and resulting in malabsorption of nutrients. Usually, CCD is clinically characterized by abdominal complaints and weight loss, but failure of statural growth may also be a prominent symptom, especially when diagnosed beyond a toddler's age. The hormonal status of these patients resembles that observed in malnutrition, with elevated GH secretion and low insulin-like growth factor-1 (IGF-1). Withdrawal of gluten from the diet constitutes removal of the pathogenetic factor and is an effective treatment for CCD. After institution of a gluten-free diet, the symptoms usually disappear quite rapidly, and most patients exhibit complete catch-up. IGF-1 will return to normal values very rapidly during the course of the treatment. If the compliance to treatment is correct and nutrition is adequate, the final height of these patients should be normal.

### Thalassemia

Growth retardation is still present in thalassemia despite efforts to prevent iron overload. In 2000, an Italian series showed that short stature (height  $< 3$ rd percentile) was present in 34% of the patients.

The mechanism of growth retardation is complex and has been attributed mainly to iron overload in the tissues and, more specifically, the endocrine glands. A direct role of chronic anemia, folic acid deficiency, has also been suggested. Failure of pubertal development has been observed in 43% of patients over 16 years of age. In some patients, a direct effect of a gonadal lesion was the main cause, whereas in other cases, pituitary insufficiency or a combination of both primary and secondary hypogonadism was demonstrated.

### Cystic Fibrosis

Growth retardation is present in cystic fibrosis (CF) and for a long time was thought to be due to malnutrition. In a large survey of approximately 5000 patients from Boston and Toronto, the population of CF patients from Toronto was taller than the population from Boston. The only difference was a better nutritional status in the Canadian cohort, a result suggesting that nutritional status was a key component of growth in this population. In fact, it has been demonstrated that by improving nutritional support, the difference in height between these U.S. and Canadian populations disappeared.

Malnutrition is frequent at diagnosis, and early nutritional intervention can improve growth velocity, height, and weight. However, nutrition is not the only factor affecting growth in CF. For example, in Denmark, despite improvement in nutritional status of CF patients, growth velocity and height did not change in the patients from 1960 to 1990 and the only parameter that improved was the body mass index (BMI). It is evident that there are other causes in growth retardation besides malnutrition per se. Chronic inflammation is probably another key factor that influences growth in these patients, as it is in many other chronic diseases. Whether systematic antibiotic courses and anti-inflammatory drugs, in addition to nutritional support, will improve final height remains to be determined.

Insufficient insulin secretion and diabetes is another cause of short stature in these patients. Abnormal growth velocity or lower BMI has been shown before the onset of diabetes. Because the anabolic action of insulin is essential in normal growth, especially

during puberty, it has been suggested that insulin insufficiency is responsible for poor growth in these patients. Controlled trials with insulin injection during the early prediabetic stage are necessary to demonstrate the effect of insulin supplementation on growth and nutrition.

## Diabetes

In the past, abnormal growth and a decrease in final stature were common in patients with uncontrolled diabetes and were components of Mauriac syndrome. A study of identical twins, discordant for diabetes prior to puberty, found that the diabetic twin was on average 5.8 cm shorter than the control twin.

Improvements in the management of diabetes have eliminated severe growth alterations. However, growth impairment has been reported in children with poorly controlled disease or improved growth after a period of better metabolic control. Longer disease duration and poorer metabolic control were associated with a shorter final stature. To a large extent, the decreased final height is related to a decrease in the pubertal growth spurt, particularly in girls. It should be kept in mind that hypothyroidism and celiac disease should be looked for in diabetic children with slow statural growth.

## Growth in Other Glucocorticoid-Treated Diseases

Information on several chronic diseases treated by glucocorticoids is not available. This is the case for Blackfan Diamond disease, dermatomyositis, Lupus, and the like. In children with asthma, inhaled corticosteroids might have a short-term effect on growth velocity. However, long-term follow-up seems to indicate that adult height is normal.

## MECHANISMS OF ABNORMAL GROWTH VELOCITY AND ADULT SHORT STATURE

### Primary Organ Lesion

Abnormal organ function is one of the major causes of growth retardation. This is the case for chronic anemia observed in children suffering from congenital anomalies of red cells, for enteric losses and malabsorption in IBD, and for reduced lung capacity and function in CF. In renal disease, growth retardation is mediated

by predominant tubular abnormalities, resulting in excessive loss of growth-promoting substances. Urinary losses of water and electrolytes are present in many congenital renal diseases. Urinary losses of amino acids, bicarbonate potassium, phosphate, or calcium are observed in complex tubular disorders such as Fanconi syndrome and cystinosis.

## Malnutrition

Energy malnutrition contributes to growth failure in children with chronic disease. Dietary intake may be inadequate because of chronic anorexia, which is often observed in these patients. Protein-calorie malnutrition is observed in protein-losing enteropathy and is present in 70% of child and adult patients with CD. Bleeding and extensive damage to intestinal cells due to chronic inflammation and ulceration tend to increase nitrogen losses and result in hypoalbuminemia. This is accompanied by malabsorption, notably of fat, due to the damage to the ileum, with bile salt malabsorption in approximately 30% of children. It is associated with the malabsorption of fat-soluble vitamins (A, D, and E). The plasma concentrations of minerals (Ca and Mg) and trace elements (notably zinc) are also low (20–30%), as are those of folic acid and vitamin B12, a consequence of steatorrhea, malabsorption, and/or protein-losing enteropathy. However, the malnutrition should be taken in a broad sense and include, for example, metabolic acidosis (observed in patients with CRF) and chronic anemia in thalassemia.

## Chronic Inflammation

Several early clinical observations suggest a direct link between factor(s) produced during chronic inflammation and growth failure. In systemic JIA, for example, impairment of linear growth is seen during periods of disease activity, with subsequent normalization of growth rate during remission. In addition, in patients with systemic JIA and growth retardation treated with GH, growth velocity during treatment appears to be inversely correlated with the intensity of inflammation. In CD, the growth defect seems to be related to the degree of disease activity rather than to steroid treatment. In CF, growth retardation appears to be related to the severity and number of episodes of pulmonary infection and not to the degree of pancreatic failure.

During the past few years, the progress in our understanding of inflammation, and especially of the

role of cytokines, has opened up new fields of investigation. Three cytokines—interleukin-1 (IL-1), tumor necrosis factor- $\alpha$  (TNF- $\alpha$ ), and IL-6—possess a particularly intense proinflammatory activity and have, for this reason, been investigated extensively within various inflammatory diseases, both acute and chronic. In CD, for example, large amount of cytokines, notably TNF- $\alpha$  and IL-1, are produced. Experimental work was carried out by de Benedetti on a murine transgenic model in which an overexpression of the IL-6 gene could be observed. As a consequence, these animals had high circulating levels of IL-6 and, interestingly, an important growth defect. The reduction in the growth rate led to an adult size that was about 30 to 50% smaller than that of nontransgenic litter mates. Interestingly, the growth defect was completely abolished by immunoneutralization of IL-6, demonstrating the direct role of IL-6 overexpression in the growth defect. The exact mechanism remains to be determined, although a role of IGF-1-binding protein 3 has been shown in transgenic animals that could result in decreased half-life and clearance of IGF-1.

Altogether, chronic inflammation plays an important role, probably variable from one disease to another (e.g., prominent in JIA and secondary in CF), but this role is now well recognized and will introduce interesting therapeutic options in the future.

## Glucocorticoids

Glucocorticoids have a general antianabolic and catabolic influence on bone and cartilage. Therefore, they have a direct action on the growth plate, suppressing multiple-gene expression. Glucocorticoids interfere with chondrocyte proliferation and matrix proteoglycans synthesis, and they increase the apoptotic rate of hypertrophic cells. In the tissues adjacent to the growth plate, glucocorticoids enhance osteoclastic activity and suppress osteoblast development and function. Thus, glucocorticoid therapy started during childhood induces severe bone loss, which may increase the risk of osteoporosis and fracture during early adulthood.

In addition, glucocorticoids interfere with the GH-IGF-1 axis and so influence growth. However, in chronic disease, several factors influence the GH-IGF-1 axis; therefore, it is difficult to dissociate the responsibility of glucocorticoid administration from other causes, such as malnutrition and chronic inflammation, that also have distinct action on growth factor synthesis or secretion.

## The GH-IGF-1 Axis

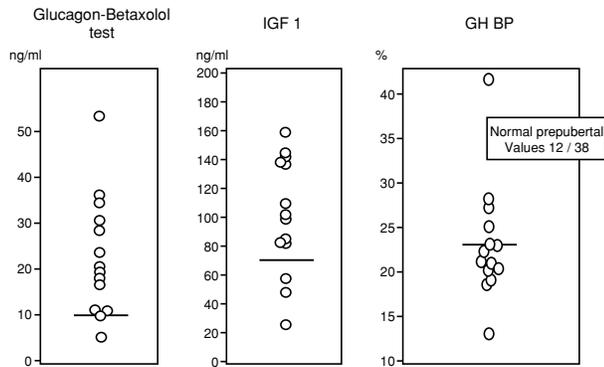
Several animal studies and *in vitro* experiments showed that chronic disease and glucocorticoid administration may influence the GH-IGF-1 axis. However, there are no major anomalies in GH secretion in patients with chronic diseases. For instance, in JIA, pharmacological stimulation of GH and sleep-related GH secretion are normal (Fig. 3), as are IGF-1 plasma concentrations and circulating IGF-binding protein. However, there is a wide heterogeneity among the patients, with some having clearly abnormal values. A similar observation was made with patients suffering from CRF.

Although glucocorticoids do not consistently reduce circulating immunoreactive IGF-1, they do inhibit IGF bioactivity in children with chronic diseases such as in nephrotic syndrome. There is some experimental evidence that glucocorticoids suppress GH receptor expression. Moreover, in JIA, circulating GH-binding protein has been found to be subnormal (Fig. 3). A similar observation has been made in patients suffering from CRF. Moreover, GH administration increases IGF-1 plasma values in both diseases. Therefore, even though glucocorticoids may induce a certain degree of GH resistance through their action on the GH receptor, GH administration at therapeutic dose can overcome this resistance. Altogether, changes in GH, GH-binding protein, IGF-1, and IGF-binding protein are minor. However, recent studies have shown that the local effect of the disease and, more important, the local effect of glucocorticoids on the growth plate seem to play a major role. The general message is that glucocorticoids lower local production of IGF-1 via IGF-1 transcription inhibitors. Unfortunately, it is impossible to study the mechanisms involved in humans, and most of the available information is from animal models.

## TREATMENT

The dramatic improvement in the metabolic conditions of several diseases has played a key role in the growth and development of chronically ill children, particularly those with CRF, CD, diabetes, and CF. However, better nutrition has not been sufficient to normalize the height of children suffering from CRF and JIA.

The important role of glucocorticoids in growth limitation and bone formation has been discussed previously. Therefore, careful administration of glucocorticoids in chronic disease is of utmost importance



**Figure 3** Plasma growth hormone (GH) after pharmacological stimulation, insulin-like growth factor-1 (IGF-1), and GH-binding protein (GH BP) in 14 children with juvenile idiopathic arthritis. Although mean plasma values are normal, there are large interindividual variations and some patients have extremely low values. Reproduced from Touati, G., Prieur, A. M., Ruiz, J. C., *et al.* (1998). Beneficial effects of one-year growth hormone administration to children with juvenile chronic arthritis on chronic steroid therapy: I. Effects on growth velocity and body composition. *J. Clin. Endocrinol. Metab.* **83**, 403–409.

in the prevention of short stature. Alternate-day steroid therapy has also been efficient in the attenuation of corticoid-induced impairment of linear growth.

The role of GH as a growth-promoting agent in several chronic diseases has been studied extensively. Significant research has been done in CRF, and this indication for GH is now recognized in a number of countries. Not only does long-term GH treatment of these children induce persistent catch-up growth, but if the treatment duration is sufficiently long, it is now recognized that patients will achieve a normal adult height (Fig. 1).

Concomitant administration of glucocorticoids in nephrotic syndrome, for example, is certainly a limitation to the anabolic action of GH when the prednisone dose is greater than 0.3 mg/kg/day.

The use of GH in several other conditions is under investigation. In JIA, for example, short-term GH treatment can improve growth velocity and body composition. Long-term studies are needed to confirm these results. The available data regarding GH therapy in chronic illness are encouraging and should be extended to other conditions where chronic disease and glucocorticoids have a serious impact on growth.

### See Also the Following Articles

Constitutional Delay of Growth and Puberty (CDGP) • Cytokine Actions, Cellular Mechanism of • Glucocorticoids,

Overview • Growth and Glucocorticoids • Growth Hormone (GH) • Growth, Normal Patterns and Constitutional Delay • Insulin-like Growth Factors • Puberty: Physical Activity and Growth • Thalassemia, Endocrine Sequelae

### Further Reading

- Agertoft, L., and Pedersen, S. (2000). Effect of long-term treatment with inhaled budesonide on adult height in children with asthma. *N. Engl. J. Med.* **343**, 1064–1069.
- Allen, D. B. (1996). Growth suppression by glucocorticoid therapy. *Endocrinol. Metab. Clinics North America* **25**, 699–717.
- Ansell, B. M., and Bywaters, E. G. L. (1956). Growth in Still's disease. *Ann. Rheum. Dis.* **15**, 295–319.
- Berchovitch, M., Bistrizter, T., Milone, S. D., *et al.* (2000). Iron deposition in anterior pituitary in homozygous beta-thalassemia: MRI evaluation and correlation with gonadal function. *J. Clin. Pediatr. Metab.* **13**, 179.
- Castille, R. G., Telander, R. L., Cooney, D. R., *et al.* (1970). Crohn's disease in children: Assessment of the progression of disease. *Pediatrics* **45**, 386–393.
- De Benedetti, F., Alonzi, T., Moretta, A., *et al.* (1997). Interleukin-6 causes growth impairment in transgenic mice through a decrease in insulin-like growth factor-I: A model for stunted growth in children with chronic inflammation. *J. Clin. Invest.* **99**, 643–650.
- De Benedetti, F., Pignatti, P., Vivarelli, M., *et al.* (2001). *In vivo* neutralization of human IL-6 (hIL-6) achieved by immunization with a hIL-6 receptor antagonist. *J. Immunol.* **166**, 4334–4340.
- Falcini, F., Tacetti, G., Trapani, S., Tafi, L., and Volpi, M. (1991). Growth retardation in juvenile chronic arthritis patients treated with steroids. *Clin. Exp. Rheumatol.* **9**(Suppl. 6), 37–40.
- Giustina, A., Angeli, A., and Canalis, E. (eds.) (2002). "Glucocorticoid-Induced Osteoporosis." Karger, Basel, Switzerland.
- Laursen, E. M., Koch, C., Petersen, J. H., and Muller, J. (1999). Secular changes in anthropometric data in cystic fibrosis patients. *Acta Paediatr.* **88**, 169–174.
- Polito, C., Strano, C. G., Olivieri, A. N., *et al.* (1997). Growth retardation in non-steroid treated juvenile rheumatoid arthritis. *Scand. J. Rheumatol.* **26**, 99–103.
- Ripa, P., Robertson, I., Cowley, D., Harris, M., Brent Masters, I., and Cotterill, A. M. (2002). The relationship between insulin secretion, the insulin-like growth factor axis, and growth in children with cystic fibrosis. *Clin. Endocrinol.* **56**, 383–389.
- Rolon, M. A., Benali, K., Munck, A., *et al.* (2001). Cystic fibrosis-related diabetes mellitus: Clinical impact of prediabetes and effects of insulin therapy. *Acta Paediatr.* **90**, 860–867.
- Schaefer, F., Seidel, C., Binding, A., Gasser, T., Largo, R. H., and Schärer, K. (1990). Pubertal growth in chronic renal failure. *Pediatr. Res.* **28**, 5–10.
- Simon, D., Fernando, C., Czernichow, P., and Prieur, A.-M. (2002). Linear growth and final height in patients with systemic juvenile idiopathic arthritis treated with long-term glucocorticoids. *J. Rheumatol.* **29**, 1296–1300.
- Sizonenko, P. C. (2002). Effects of inhaled or nasal glucocorticoids on adrenal function and growth. *J. Pediatr. Endocrinol. Metab.* **15**, 5–26.
- Zak, M., Muller, J., and Pedersen, F. K. (1999). Final height, arm-span, subischial leg length, and body proportions in juvenile chronic arthritis: A long-term follow-up study. *Horm. Res.* **52**, 80–86.

Table I Commonly Used Glucocorticoid Preparations

Oral	Equivalent adult dose (mg)	Potency		Relative affinity for glucocorticoid receptor	Half-life (h)
		Anti-inflammatory	Sodium retaining		
Cortisone	25.0	0.8	0.8	0.01	8–12
Cortisol	20	1	1	1	8–12
Prednisone	7.5	4	0.8	0.05	12–36
Prednisolone	5	4	0.8	2.2	12–36
Methylprednisolone	4	5	0	11.9	12–36
Triamcinolone	4	5	0	1.9	12–36
Betamethasone	0.75	30	0	5.4	36–54
Dexamethasone	0.75	35	0	7.1	36–54

Inhaled/Nasal	Relative vasoconstrictor potency	Half-life (h)	Licensed dose limit ( $\mu\text{g}/\text{d}$ )	
			Inhaled	Nasal
Fluticasone propionate	1.2	14.4	200	200
Budesonide	1.0	2.3	800	—
Flunisolide	0.3	2.0	—	100
Beclomethasone dipropionate	0.6	15.0	400	—
Triamcinolone acetonide	0.4	3.1	—	55

## GENERAL CONSIDERATIONS

### Variability in Glucocorticoid Sensitivity

Variability in the sensitivity to glucocorticoids across individuals is illustrated physiologically by the wide range in basal early-morning plasma cortisol concentrations in the normal population and pharmacologically by the response of patients treated with glucocorticoids regarding both efficacy and side effects.

### General Influence of Glucocorticoids on the Growth Rate, Bone Maturation, and Final Height

Excess endogenous cortisol and exogenous synthetic glucocorticoids over several weeks or more will lead to impairment of linear growth along with weight gain. The greatest detriment is likely to occur when growth is normally most rapid during the infancy and pubertal phases. The growth impairment is associated with a variable delay in skeletal maturation, and final height attained is influenced by the rate of growth relative to the rate of skeletal maturation. A transient period of impaired linear growth and comparable retardation in bone age favors good “catch-up” growth and attainment of normal predicted adult height. However, a blunted pubertal growth spurt and

permanent stunting may occur with long-term treatment (6 months or longer).

Growth impairment in glucocorticoid-treated children and the potential for catch-up growth are influenced by the duration of treatment, dose, administration regimen, and potency of the glucocorticoid preparation (Table II). In addition, the effects of glucocorticoids may be difficult to distinguish from the natural history of the underlying disease and other disease factors that might adversely affect growth, especially considering that those most severely affected are more likely to receive glucocorticoids.

Table II Approaches to Minimize Adverse Effects of Nonphysiological Doses of Systemic Glucocorticoid Treatment

- Use only for well-established indications.
- Use the lowest possible dose that will control the primary symptoms.
- Avoid prolonged treatment (>3 months).
- Give alternate-day regimen if prolonged systemic treatment is required (growth suppression is likely with cortisone  $\geq 30\text{ mg}/\text{m}^2$  daily, prednisone  $\geq 4\text{ mg}/\text{m}^2$  daily, betamethasone  $\geq 0.6\text{ mg}/\text{m}^2$  daily).
- Use concomitantly with glucocorticoid-sparing agents (e.g., cyclosporin-A for nephrotic syndrome).
- Avoid during infancy and/or puberty.

### Concomitant Effects on the Hypothalamic–Pituitary–Adrenal Axis, Body Composition, and Bone Metabolism

In addition to inhibiting linear growth, supraphysiological doses of oral glucocorticoids and high doses of locally administered glucocorticoids can suppress hypothalamic–pituitary–adrenal (HPA) function and may reduce bone mass. It is not clear exactly how these major side effects relate to each other in terms of dose–response and whether different glucocorticoid preparations have the same hierarchy of systemic effects.

## GLUCOCORTICOID THERAPY IN SPECIFIC DISEASES

### Glucocorticoid Replacement for Abnormal Hypothalamic–Pituitary–Adrenal Axis

#### ***Congenital Adrenal Hyperplasia and Glucocorticoid Substitution Treatment***

Substitution treatment with glucocorticoids is essential in the management of congenital adrenal hyperplasia (CAH) to suppress excessive hypothalamic corticotropin-releasing hormone (CRH) and pituitary adrenocorticotropic hormone (ACTH) secretion, reduce circulating levels of adrenal androgens, and provide physiological replacement. This is achieved with hydrocortisone (10–15 mg/m<sup>2</sup>/day in two or three doses, i.e., above the physiological cortisol production rate). Overtreatment, especially during the first 2 years of life, and undertreatment can result in abnormal growth during childhood and in failure to reach adult height potential. Potential future approaches with a reduced hydrocortisone dose or alternative glucocorticoid preparations along with other treatments, such as an antiandrogen (flutamide) or an aromatase inhibitor (testolactone) and luteinizing hormone-releasing hormone (LHRH) agonist with or without growth hormone (GH), may improve growth and adult height in patients with CAH.

#### ***Primary Adrenal Insufficiency***

The most common cause of primary adrenal insufficiency during childhood is autoimmune disease. Height is below average but not usually less than the 3rd percentile. Growth is normal with adequate replacement, and although the onset of puberty may be delayed, pubertal development is normal.

#### ***Familial Glucocorticoid Deficiency (ACTH Resistance)***

These autosomal-recessive disorders are characterized by a failure of the adrenal cortex to respond to ACTH but normal mineralocorticoid secretion. Tall stature and advanced skeletal maturation have been described despite glucocorticoid deficiency. Physiological replacement with glucocorticoids allows normal growth.

#### ***Hypopituitarism***

Deficiency of GH, ACTH, TSH, and gonadotropins contribute to the growth failure and the delayed or lack of pubertal development seen in hypopituitarism. As with other glucocorticoid-deficient states, the lowest dose of hydrocortisone necessary to prevent the symptoms of adrenal failure should be used to avoid suppression of growth.

### Pharmacotherapy in Diseases That Do Not Primarily Involve the Hypothalamic–Pituitary–Adrenal Axis

#### ***Atopic Diseases***

##### ***Asthma and Inhaled Glucocorticoid Treatment***

Impaired growth and a higher prevalence of short stature in children with atopic disease were reported well before the availability of glucocorticoids (8–25% of children with asthma not treated with glucocorticoids and up to 45% of those treated with daily oral glucocorticoids had a height standard deviation score [SDS] ≤ -2). Independent of glucocorticoid treatment, short stature in children with asthma is associated with delayed bone age, delay in the onset of puberty, and delay in the pubertal growth spurt by approximately 1.3 years. Although the mean adult height attained is normal, up to 20% of those with severe unremitting asthma have final height less than the 10th percentile. The degree of growth impairment is greater in those with more severe asthma and with onset of asthma before 3 years of age.

Inhaled glucocorticoids are the treatment of choice for childhood asthma, and dose–response studies suggest that doses of less than 400 µg/d of any inhaled glucocorticoid are beneficial in controlling symptoms. Higher doses confer little added advantage, but the potential for adverse effects increases dramatically. Lower leg length assessed by knemometry is a highly sensitive measure of the dose-dependent systemic effects of inhaled glucocorticoids over days to weeks (short term) but does not predict longer term growth or final adult height. The potential for linear

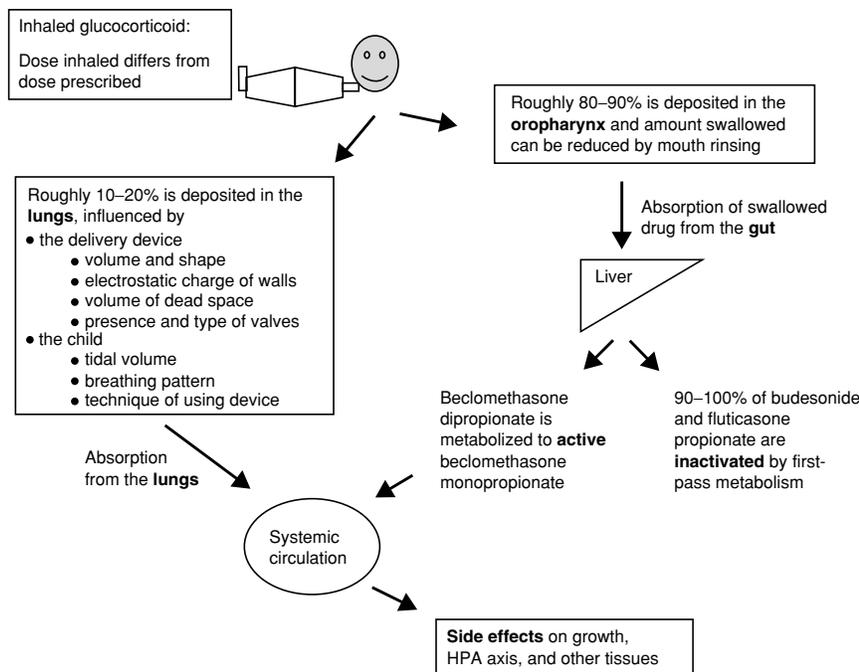


Figure 1 The pharmacokinetics of inhaled glucocorticoids.

growth impairment with inhaled glucocorticoid treatment depends on systemic bioavailability (Fig. 1), which in turn is influenced by the delivery device, inhalation technique, prescribed dose, lipophilicity, pharmacokinetics of the glucocorticoid preparation, and severity of asthma. Randomized controlled trials help to distinguish the effects on linear growth of inhaled glucocorticoid treatment from those of the asthma itself. At doses of at least 400  $\mu\text{g}/\text{day}$ , beclomethasone dipropionate suppresses linear growth by approximately 1.5 cm/year but does not necessarily suppress the HPA axis. Although it is advisable to use the minimum dose necessary to control symptoms, undertreatment that compromises asthma control may also impair linear growth. In addition, there is no evidence of impairment in final adult height with prolonged inhaled or intranasal glucocorticoids in recommended doses.

Growth failure associated with increased body fat with centripetal distribution and reduced lean mass indicative of iatrogenic Cushing's syndrome has been reported in patients on high doses of inhaled glucocorticoids. Paradoxically, poor linear growth and poor weight gain associated with endogenous cortisol insufficiency but without these well-known peripheral effects of excess exogenous glucocorticoids have also been described in a small number of children treated with inhaled glucocorticoids.

#### *Chronic Rhinitis and Intranasal Glucocorticoid Treatment*

Intranasal glucocorticoids are increasingly being used as a first-line treatment for conditions associated with chronic mucosal inflammation. Systemic effects, including growth impairment, can occur and depend on the dose and duration of treatment. As with inhaled treatment, growth suppression may occur without adverse effects on the HPA axis. It has been observed for lower leg growth and linear growth, for example, with budesonide nasal spray (400  $\mu\text{g}/\text{day}$ ) and beclomethasone dipropionate (>160  $\mu\text{g}/\text{day}$  for 6 weeks to 12 months) but not with mometasone furoate (100  $\mu\text{g}/\text{day}$ ). Like fluticasone propionate, mometasone furoate has reduced systemic bioavailability when absorbed from the gut, owing to complete first-pass hepatic metabolism. There are no long-term studies of growth in children treated with intranasal glucocorticoids.

#### *Atopic Dermatitis and Topically Applied Glucocorticoid Treatment*

A number of factors influence percutaneous absorption of topically applied glucocorticoids and the potential for systemic effects such as the formulation, method of application, sites treated, and severity of dermatitis. Growth retardation in association with Cushing's syndrome and HPA suppression, with catch-up growth when treatment was stopped or

reduced, has been reported in isolated cases treated with topical glucocorticoids. There are no randomized controlled trials on the systemic effects of topical glucocorticoid treatment in children and, thus, no studies that clearly separate the contribution made by such treatment from other disease factors that might adversely affect growth. Impaired growth in children with atopic dermatitis and a higher prevalence of atopic dermatitis among children with shorter stature than those with normal stature have been reported (up to 10% of children with atopic dermatitis have height <3rd percentile). It is likely that growth in children with atopic dermatitis may follow a pattern consistent with constitutional delay independent of glucocorticoid treatment, as observed in asthma and other chronic diseases. This is supported by observations of deviation of linear growth from normal in prepubertal children with atopic dermatitis, a greater delay in bone age as these children approach the teenage years, and attainment of adult height within the normal range (mean heights of males and females 162 and 175 cm, respectively).

#### **Inflammatory Conditions**

Growth failure, influenced by complex disease-related factors, is common in inflammatory conditions such as Crohn's disease and juvenile chronic arthritis. By controlling the inflammatory process in Crohn's disease, glucocorticoid treatment may facilitate growth, and height velocity has been found to increase during treatment and follow-up (from mean  $-0.8$  to  $1.0$  SDS). However, prolonged systemic glucocorticoid treatment can exacerbate linear growth inhibition, reduce final adult height attained, and alter body composition. For localized disease, topical treatment offers the potential for minimizing adverse effects (e.g., rectal budesonide or hydrocortisone for Crohn's disease, intra-articular triamcinolone hexacetonide for mono or oligoarticular arthritis).

#### **Conditions Requiring Immunosuppression**

Prednisolone is the most widely used treatment for nephrotic syndrome. Children with frequently relapsing or steroid-dependent nephrotic syndrome are most likely to have impaired linear growth. Catch-up growth occurs after stopping prednisolone, and final height attained is satisfactory provided that high-dose prolonged maintenance therapy is avoided.

## **STRATEGIES TO OVERCOME GLUCOCORTICOID-INDUCED GROWTH FAILURE**

Short-term trials of recombinant human growth hormone (hGH,  $0.32$ – $0.47$  mg/kg/week for 1–10 years) in children with nephrotic syndrome and juvenile chronic arthritis reveal catch-up growth in the former (mean height SDS  $-1.4$  before treatment and  $-0.3$  after treatment) and prevention of further decline in height in the latter (mean height SDS  $-4.1$  before treatment, with height velocity increasing from  $1.9$  to  $5.4$  cm/year during treatment). Serum insulin-like growth factor-I (IGF-I) and IGF binding proteins-3 (BP-3) levels increase during hGH treatment but fall to pretreatment values when treatment is stopped. Treatment with hGH improves growth and final height in chronic renal failure as well as posttransplantation (mean adult height SDS  $-1.7$  in males and  $-1.3$  in females vs pre-hGH treatment height SDS of  $-3.2$  in males and  $-2.5$  in females). For other conditions, long-term studies are required to determine whether pubertal maturation is influenced by hGH or interferes with the growth-promoting effect of hGH and final adult height.

## **INTERACTIONS BETWEEN GLUCOCORTICOID AND THE GROWTH HORMONE AXIS**

The mechanisms that underpin glucocorticoid-induced growth suppression are complex, involving interactions at multiple levels with those hormones that mediate growth. The actions of glucocorticoids are delineated in terms of their effects at the hypothalamic–pituitary level as well as their actions in peripheral growing tissues.

### **The Growth Hormone Axis**

#### **Growth Hormone Secretion**

The principal hormonal regulators of childhood growth are GH and thyroxine, with a contribution from sex steroids during puberty. The secretion of GH from pituitary somatotrophs is under complex control, not only from hypothalamic factors (themselves regulated by input from higher cortical centers) but also from negative feedback. GH production and release are regulated positively by GH-releasing hormone from neurons in the arcuate nucleus and negatively by somatostatin from the periventricular

nucleus. A third stimulatory factor has recently been recognized, namely Ghrelin, the endogenous ligand for GH secretagogue receptors, located in the hypothalamus and pituitary. The interplay among these factors leads to the generation of a pulsatile pattern of GH secretion. Approximately half of the GH released into the circulation is bound to a high affinity-binding protein that is homologous to the extracellular domain of the GH receptor. This binding protein appears to be generated by proteolytic cleavage at the cell membrane of the extracellular domain of the GH receptor. The role of GH binding protein (GHBP) to GH action is not clear, but it prolongs the half-life of GH.

### **GH and IGF-I in Peripheral Tissues**

To exert its cellular actions, GH must bind sequentially to two GH receptors to form a trimeric complex. GH receptors are found in high concentration in the liver but are also present in many tissues throughout the body. Of particular relevance to growth are the GH receptors located on growth plate chondrocytes. Activation of GH receptors leads to changes in expression of a range of genes, regulating growth, cell differentiation, and metabolism. One such gene is IGF-I. It is the local production of IGF-I in the growth plate that is most significant to growth. IGF-I binds to its receptor and, like GH, activates signaling pathways that lead to changes in gene expression and ultimately to growth. IGF-I is also present in the circulation, secreted primarily from the liver. Here it associates with a number of specific binding proteins (IGFBPs), most commonly IGFBP-3. It can be transported into tissues, where the local IGFBP milieu is specific to that tissue. Binding proteins can act to enhance or inhibit IGF-I actions.

### **Glucocorticoid Effects on the GH Axis**

The interactions between glucocorticoids and this GH-IGFBP-IGF-I axis are complex. Glucocorticoids can affect all levels of the axis—at the hypothalamus, pituitary, liver, and growth plate—influencing individual components such as the GH receptor, intracellular signaling, the IGF-I gene, the IGF-I receptor, and IGFBP expression. In children treated with potent high doses of glucocorticoids, the net effect is to inhibit growth, generating a state of GH/IGF-I insensitivity. However, not all of these interactions are negative.

### **Positive Effects of Glucocorticoids**

Time of exposure to glucocorticoids is most important in that acute administration can enhance GH secretion. Dexamethasone can also increase liver and growth plate GH receptor expression *in vivo*. Glucocorticoid response elements are found in the promoter region of the GH gene along with thyroid hormone response elements. In fact, glucocorticoids enhance the thyroid hormone effect on GH gene expression. Transgenic mice, overexpressing the GH gene, secrete more GH when treated with glucocorticoids. In addition, the intracellular glucocorticoid-receptor complex interacts with one of a family of signal transducers and activators of transcription (STAT-5) proteins that is activated by GH to enhance the GH effect on STAT-5-dependent gene expression. Therefore, there is an important synergy between glucocorticoids and the GH axis. Glucocorticoids in physiological amounts are important to normal growth, and it is the failure of these positive interactions that may contribute to the growth failure of children with glucocorticoid deficiency.

### **Inhibitory Effects of Glucocorticoids**

However, the dominant clinical effect of high-dose, long-term glucocorticoids is growth suppression. At the hypothalamic level, glucocorticoids increase somatostatin tone. At low concentrations, they enhance GHRH secretion; however, at higher concentrations, they cause inhibition. The overall effect is a reduction in GH release following stimulation with, for instance, GH-releasing hormone or clonidine and attenuation of GH pulse amplitude in physiological GH secretion profiles. At the peripheral level, using rat chondrocytes *in vitro*, dexamethasone reduces GH receptor number and, hence, GH binding. GHBP is also reduced in those chronically treated with glucocorticoids. Growth of rat chondrocytes, both basally and in response to GH or IGF-I, is reduced by dexamethasone. Local IGF-I production in response to GH is lowered by dexamethasone, and a further reduction in the effectiveness of local IGF-I results from a decrease in IGF-I receptor expression.

In the circulation, IGF-I levels do not appear to be consistently altered; however, changes in binding proteins have been identified. In Cushing's syndrome, IGFBP-3 is slightly increased, IGFBP-1 is unchanged, and IGFBP-2 is elevated. In osteoblast cultures, expression of IGFBP-1, -3, -4, and -5 is reduced by glucocorticoid. IGFBP-5 can potentiate IGF-I action in bone; therefore, its reduction may inhibit IGF-I action.

**Table III Possible Effects of Chronic Exposure to High-Dose Glucocorticoids on the GH-IGF-I Axis**

- GH release from the pituitary gland is reduced, primarily by an increase in somatostatin.
- Chondrocyte GH receptor number is decreased, but a possible increase in hepatic GH receptor may occur.
- GH-binding protein, released into the circulation from proteolysis of membrane-bound GH receptors, is reduced.
- Local tissue IGF-I production is reduced, although circulating levels do not consistently change.
- IGF-BPs in the circulation can be altered.
- IGF-I receptor expression in chondrocytes is reduced.

Glucocorticoids, when present in excess, are acting throughout the GH-IGF axis to suppress growth-promoting activity (see summary of effects in [Table III](#)). Other mechanisms may be contributing. For instance, glucocorticoids down-regulate type I collagen gene expression and up-regulate collagenase 3 expression. Thus, the synthesis of important matrix proteins, necessary for normal skeletal growth and mineralization, and their degradation by proteases are facilitated by glucocorticoids. In addition, effects of glucocorticoids on other local growth factors in bone may contribute to growth failure. Nevertheless, despite this multilevel inhibition, exogenous GH has been used successfully to reverse steroid-induced growth failure.

## See Also the Following Articles

Congenital Adrenal Hyperplasia, Prenatal Diagnosis and Therapy • Glucocorticoids, Overview • Growth and Chronic Disease • Growth Hormone (GH) • Growth, Normal Patterns and Constitutional Delay • Hypopituitarism • Skeletal Development During Childhood and Adolescence

## Further Reading

- Canalis, E. (1998). Inhibitory actions of glucocorticoids on skeletal growth: Is local IGF-I to blame? [editorial]. *Endocrinology* **139**, 3041–3042.
- Hargitai, G., Sólyom, J., Battelino, T., Lebl, J., Pribilincová, Z., Hauspie, R., Kovács, J., Waldhauser, F., Frisch, H., and the MEWPE-CAH Study Group. (2001). Growth patterns and final height in congenital adrenal hyperplasia due to classical 21-hydroxylase deficiency. *Horm. Res.* **55**, 161–171.
- Heinrichs, C., Yanovski, J. A., Roth, A. H., Yu, Y. M., Domene, H. M., Yano, K., Cutler, G. B., and Baron, J. (1994). Dexamethasone increases GH receptor mRNA levels in liver and growth plate. *Endocrinology* **135**, 1113–1118.
- Jux, C., Leiber, K., Hugel, U., Blum, W., Ohlsson, C., Klaus, G., and Mehls, O. (1998). Dexamethasone impairs GH-stimulated growth by suppression of local IGF-I production and expression in cultured rat chondrocytes. *Endocrinology* **139**, 3296–3305.
- Loke, K. Y., Yap, H. K., Zhou, X., Tan, S. P., Chao, S. M., and Lee, K. O. (1997). Efficacy and safety of one year of growth hormone therapy in steroid-dependent nephrotic syndrome. *J. Pediatr.* **130**, 793–799.
- Sizonenko, P. C. (2002). Effects of inhaled or nasal glucocorticoids on adrenal function and growth. *J. Pediatr. Endocrinol. Metab.* **15**, 5–26.

receptor gene transcription via protein–DNA and protein–protein interactions.

### TRANSCRIPTIONAL REGULATORY REGIONS OF GROWTH FACTOR RECEPTOR GENES

Transcription initiation is a key regulatory step in the control of gene expression. Formation of a pre-initiation complex on GC-rich TATA-less promoters involves the interplay of many transcription factors and a combinatorial array of *cis*-regulatory DNA elements. A number of growth factor receptor gene promoter regions have been cloned and sequenced to identify potential transcription factor-binding motifs. Most of them lack consensus TATA- and CCAAT-boxes and are GC-rich in their upstream regulatory region. The rat epidermal growth factor receptor (EGFR), rat transforming growth factor- $\beta$  type III (TGF- $\beta$ RIII), and human FGFR3 promoters are 64, 69, and 82% GC-rich, respectively. Structural analyses of the promoters of the human and rat IGF-IR and human TGF- $\beta$ RIII genes revealed a lack of consensus TATA and CCAAT motifs in the proximal promoter regions. Similarly, the human EGFR, hepatocyte growth factor receptor (HGFR), FGFR1, and FGFR3 gene promoters also lack TATA elements. Some growth factor receptors that lack a TATA-box do contain a functional CCAAT-box. For example, the platelet-derived growth factor  $\beta$  receptor (PDGF $\beta$ R) promoter contains a functional CCAAT motif located 60 bp upstream from the start of transcription. Similarly, the TGF- $\beta$ RII promoter contains two CCAAT-boxes at –440 to –444 bp and at –423 to –427 bp.

Many growth factor receptor promoters, whether they contain functional or nonfunctional CCAAT sequences, do contain multiple Sp transcription factor-binding sites in the GC-rich region upstream of their major transcription start sites. Sp transcription factors constitute a family of zinc-finger DNA-binding proteins that regulate a variety of genes involved in cell growth and differentiation. Sp factors bind to the consensus GC-boxes (GGGCGG), which are often located near the transcriptional start site. The human TGF- $\beta$ RII promoter contains two Sp-binding sites located 25 and 143 bp upstream from the start of transcription. The human and rat IGF-IR promoters contain at least five Sp-binding sites within the proximal 450 bp. Similarly, the human FGFR4 and mouse FGFR3 regulatory regions contain at least five

Sp-binding sites, the majority of which are located within the proximal 100 bp.

In addition to the canonical Sp factor-binding site or GC-box, additional Sp factor-binding sequences have been identified in promoters of growth factor receptor genes. Typically, these sequences are composed of a “GT motif” (GGGTTGTGGC) or they have a homopurine/homopyrimidine motif (CCTCCTCCTCCTCGGCCTCCTCCCC) and constitute a “CT element.” These latter motifs have been identified in IGF-I, EGF, and vascular endothelial growth factor (VEGF) receptor promoters. Such CT elements are also located in the chicken FGFR1 promoter, which is structurally and functionally divided into proximal and distal regions. The distal region is located more than 1 kb upstream from the start of the transcription initiation site. The proximal region contains three Sp1 sites and the distal region contains two Sp1 sites. Mutation of the FGFR1 sequence, CTCCCCTC, located 23 bp upstream from the start of transcription reduced FGFR1 promoter activity in proliferating skeletal myoblasts to basal levels. Electromobility shift and Southwestern blot analyses have confirmed that this site binds the Sp1 transcription factor. In addition, another four Sp factor-binding sites with the sequence, CTGCCC, are also required for maximal FGFR1 promoter activity. Mutation of the two proximal CTGCCC sites at –42 and –54 bp reduced promoter activity by 50 and 79%, respectively. Interestingly, deletion or mutation of either of the two CTGCCC sites located more than 1 kb upstream from the proximal sites also fully abrogated promoter activity. These studies indicate that Sp1 is a key regulatory factor of the chicken FGFR1 promoter.

Several other promoters, such as human TGF- $\beta$ RII and rat TGF- $\beta$ RIII, contain CT elements located within 500 bp of the transcription initiation site. Similarly, mouse FGFR2, human IGF-IR, and HGFR promoters also contain these elements within the proximal 25 bp regulatory region. It is likely that the CT elements in these promoters have a function similar to that observed in the chicken FGFR1.

In addition to the multiple Sp sites, activator protein 1 (AP-1) and AP-2 consensus binding sites are also present in the GC-rich region of almost all growth factor receptor promoters, suggesting that these sites are of functional significance. AP-1 and AP-2 factors, composed of *c-fos* and *Jun* subunits, often direct transcription in proliferative cellular response to extracellular signals. For example, the rat TGF- $\beta$ RIII promoter contains two AP-1-binding sites located at –511 to –480 and –464 to –434 bp and these sites are necessary for activation of the promoter. Nonetheless,

overexpression of AP-2 by transfection did not further stimulate TGF- $\beta$ RIII promoter activity, suggesting that the endogenous AP-2 may be sufficient for TGF- $\beta$ RIII expression or that the complex interactions between AP-2 and Sp1 (which is located in close proximity) may govern overall TGF- $\beta$ RIII expression. Interestingly, human TGF- $\beta$ RII, chicken FGFR1, VEGFR2, human IGF-IR, HGFR, and mouse and human FGFR3 and FGFR4 promoters also contain AP-1- and AP-2-binding sites in close proximity to Sp-binding sites but their functional importance is poorly understood. Therefore, it seems reasonable to expect that AP-1 and AP-2 may be activators of other growth factor receptor promoters as well and that AP-1/AP-2-mediated promoter activation acts in concert with Sp factor transcriptional activation.

## TRANSCRIPTIONAL ACTIVATION OF GROWTH FACTOR RECEPTOR GENES

Transcriptional activation is one of the means of regulating gene expression in many cell types and is governed by complexes of transcription factors. The lack of consensus TATA- and CCAAT-boxes in the majority of analyzed growth factor receptor gene promoters suggests that other DNA-binding sites present in the promoter region and their interactions with site-specific transcription factors are required for basal as well as heightened transcriptional activity. For the few growth factor receptor promoters that do contain consensus CCAAT-boxes, most notably PDGFR $\alpha$  and PDGFR $\beta$ , these sites bind positive transcription factors. The PDGFR $\alpha$  CCAAT element binds the CCAAT/enhancer-binding protein  $\delta$  and the TGF- $\beta$ R CCAAT element binds the nuclear factor Y (NFY) transcription factor. Numerous studies have implicated Sp transcription factors in the positive regulation of promoter activities for these growth factor receptor genes. It has been proposed that such multiple Sp-binding sites and associated protein(s) may stabilize the transcriptional machinery and establish a site of transcription initiation in promoters without TATA elements. Sp1 activates the promoters of TGF- $\beta$ RI, TGF- $\beta$ RII, TGF- $\beta$ RIII, FGFR1–4, IGF-IR, EGFR, and HGFR. Most of these promoters contain 1 to 11 Sp sites in their regulatory region.

Positive transcriptional activity is conferred to the FGFR1 promoter during skeletal myoblast proliferation. Occupancy of an increasing number of Sp sites in the FGFR1 promoter additively increases transcriptional activity. However, the order of

Sp-binding sites in relation to the start of transcription does not necessarily correlate with activity in regulating transcription. The chicken FGFR1 promoter contains three Sp sites in the proximal region, between –60 to –23 bp from the start of transcription. Mutation in the most proximal region (–23 Sp site) abolished promoter activity, indicating that this site confers most of the promoter activity. Mutation of the other two sites also reduced promoter activity to a lesser extent. Moreover, two additional Sp sites located 1 kb upstream from these proximal sites are also required for transcriptional activity of the FGFR1 gene promoter in proliferating myoblasts. These results suggest that Sp1–Sp1 protein interactions may induce looping of the intervening promoter sequences to create protein–protein complexes formed by distant *cis*-elements.

Other members of the Sp family, such as Sp2, Sp3, and Sp4, are also involved in the regulation of growth factor receptor genes. Like Sp1, both Sp3 and Sp4 can bind to GC-box or GT motifs, whereas Sp2 has a much weaker binding affinity for GT motifs. Sp2 and Sp4 often function as transcription activators, whereas the function of Sp3 is highly variable. Sp3 is a bifunctional transcription factor that can either activate or repress transcription. It has discrete domains for both activation and repression and therefore either activates or represses gene transcription depending on the cellular context. For instance, Sp3 augmented Sp1-mediated transcription of the FGFR1 gene promoter in chicken myoblasts. However, Sp3 itself was not able to activate the FGFR1 promoter. Therefore, Sp3 may function as a transcriptional coactivator of growth factor receptor promoters, in particular molecular and cellular contexts defined by the presence and activity of other interacting activators.

In many cases, Sp1 interacts with transcriptional factors other than Sp family members and synergistically activates growth factor receptor gene expression. The mouse TGF- $\beta$ RIII promoter has been cloned and its sequence showed putative binding sites for Sp1, Smad3, Smad4, myogenic transcriptional regulators such as MyoD and myocyte enhancer factor 2 (MEF2), and retinoic acid receptor within a 2100 bp segment. All these factors are involved in the up-regulation of TGF- $\beta$ RII gene expression in the mouse myogenic cell line, C<sub>2</sub>C<sub>12</sub>.

Other transcription factors are also involved in the transcriptional activation of growth factor receptor genes and they are mostly found in cancer cells. For instance, the murine TGF- $\beta$ RII gene promoter contains two conserved Ets-binding sites and mE1f-3, a member of the Ets family, plays a key role in the activation of the promoter in mouse embryonal

carcinoma cells. Similarly, AP-1 binds to at least seven sites in the EGFR promoter region and activates expression of the EGFR gene, which may contribute to cancer cell progression. Transcription factor Pax1 also acts as a transcriptional activator of the PDGFR $\alpha$  gene in differentiated Tera-2 human embryonic carcinoma cells. Many growth factor receptor gene promoters have been analyzed for interactions with Sp1 by site-specific binding assays and functional assays of expressed protein. However, the functional importance of transcription factors other than Sp factors binding to these promoters is poorly understood.

### TRANSCRIPTIONAL REPRESSION OF GROWTH FACTOR RECEPTOR GENES

Negative transcriptional regulators also control the expression of growth factor receptor genes for normal cellular function. Identification of negative regulators of growth factor receptor gene promoters has been more elusive than that of positive regulators. Nevertheless, several direct repressors of growth factor receptor promoters have been identified. The transcription factor GC factor (GCF) and its close relative GCF2 have been shown to function as transcriptional repressors of EGFR and IGF-IR promoters. GCF was initially identified as a factor that binds GC-rich elements. Since most of the growth factor receptor promoters are GC-rich, it is possible that GCF-binding sites would exist and repress their gene expression in a variety of cellular contexts.

Most growth factor receptor genes are transcriptionally activated by the Sp1 transcription factor and therefore down-regulation of these genes may occur by direct inhibition of Sp1. In many cancer cells, tumor suppressor genes may act as direct negative regulators by binding to promoter regions of these growth factor receptor genes and reduce transcription, thereby controlling receptor-specific cell signaling and/or cell proliferation. For instance, in prostate cancer cells, Wilm's tumor suppressor 1, a member of the Egr transcription family, directly binds to the IGF-IR promoter region to down-regulate its activity. In addition, Egr-1 itself interacts with an Sp-binding site in the TGF- $\beta$ RII promoter, causing transcriptional repression, possibly by interfering with Sp1 binding to DNA.

Given the number of Sp transcription factor-binding sites typically located in growth factor receptor promoters, whether GC or CT elements, negative as well as positive regulation may be mediated via these sites. Sp3 is a likely candidate as a transcriptional repressor

interacting with the Sp sites. Though Sp1 and Sp3 share very similar structural features, opposing functions were observed in the regulation of growth factor receptor genes. For example, Sp3 was first considered to be a transcriptional repressor as it interfered with Sp1 binding to the promoter region of TGF- $\beta$ R. However, other evidence indicates that Sp3 can function as a coactivator with Sp1 in the regulation of the FGFR1 promoter. This dual regulatory capability of Sp3 is due to the presence of a broad activation domain and a more narrowly defined repressor domain, both of which are likely to be selectively functional based on the molecular and cellular contexts of promoter activity. Careful analysis of protein-DNA interactions of Sp-binding sites in other growth factor receptor promoters may further elucidate the role of Sp family members in such dual regulation of transcription.

Other Sp-like transcription factors that bind Sp-binding sites include basic transcriptional elements (BTEB1, BTEB2, and BTEB3), core promoter-binding protein, Krüppel-like factors, and TGF- $\beta$ -inducible early genes (TIEG1 and TIEG2). These factors compete with Sp factors in binding to the Sp sites of the promoters and either synergistically activate or repress Sp-mediated transcription of many genes involved in cell growth and differentiation. Therefore, further sequence analysis and functional interactions with Sp-like factors involving canonical and noncanonical Sp sites will likely reveal important molecular mechanisms by which growth factor receptor promoters are regulated.

### REGULATORY CONTROL OF GROWTH FACTOR RECEPTOR GENES VIA PROTEIN-PROTEIN INTERACTIONS

In normal and cancer cells, growth factor receptor genes are positively and negatively regulated via protein-protein interactions among transcription factors. Positive transcription of the PDGFR $\beta$  gene occurs via the nuclear factor NFY, which binds to a CCAAT-box located 60 bp upstream of the initiation site. In addition, two Sp1 sites located in close proximity to the CCAAT-box are also essential for synergistic activation of this promoter. Deletion of the two Sp1 sites resulted in 50% reduction of promoter activity and Sp1 overexpression did not increase promoter activity. Furthermore, transfection of NFY failed to enhance transcriptional activity when the Sp1 sites were deleted from the promoter, suggesting an important role for Sp1 sites in NFY-controlled transcription. In contrast, negative regulation of the PDGF $\beta$ R promoter

also occurs via the NFY transcription factor. C-Myc interacts with NFY and interferes with its binding to the promoter. Other evidence also suggests that the transcription factor p73 (a p53 homologue) represses transcription of the PDGF $\beta$ R gene by interaction with NFY. Therefore, combinatorial interactions between positive and negative regulatory factors regulate the expression of the PDGF $\beta$ R gene.

In cancer cells, several transcription factors possessing tumor suppressor activity negatively modulate Sp1-mediated transcription through protein–protein interactions. For instance, BRCA1 (breast cancer 1) repressed the activity of the IGF-IR promoter in a number of breast cancer-derived cell lines. Electromobility shift assay results indicated that BRCA1 did not exhibit any specific DNA binding to the IGF-IR promoter. However, it repressed promoter activity by preventing Sp1 from binding to the promoter. Similarly, promyelocytic leukemia protein interfered with Sp1 binding to DNA and suppressed EGFR promoter activity. In another instance, p53 also repressed Sp1-mediated transcription of IGF-IR in Saos-2 osteosarcoma cells.

Decreased promoter activity of growth factor receptor genes may also be regulated by a reduction in transcriptional activators as cells cease proliferation. As skeletal muscle myoblasts exit the cell cycle and differentiate into multinucleated muscle fibers, FGFR1 gene expression and signaling decline. Simultaneously with cell differentiation, Sp1 and Sp3 protein levels also decline, indicating that regulation of this promoter is highly dependent on the expression and availability of both Sp1 and Sp3. Similar mechanisms of Sp factor deficiency appear to repress growth factor receptor gene expression in several human cancer cell lines. The TGF- $\beta$ RI promoter in the GEO human colon carcinoma cell line is inactivated due to reduced Sp transcription factors. Therefore, developmentally regulated and cell type-specific expression of these factors can significantly affect growth factor receptor promoter activity and ultimately affect the capacity of cells to proliferate and differentiate.

In conclusion, most studies have indicated that Sp1 is the key regulator of many growth factor receptor genes. Though other Sp-like proteins bind to the same *cis*-elements, how these proteins exert their functions

is not understood. At one level, specificity can follow from tissue-restricted expression. However, not only are a number of these factors expressed in multiple tissues, most cells express more than one factor at a time. As a result, the second level of specificity may occur via differential affinity in DNA binding or via their distinct activation and repression domains that mediate protein–protein interactions. Future studies in the involvement of transcription factors other than Sp1 in the regulation of growth factor receptor genes will provide more insight into the molecular mechanisms that govern the normal expression of these factors and provide potential therapeutic means to prevent disease.

### See Also the Following Articles

EGF and Related Growth Factors • Fibroblast Growth Factor (FGF) • GI Hormones as Growth Factors • Hepatocyte Growth Factor • Insulin and Insulin-like Growth Factors, Evolution of • Insulin-like Growth Factors • Peptide Hormones and Growth Factors: Cellular Signaling Mechanisms • Platelet-Derived Growth Factor (PDGF)

### Further Reading

- DiMario, J. X. (2002). Activation and repression of growth factor receptor gene transcription. *Int. J. Mol. Med.* **10**, 65–71.
- Dynan, W. S., and Tjian, R. (1983). The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. *Cell* **35**, 79–87.
- Itoh, N., Mima, T., and Mikawa, T. (1996). Loss of fibroblast growth factor receptor is necessary for terminal differentiation of embryonic limb muscle. *Development* **122**, 291–300.
- Johnson, A. C. (1996). Activation of epidermal growth factor receptor gene transcription by phorbol 12-myristate 13-acetate is mediated by activator protein 2. *J. Biol. Chem.* **271**, 3033–3038.
- Molander, C., Hackzell, A., Ohta, M., Izumi, H., and Funai, K. (2001). Sp1 is a key regulator of the PDGF  $\beta$ -receptor transcription. *Mol. Biol. Rep.* **28**, 223–233.
- Parakati, R., and DiMario, J. X. (2002). Sp1- and Sp3-mediated transcriptional regulation of the fibroblast growth factor receptor 1 gene in chicken skeletal muscle cells. *J. Biol. Chem.* **277**, 9278–9285.
- Patel, S. G., and DiMario, J. X. (2001). Two distal Sp1-binding *cis*-elements regulate fibroblast growth factor receptor 1 (FGFR1) gene expression in myoblast. *Gene* **270**, 171–180.
- Scata, K. A., Bernard, D. W., Fox, J., and Swain, J. L. (1999). FGF receptor availability receptor availability regulates skeletal myogenesis. *Exp. Cell Res.* **250**, 10–21.

Table 1 Growth Hormone Deficiency: Genetic Causes

Disorder	OMIM no.	Gene	Mode	Endocrine features	Comments
Achondroplasia	100800	FGFR3	AD	Decreased GH	Skeletal dysplasia, GHD due to sleep apnea
Borjeson-Forssman-Lehmann syndrome	301900	SOX3 (?)	XL	Decreased GH	
CHARGE syndrome	214800	22q11	Sporadic	Decreased GH	
Biodeficient GH	139250	GH1	?	Inactive GH	Responds to GH
Combined pituitary hormone deficiency					
HESX1	182230	HESX1	AR	CPHD	Absent corpus callosum and septo-optic dysplasia
LHX3	262600	LHX3	AR	CPHD	Abnormal cervical spine
PIT1	173110	PIT1	AR, AD	CPHD	GH, thyroid, & PRL deficiencies
PROP1	601538	PROP1	AR	CPHD	GH, thyroid, PRL, ACTH, and gonadotropin deficiencies
X-Linked	312000		XL	CPHD	Duplication of DXS102
Fanconi anemia	227650	FANCA	AR	Decreased GH	Heterogeneous defects
	227660	FANCB			
	227645	FANCC			
	227646	FANCD			
	600901	FANCE			
GHRHR defects	139191	GHRHR	AR	Decreased GH	Responds to GH
Hemochromatosis	235200	HFE	AR	CPHD	
	602390	HFE2			
Hemoglobinopathies	141900	HBB	AD	Occasional CPHD	
Histiocytosis X	246400	?	AR	Decreased AVP & GH	

following: (1) in combined pituitary hormone deficiency (CPHD), defects in the genes HESX1, LHX3, PROP1, and PIT1 are found; (2) in isolated GH deficiency (IGHD), there are defects in GH1 and GHRHR; and (3) in GH resistance, there are defects in the genes GHR and IGF-I.

## ISOLATED GH DEFICIENCY

Estimates of the frequency of GH deficiency (GHD) range from 1/4000 to 1/10,000 in various studies. The causes of GHD include central nervous system (CNS) insults or defects such as cerebral edema, chromosome anomalies, histiocytosis, infections, radiation, septo-optic dysplasia, and trauma or tumors that affect the hypothalamus or pituitary. Whereas most GHD patients are the only member of their families to be affected, estimates of the proportion of cases that have an affected parent, sibling, or child range from 3 to 30% in different studies. This familial clustering suggests that having a close relative affected conveys substantial relative risk and that a significant proportion of GHD cases may have a genetic basis. The genetic contributions to a series of familial

defects in the GH pathway are discussed in the following sections.

## Clinical Features

An intact GH pathway is needed throughout childhood to maintain normal growth. Concomitant or combined deficiencies of other pituitary hormones [luteinizing hormone (LH), follicle-stimulating hormone (FSH), thyroid-stimulating hormone (TSH), and/or adrenocorticotrophic hormone (ACTH)] in addition to GH deficiency constitute combined pituitary hormone deficiency or panhypopituitary dwarfism. The combination of GH and these additional hormone deficiencies often causes more severe retardation of growth, and skeletal maturation and spontaneous puberty may not occur.

## Diagnosis

Short stature, delayed growth, and delayed skeletal maturation occur with GH pathway defects. Since these signs can also be associated with systemic illnesses, individuals suspected of having GHD should be evaluated for systemic diseases before



IGHD II mutation that does not cause abnormal splicing is a G-to-A transition that results in an Arg to His substitution at residue 183 (Arg183His) of the GH molecule. This substitution is thought to alter the intracellular processing of the GH molecule by binding to zinc, thereby deranging the zinc-associated presecretory packaging of GH.

### **IGHD III**

A third form of IGHD called IGHD III (OMIM No. 307200) has an X-linked mode of inheritance and there have been distinct clinical findings in different families. In some families, all cases have agammaglobulinemia associated with their IGHD, whereas in other families, all cases have only IGHD. This suggests that contiguous gene defects on the long arm of the X chromosome may cause some IGHD III cases. Duriez *et al.* reported that X-linked agammaglobulinemia and IGHD are caused by a mutation in the Bruton's tyrosine kinase (BTK) gene.

Laumonnier *et al.* studied the SOX3 gene in families with X-linked mental retardation, where the causative gene had been mapped to Xq26-q27. They showed that the SOX3 gene maps to Xq26.3 and was involved in a large family in which affected individuals had mental retardation and IGHD (OMIM Nos. 300123 and 313430; [Table I](#)). The mutation was an in-frame duplication of 33 bp encoding 11 alanines in a polyalanine tract of the SOX3 gene. The expression pattern during neural and pituitary development suggested that dysfunction of the SOX3 gene caused by this polyalanine expansion might disturb transcription pathways and the regulation of genes involved in pituitary development.

### **Biodefective GH**

There have been a number of reports of patients with the clinical features of IGHD who achieved normal plasma immunoactive GH levels following GH provocative or stimulation tests, but low levels of somatomedin (OMIM No. 139250; [Table I](#)). Less GH was detected by radioreceptor assay than by RIA analysis in some studies. In view of these patients' clinical syndrome of IGHD, their apparently normal plasma concentrations of GH, their low basal somatomedin levels, and their normal response to exogenous GH, individuals with bioinactive GH are thought to secrete a biologically inert GH. Takahashi *et al.* identified a C-to-T transition at codon 77, which results in an Arg to Cys substitution in the GH1 gene of a subject diagnosed with bioinactive GH.

### **GHRH Receptor Defects**

A variety of mutations have been detected in the human GHRHR gene in individuals with IGHD (OMIM No. 139191). In a kindred with a nonsense mutation, affected family members had poor growth since infancy and were extremely short. They failed to produce GH in response to standard provocative tests and had good responses to GH replacement. Cases were homozygous for a G-to-T transversion that caused a premature termination mutation (Glu72Stop). Salvatori *et al.* described a large Brazilian family with many family members with IGHD due to an intronic G-to-T transition that destroys the 5' splice site of IVS1 of the GHRHR gene.

## **COMBINED PITUITARY HORMONE DEFICIENCY**

Cases with CPHD vary in their clinical findings because they have deficiencies of varying severity of one or more of the other pituitary tropic hormones (ACTH, FSH, LH, or TSH) in addition to GHD (OMIM No. 262600). Whereas most cases of CPHD are sporadic, a variety of familial forms that can have autosomal-recessive, autosomal-dominant, or X-linked modes of inheritance are known.

### **HESX1 Mutations**

HESX1 is expressed in the thickened layer of oral ectoderm that gives rise to Rathke's pouch, the primordium of the anterior pituitary. Down-regulation of HESX1 coincides with the differentiation of pituitary-specific cell types. Dattani *et al.* found a missense HESX1 mutation (Arg53Cys) in a homozygous state in a brother and sister with septo-optic dysplasia, agenesis of the corpus callosum, and CPHD (OMIM No. 182230).

### **LHX3 Mutations**

Murine Lhx3 mRNA accumulates in Rathke's pouch, the primordium of the pituitary, and may be involved in the differentiation of pituitary cells. Netchine *et al.* identified two families with CPHD (OMIM No. 262600) caused by mutations in the LHX3 gene. The phenotype associated with these mutations included the following: (1) severe growth retardation, (2) complete deficiency of all but one of the anterior pituitary hormones (ACTH), (3) elevated and anteverted shoulders with a short neck associated with severe restriction of rotation of the cervical spine, and (4) an

enlarged anterior pituitary. The authors concluded that LHX3 is required for the proper development of all anterior pituitary cell types except corticotropes and that the rigid cervical spine phenotype is consistent with a function of LHX3 in the proper development of extrapituitary structures as well.

### PIT1 Mutations

Defects in the PIT1 gene cause familial CPHD cases, which have a different phenotype (OMIM No. 173110). PIT1 is an anterior pituitary-specific transcription factor that regulates the expression of GH, prolactin (PRL), and TSH. PIT1 is also required for pituitary cellular differentiation and function. PIT1 has functional domains that enable the transactivation of other genes including GH, PRL, and TSH or binding to these genes. At least six different PIT1 mutations causing autosomal-recessive CPHD and two others causing autosomal-dominant CPHD have been found in humans in a subtype of panhypopituitary dwarfism associated with GH, PRL, and TSH deficiency (see [Table I](#)).

### PROPI Mutations

PROPI is a pituitary-specific homeodomain factor that is required for the development of somatotropes, lactotropes, and thyrotropes of the anterior pituitary and for the expression of PIT1. Multiple PROPI gene mutations cause an autosomal-recessive CPHD that has a third phenotype in humans (OMIM No. 601538; [Table I](#)). In addition to deficiencies of GH, PRL, and TSH that are seen in those with PIT1 defects, subjects with PROPI defects also have deficiencies of LH and FSH, which prevent the onset of spontaneous puberty and, in some cases, ACTH deficiency in later life. The various PROPI mutations include (1) a C-to-T transition at codon 120, which encodes a TGC (Arg) to CGC (Cys) substitution; (2) a T-to-A transversion that encodes a TTC (Phe) to ATC (Ile) substitution at codon 117; and (3) a 2 bp AG deletion in codon 101 (101delAG) that causes a frameshift and results in a premature stop at codon 109. The resulting protein products from all three of these different PROPI mutations have greatly reduced DNA-binding and transactivation abilities. The 101delAG is a recurring mutation that is estimated to occur in 55% of familial and 12% of sporadic CPHD cases. A fourth PROPI mutation is a 2 bp GA deletion at codon 51 (51delGA). Like the 101delAG mutation, the 51delGA mutation causes a frameshift that results in a premature stop codon.

This mutation was found in 12% of familial and 21% of sporadic CPHD cases.

### X-Linked CPHD

Lagerstrom-Fermer *et al.* described a family that included affected males suffering from variable degrees of CPHD (OMIM No. 312000). Some affected males who died during the first day of life had postmortem findings of hypoadrenalism, presumed to be due to CPHD. Others had variable combinations of hypothyroidism, delayed pubertal development, and short stature due to GHD. All surviving patients exhibited mild to moderate mental retardation. They found linkage with markers in the Xq25–q26 region. Furthermore, they found an apparent extra copy of the marker DXS102 in affected males and heterozygous carrier females, suggesting that a segment including this marker was duplicated.

## MENDELIAN DISORDERS WITH ENDOCRINE ABNORMALITIES

A variety of Mendelian disorders have among their pleiotropic effects endocrine abnormalities. These disorders include some abnormalities, such as achondroplasia, that have a single common mutation and others, such as hemoglobinopathies, that are caused by heterogeneous mutations.

### Achondroplasia with Obstructive Sleep Apnea

Achondroplasia is a common skeletal dysplasia in which the dwarfism is due to an abnormality in endochondral ossification (see OMIM No. 100800; [Table I](#)). Up to 10% of patients with achondroplasia have been reported to have serious respiratory complications. Some with achondroplasia and obstructive sleep apnea have low growth hormone secretion during sleep as a contributing cause of their growth retardation.

### Borjeson-Forssman-Lehmann Syndrome

The X-linked Borjeson-Forssman-Lehmann syndrome is characterized by short stature, hypogonadism, hypotonia, severe mental deficiency, and coarse facial appearance with a prominent brow ridge and large ears in affected males (see OMIM No. 301900; [Table I](#)). Markedly deficient GH responses to arginine and L-DOPA as well as low somatomedin C levels have been documented in affected individuals.

## CHARGE Association

CHARGE is an acronym that describes a nonrandom association of anomalies: *colobomas* of the eye; *heart* disease; *atresia* of the choanae; *retarded* growth and development and/or CNS anomalies; *genital* hypoplasia; and *ear* anomalies or deafness (see OMIM No. 214800; [Table I](#)). Growth retardation, which is usually of postnatal onset, and hypogonadism are prominent features of the CHARGE syndrome and may well be due to hypothalamic defects.

## Fanconi Anemia

Fanconi's syndrome is an autosomal-recessive disorder characterized by chronic pancytopenia with bone marrow hypoplasia, abnormal pigmentation, upper limb malformations, kidney anomalies, growth retardation, small genitalia, and increased frequency of chromosomal breaks in cultured lymphocytes (see OMIM Nos. 227650, 227660, 227645, 227646, and 600901; [Table I](#)). A number of investigators have documented GH deficiency in patients with Fanconi anemia and administration of GH resulted in excellent short-term and long-term responses in most of these patients.

## Hemochromatosis

Both male hypogonadism and pituitary hemosiderosis can occur in hemochromatosis (see OMIM Nos. 235200 and 602390; [Table I](#)) and abnormalities have also been found in gonadotropin, cortisol, GH, PRL, and TSH secretion.

## Hemoglobinopathies

There are well-documented cases of acquired pituitary insufficiency occurring in adults with hemoglobinopathies, presumably secondary to infarction of the gland (see OMIM No. 141900; [Table I](#)).

## Histiocytosis X

Histiocytosis X (also known as Letterer-Siwe disease, Hand-Schuller-Christian disease, or eosinophilic granuloma) is characterized by foamy histiocyte infiltration in many areas of the body, including the hypothalamus. When the histiocytic infiltration involves the hypothalamus, prepubertal growth retardation associated with GH deficiency and diabetes insipidus frequently occur (see OMIM No. 246400; [Table I](#)).

## Neurofibromatosis Type I

A variety of endocrine disturbances have been reported in patients with neurofibromatosis, which has an autosomal-dominant mode of inheritance (see OMIM No. 162200; [Table I](#)). The most common associated endocrine disorder in children with neurofibromatosis type 1 (NF1) is sexual precocity, whereas the most common associated endocrine disorder in adults is pheochromocytoma. Marked growth retardation and GH deficiency have also been reported.

## Pallister-Hall Syndrome

This neonatally lethal malformation syndrome consists of hypothalamic hamartoblastoma, hypopituitarism, postaxial polydactyly, and imperforate anus (see OMIM No. 146510; [Table I](#)). An anterior pituitary gland was absent in all cases. The posterior pituitary was absent in the majority of cases.

## Rieger's Syndrome

Rieger's syndrome (also known as iris-dental dysplasia) is an autosomal-dominant disorder associated with malformation of the iris, pupillary anomalies, and hypoplasia of the teeth, with or without maxillary hypoplasia. A large family in which multiple individuals had both Rieger's syndrome and IGHD (see OMIM No. 180500; [Table I](#)) has been described. Siblings of the proband had Rieger's syndrome with normal pituitary function, but GHD was not found in any member of the family who did not have Rieger's syndrome. One subject who was treated with GH exhibited substantial enhancement of his rate of growth.

## CHROMOSOMAL DISORDERS WITH ENDOCRINE ABNORMALITIES

A large variety of chromosomal abnormalities are associated with endocrine disorders. These chromosomal anomalies can affect a variety of autosomes as well as the sex chromosomes.

## Autosomal

GH deficiency has been described with 18p- and 20p- chromosomal deletions (see OMIM No. 146390; [Table II](#)). In addition, molecular detection of deletions of 17q (i.e., the GH gene cluster region) has been documented in IGHD IA. Since the gene for

GHRH has been mapped to 20p, the GH deficiency in 20p– could result from either the deletion of the GHRH gene or a developmental anomaly of the hypothalamus. Patients with 18p– have been reported with hypopituitarism and solitary central maxillary incisor, suggesting that the pituitary insufficiency in 18p– may be due to a structural malformation of the hypothalamus. GHD has been rarely associated with 47,XXY, 49,XXXXY, and ring 5. Finally a search of the London Dysmorphology Database at <http://www.hgmp.mrc.ac.uk/DHMHHD/view.html> identifies a series of partial chromosome deletions or duplications that are associated with short stature and pituitary abnormalities (see Table II). These include the deletions del(4)pter–p16, del(7)q32–qter, del(13)q22–qter, del(14)q22–q23, del(18)p, del(18)q21–qter, and del(22)pter–q11 and the duplications dup(1)q25–q32, dup(9)p, dup(9)pter–q22, and dup(11)q23–qter.

### Turner's Syndrome (45,X)

Although GH secretion has been reported to be normal, or paradoxically increased, in most patients with gonadal dysgenesis, pituitary insufficiency has been reported in several patients (see Table II). Ross *et al.* studied GH secretion in 30 patients with Turner's syndrome and found no differences in mean GH concentration or peak amplitudes throughout the day and night between patients less than 8 years of age and controls. Patients over 9 years of age had lower

mean GH levels and peak amplitudes. Reduced plasma somatomedin C levels and delayed bone age were found in patients of all ages. These abnormalities in GH secretion in Turner's syndrome are probably secondary to the absence of sex hormones during adolescence.

### TREATMENT OF PITUITARY GENE MUTATIONS

Recombinant-derived GH is widely available but must be given by subcutaneous injection. To obtain an optimal outcome, children with GHD should be started on replacement therapy as soon as their diagnosis is established. The dosage increases with increasing body weight to a maximum during puberty and is usually discontinued by ~17 years of age. Disorders in which GH treatment is of proven efficacy include GHD, either isolated or in association with CPHD, and Turner's syndrome. The clinical responses of individuals with IGHD or CPHD to GH replacement therapy vary depending on (1) the severity and age at which treatment is begun, (2) the recognition and response to treatment of associated deficiencies such as thyroid hormone deficiency, and (3) whether treatment is complicated by the development of anti-GH antibodies. The outcome of Turner's syndrome subjects varies with the severity of their (1) short stature, (2) chromosomal complement, and (3) age when treatment began.

Table II Chromosomal Disorders with Growth Hormone Deficiency

Disorder	Cytogenetic change	Endocrine features	Comments
Autosomes			
Deletions	del(4)pter–p16 del(7)q32–qter del(13)q22–qter, del(14)q22–q23 del(18)p del(18)q21–qter del(20)p del(22)pter–q11	Pituitary abnormalities with decreased GH, CPHD, or other deficiencies	Replace appropriate hormone(s)
Duplications	dup(1)q25–q32 dup(9)p dup(9)pter–q22 dup(11)q23–qter	Pituitary abnormalities with decreased GH, CPHD, or other deficiencies	Replace appropriate hormone(s)
Rings	Ring 5	Decreased GH or CPHD	Replace appropriate hormone(s)
Sex chromosomes	45,X 47,XXY 49,XXXXY	Decreased GH or CPHD	Replace appropriate hormone(s); GH & estrogen for 45,X

## See Also the Following Articles

Constitutional Delay of Growth and Puberty (CDGP) • Gigantism: Excess of Growth Hormone • Growth and Chronic Disease • Growth Hormone (GH) • Growth Hormone Insensitivity • Growth Hormone-Binding Proteins • Postnatal Normal Growth and Its Endocrine Regulation • Short Stature and Chromosomal Abnormalities • Turner Syndrome

## Further Reading

- Cogan, J. D., and Phillips, J. A., III (2001). Inherited defects in growth hormone synthesis and action. *In* "The Metabolic and Molecular Bases of Inherited Disease" (C. R. Scriver, A. L. Beaudet, W. S. Sly, and D. Valle, eds.), 8th ed., pp. 4159–4180. McGraw-Hill, New York.
- Cogan, J. D., Wu, W., Phillips, J. A., III, Arnhold, I. J. P., Agapito, A., Fofanova, O. V., Osorio, M. G. F., Bircan, I., Moreno, A., and Mendonca, B. B. (1998). The PROP1 2-bp deletion is a common cause of CPHD. *J. Clin. Endocrinol. Metab.* **83**, 3346–3349.
- Dattani, M. T., Martinez-Barbera, J.-P., Thomas, P. Q., Brickman, J. M., Gupta, R., Martensson, I.-L., Toresson, H., Fox, M., Wales, J. K. H., Hindmarsh, P. C., Krauss, S., Beddington, R. S. P., and Robinson, I. C. A. F. (1998). Mutations in the homeobox gene HESX1/Hesx1 associated with septo-optic dysplasia in human and mouse. *Nat. Genet.* **19**, 125–133.
- Duriez, B., Duquesnoy, P., Dastot, F., Bougneres, P., Amselem, S., and Goossens, M. (1994). An exon-skipping mutation in the btk gene of a patient with X-linked agammaglobulinemia and isolated growth hormone deficiency. *FEBS Lett.* **346**, 165–170.
- Fofanova, O., Takamura, N., Kinoshita, E., Parks, J. S., Brown, M. R., Peterkova, V. A., Evgrafov, O. V., Goncharov, N. P., Bulatov, A. A., Dedov, I. I., and Yamashita, S. (1998). Compound heterozygous deletion of the PROP-1 gene in children with combined pituitary hormone deficiency. *J. Clin. Endocrinol. Metab.* **83**, 2601–2604.
- Lagerstrom-Fermer, M., Sundvall, M., Johnsen, E., *et al.* (1997). X-linked recessive panhypopituitarism associated with a regional duplication in Xq25–q26. *Am. J. Hum. Genet.* **60**, 910–916.
- Laumonnier, F., Ronce, N., Hamel, B. C. J., Thomas, P., Lespinasse, J., Raynaud, M., Paringaux, C., van Bokhoven, H., Kalscheuer, V., Fryns, J.-P., Chelly, J., Moraine, C., and Briault, S. (2002). Transcription factor SOX3 is involved in X-linked mental retardation with growth hormone deficiency. *Am. J. Hum. Genet.* **71**, 1450–1455.
- Netchine, I., Sobrier, M.-L., Krude, H., Schnabel, D., Maghnie, M., Marcos, E., Duriez, B., Cacheux, V., Moers, A. V., Goossens, M., Gruters, A., and Amselem, S. (2000). Mutations in LHX3 result in a new syndrome revealed by combined pituitary hormone deficiency. *Nat. Genet.* **25**, 182–186.
- Online Mendelian Inheritance in Man (OMIM). (2000). McKusick–Nathans Institute for Genetic Medicine, Johns Hopkins University (Baltimore, MD) and National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health (Bethesda, MD). World Wide Web URL: <http://www.ncbi.nlm.nih.gov/omim>.
- Rimoin, D. L., and Phillips, J. A., III (1997). Genetic disorders of the pituitary gland. *In* "Principles and Practice of Medical Genetics" (D. L. Rimoin, J. M. Connor, and R. E. Pyeritz, eds.), 3rd ed., pp. 1331–1364. Churchill Livingstone, New York.
- Ross, J., Long, L., Loriaux, D., and Cutler, G. (1985). Growth hormone secretory dynamics in Turner syndrome. *J. Pediatr.* **106**, 202.
- Salvatori, R., Gondo, R. G., de Aguiar Oliveira, M. H., Phillips, J. A., III, Souza, A. H., Hayashida, C. Y., Toledo, S. P., Conceicao, M. M., Prince, M., Baumann, G., and Levine, M. A. (1999). Familial isolated growth hormone deficiency due to a novel mutation in the growth hormone-releasing hormone receptor. *J. Clin. Endocrinol. Metab.* **84**, 917–923.
- Shohat, M., Hermon, V., Melmed, S., *et al.* (1991). Deletion of 20p11.23–pter with growth hormone neurosecretory disorder but normal growth hormone releasing hormone genes. *Am. J. Med. Genet.* **39**, 56–63.
- Takahashi, Y., Kaji, H., Okimura, Y., Goji, K., Abe, H., and Chihara, K. (1996). Brief report: Short stature caused by a mutant growth hormone. *N. Engl. J. Med.* **334**, 432–436.
- Vnencak-Jones, C. L., Phillips, J. A., III, Chen, E. Y., and Seeburg, P. H. (1988). Molecular basis of human growth hormone gene deletions. *Proc. Natl. Acad. Sci. USA* **85**, 5615–5619.
- Wajnrajch, M. P., Gertner, J. M., Mullis, P. E., Deladoey, J., Cogan, J. D., Lekhakula, , Kim, S., Dannies, P. S., Saenger, P., Moshang, T., Phillips, J. A., III, and Leibel, R. L. (2000). Arg183His, a new mutational "hot-spot" in the growth hormone gene causing isolated GH deficiency type II. *J. Endocr. Genet.* **1**, 125–135.
- Wu, W., Cogan, J. D., Pfaffle, R. W., Dasen, J. S., Frisch, H., O'Connell, S. M., Flynn, S. E., Brown, M. R., Mullis, P. E., Parks, J. S., Phillips, J. A., III, and Rosenfeld, M. G. (1998). Mutations in PROP1 cause familial combined pituitary hormone deficiency. *Nat. Genet.* **18**, 147–149.

## GH ISOFORMS

Several molecular forms of hGH exist. The most abundant form (except during pregnancy) is pituitary GH-1, also known as GH-N or 22-kDa GH. GH-2 (also known as GH-V or placental GH) differs from pituitary GH-1 at 13 of the 191 amino acid residues. In addition, GH-2 contains an N-linked glycosylation site, and indeed, GH-2 exists as both a glycosylated and a nonglycosylated form. During pregnancy, placental GH progressively supplants pituitary GH in the maternal circulation. The *GH-1* gene also generates an internally deleted GH (lacking 15 internal amino acids and named 20-kDa GH) through alternative mRNA splicing. A secondary splice site within exon 3 is used for that purpose. This 20-kDa variant accounts for 5 to 10% of GH production. A second postulated splice variant arises from complete exon 3 skipping, owing to the relatively weak splice donor site in intron 3. Although the mRNA for this so-called 17.5-kDa GH has been demonstrated, it is not entirely clear whether the corresponding protein is produced in significant amounts in normal individuals. Additional GH molecular variants arise from posttranslational processing. These include glycosylated (especially placental GH), acetylated, deamidated, and (in some animals) phosphorylated GHs. Further molecular heterogeneity results from oligomeric GH forms (up to at least pentameric GH), with both noncovalently associated oligomers and disulfide-linked oligomers. The biological roles of the GH variants are largely unknown. They vary in biological activity, with 22-kDa GH-1 and GH-2 being the most biologically active.

## REGULATION OF GH PRODUCTION AND SECRETION

Expression of the *GH-1* gene in pituitary somatotropes is positively regulated by GH-releasing hormone (GHRH), glucocorticoids, and (in rodents) thyroid hormones. It is negatively regulated by insulin-like growth factor-1 (IGF-1) and somatostatin. GHRH is also very important for normal somatotrope cell proliferation during pituitary development. GH is stored in secretory granules and released in response to GHRH; its release is inhibited by somatostatin. Ghrelin or related GH secretagogues, in pharmacological amounts, are potent releasers of GH, but the physiological role of ghrelin in GH regulation appears to be minor at best. The various GH isoforms appear to be cosecreted, and no isoform-specific stimulus has been identified.

GH secretion is under complex hypothalamic control by at least two hypophysiotropic hormones: GHRH (stimulating) and somatostatin (inhibitory). Secretion is pulsatile, with a marked ultradian rhythm of pulses of widely varying amplitude occurring every 1 to 2 h. The highest secretory pulses are linked to slow wave sleep and typically occur during the first 2 h after sleep onset. However, major pulses can also occur at other times due to stimuli such as stress, exercise, pain, and other acute events. Feedback inhibition of GH secretion is provided by IGF-1 as well as by GH itself. Daily GH production is high and largely unregulated during fetal and neonatal life, falls after birth under the regulatory influence of IGF-1, is again high during puberty, and falls progressively thereafter by about 15% per decade. Physiological and metabolic regulators of GH secretion are nutrition, estrogen (stimulatory), glucocorticoids (the inhibitory effect at the hypothalamic level predominates over the stimulatory effect at the pituitary level), and thyroid hormone (stimulatory, especially in rodents). In humans, undernutrition stimulates GH production, whereas overnutrition dampens GH production (the opposite is generally true in rodents). There is marked sexual dimorphism in the GH secretory rhythm, with women showing generally higher values, higher basal secretion, and more "noisy" circadian profiles. This difference can be largely attributed to an estrogen effect. In rats, the male secretion pattern is characterized by very high pulses interrupted by nearly complete quiescence, whereas the female secretion pattern is characterized by lower pulses but higher tonic secretion between pulses.

In contrast to this complex regulation of GH production in the pituitary gland, the expression and secretion of placental GH during pregnancy appear to be constitutive and largely unregulated.

## BLOOD TRANSPORT OF GH

GH circulates in blood partly bound to two GH-binding proteins (GHBPs). The principal carrier is the high-affinity GHBP that corresponds to the soluble ectodomain of the GH receptor (GHR). In humans and many other species, the GHBP is produced by proteolytic cleavage of the GHR in its juxtamembranous stem region by a metalloproteinase(s) of the ADAM family (TACE = TNF- $\alpha$ -converting enzyme and perhaps other ADAMs), with shedding of the ectodomain from the cell as the GHBP. In rodents, the GHBP is generated from the GHR gene as an alternative mRNA splice product. Under basal conditions, approximately half of circulating

GH is bound to this GHBP. A minority of GH (~5%) is bound to another, low-affinity GHBP that appears to correspond to a modified form of  $\alpha_2$ -macroglobulin. The GHBPs prolong the plasma half-life of GH and act to provide a circulating GH reservoir. In addition, the high-affinity GHBP is a modulator of GH action at the cellular level. The plasma half-life of GH in humans is 15 to 20 min and is eliminated primarily through glomerular filtration and GHR-mediated clearance.

## THE GHR AND GH SIGNAL TRANSDUCTION

The GHR is a 620-amino acid single-chain glycoprotein that belongs to the family of cytokine receptors. It has a large, 246-residue extracellular domain, a single-transmembrane domain, and a 350-amino acid cytoplasmic domain. The ectodomain is folded into two subdomains; the amino-terminal subdomain 1 contains the GH-binding site, and the carboxy-terminal subdomain 2 is involved in receptor dimerization. A short (~10 residue) linear stem region between subdomain 2 and the transmembrane domain serves as the substrate for cleavage by TACE to yield the GHBP. The cytoplasmic domain contains several features involved in hormone signaling. Among those, a membrane-proximal proline-rich region (Box 1) is most important; it serves as a docking site for Janus kinase 2 (Jak 2). Another important region in the cytoplasmic portion is the internalization domain.

The single *GHR* gene resides on the short arm of human chromosome 5 (5p13–p12), spans more than 156 kb, and consists of at least 10 exons, of which exons 2 to 10 encode the protein. The 22-residue portion encoded by exon 3 is not necessary for function, and inclusion or exclusion of exon 3 occurs as a form of *GHR* polymorphism. Several alternative exons 1 can be used to express the GHR in a tissue- or metabolic state-specific manner. They are only incompletely characterized. Two truncated splice variants of the GHR are known; they lack most of the cytoplasmic domain, represent a small part (~1–5%) of the total GHR complement, and have an unknown function. The GHR is expressed ubiquitously, although the level of expression varies among tissues. The liver is the tissue with the highest GHR content.

GH binding to the GHR results in GHR dimerization as two GHRs associate sequentially with the two binding sites on GH. (Some evidence also suggests that the GHR is predimerized and then conformationally changed by GH binding.) GHR

dimerization is followed by initiation of a signal transduction cascade by recruitment of Jak 2 to the GHR cytoplasmic domain and phosphorylation of both Jak 2 and the GHR. Following this initial step, several downstream signal transduction pathways are activated. They include, most prominently, the Jak–Stat (especially Stat 5b) pathway, but also the mitogen-activated protein (MAP) kinase pathway, the insulin receptor substrate (IRS)/phosphatidylinositol 3-kinase (PI3K) pathway, the phospholipase C (PLC)/protein kinase C (PKC) pathway, and probably other signaling pathways. A detailed discussion of these signaling cascades is beyond the scope of this article. The various pathways may subserve different components of the GH bioactivity spectrum. For example, the growth-promoting/IGF-1-generating action is principally mediated by signaling through the Jak–Stat pathway, whereas some metabolic activities are mediated by signaling through PI3K. A whole host of genes are activated in response to GH signaling. Among those best recognized are serine protease inhibitor 2.1 (Spi 2.1), c-Fos, and IGF-1. GH signaling also has nongenomic effects in the cell, such as enhanced glucose transport.

GH signaling is negatively regulated by several mechanisms to prevent runaway cellular stimulation, with some acting in a classic feedback loop. Among the factors inhibiting GH signaling once it has started are induction of SOCS (suppressors of cytokine signaling) proteins and CIS (cytokine-inducible SH2 protein), activation of phosphatases, GHR down-regulation through GHR internalization, and perhaps GHR inactivation by proteolytic GHR decapitation/GHBP shedding.

The GHBP, in addition to its role in the circulation, acts as a local modulator of GH action by competing with GHRs for ligand and probably by forming biologically unproductive GHR/GHBP dimers at the cell surface. Thus, at the local tissue level, GHBP acts primarily as an inhibitor of GH action.

## BIOLOGICAL ACTIONS OF GH

Many, but not all, of the activities of GH are mediated by its second messenger, IGF-1. IGF-1 is generated in response to GH in liver and many other target tissues. It is primarily responsible for the growth-promoting action, serving as both a mitogen and a metabolically active hormone. For some GH actions, it is still not clear whether they are mediated by IGF-1, by GH directly, or by both. [Table I](#) lists the principal bioactivities of GH. The protean and diverse manifestations of GH action are evident from the table. All

**Table 1 Biological Activities of GH**

A. Direct GH effects
Lipolysis
Prechondrocyte differentiation
Preadipocyte differentiation
Amino acid transport in muscle
Glucose transport (acute hypoglycemic effect, so-called “insulin-like effect”)
IGF-1 generation, IGFBP3 and ALS generation
Hypothalamic somatostatin secretion (short-loop feed back)
Milk production (human GH only)
B. IGF-1 mediated effects
Sulphate and thymidine incorporation into growth cartilage
Clonal expansion of chondrocytes
Linear bone growth
DNA and RNA synthesis
Renal Na and phosphate retention
C. Unknown or combined GH/IGF-1 effects
Nitrogen retention
Somatic growth <sup>a</sup>
Insulin antagonism
Beta cell hyperplasia
Erythropoiesis
Immune system stimulation

<sup>a</sup>Somatic growth involves not only bone elongation, but also concomitant muscle and organ growth. Somatic growth is possible with IGF-1 alone, but differs both quantitatively and qualitatively from that induced by GH.

tissues and organs are targets for GH action; hence, the term “somatotropin” describes this hormone more aptly than does the term “growth hormone.” The most prominent net effects of GH actions *in vivo* are somatic growth, loss or maintenance of fat mass, muscle anabolism, increase or maintenance of bone mineral density, and insulin antagonism.

## IGF-1 AND IGF-BINDING PROTEINS

IGF-1 and the related IGF-2 are 70- and 67-amino acid, proinsulin-like polypeptides of about 7.5 kDa, respectively, that are induced by GH (IGF-1 and IGF-2) and mediate many of the GH actions. IGF-1, rather than IGF-2, is the principal player in the GH-IGF axis. Circulating IGFs are bound to six IGF-binding proteins (IGFBPs). Foremost among those is IGFBP3, which complexes the great majority of IGFs in a 150-kDa ternary complex composed of IGF, IGFBP3, and another protein called acid-labile subunit (ALS). Both IGFBP3 and ALS are GH inducible and GH dependent. The ternary complex acts to

retain IGFs in the circulation, thereby prolonging their half-life (~20 h) and modulating their access to tissues and *in vivo* bioactivity. The other IGFBPs bind smaller amounts of IGFs and are of lesser importance as IGF transport proteins. They play a role in local regulation of IGF action in tissues. IGFBP1 is inversely regulated by insulin (i.e., insulin down-regulates IGFBP1) and acts as a circulating inhibitor of IGF action by complexing IGFs. IGF-1 acts through its own receptor (the type 1 IGF receptor), which is structurally related to the insulin receptor and expressed widely.

## DISEASE STATES WITH ABNORMALITIES IN THE GH-IGF AXIS

### GH Deficiency

GH deficiency can result from a variety of causes such as genetic defects, birth trauma, and organic lesions affecting the pituitary or the hypothalamus. The most common cause is a pituitary or hypothalamic tumor resulting in destruction of somatotropes or GHRH-producing neurons as well as interruption of hypothalamo-pituitary communication. In patients with pituitary tumors, GH deficiency usually develops before the other pituitary hormones are compromised. Genetic causes include inactivating mutations in the *GH-1* gene, the *GHRH receptor* gene, or genes involved in pituitary development such as *PROP-1*, *PIT-1*, *HESX1*, *PITX2*, *LHX3*, and *LHX4*. One relatively common type of GH deficiency is called “idiopathic,” meaning that no cause is evident. In general, this is a diagnosis of exclusion in a child with poor growth and subnormal serum GH levels. The delineation between this entity and normal variation of growth patterns is difficult.

The clinical manifestations of GH deficiency depend, in part, on whether it develops during childhood or during adult life. Childhood GH deficiency is characterized by growth retardation, a feature that does not apply to adults. Other clinical manifestations of GH deficiency include moderately delayed puberty, childhood hypoglycemia, increased adiposity (especially visceral fat), osteopenia, decreased lean body mass, low exercise tolerance and stamina, extracellular volume depletion, impaired psychosocial functioning, and decreased quality of life. In humans, immune function is not sufficiently affected to be clinically relevant. Except for growth retardation and hypoglycemia, the manifestations of GH deficiency are relatively subtle and not readily recognized unless sought out specifically. Diagnosis is suspected in the proper

clinical setting and is confirmed by the failure of serum GH to rise in response to standard pharmacological stimuli. Typical biochemical features include low levels of serum IGF-1, IGFBP3, and ALS as well as an elevated level of IGFBP2. Treatment consists of replacement therapy with recombinant GH.

Deficiency of the placental GH appears to have no adverse effects on either mother or fetus, as has been learned from naturally occurring cases with a deletion of the *GH-2* gene.

### GH Insensitivity

GH insensitivity or resistance shows a clinical picture that is similar, although not identical, to GH deficiency. There is no shortage of GH; rather, there is inability of GH to act. In its most extreme form, GH insensitivity is a genetic syndrome (Laron syndrome) caused by inactivating mutations in the *GHR* gene. Patients show all of the physical manifestations of GH deficiency, but serum GH levels are high, while IGF-1, IGFBP3, and ALS levels are low. In most, but not all, cases, the serum GHBP level is low or undetectable (the type of mutation in the GHR determines the presence or absence of GHBP). The phenotype is one of complete absence of GH activity and resembles that of patients with the most severe degrees of GH deficiency. Diagnosis is suspected in the proper familial setting by measuring serum GH, IGF-1, and GHBP; it is usually confirmed by genetic analysis. A unique form of GH insensitivity was reported in a single patient who carried a partial deletion of the *IGF-1* gene. Treatment consists of IGF-1 replacement therapy.

A mild to moderate form of acquired GH resistance is frequently seen in catabolic states. This appears to be an adaptive response, with low serum IGF-1 and elevated GH levels. The nutritional deprivation that frequently accompanies such conditions explains part, but not all, of this phenomenon. This is a reversible derangement in the GH-IGF axis that returns to normal when the underlying disease process is corrected. Treatment with GH is not recommended because GH therapy, in severe cases of illness, has been associated with increased mortality. However, one partially GH-resistant condition where GH treatment is beneficial is chronic renal failure.

### GH Excess

Overabundance of GH secretion leads to a condition called acromegaly in adults. When it occurs during

childhood, it leads to gigantism. In the latter case, overall somatic growth is accelerated; in the former case, there is only acral overgrowth in hands, feet, and facial structures. The condition is typically caused by a GH-producing pituitary adenoma. At least half of these adenomas have somatic mutations in the  $G_s\alpha$ -subunit of the signal-transducing G protein. (GHRH normally signals through this pathway.) The mutation renders the G protein constitutively active and leads to tumor formation and unchecked GH production. A germ line variant of this type of G-protein activation is seen in McCune-Albright syndrome, which has as one of its manifestations the occurrence of acromegaly. In rare cases, acromegaly can be caused by overproduction of GHRH, either by a eutopic source (e.g., a hypothalamic lesion) or (more frequently) in an ectopic site by tumors of neuroendocrine lineage (e.g., carcinoids, islet cell tumors). The high levels of systemic unregulated GHRH lead to somatotrope hyperplasia and GH overproduction. Only one convincing case of ectopic production of GH itself—by an islet cell tumor—has been reported. The clinical aspects of acromegaly include soft tissue swelling and bony overgrowth in hands and feet, prognathia and frontal bossing, dental malocclusion, a general coarsening of facial features and body habitus, general organomegaly, hypertension, carbohydrate intolerance or diabetes mellitus, and increased cardiovascular morbidity and mortality. The diagnosis is typically delayed because of the insidious onset of clinical signs. Biochemical findings include increased serum GH and IGF-1 levels. The diagnosis is supported by demonstrating a pituitary tumor and confirmed by showing that serum GH is not normally suppressible by administration of glucose (an oral or intravenous glucose tolerance test). In cases of ectopic GHRH production, a high GHRH level in peripheral blood is diagnostic. Treatment of acromegaly consists of surgical removal of the pituitary adenoma (or, rarely, the ectopic tumor). Surgical resection is frequently not curative, especially when the adenoma exceeds the confines of the sella turcica. In such cases, radiation or medical therapy can be used. Medical treatment in the form of the long-acting somatostatin analogue octreotide (or lanreotide) is often effective in reducing GH secretion toward or into the normal range. Another form of medical treatment is available in the form of the GH antagonist pegvisomant, which blocks GH action at the GHR by preventing receptor dimerization. The aim of therapy is to lower serum IGF-1 levels to the normal range. Effective treatment of acromegaly is successful in reversing the soft tissue changes and metabolic

derangements but is only partially effective in reversing the bony changes. Thus, early diagnosis and curative intervention are of paramount importance. Early diagnosis is linked to cure rate because the size of the pituitary tumor largely determines the chance for a surgical cure.

## DIAGNOSTICS

GH deficiency is diagnosed by pituitary stimulation test because a random serum GH level is largely uninformative due to the normally pulsatile nature of GH secretion. Standard and reliable pharmacological tests used for stimulation of GH secretion are: insulin hypoglycemia, GHRH-arginine infusion, and possibly GHRH-ghrelin or GHRH-hexarelin injection. Other useful but less reliable tests include arginine alone, GHRH alone, L-dopa, glucagon, and clonidine. Clonidine appears to be a more potent secretagogue in children than in adults, where it is considered a weak stimulus. The absolute GH response to these provocative stimuli (in terms of serum GH levels achieved) varies depending on the GH assay used. Typically, monoclonal immunoassays yield lower GH values than do polyclonal assays. Diagnostic guidelines have been published by the GH Research Society and by the American Association of Clinical Endocrinologists.

Acromegaly is readily diagnosed in the right clinical setting by an elevated serum IGF-1 level and a high GH level. (The latter is not diagnostic by itself due to the normally pulsatile GH secretion.) The confirmatory test, both in overt acromegaly and especially after surgical treatment, is the glucose tolerance test. Serum GH should fall below 1 ng/ml after glucose. In addition, IGF-1 should be within the normal age-adjusted range. Frequently, evidence for low-grade acromegaly persists even in cases where these criteria are met as GH secretory dynamics remain disordered when examined carefully. For practical clinical purposes, an IGF-1 level in the mid-normal range is a reasonable criterion for a cure.

## GH AS A THERAPEUTIC AGENT

Human GH was extracted from cadaveric human pituitaries until the mid-1980s because there was no other source of GH that was biologically active in humans. Only children were recipients of this GH because supplies were limited by necessity. The occurrence of cases of Creutzfeldt-Jacob disease attributed to hGH contaminated with prions and the

simultaneous arrival of recombinant hGH in unlimited quantities resulted in a universal switch to recombinant hGH. Supplies became sufficient to treat adult patients with GH deficiency as well as conditions not associated with GH deficiency (e.g., chronic renal failure, Turner syndrome). Several pharmaceutical companies manufacture 22-kDa hGH, which is highly effective despite lacking all the other GH isoforms that are normally produced by the pituitary. Indications for GH continue to expand beyond classical GH deficiency, and much has been learned about GH biology from the availability of large quantities of chemically defined, highly pure GH preparations as well as from clinical trials and more intense scrutiny of the manifestations of GH deficiency and the effects of GH replacement therapy.

## Acknowledgments

This research was supported in part by a grant from the National Science Foundation and a Merit Review grant from the Department of Veterans Affairs.

## See Also the Following Articles

Acromegaly, Clinical Features of • Acromegaly, Diagnosis of • Aging: Muscle • Gigantism: Excess of Growth Hormone • Growth and Chronic Disease • Growth and Glucocorticoids • Growth Hormone Insensitivity • Growth Hormone-Binding Proteins • Insulin-like Growth Factors • Lipoprotein(a) • Natural and Synthetic Growth Hormone Secretagogues • Pituitary Gland Anatomy and Embryology

## Further Reading

- Anonymous. (2000). Consensus guidelines for the diagnosis and treatment of growth hormone (GH) deficiency in childhood and adolescence: summary statement of the GH Research Society. *GH Research Society. J. Clin. Endocrinol. Metab.* **85**, 3990–3993.
- Baumann, G. (1991). Growth hormone heterogeneity: Genes, isohormones, variants, and binding proteins. *Endocrine Rev.* **12**, 424–449.
- Baumann, G. (2001). Growth hormone binding protein 2001. *J. Pediatr. Endocrinol. Metab.* **14**, 355–375.
- Baumann, G. (2002). Genetic characterization of growth hormone deficiency and resistance: Implications for treatment with recombinant growth hormone. *Am. J. Pharmacogenomics* **2**, 93–111.
- Baumann, G., and Frank, S. J. (2002). Metalloproteinases and the modulation of growth hormone signalling. *J. Endocrinol.* **174**, 361–368.
- Carroll, P. V., Christ, E. R., Bengtsson, B. A., Carlsson, L., Christiansen, J. S., Clemmons, D., Hintz, R., Ho, K., Laron, Z., Sizonenko, P., Sonksen, P. H., Tanaka, T., and Thorner, M. (1998). Growth hormone deficiency in adulthood and the effects of growth hormone replacement: A review—Growth Hormone

- Research Society Scientific Committee. *J. Clin. Endocrinol. Metab.* **83**, 382–395.
- Cunningham, B. C., Ultsch, M., De Vos, A. M., Mulkerrin, M. G., Clauser, K. R., and Wells, J. A. (1991). Dimerization of the extracellular domain of the human growth hormone receptor by a single hormone molecule. *Science* **254**, 821–825.
- Giustina, A., and Veldhuis, J. D. (1998). Pathophysiology of the neuroregulation of growth hormone secretion in experimental animals and the human. *Endocrine Rev.* **19**, 717–797.
- Herrington, J., and Carter-Su, C. (2001). Signaling pathways activated by the growth hormone receptor. *Trends Endocrinol. Metab.* **12**, 252–257.
- Kelly, P. A., Djiane, J., Postel-Vinay, M. C., and Edey, M. (1991). The prolactin/growth hormone receptor family. *Endocrine Rev.* **12**, 235–251.
- Rosenfeld, R. G., Rosenbloom, A. L., and Guevara-Aguirre, J. (1994). Growth hormone (GH) insensitivity due to primary GH receptor deficiency. *Endocrine Rev.* **15**, 369–390.

**Table I Primary Classification of GH Insensitivity**

Primary GH insensitivity (hereditary defects)
<ul style="list-style-type: none"> <li>• GH receptor defect (may be positive or negative for GH-binding protein) <ul style="list-style-type: none"> <li>- Extracellular mutation</li> <li>- Cytosolic mutation</li> <li>- Intracellular</li> </ul> </li> <li>• GH signal transduction defect (distal to cytoplasmic domain of GH receptor)</li> <li>• IGF-1 synthetic defect</li> <li>• IGF-1 gene deletion</li> <li>• IGF-1 transport defect</li> <li>• IGF-1 receptor defect</li> <li>• Bioinactive GH molecule</li> </ul>
Secondary GH insensitivity (acquired defects)
<ul style="list-style-type: none"> <li>• Circulating antibodies to GH that inhibit GH action</li> <li>• Antibodies to the GH receptor</li> <li>• GH insensitivity caused by malnutrition, liver disease, catabolic states, and the like</li> <li>• Other conditions that cause GH insensitivity</li> </ul>

–10 SD. In the large series of Laron syndrome patients from the Middle East reported by Laron's group, final height ranged from –4 to –8 SD. In the series reported from Ecuador studied by Guevara-Aguirre and Rosenfeld's group, final height ranged from –6.8 to –9.6 SD.

GH secretion is increased and is associated with extremely low or undetectable levels of the GH-dependent peptides IGF-1, IGF-binding protein-3 (IGFBP-3), and acid labile subunit (ALS). The IGF-1 generation test, designed originally to predict growth response to GH therapy, shows a notable absence of increase in the preceding three peptides, confirming the presence of insensitivity to both endogenous and exogenous GH. A score of criteria for the diagnosis of GH insensitivity were proposed by Blum and colleagues (Table II).

### PARTIAL GH INSENSITIVITY

During the early 1990s, the availability of recombinant IGF-1 for treatment led to the collection and central

**Table II Scoring System for the Diagnosis of GHIS**

Test	Parameter	Criterion	Score
Auxology	Height	<–3 SDS	1
Basal GH	GH	>2.5 ng/ml	1
Basal IGFs	IGF-1	<50 µg/L	1
	IGFBP-3	<–2 SDS	1
IGF-1 Generation	Change in IGF-1	<15 µg/L	1
	Change in IGFBP-3	<0.4 mg/L	1
GH Binding	Percentage GH bound	<10%	1

*Note.* A total of ≥ 5 points indicates GHIS.

analysis of a large, predominantly European population of patients with GH insensitivity. This series consisted of 82 patients, each with the diagnostic criteria of GH insensitivity (Table I) but with varying degrees of phenotypic and biochemical abnormality. It became clear that GH insensitivity was not a homogenous state but rather a highly heterogeneous clinical state. Height in this population ranged from –2.2 to –10.4 SDS, and serum IGFBP-3 ranged from –1.4 to –14.7 SDS.

One analysis of the European population of GH insensitivity patients classified the patients into those with classical Laron syndrome features and those with atypical features such as normal facial appearance. Clinical and biochemical details of the atypical and classical patients are shown in Tables III and IV, respectively. It can be concluded that the disorder of atypical or partial GH insensitivity does exist. Such patients might not have the typical clinical features of Laron syndrome and, hence, may present as idiopathic short stature (ISS).

There have been several reports of children with ISS having clinical, biochemical, and molecular evidence of GH insensitivity. However, it appears likely that mild GH insensitivity is only a rare cause of ISS. Nevertheless, the IGF-1 generation test, using either the standard dose of GH (0.033 mg/kg/day × 4) or a lower dose, with measurement of IGF-1 and IGFBP-3, may be of value in identifying such patients.

**Table III Clinical Details of Atypical and Classical GHIS Patients**

	n	Age (years)	Sex (M/F)	Height	Body weight (g)	Body length (cm)
Atypical GHIS	9	7.8 ± 4.1	7/2	–4.0 ± 1.4	3445 ± 245	51.8 ± 2.5
Classical GHIS	50	8.6 ± 4.6	24/26	–6.7 ± 1.4	3290 ± 374	48.3 ± 1.7
				P<0.001		

Table IV Biochemical Details of Patients with Atypical and Classical GHIS

	Basal GH (ng/ml)	Basal IGF-1 ( $\mu$ g/L)	IGF-1 post-GH ( $\mu$ g/L)	IGFBP-3 ( $\mu$ g/L)	Normal GHBP (>10%)	IGF-2 ( $\mu$ g/L)	IGFBP-1 ( $\mu$ g/L)
Atypical	18 (1.1–69.5)	22 (<20–53)	26 (<20–82)	776 (213–1670)	28 (4–52)	148 (39–283)	165 (28–521)
Classical	36 (1–135)	<20 (<20–135)	<20 (<20–62)	334 (95–170)	0 (0–46)	84 (26–205)	153 (39–489)

Note. Values are expressed as means and ranges (in parentheses). Levels of IGF-1, IGFBP-3, IGF-1, and IGFBP-1 are age dependent. Standard deviation scores for IGFBP-3 are outlined in the text.

## MOLECULAR BASIS OF GH INSENSITIVITY

The demonstration by the group of Laron that the basic defect in Laron syndrome, an autosomal recessively inherited disorder, was situated at the level of the GH receptor orientated scientists to characterize the molecular basis of GHR dysfunction. It is now established that the majority of GHR mutations are situated in exons coding for the extracellular domain of the receptor. This was confirmed in the genetic analysis of the European GH insensitivity population (Fig. 1). More recently, new mutations are being described, extending the mechanism of GH receptor dysfunction. Examples of these are splice site mutations affecting receptor dimerization, intronic mutations associated with pseudoexon formation, and dominantly inherited heterozygous mutations with dominant negative action on GHR function. Most recently, a mutation of the STAT5 gene, causing impaired intracellular signal transduction, was identified by Rosenfeld's group.

## IGF-1 GENE DELETION

An example of IGF-1 deficiency, not linked to the GHR, was reported by Woods and colleagues in

1996. The patient had severe intrauterine growth retardation and postnatal growth failure. He also had a dysmorphic phenotype with microcephaly, sensorineural deafness, and mild mental retardation. Serum IGF-1 was undetectable, IGFBP-3 was normal, and GH was increased. Molecular analysis revealed partial deletion of the IGF-1 gene, with absence of exons 4 and 5. This patient also has severe insulin resistance, secondary to high GH levels, acting through an intact GH receptor. Treatment with recombinant IGF-1 suppressed GH and corrected the insulin resistance.

## ACQUIRED GH INSENSITIVITY

GH insensitivity is seen in association with a number of chronic disorders that have mostly been studied in adults. Predominant among these are disorders of acute illness such as patients receiving intensive care or undergoing major surgery. Serum levels of IGF-1 are reduced and GH is elevated, indicating GH insensitivity. Other examples are chronic liver disease, where there may be impaired synthesis of IGF-1, and malnutrition, where IGF-1 generation is sensitive to calorie intake. Two chronic disorders

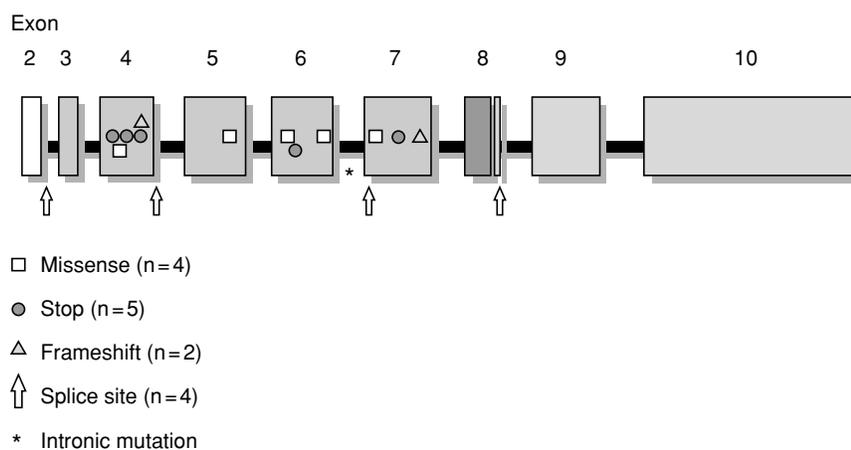


Figure 1 GHR mutations in 21 families with GH insensitivity from the European series. From Woods, K. A., Dastot, F., Preece, M. A., Clark, A. J. L., Postel-Vinay, M.-C., Chatelain, P. G., Ranke, M. B., Rosenfeld, R. G., Amselem, S., and Savage, M. O. (1997). Phenotype-genotype relationships in growth hormone insensitivity syndrome. *J. Clin. Endocrinol. Metab.* 82, 3529–3535, © The Endocrine Society, with permission.

affecting children and adolescents, both associated with GH insensitivity, are type 1 diabetes mellitus and Crohn's disease. In diabetes, GH secretion is elevated, particularly during adolescence, contributing to the insulin resistance, which is well recognized in adolescent patients. Dunger and colleagues showed effective reduction of insulin resistance and improvement in glycemic control during treatment with recombinant IGF-1.

In Crohn's disease, there is usually a combination of chronic malnutrition and inflammation. The malnutrition causes IGF-1 deficiency, and the inflammation, mediated through the effect of cytokines (which also impair IGF-1 production), exacerbates the GH insensitive state. A similar situation is seen in the chronic inflammatory state of patients with juvenile chronic arthritis.

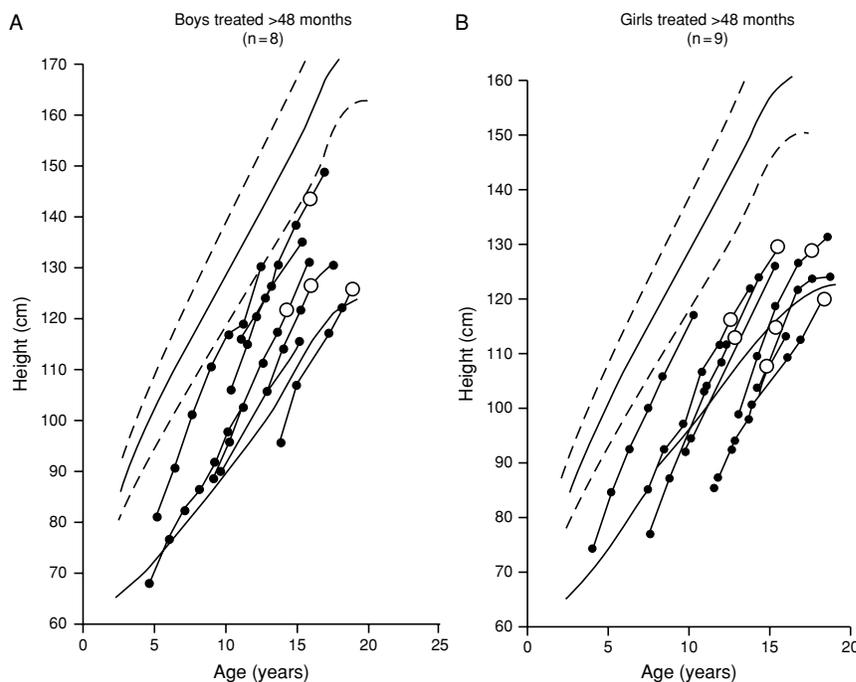
## TREATMENT OF GH INSENSITIVITY WITH IGF-1

GH insensitivity is associated with deficiency of IGF-1 and IGFBP-3. Because GH treatment is ineffective, IGF-1 is the only available therapy for

this group of disorders. Early trials during the late 1980s demonstrated a clear growth-promoting effect of recombinant IGF-1 given by subcutaneous injection in twice-daily doses of 40 to 120  $\mu\text{g}/\text{kg}$ . Laron showed similar responses using IGF-1 (150  $\mu\text{g}/\text{kg}$ ) once daily. A compilation of results of treatment trials from 1993 to 1997 showed an increase in mean growth velocity during the first year of treatment, from 3.6 cm/year (range = 2.9–4.7) to 8.4 cm/year (range = 7.2–9.3).

IGF-1 does not increase growth in GH insensitivity to the same extent as does GH therapy in severely GH-deficient children. This may be related to rapid clearance of free IGF-1 because of the preexisting IGFBP-3 deficiency. A potential solution to this problem would be the use of the IGF-1/IGFBP-3 complex, an equimolar preparation that has been shown to be effective in the treatment of the GH insensitivity of adolescent diabetes mellitus.

However, Ranke and colleagues demonstrated the long-term benefit of IGF-1 therapy on growth in children with GH insensitivity (Fig. 2). IGF-1 has been in short supply, resulting in limited opportunity to establish the optimal treatment regimens. The



**Figure 2** Height development in boys (A) and girls (B) during IGF-1 treatment in relationship to normal standards (Tanner) and (median) untreated patients with Laron syndrome. Open circles mark the onset of puberty. Ranke, M. B., Savage, M. O., Chatelain, P. G., Preece, M. A., Rosenfeld, R. G., and Wilton, P. (1999). Treatment of growth hormone insensitivity syndrome with insulin-like growth factor-I: Long-term results of the European Multicentre Study. *Horm. Res.* 51, 128–134, S. Karger AG, Basel, with permission.

majority of children with GH insensitivity have not been able to obtain access to IGF-1 treatment. It is hoped that this situation will improve significantly during the next decade.

### See Also the Following Articles

Diabetes, Type 2 • Growth Hormone (GH) • Growth Hormone-Binding Proteins • Insulin-like Growth Factors

### Further Reading

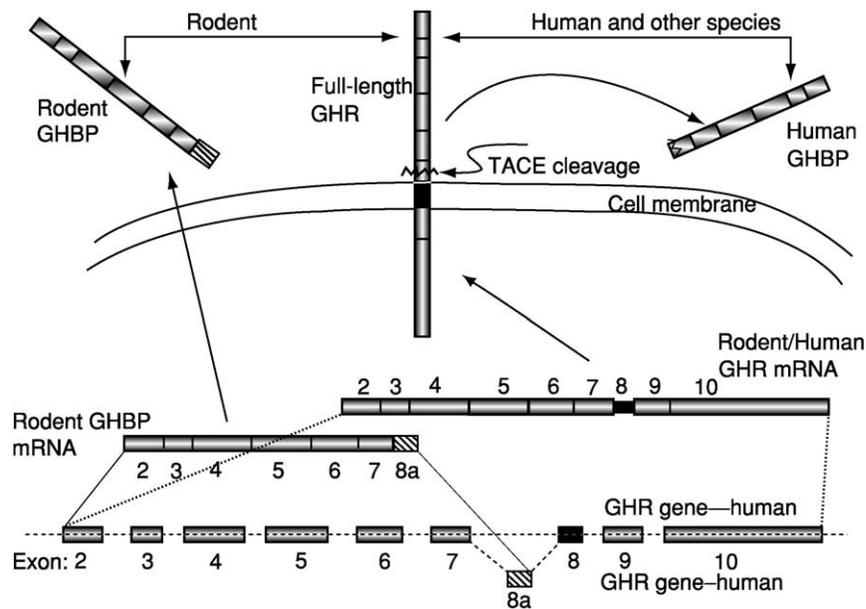
- Ayling, R. M., Ross, R., Towner, P., von Laue, S., Finidori, J., Moutoussamy, S., Buchanan, C. R., Clayton, P. E., and Norman, M. R. (1997). A dominant-negative mutation of the growth hormone receptor causes familial short stature. *Nat. Genet.* **16**, 13–14.
- Backeljaw, P. E., and Underwood, L. E., for the GHIS Collaborative Group. (2001). Therapy for 6.5–7.5 years with recombinant insulin-like growth factor I in children with growth hormone insensitivity syndrome: A clinical research center study. *J. Clin. Endocrinol. Metab.* **86**, 1504–1510.
- Kofoed, E. M., Hwa, V., Little, B., Woods, K. A., Buckway, C. K., Tsubacki, J., Pratt, K. L., Bezrodnik, L., Jasper, H., Tepper, A., Heinrich, J. J., and Rosenfeld, R. G. (2003). Growth hormone insensitivity associated with a STAT5b mutation. *New Engl. J. Med.* **349**, 1139–1147.
- Laron, Z. (1999). Laron syndrome primary growth hormone resistance. In “Contemporary Endocrinology: Hormone Resistance Syndromes” (J. L. Jameson, ed.), pp. 17–37. Humana Press, Totowa, NJ.
- Laron, Z., Pertzalan, A., and Mannheimer, S. (1966). Genetic pituitary dwarfism with high serum concentration of growth hormone: A new inborn error of metabolism? *Isr. J. Med. Sci.* **2**, 152–155.
- Ranke, M. B., Savage, M. O., Chatelain, P. G., Preece, M. A., Rosenfeld, R. G., and Wilton, P. (1999). Treatment of growth hormone insensitivity syndrome with insulin-like growth factor-I: Long-term results of the European Multicentre Study. *Horm. Res.* **51**, 128–134.
- Rosenbloom, A. L. (1999). IGF-I treatment of growth hormone insensitivity. In “The IGF System” (R. G. Rosenfeld and C. T. Roberts, eds.), pp. 739–770. Humana Press, Totowa, NJ.
- Rosenbloom, A. L., and Guevara-Aguirre, J. (1998). Lessons from the genetics of Laron syndrome. *Trends Endocrinol. Metab.* **9**, 276–283.
- Rosenfeld, R. G., Rosenbloom, A. L., and Guevara-Aguirre, J. (1994). Growth hormone (GH) insensitivity due to primary GH receptor deficiency. *Endocrine Rev.* **15**, 369–390.
- Woods, K. A., Dastot, F., Preece, M. A., Clark, A. J. L., Postel-Vinay, M.-C., Chatelain, P. G., Ranke, M. B., Rosenfeld, R. G., Amselem, S., and Savage, M. O. (1997). Phenotype–Genotype relationships in growth hormone insensitivity syndrome. *J. Clin. Endocrinol. Metab.* **82**, 3529–3535.
- Woods, K. A., and Savage, M. O. (2000). IGF deficiency. In “IGF in Health and Disease” (R. G. Rosenfeld and C. T. Roberts, eds.), pp. 651–674. Humana Press, Totowa, NJ.

## BIOCHEMICAL AND MOLECULAR CHARACTERIZATION

Binding studies using radioactively labeled GH coupled with a variety of chromatographic and immunological techniques have reported significant heterogeneity (number and size) in GHBP isoforms present in the sera of many different vertebrates. A number of these isoforms are of low abundance, low specificity, and (often) low affinity, and their physiological importance, if any, is currently unknown. By far, the major focus of attention has been on the characteristic high (nanomolar)-affinity GHBP that, in most species, has a molecular mass of approximately 60 kDa. In rodents, it appears to be smaller in size (35–45 kDa), probably due to lower levels of glycosylation. This high-affinity GHBP is, structurally, closely related to the extracellular GH-binding domain of the cellular receptor for GH (GHR) and is generated by two distinct, apparently species-specific, mechanisms (Fig. 1).

There is a single GHR gene in all species, and it consists of a number of coding exons (numbered 2–10) for which exon 2 represents the signal peptide, exons 3 to 7 represent the extracellular (ecto)domain (with exons 4–6 being involved in the GH-binding site), exon 8 represents the transmembrane-spanning domain, and exon 9 and part of exon 10 represent the

intracellular (cytoplasmic) signaling domain. In rodents, but no other species known to date, an additional small exon (8a) has been identified, situated between exons 7 and 8. Transcription of the GHR gene leads to a single primary mRNA transcript in humans and other nonrodent species, constituting exons 2 to 10. However, in rodents, because of alternative splicing mechanisms, two distinct mRNAs are produced. The first is comparable to that in other species (exons 2–10) and codes for the full-length GHR. The second is a truncated form that codes for a shorter protein, the GHBP, which has an identical extracellular GH-binding domain (exons 2–7) but no GHR transmembrane (hydrophobic) domain (exon 8) or intracellular signaling domain (exon 9 and part of exon 10). Instead, exons 8 to 10 are replaced with exon 8a, which codes for a short 17- to 25-amino acid hydrophilic sequence, commonly referred to as the “hydrophilic tail.” The absence of the hydrophobic transmembrane domain and the substitution of the hydrophilic tail provide a mechanism for the rodent GHBP to be secreted directly from the cell of synthesis. No equivalent mechanism has been identified in human or other species despite considerable effort on the part of a number of researchers. There have been various alternatively spliced GHR mRNA transcripts reported in human cells, some of which lead to truncated membrane-bound GHR isoforms (e.g., exon 3



**Figure 1** Diagrammatic representation of the GHR gene and the alternative mechanisms for generation of full-length GHR and the soluble truncated GHBP in both rodents and humans (and other species). The rodent GHR gene differs from the human gene essentially due to inclusion of an additional coding exon (8a, ▨). The classical transmembrane domain of the full-length GHR is denoted by exon 8 (■).

deleted forms and exon 9/10 truncations), but none of these leads to direct expression of a unique GHBP mRNA.

Given the absence of a specific GHBP mRNA in humans and species other than rodents, a quite distinct mechanism has been identified to account for the production of GHBP in these species. The GHBP is produced as a result of a specific posttranslational proteolytic cleavage of the ectodomain of the cell membrane-embedded GHR. There is some evidence that cleavage may be influenced by intracellular mechanisms such as endosomal proteases, but generally it is thought to be an extracellular event. Cleavage occurs through the action of one or more metalloproteinases, and recent evidence indicates that TNF $\alpha$ -converting enzyme (TACE), the enzyme responsible for membrane cleavage of the growth factor TNF $\alpha$ , is also involved in GHBP generation. TACE (also known as ADAM-17) is a member of the large ADAM family of zinc-dependent metalloproteinases or "sheddas." The cleavage site in the extracellular domain of the GHR, and therefore the precise C-terminal sequence of the GHBP, is currently not known, but it is almost certainly very close to the transmembrane insertion site.

Like the GHR, the human GHBP is quite heavily glycosylated, although there is little evidence to suggest that glycosylation is necessary for GH binding. The GHBP is immunologically identical to the extracellular domain of the GHR and so can be immunoprecipitated and/or detected using N-terminally directed GHR antisera. Therefore, several immunologically based assay systems have been developed for the direct measurement of both "free" (unoccupied by GH) and "total" human GHBP.

Based on studies primarily in rodents and rabbits, it is known that most epithelial, endothelial, and mesenchymal cell types express the GHR and also produce GHBP. Although first identified and primarily studied in serum, the GHBP is found in secretions other than serum such as milk, urine, and follicular fluid. Liver is perhaps the major source of circulating GHBP, but tissues such as adipose, kidney, placenta, breast, ovary, brain, muscle, and bone all are significant producers. In adipose tissue, significant amounts of GHBP remain attached via an unknown linkage mechanism to the adipocyte cell membrane. The functional implications of this are not clear but may facilitate modulation of GH action on adipose tissue. A major store of GHBP is also found within the cytosolic compartment of several rodent (and rabbit) tissues, and this may arise from *de novo* synthesis and/or via internalization and intracellular processing of

externalized GHR/GHBP. Of particular interest is the observation that GHBP has also been detected, by direct binding studies and immunohistological examination, in the nucleus of several cell types. The mechanism for nuclear localization is unknown, but it is likely to be via facilitated transport in association with another nuclear-targeting protein. The GHBP itself has no recognizable nuclear localization motif. Nonetheless, there is emerging evidence to suggest that the GHBP may have the capacity to bind to nuclear DNA and thereby influence nuclear transcriptional activity. If this is proven, it will revolutionize our thinking about the potential physiological role and importance of the GHBP.

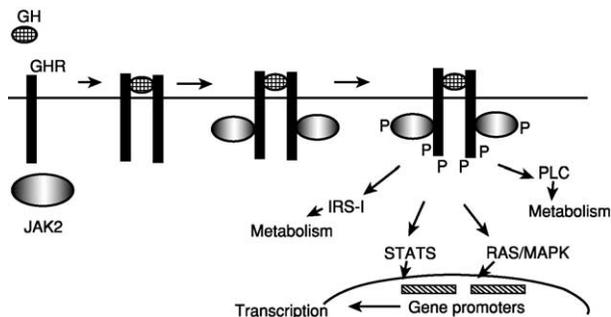
## FUNCTIONAL ROLES AND IMPLICATIONS FOR GH ACTION

In addition to the hypothesis proposed earlier regarding potential nuclear roles for GHBP, there is strong evidence to indicate that the GHBP is functionally important and has quite varied roles. Indirect evidence for physiological relevance and roles distinct from those of the GHR comes from the now quite extensive work demonstrating in rodents, at least, that there is differential regulation of the GHBP and GHR mRNA expression. Given that both are independently synthesized from distinct mRNAs, the differential regulation is not surprising. Several studies primarily, but not exclusively, examining hepatic GHBP production have indicated its sex, pregnancy, GH, steroid, and ontogenic dependence. In humans, of course, such definitive studies are more difficult to conduct (due to tissue availability) and interpret because there is a much closer link between regulation and, therefore, levels of GHBP and its parent or precursor molecule, the GHR. Indeed, one of the indicated advantages of being able to readily measure GHBP in serum is that it may well reflect closely the status (abundance and binding affinity/capacity) of the generally nonaccessible human tissue GHR. In this context, numerous studies have measured the levels of GHBP in various physiological and pathophysiological states and have drawn conclusions about GHR status and the likely impact on tissue GH responsiveness. However, it is important to keep in mind that indirect regulatory mechanisms, such as those that modulate abundance and/or activity of the GHR cleavage enzyme TACE, may lead to differential regulation of tissue GHR and soluble GHBP in humans.

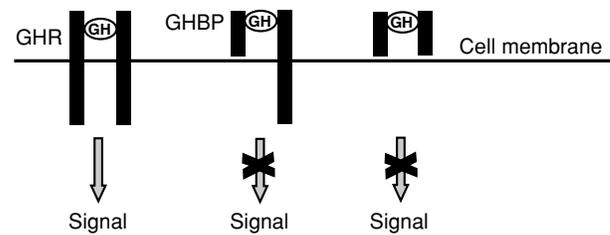
The roles of GHBP can be classified into two primary groups: endocrine roles in the circulation

and paracrine/autocrine roles in the immediate extracellular environment of GH-responsive cells. In the circulation, the high-affinity GHBP binds approximately 50% of circulating GH under normal conditions. There is significant unoccupied GHBP in the serum under these conditions, but this can be readily saturated in the face of high levels of GH such as in acromegaly. The effect of binding GH in serum is to increase the half-life of circulating GH via modulation of GH clearance and/or degradation and to influence its capacity for transcapillary transfer out of the vasculature and into the extracellular environment of target cells. Both of these roles would have a subsequent influence on the responsiveness of tissue to GH by regulating the availability of GH to its target cell receptors. Overall, the GHBP provides a circulating reservoir of GH that can serve as a readily releasable pool of GH in the face of diminished supply. Hence, one would expect the endocrine effects of GHBP to be a positive regulator of GH action.

However, the GHBP is also known to have a distinct and influential paracrine/autocrine role in the immediate extracellular environment of secreting cells. Studies *in vitro* have shown that addition of exogenous purified GHBP to cells in culture can inhibit the effects of added GH on those cells. This may be due in part to binding and sequestration of GH away from interaction with the cell's GHR. On the other hand, an additional and more important explanation is now known as a result of detailed understanding of the mechanisms involved in GH activation of its tissue GHR (Fig. 2). When GH binds to its receptor, it results in dimerization of the



**Figure 2** Diagrammatic representation of the sequential GH-induced dimerization of the GHR, the recruitment of the cytoplasmic tyrosine kinase JAK2, the activation of GHR through phosphorylation reactions, and the initiation of various cytoplasmic signaling pathways. For GH, the primary signaling pathways involve phosphorylation cascades using the classical phospholipase C (PLC), insulin receptor substrate-1 (IRS-1), signal transducer and activator of transcription (STAT), and the small G-protein RAS/mitogen activated protein kinase (MAPK) pathways.



**Figure 3** Models for the formation of homo- and heterodimers resulting from the GH-induced dimerization of mixed populations of full-length membrane-bound GHR and truncated soluble GHBP. Based on experimental evidence, only the GHR homodimer is likely to initiate GH-induced intracellular signaling reactions.

GHR in a  $1 \times \text{GH}:2 \times \text{GHR}$  stoichiometry. Dimerization is essential for the activation of the GHR. This activation occurs through the recruitment of the intracellular tyrosine kinase JAK2, the cross-phosphorylation of the twin cytoplasmic domains of the GHR dimer, and the subsequent docking and activation of additional SH2- and SH3-containing cytoplasmic kinases, which then initiate an intracellular phosphorylation cascade leading to expression of a multitude of GH actions. In the presence of GHBP, which binds GH but clearly lacks any cytoplasmic "activation" domains, the obvious possibility exists that a variety of homo- and hetero-GHR/GHBP dimers may form (Fig. 3), depending on the relative available concentrations of each component. Formation of GHBP homodimers or GHBP/GHR heterodimers, with no or only one GHR cytoplasmic domain present, will not be sufficient to recruit and activate JAK2. Hence, one would predict a reduction or blockade of GH action in these circumstances. Experimental evidence indeed shows that overexpression of GHBP in wild-type, GHR-containing cell lines can act as a dominant negative influence on GH action due to the formation of inactive GHBP:GH:GHR (short:long) heterodimers. Thus, the GHBP adds another dimension to the already complex and varied mechanisms employed in the regulation of GH action.

## RELATIONSHIP OF GHBP WITH GH RESPONSIVENESS IN HUMAN DISEASE STATES

As indicated previously, in humans, the general consensus is that circulating levels and binding ability of GHBP reflect the abundance and function of the tissue GHR. In turn, this often reflects the degree of responsiveness or resistance to GH action in particular clinical settings.

GH treatment is now used widely in a variety of clinical conditions, and numerous studies have investigated the possible predictive nature of GHBP with respect to correlation with GH responsiveness. In GH deficiency, there is a general positive correlation between serum GHBP and GH response as measured by change in growth rate or, acutely, changes in IGF-I levels. Growth responses to GH therapy may be influenced by the frequency of GH injection such that more frequent, smaller doses (mimicking the natural pulsatility of GH secretion) lead to better growth responses than do larger, more infrequent doses. A few studies have examined in some detail whether this increased GH response is associated with a corresponding change in GHBP based on the knowledge that, at least in rodents, GHBP expression is regulated by GH. However, there is little consensus as to whether the increased growth responses due to the mode of GH delivery are reflected or predicted by similar changes in serum GHBP activity.

Children with idiopathic short stature often have slightly reduced levels of GHBP, reflecting a subtle endogenous resistance to GH. Similarly, African pygmies also show reduced GHBP together with apparent GH resistance during puberty. However, perhaps the most clear-cut correlation between GHBP and GH responsiveness occurs in cases of GH insensitivity syndrome (GHIS or Laron dwarfism). These children are severely growth retarded due to genetic mutations in the GHR. Most are completely resistant to GH treatment. The majority of mutations occur in the extracellular binding domain of the GHR and prevent the binding of GH to the receptor. Hence, the same defect occurs within the proteolytically cleaved GHBP, which exhibits complete failure to bind GH and so is undetectable by “functional” assays, although the inactive protein can be detected immunologically in many GHIS sera. In a few reported cases of GHIS, however, active GHBP was detected in serum, and this led researchers to examine and identify additional mutations in the GHR that occurred not in the GH-binding domain but rather in other domains such as the cytoplasmic signaling domain and the region required for stabilization of receptor dimerization. Thus, although these GHRs and the GHBP would still bind GH normally, the receptor itself would be nonfunctional and would result in severe GH resistance.

There are a host of other pathophysiological states that exhibit various degrees of GH responsiveness. With respect to decreased responsiveness, such states include metabolic disturbances such as malnutrition

and poorly controlled diabetes as well as thyroid, renal, and liver disease. In each of these cases, there are many studies indicating that levels of GHBP are low, consistent with the degree of GH resistance. Conversely, in obesity, one of the rather less common situations of increased GH responsiveness, GHBP levels are reported to be significantly higher than normal. Again, each of these examples supports the argument that the levels of circulating GHBP in general reflect the abundance and activity of the tissue GHR and, hence, GH responsiveness.

This is not always the case, however, as illustrated by the rare GHIS example cited earlier and perhaps by observations in pregnancy. Although pregnancy in the rat shows a quite dramatic increase in expression of the GHBP, the situation is somewhat less marked and more varied in human pregnancy. Indeed, a rather complex biphasic effect on maternal GHBP levels is seen during human pregnancy—an initial rise followed by a decrease as gestational age advances. This pattern is particularly in contrast to that of placental GH, which shows a rather dramatic rise across the final two trimesters. The precise role of placental GH during pregnancy is not known; however, it is likely to play a significant metabolic role in the delivery and regulation of nutrients and metabolites to the placenta and the growing fetus, and in this context one might expect the sensitivity of the system to GH to rise. It is not known whether the decrease in maternal GHBP reflects a concomitant change in the expression and activity levels of the placental GHR; however, this may represent one situation, compounded by the complex milieu of interacting systems (e.g., maternal, fetal, placental) and pregnancy-related regulatory factors, where a tight linkage between GHBP and GHR expression is lost.

Clearly, there is still much to learn about the clinical relevance of the GHBP and whether it is a useful and effective marker of tissue GHR status in humans. Nonetheless, evidence to date suggests that the GHBP has potentially very important physiological sequelae with respect to GH action in both normal and abnormal clinical conditions. Continued exploration of the endocrine and paracrine/autocrine effects of GHBP is paramount to gaining a full understanding of the complex and diverse mechanisms involved in regulating GH action *in vivo*.

### See Also the Following Articles

Growth Hormone Deficiency, Genetic • Growth Hormone (GH) • Growth Hormone Insensitivity • Insulin-like Growth Factors

## Further Reading

- Amit, T., Youdim, M. B., and Hochberg, Z. (2000). Clinical review 112: Does serum growth hormone (GH) binding protein reflect human GH receptor function? *J. Clin. Endocrinol. Metab.* **85**, 927–932.
- Baumann, G. (2001). Growth hormone binding protein 2001. *J. Pediatr. Endocrinol. Metab.* **14**, 355–375.
- Baxter, R. C. (2000). Insulin-like growth factor (IGF)-binding proteins: Interactions with IGFs and intrinsic bioactivities. *Am. J. Physiol. Endocrinol. Metab.* **278**, E967–E976.
- McIntyre, H. D., Serek, R., Crane, D. I., Veveris-Lowe, T., Parry, A., Johnson, S., Leung, K. C., Ho, K. K., Bougoussa, M., Hennen, G., Igout, A., Chan, F. Y., Cowley, D., Cotterill, A., and Barnard, R. (2000). Placental growth hormone (GH), GH-binding protein, and insulin-like growth factor axis in normal, growth-retarded, and diabetic pregnancies: Correlations with fetal growth. *J. Clin. Endocrinol. Metab.* **85**, 1143–1150.
- Ross, R. J. (2000). GH, IGF-I, and binding proteins in altered nutritional states. *Int. J. Obes. Related Metab. Disorders* **24**(Suppl 2), S92–S95.
- Zhang, Y., Jiang, J., Black, R. A., Baumann, G., and Frank, S. J. (2000). Tumor necrosis factor-alpha converting enzyme (TACE) is a growth hormone binding protein (GHBP) sheddase: The metalloprotease TACE/ADAM-17 is critical for (PMA-induced) GH receptor proteolysis and GHBP generation. *Endocrinology* **141**, 4342–4348.
- Zhu, T., Goh, E. L., Graichen, R., Ling, L., and Lobie, P. E. (2001). Signal transduction via the growth hormone receptor. *Cell Signalling* **13**, 599–616.

Many of the hormonal components of growth are controlled via the hypothalamic–pituitary axis. Thus, the secretion of GH, leading to the secretion of IGFs, is under the control of hypothalamic neuropeptide hormones known as growth hormone-releasing factor (GRF) and somatostatin or somatotropin release-inhibiting factor (SRIF). Similarly, the secretion of thyrotropin-simulating hormone (TSH), follicle-stimulating hormone (FSH), luteinizing hormone (LH), and adrenocorticotrophic hormone (ACTH), all of which play major roles in the control of growth, is itself under the control of hypothalamic neurohormones. In addition, the secretion of insulin, which is not directly under pituitary control, also plays an important role in the control of cellular growth. Of course, many nonhormonal factors play major roles in the control of growth. Foremost among these is nutrition given that no cell could grow or divide without substrates. There are also underlying genetic factors that determine the potential for growth, not only for the whole organism but also for each organ and tissue. In addition, there are a host of tissue growth factors, such as epidermal growth factor (EGF) and nerve growth factor (NGF), whose roles in the overall control of growth are only beginning to be understood. The interaction of these hormonal and nonhormonal factors leads to the normal growth and development of children, and abnormalities of these mechanisms lead to disorders in growth that present themselves to a physician. A thorough understanding of the factors controlling growth is crucial for a rational approach to the diagnosis and management of growth disorders.

## Growth Hormone

GH is a 191-amino acid polypeptide secreted by special cells in the anterior pituitary gland called somatotropes. Neurohormones control these cells through the hypophyseal portal system, and both positive effectors of GH secretion and an inhibitor of GH secretion exist. Current data indicate that the combination of GRF and ghrelin (growth hormone-releasing peptide [GHRP]) is responsible for the rapid pulses of GH secretion seen mainly during the nighttime hours and in response to meals and exercise, whereas the inhibitor of GH secretion, somatostatin (SRIF), appears to be responsible for the underlying tone of GH secretion. There are two major forms of GH synthesized, stored, and secreted by the somatotropes. The most abundant form (approximately 90%) is the 22-kDa molecular weight form. In addition, there is an alternative splicing of the GH messenger RNA

that leads to a minor 20-kDa form of GH whose physiological role is unknown.

The secretion of GH is followed by a complex chain of events that ultimately leads to the stimulation of growth at the endocrine level. After GH is released into the bloodstream, a large proportion of it binds to a specific GH-binding protein. It is now clear that the human GH-binding protein represents the extracellular portion of the GH receptor site. Like many receptors for polypeptide hormones, the receptor site for GH consists of three polypeptide domains: an extracellular domain that contains the three-dimensional structure necessary for the recognition of the hormone, a transmembrane domain that is highly hydrophobic, and an intracellular domain that contains the three-dimensional structure necessary to lead to the biological action within the cell. In humans, it appears that there is proteolytic cleavage of the extracellular domain of the GH receptor that leads to a soluble circulating polypeptide that plays a major role in protein binding of GH in the serum. The exact biological role of this binding protein is unclear; however, it does serve as a useful index to the level of GH receptors in the body. GH receptors are widely distributed throughout tissues in the body, and it is the binding of the GH ligand to these GH receptors that leads to the biological events within the cells themselves. Of the many direct biological events that are ascribed to GH action, one of the most important ones is clearly control of the production of yet another polypeptide hormone known as IGF-1.

## Insulin-like Growth Factors

IGF mRNA is widely distributed throughout the body and is under GH control not only in the liver but also in other tissues. IGFs circulate tightly bound to proteins known as insulin-like growth factor-binding proteins (IGFBPs). In most normal circumstances, there are few or no free IGF polypeptides circulating. It is now known that a total of six distinct but homologous IGFBPs exist. By far the most abundant form of IGFBP in normal serum is IGFBP-3. This is a glycosylated protein of approximately 40 kDa molecular weight. The combined IGF-peptide/IGFBP-3 complex binds to yet another protein known as the acid labile subunit (ALS) to form a 150-kDa, three-subunit protein complex that is the major circulating form of the IGFs. This 150-kDa complex contains well over 90% of the total IGF in the serum in most circumstances. A smaller proportion of the IGFs in serum circulate bound to the other forms of IGFBPs (IGFBP-1,-2,-4,-5, and-6). In addition, these other

IGFBPs are major components in certain bodily fluids such as IGFBP-1 in amniotic fluid, IGFBP-2 in joint fluid, and IGFBP-5 in cerebrospinal fluid. The exact biological purpose of this complex system of controlling the amount and distribution of IGF in the serum and extracellular fluid spaces is unclear. However, it does result in the prolongation of the circulating time of IGFs in plasma and may play a major role in modulating the delivery of the IGF peptides to the receptor sites on cells. Most of the observed biological actions of GH on the control of growth appear to be subserved by the chain of action initiated by GH. This leads to the production of IGF-1 and its delivery to the IGF type 1 receptor sites in the tissues. It is this chain of action that has been most clearly linked to the anabolic and growth-promoting events that we associate with GH in the whole organism. Any abnormality in this complex chain of events can lead to growth failure.

### Thyroid Hormone

Thyroid hormone also plays an extremely important role in the control of growth. Like the GH-IGF axis, there is a hypothalamic mechanism of control with the neurohormone thyrotropin-releasing hormone (TRH) being secreted by specialized neurons into the hypophyseal portal system leading to the release of the pituitary hormone TSH. TSH then leads to the production and release of the final hormone in the chain: thyroid hormone (thyroxine [ $T_4$ ] and triiodothyronine [ $T_3$ ]). Like the IGFs, the thyroid hormones are nearly totally bound to plasma proteins that control the half-life and tissue delivery of the thyroid hormones. The thyroid hormone receptor site that is widely distributed throughout cells in the body subserves thyroid hormone action. This receptor site is located intracellularly, similar to that of steroid hormones, and the ligand- $T_3$  complex binds to chromatin and DNA to initiate mRNA synthesis. Without thyroid hormone action through its receptor sites, GH and IGFs are unable to stimulate anabolic and growth responses. Furthermore, there is a close interaction of thyroid hormone with GH secretion. In the presence of hypothyroidism, GH secretion from the pituitary gland is decreased in response to both pharmacological and physiological stimulation. Therefore, the normality of thyroid hormone secretion must be proven before any attempt to assess GH secretion in a short patient.

### Sex Hormones

The sex hormones, estrogens and androgens, play an extremely important role in the control of growth

during puberty. The secretion of these hormones is relatively low after the fetal and perinatal periods of life until there is an increase in the gonadotropins (LH and FSH) prior to the clinical onset of puberty. It is the increased secretion of estrogens and androgens under the control of the pituitary gonadotropins that leads to the development of the secondary sex characteristics that are so closely associated with the events of puberty. Estrogens and androgens have been associated with growth-stimulating actions directly as well as indirectly. Both androgens and estrogens have been shown to influence the secretion of GH at the time of puberty. This increase in GH secretion at the time of puberty is thought to play a major part in the pubertal growth spurt. In addition to the direct and indirect roles that androgens and estrogens play in the stimulation of growth, estrogens play a major role in the maturation of bones and the ultimate disappearance of the epiphyseal centers. It is this event that results in the cessation of linear growth at the end of puberty. Like the thyroid hormones, the sex hormones appear to act by diffusing into the cell and binding to the specific receptor sites of the steroid/TSH category. The binding of androgens and estrogens to their respective receptor sites initiates changes in the three-dimensional structure that allow the interaction of these activated receptors with chromatin and the chromosomal DNA. This interaction leads to the production of specific mRNA and the secretion of specific proteins, leading to the biological action observed. The evaluation of sex hormone secretion plays an important part in the evaluation of disorders of size, particularly during the peripubertal period.

### Adrenal Hormones

In addition to the other hormonal factors reviewed, it is clear that adrenal hormones play a role in the control of growth. Adrenal androgens appear to have an important adjunctive role in the events of puberty and act in concert with androgens from the testes and ovary to lead to growth stimulation, by both direct and indirect means, and the maturation of the bones. In addition, the adrenal glucocorticoids appear to have a bimodal action in the control of growth. A minimal level of glucocorticoids is necessary for cells to function, grow, and divide. On the other hand, many observations have shown that even a slight excess of glucocorticoids makes cells in the body unresponsive to the other hormonal growth-stimulating agents and leads to slowing of linear growth.

## Other Factors

There are many extremely important nonhumoral factors controlling growth. Without adequate food intake, and therefore without adequate substrates for cells, no effective growth can occur. This is clearly seen in the clinical disorders of kwashiorkor and marasmus as well as in more subtle clinical disorders in nutrition. Genetics also has an extremely important role in determining the potential for growth. The differences in stature among humans depend to a large extent on differing genetic constitutions. The best, but a relatively crude, way in which to determine these underlying genetic factors is parental and familial heights. Tissue growth factors such as EGF play a critical role in the control of growth in the organism. This is best illustrated by achondroplasia, which is now known to be caused by a defect in the gene for the EGF receptor. Finally, the role of systemic disease in the growth of children and adolescents cannot be overemphasized. Even in the presence of normal hormonal, nutritional, and genetic factors, underlying systemic disease such as inflammatory bowel disease can interfere with growth potential. The existence of short stature in a patient is an important clue to the physician as to the underlying condition of the patient.

## SECULAR TRENDS IN MATURATION AND FINAL HEIGHT

The expected range of the growth rate for infants and children, timing of puberty, and final adult height is commonly thought to be immutable. However, an examination of the past history of growth and development shows that there have been very large variations in the rate of “normal” growth, puberty, and final height that must have been due to environmental changes, not genetics. These are known as the secular trends. There is a large body of data establishing that in many regions of the world, the average age of menses for women has decreased markedly over the past two centuries at the same time as the adult height of women has been increasing. In addition to these data in females, there are less well-known data in boys, including data as disparate as those from a royal boys school in 18th-century Germany and height records of male slaves in the American South in the 19th century, illustrating that a similar process of earlier puberty and taller final stature has occurred over the past two centuries in men. The environmental reasons for this appear to be a combination of better nutrition and sanitation, particularly during infancy, and control of infectious diseases. One of the most remarkable

illustrations of a secular trend occurred in Japan after World War II. Coincident with the introduction of more protein and calories in the Japanese diet, especially during infancy, the average adult height of Japanese men has increased by more than 25 cm, and a similar phenomenon has occurred in Japanese women. This fact can be vividly illustrated by a ride on a Tokyo subway, where often the older passengers are much shorter than the younger passengers. In some areas of the world such as the United States, this process has slowed down or even ceased. In other parts of the world, the secular trend in growth and development is continuing today, or in some areas may even be going backward toward slower maturation and shorter adult height, due to deteriorating health and nutrition. The absolute worst time to grow up in England, as judged by the height of army recruits, was during the early stages of the Industrial Revolution and the Napoleonic Wars. Children we now consider to have CDG would have been average or even tall in another time and place. The syndrome of constitutional delay is the perfect example of the relativity of human development; it depends on the standard of comparison. Most of the children who present with CGD do not have an underlying hormonal or genetic disease. However, hidden within this large number of the basically normal children are a few who have clinically important disorders that the physician must search for diligently.

## DISTINCTION FROM PATHOLOGICAL CAUSES OF SHORT STATURE

The pathological causes of short stature are outlined in [Table I](#). Clinical abnormalities of GH secretion or action are clearly associated with short stature. The most apparent of these is GH deficiency, in which there is either a disorder of the hypothalamic control of GH secretion or an inability of the pituitary itself to secrete GH. The consequences of classical GH deficiency are reflected not only in low levels of circulating GH, as assessed by both physiological and pharmacological stimulation, but also in extremely low levels of IGF-1 and a consequent decrease in the growth rate. These children almost invariably respond to GH treatment with a marked increase in growth rates. An interesting but rare disorder of the chain of GH action is Laron syndrome (also known as GH insensitivity syndrome). This disorder, originally described in Ashkenazi Jews, is now known to be more widely distributed both ethnically and geographically. Laron syndrome is a group of genetic disorders of the

**Table 1** Causes of Short Stature

Non-pathological short stature
Constitutional delay
Familial
Nutritional
GH related causes
Growth hormone deficiency
Growth hormone resistance syndrome (Laron syndrome)
Hypothyroidism
Sex hormone related causes
Delayed puberty
Hypogonadotropic hypogonadism (Kallmann's syndrome)
Glucocorticoid excess
Cushing's disease
Pharmacological administration
Genetic causes
1. Chromosomal
Turner syndrome (XO and variants)
Down syndrome (21 trisomy)
2. Syndromes
Pseudohypoparathyroidism
Prader-Willi syndrome
Lawrence-Moon-Biedl syndrome
Skeletal dysplasias
Miscellaneous syndromes (e.g., Russell-Silver, Seckel)

GH receptor that make the GH receptor unable to bind GH. This results in an increase in circulating GH levels (due to lack of IGF feedback), a decrease in the level of circulating GH-binding protein derived from the GH receptor, markedly decreased levels of IGF-1, and extremely poor growth rates. Children with Laron syndrome will respond to treatment with synthetic IGF-1 with an increased rate of growth, thereby validating the somatomedin hypothesis. It is possible that more subtle abnormalities of the GH-binding protein or GH receptor are underlying causes in some of the undiagnosed cases of poor growth.

Because of the important role of thyroid hormone in both the secretion and endocrine actions of GH, it is clear that any level of hypothyroidism would lead to a decrease in growth rate and eventually to short stature. There is a remarkable decrease in growth rate seen with severe hypothyroidism, and a rapid increase in growth rate is seen after thyroid therapy is instituted. Some patients with severe hypothyroidism will actually present with a peculiar form of precocious puberty known as "overlap syndrome," in which there is a premature activation of gonadotropins apparently associated with the high levels of TRH secretion by the hypothalamus. On the other hand, the treatment for hypothyroidism can lead to the rapid onset of true

puberty, which in turn leads to an accelerated advance in bone age and may compromise final height despite the institution of thyroid therapy.

Short stature relative to that expected for age can also occur as a result of inadequate sex hormone secretion. This is most commonly seen in boys who have delayed puberty, but some girls will also have a significant delay in puberty. Treatment with short courses of androgen has been used in boys with severely delayed puberty. Many times, this delay in puberty is physiological; however, it can also be associated with hypogonadotropic hypogonadism in both boys and girls. These patients may present at the early pubertal age with short stature that is due to the lack of sex hormone secretion and pubertal growth spurt, even though their final adult height might not be short. A relatively common form of hypogonadotropic hypogonadism in both boys and girls is Kallmann's syndrome, in which the hypogonadotropic hypogonadism is associated with anosmia. These patients can be treated with androgens or estrogens as appropriate.

Short stature can also be associated with hypersecretion of glucocorticoids caused by an increase in ACTH secretion (Cushing's disease) or the hypersecretion of glucocorticoids by a functioning adrenal tumor. These patients need to be treated by the control of their excess glucocorticoid secretion. The most common cause of glucocorticoid excess leading to growth failure is in association with the treatment of steroid-responsive diseases with pharmacological doses of glucocorticoids. In these patients, the obvious treatment of decreasing or stopping the glucocorticoid therapy may in practice be very difficult to accomplish because of exacerbation of the underlying disease state.

In girls, the most common genetic disorder leading to short stature is Turner syndrome. The mechanism by which Turner syndrome leads to short stature is not yet clear, but it does not appear to be directly GH related. These girls have relatively poor growth rates during childhood, leading to a progressive decrease in their average height when compared with normal girls of the same age. In addition, they fail to enter puberty due to the ovarian failure associated with Turner syndrome and, therefore, achieve very short adult stature with an average height of approximately 140 cm. The use of GH treatment has seen some success in increasing growth rates and final height in girls with Turner syndrome. Although the vast majority of these children will have delayed puberty, a significant percentage of them do have enough estrogen secretion from their ovaries to develop secondary sex characteristics and even to menstruate. Thus, any

girl who is pathologically short with no other established diagnosis should always obtain chromosomes, irrespective of the presence or absence of secondary sex characteristics. Other genetic causes of severe short stature include Down syndrome, pseudohypoparathyroidism, Lawrence–Moon–Biedl syndrome, Weaver syndrome, and Prader–Willi syndrome.

## MANAGEMENT

For the patient with CDG, the following points are important to keep in mind:

1. The diagnosis is one of exclusion of other problems.
2. There is often a familial history of this growth pattern.
3. The child has short stature, but normal growth velocity after 3 years of age, and absent or early sexual development.
4. There is modest bone age delay (within 75% of the chronological age).
5. After careful investigation, it is determined that the child has an absence of systemic or endocrine disease.

After establishing the diagnosis of CDG by excluding pathological causes of short stature, no pharmacological treatment is usually necessary. The parents and patient should be thoroughly educated about normal patterns of growth and development and should be given realistic projections of the expected range of adult height. The patient should be followed for assessment of the growth rate and pubertal development. If the growth rate is abnormally low or if

puberty is markedly delayed, more detailed biochemical and radiological investigations may become indicated. In some instances in CDG, pubertal development may be so late that it becomes a major psychological problem to the patient, and so giving hormonal therapy to induce the onset of puberty may become indicated.

## See Also the Following Articles

Body Composition During Growth • Body Proportions • Growth and Chronic Disease • Growth Hormone (GH) • Growth Hormone Insensitivity • Insulin-like Growth Factors • Intrauterine Growth Retardation • Postnatal Normal Growth and Its Endocrine Regulation • Puberty: Physical Activity and Growth • Skeletal Development During Childhood and Adolescence • Turner Syndrome

## Further Reading

- Bideci, A., Cinaz, P., Hasanoglu, A., and Tumer, L. (2002). Leptin, insulin-like growth factor (IGF)-I, and IGF binding protein-3 levels in children with constitutional delay of growth. *J. Pediatr. Endocrinol. Metab.* **15**, 41–46.
- Bierich, J. R. (1992). Constitutional delay of growth and adolescence. *Baillieres Clin. Endocrinol. Metab.* **6**, 573–588.
- Hintz, R. L. (2001). Disorders of growth. In “Pediatric and Adolescent Gynecology” (P. Lee and J. Sanfilippo, eds.), 2nd ed., pp. 41–50. W. B. Saunders, Philadelphia.
- Hintz, R. L. (2001). Disorders of size and shape. In “Clinical Paediatric Endocrinology” (C. G. D. Brook and P. C. Hindmarsh, eds.), 4th ed., pp. 124–139. Blackwell, London.
- Hintz, R. L. (2001). Final height prediction in constitutional growth delay. *J. Pediatr. Endocrinol. Metab.* **14**, 1535–1540.
- Horner, J. M., Thorsson, A. V., and Hintz, R. L. (1978). Growth deceleration patterns in constitutional short stature: An aid to diagnosis. *Pediatrics* **62**, 529–534.

**Table I** Ligands, Human Gene Loci, Human and Mouse Gene Symbols, and Tissue Expression of One Transmembrane-Segment Forms of Guanylyl Cyclase

Cyclase	Ligand(s)	Human locus	Human (mouse) gene symbol	Expression in normal tissues <sup>a</sup>
GC-A (NPR-A)	ANP BNP	1q21-q22	<i>NPR1</i> ( <i>Npr1</i> )	Cerebrum, cerebellum, brainstem, pituitary gland, ocular ciliary body, retina, chochlea, thymus, <u>lung</u> , <u>heart</u> , <u>aorta</u> , <u>kidney</u> , <u>adrenal gland</u> , liver, spleen, small intestine, colon, <u>adipose tissue</u> , testis, ovary, uterus, placenta
GC-B (NPR-B)	CNP	9p12-p21	<i>NPR2</i> ( <i>Npr2</i> )	Cerebrum, <u>cerebellum</u> , brainstem, <u>pituitary gland</u> , <u>pineal gland</u> , ocular ciliary epithelium, thymus, trachea, <u>lung</u> , aorta, heart, kidney, adrenal gland, liver, spleen, stomach, small intestine, colon, skeletal muscle, <u>bone growth plate</u> , testis, ovary, <u>oviduct</u> , uterus, placenta
GC-C	STa Guanylin Uroguanylin Lymphoguanylin	12p12	<i>GUCY2C</i> ( <i>Gucy2c</i> )	<u>Small intestine</u> , <u>colon</u> , kidney
GC-D	?	11p15.4 or 11q13-q14.1 <sup>b</sup>	( <i>Gucy2d</i> )	Olfactory neuroepithelium
GC-E	?	17p13.1	<i>GUCY2D</i> ( <i>Gucy2e</i> )	Retina, pineal gland
GC-F	?	Xq22	<i>GUCY2F</i> ( <i>Gucy2f</i> )	Retina
GC-G	?	10q24-q26		Lung, skeletal muscle, kidney, small intestine

Note. Underlines indicate high expression of the gene.

<sup>a</sup>Determined by Northern blotting or RT-PCR.

<sup>b</sup>Expected by syteny between mouse and human chromosomes.

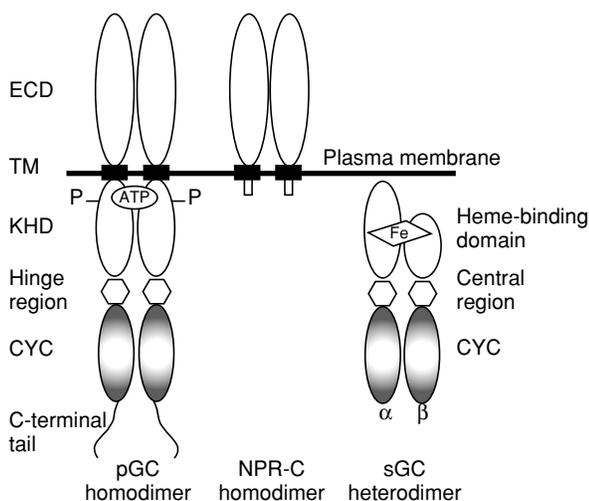
buried within the GC-A ECD, consistent with observations that ANP binding to GC-A appears to depend on Cl<sup>-</sup> concentration.

The KHD has substantial amino acid sequence identity to the catalytic domain of protein kinases, but no intrinsic protein kinase activity has been

reported. The KHD has been proposed to function as an inhibitory regulator of the CYC given that deletion of the KHD from GC-A and GC-B renders constitutively active enzymes. GC-A and GC-B KHDs contain adenosine triphosphate (ATP)-binding motif-like sequences (GxGxxxG and LxGxxxG, respectively), and it has been suggested that ATP binding to these sequences is essential for ligand-induced activation of GC-A and GC-B.

Two other potential subdomains have been defined in pGCs: the hinge region and the C-terminal tail (Fig. 1). The hinge region is a putative, amphipathic  $\alpha$ -helical region connecting the KHD and the CYC, and it has been suggested to function in dimerization of pGCs. GC-C, GC-D, GC-E, and GC-F have extended C-terminal tails. GC-C, having the longest tail, appears to associate with the cytoskeleton through this subdomain.

Soluble GCs possess two subunits,  $\alpha$  and  $\beta$ , and two isoforms have been identified for each subunit:  $\alpha_1$ ,  $\alpha_2$ ,  $\beta_1$ , and  $\beta_2$ . In the human, two subunits named  $\alpha_3$  and  $\beta_3$  are actually the orthologues of rat  $\alpha_1$  and  $\beta_1$  (Table II). Soluble GCs contain a heme-binding domain within the N terminus and the CYC near the C terminus (Fig. 1). Heterodimer formation is essential for full sGC activity; one heme molecule is bound



**Figure 1** The structure of guanylyl cyclases. “P”, “ATP” in the ellipse, and “Fe” in the rectangle indicate phosphorylation, ATP binding, and a heme moiety, respectively.

Table II Human Gene Loci, Human and Mouse Gene Symbols, and Tissue Expression of No Transmembrane-Segment Forms of Guanylyl Cyclase

Cyclase	Human locus	Human (mouse) gene symbol	Expression in normal tissues <sup>a</sup>
$\alpha_1$	4q31.1–q31.2	<i>GUCY1A3</i> ( <i>Gucy1a3</i> )	Cerebrum, cerebellum, spinal cord, pituitary gland, retina, salivary gland, thymus, thyroid gland, trachea, lung, <u>heart</u> , aorta, kidney, adrenal gland, liver, spleen, stomach, pancreas, small intestine, colon, bladder, skeletal muscle, bone marrow, lymph node, peripheral lymphocyte, mammary gland, prostate, testis, corpus carvanosum, ovary, uterus, placenta
$\beta_1$	4q31.1–q33	<i>GUCY1B3</i> ( <i>Gucy1b3</i> )	<u>Brain</u> (not divided), spinal cord, pituitary gland, retina, salivary gland, thymus, trachea, <u>lung</u> , heart, aorta, kidney, adrenal gland, liver, spleen, stomach, pancreas, small intestine, colon, bladder, skeletal muscle, bone marrow, lymph node, peripheral lymphocyte, mammary gland, prostate, testis, corpus carvanosum, ovary, uterus, placenta
$\alpha_2$	11q21–q22	<i>GUCY1A2</i> ( <i>Gucy1a2</i> )	<u>Brain</u> (not divided), retina, lung, kidney, liver, spleen, <u>pancreas</u> , colon, skeletal muscle, testis, corpus carvanosum, <u>uterus</u> , <u>placenta</u>
$\beta_2$	13q14.3	<i>GUCY1B2</i> ( <i>Gucy1b2</i> )	<u>Kidney</u> , liver, corpus carvanosum

Note. Underlines indicate high expression of the gene.

<sup>a</sup>Determined by Northern blotting or RT-PCR.

per heterodimer and serves as the nitric oxide (NO)-binding site. The central region (i.e., the region between the heme-binding domain and the CYC) of sGC isoforms is amphipathic  $\alpha$ -helical and could function as a dimerization domain.  $\alpha_1\beta_1$  and  $\alpha_2\beta_1$  heterodimers have been detected *in vivo*. NO-induced cGMP generation of  $\alpha_2\beta_1$  or  $\alpha_1\beta_2$  was reportedly one-third of  $\alpha_1\beta_1$  *in vitro*, and a dominant negative effect of  $\beta_2$  isoform on  $\alpha_1\beta_1$  activity has been suggested.

Expression of GCs in various tissues has been examined in humans, monkeys, rats, and mice (Tables I and II). GC-A and GC-B are expressed widely but not equally throughout the body. GC-A is abundant in the cardiovascular system (kidney, heart, and aorta), whereas GC-B expression is high (greater than GC-A expression) in the cerebellum, pituitary gland, pineal gland, and growth plate of bone. GC-C is found mostly in the intestine and colon, but GC-C mRNA is detected in the rat nephron by reverse transcriptase–polymerase chain reaction (RT-PCR). GC-C mRNA is detected also in the regenerating liver and in metastatic tumors of colon cancer. GC-D is expressed in a limited number of sensory neurons in the olfactory neuroepithelium. GC-E and GC-F are expressed in the outer segment of photoreceptor cells in the retina. GC-G is expressed in the lung, skeletal muscle, jejunum, and kidney.

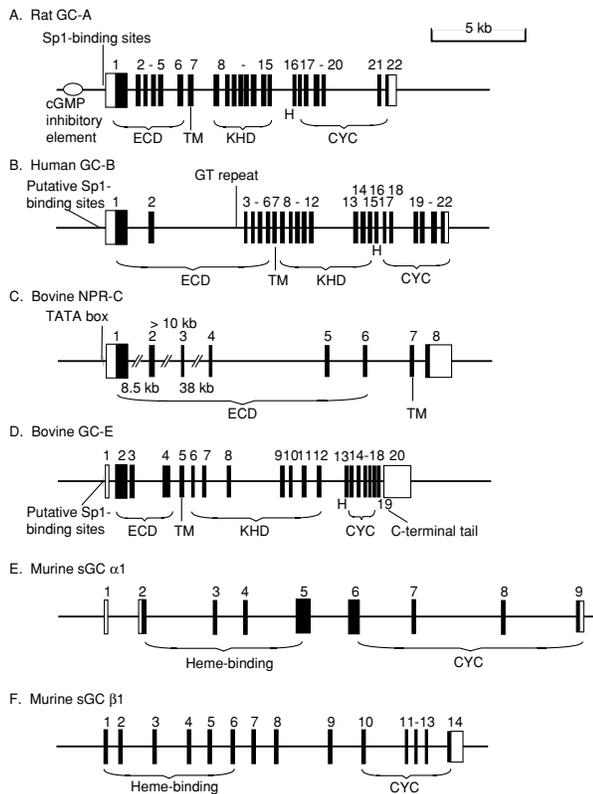
Soluble GC $\alpha_1$  and  $\beta_1$  isoforms are expressed almost ubiquitously (Table II).  $\alpha_2$  isoform expression is detected in various but more limited tissues than are  $\alpha_1$  and  $\beta_1$  isoforms.  $\beta_2$  isoform expression is found in the kidney and liver and in human gastric carcinomas (Table II).

## CHROMOSOMAL LOCALIZATION AND ORGANIZATION OF GC GENES

As shown in Table I, pGC genes are located on separate chromosomes. The genes for sGC  $\alpha_1$  and  $\beta_1$  isoforms are located on the same chromosome in mice and humans, with the other isoform genes ( $\alpha_2$  and  $\beta_2$ ) located on separate chromosomes (Table II).

Rat sGC  $\alpha_1$  and  $\beta_1$  loci are located on chromosome 2, closely linked to Na<sup>+</sup>, K<sup>+</sup>-ATPase  $\alpha_1$  isoform, and calmodulin-dependent protein kinase II  $\delta$  loci, which have been reported to flank a quantitative trait locus for blood pressure in Dahl rats. The rat sGC  $\beta_2$  locus located on chromosome 5 also cosegregates with the systolic blood pressure in Dahl rats.

Structures of GC-A, GC-B, and GC-E genes have been reported along with the NPR-C genomic structure. The rat GC-A gene of approximately 17.5 kb consists of 22 exons and 21 introns. Each functional domain is encoded by a group of exons; the ECD, the TM, the KHD, the hinge region, and the CYC are encoded by exons 1–6, 7, 8–15, 16, and 17–22, respectively (Fig. 2A). The structure is conserved in the human GC-B gene (Fig. 2B). The human GC-B gene contains a GT microsatellite repeat polymorphism within the second intron, and the (GT)<sub>11</sub> allele significantly associates with essential hypertension. The bovine NPR-C gene shares quite similar exon–intron organization with the ECD-encoding regions of GC-A and GC-B genes except for greater intron length (Fig. 2C). The overall structure of the bovine GC-E gene is similar to the overall structures of the GC-A and GC-B genes: an additional intron exists in each of the 5'- and 3'-noncoding



**Figure 2** The structure of guanylyl cyclase (GC) genes. The genomic structure is shown for rat GC-A (A), human GC-B (B), bovine NPR-C (C), bovine GC-E (D), murine sGC  $\alpha_1$  (E), and murine sGC  $\beta_1$  (F). Boxes with numbers represent exons; coding regions are filled. H, the hinge region.

regions; 3 introns, 1 intron, and 1 intron are missing from regions encoding the ECD, the KHD, and the CYC, respectively; and the C-terminal tail is encoded in a separate exon (Fig. 2D). Three murine sensory organ pGC genes (GC-D, GC-E, and GC-F) share identical exon-intron boundaries, at least within the region encoding the ECD. At least some intron positions of the murine GC-G gene are the same as those of the NPR genes.

The murine sGC  $\alpha_1$  isoform gene is at least 26.4 kb and consists of 9 exons and 8 introns (Fig. 2E), and the murine  $\beta_1$  isoform gene consists of 22 kb and contains 14 exons and 13 introns (Fig. 2F). The structures are quite different from those of pGC genes.

The promoter region has been sequenced for the GC-A, GC-B, GC-C, GC-E, and GC-F genes. These genes have no typical TATA boxes. The NPR-C gene contains a TATA box in the promoter region (Fig. 2C) but is not essential for gene transcription. The rat GC-A gene contains three Sp1-binding sites that appear to be essential for basal transcription of the

gene in the proximal promoter region and a cGMP-inhibitory element in the distal promoter region (Fig. 2A). The human GC-C gene promoter contains a binding site for hepatocyte nuclear factor (HNF)-4, a key regulator of intestinal-specific gene expression.

## REGULATION OF GC ACTIVITY

### Regulation of Gene Expression

#### Particulate GCs

Some reports indicate that natriuretic peptides and transforming growth factor- $\beta$  attenuate gene expression of GC-A and GC-B in isolated cell lines. It has been suggested that ANP suppresses GC-A gene expression through a cGMP pathway, consistent with the presence of a cGMP-inhibitory element in the GC-A gene promoter.

Signaling through GC-A or GC-B also may be modulated by the NPR-C expression given that NPR-C participates in ligand clearance. The  $\beta_2$ -adrenoceptor/cyclic adenosine 3',5'-monophosphate (cAMP) pathway or GC-B/cGMP pathway down-regulates NPR-C gene expression. Fibroblast growth factors (FGF-1 and FGF-2) and platelet-derived growth factor (PDGF)-BB, which activate receptor tyrosine kinases and subsequently mitogen-activated protein (MAP) kinases, inhibit NPR-C gene expression in rat pulmonary artery smooth muscle cells (SMCs).

Prolonged exposure of human colonic carcinoma T84 cells to 4 $\beta$ -phorbol 12-myristate 13-acetate, a potent protein kinase C (PKC) activator, results in a decrease of GC-C gene expression. The element responsible for the down-regulation has been located in a 129-bp region of the human GC-C gene promoter, which contains an HNF-4-binding site.

#### Soluble GC

NO and nerve growth factor have been reported to decrease sGC  $\alpha_1$  and  $\beta_1$  isoform mRNA levels in cultured rat vascular SMCs and rat pheochromocytoma PC12 cells, respectively.

Stimulation of cAMP/protein kinase A pathways has been shown to decrease mRNA levels of both  $\alpha_1$  and  $\beta_1$  isoforms and  $\alpha_1$  protein levels in rat aortic SMCs. Lipopolysaccharide and interleukin-1 $\beta$  reportedly attenuate sGC  $\alpha_1$  gene expression in rat vascular SMCs and  $\beta_1$  gene expression in rat cerebellar astrocyte-enriched primary cultures. It is also reported that estradiol reduces sGC  $\alpha_1$  and  $\beta_1$  mRNA levels within 1 h in the rat uterus.

## Regulation by Posttranslational Modification

### **GC-A and GC-B**

Receptor desensitization by posttranslational modification of the receptor molecule has been proposed for both GC-A and GC-B, where ANP and CNP desensitize GC-A and GC-B, respectively. Both GC-A and GC-B are phosphorylated on specific serine and threonine residues in the KHD, and ligand-induced desensitization is associated with dephosphorylation of the phosphoserine and phosphothreonine residues.

Desensitization of GC-A and GC-B by signaling molecules other than ligands has been reported as well. This type of desensitization can also generally be correlated with dephosphorylation of phosphoserine and phosphothreonine residues. Angiotensin II (AII) and arginine vasopressin can attenuate ANP-induced cGMP generation in rat vascular SMCs, glomerular mesangial cells, and adrenocortical carcinoma cell lines. Endothelins (ETs) can attenuate ANP- and CNP-induced cGMP generations in rat vascular SMCs and murine astrocytes, respectively. AII and ETs appear to cause this desensitization through stimulation of PKC.

Serum, PDGF-BB, and FGF-2 desensitize GC-B rapidly (within a few minutes) and extensively. In various kinds of cells, these agents activate MAP kinases and subsequently proliferation. It has been shown that this sequence of events can, in turn, be inhibited by CNP or ANP via cGMP-dependent mechanisms.

### **GC-C**

STa and uroguanylin cause attenuation of subsequent STa-induced cGMP generation in human colonic carcinoma cell lines (T84 and Caco2). A significant portion of the refractoriness seems to be caused by desensitization of GC-C. In clear contrast to GC-A and GC-B, phosphorylation of the C-terminal tail by PKC appears to enhance GC-C activity.

### **Soluble GCs**

Both very rapid (within 10 s) and slow (within 2 h) desensitization have been reported for sGCs. The mechanism of the very rapid desensitization has not been elucidated, but a cellular factor appears to be essential for this phenomenon. The mechanism of the slow desensitization also has not been revealed, but oxidation of the enzyme by free radicals derived from NO is suggested as being responsible.

## Regulation by Binding Proteins

### **GC-A and GC-B**

Protein phosphatase-5 and heat-shock protein 90 have been reported to bind to the GC-A KHD, but neither has been shown to regulate GC-A activity. Visinin-like protein-1 (VILIP-1), a member of the intracellular neuronal calcium sensor (NCS) family, was shown to bind to GC-A and GC-B in the CYC and can enhance ANP- and CNP-induced cGMP generation in PC12 pheochromocytoma and C6 astrocytoma cell lines, respectively. A guanylyl cyclase regulatory protein (GCRP) was also identified in the rat brain as a protein that binds to GC-A and enhances ANP-induced GC-A activation. The physiological significance of VILIP-1 and GCRP, however, has not been elucidated.

### **GC-E and GC-F**

GC-activating proteins (GCAP-1, GCAP-2, and GCAP-3) have been identified as regulatory proteins for GC-E and GC-F; they also belong to the NCS family. A GCAP-binding site has been identified within the GC-E CYC. Bleaching of rhodopsin by light results in activation of cGMP-specific phosphodiesterase (PDE6) mediated by transducin. The PDE6 decreases intracellular cGMP, causing closure of the cGMP-gated  $Ca^{2+}$  channel and subsequently hyperpolarization, which leads to visual sensations. In response to the resultant decrease in intracellular  $Ca^{2+}$ , GCAPs activate GC-E and GC-F to increase cGMP levels and restore the dark state. Targeted disruption of both GCAP-1 and GCAP-2 genes reportedly eliminates the  $Ca^{2+}$ -dependent GC regulation and reduces the capability of photoreceptors to adjust their sensitivity to light.

## Other Mechanisms

Complexes of ANP and GC-A are internalized after ANP binding to GC-A in murine Leydig tumor cells. This could explain a portion of the ANP-induced down-regulation of GC-A activity. Some studies have suggested that activation of cGMP-PDE (PDE5) or GC-C internalization by STa explain, at least to some extent, the STa-induced refractoriness to subsequent STa stimulation in T84 cells.

## PHYSIOLOGICAL ROLES OF GCS

### **GC-A**

GC-A null mice show blood pressure elevation independent of dietary salt intake and prominent cardiac

hypertrophy with cardiac fibrosis. Systolic blood pressure decreases as GC-A gene copy number is increased from 0 to 4 by targeted disruption and duplication of the gene. Intravenous infusion of ANP or isoconotic solution, which leads to an increase of plasma ANP concentrations, does not cause an increase of water and sodium excretion in GC-A null mice, whereas it has significant effects in wild-type mice. Thus, it has been suggested that GC-A plays important physiological roles in blood pressure regulation, diuresis, and natriuresis.

The GC-A system also functions as a local regulatory system to directly inhibit myocyte hypertrophy and fibrosis in the cardiac ventricle. The cardiac myocyte-specific overexpression of GC-A does not alter blood pressure and heart rate, but it reduces myocyte size in ventricles of both wild-type and GC-A null mice. Normalization of blood pressure by chronic treatment with antihypertensive drugs reportedly fails to attenuate left ventricular hypertrophy in GC-A null mice.

ANP inhibits aldosterone secretion from bovine adrenal glomerulosa cells via activation of cGMP-stimulated cAMP-PDE. GC-A null mice show 1.5- to 2-fold higher plasma aldosterone concentrations than do wild-type mice on a low-salt (0.008% NaCl) or a normal-salt (1% NaCl) diet, whereas a high-salt (8% NaCl) diet reduces plasma aldosterone levels to under the detectable limit in both genotypes. Plasma renin activities in GC-A null mice are 20 to 50% of those in wild-type mice on the normal-salt diet. An increase in circulating aldosterone levels due to loss of suppression by the ANP/GC-A system may contribute to hypertension with cardiac hypertrophy and fibrosis and to suppressed plasma renin activities.

Another endocrinological role of GC-A has been suggested. Serum testosterone concentrations are elevated, and ANP-induced testosterone secretion from Leydig cells is augmented, as GC-A gene copy number increases.

## GC-B

Studies on knockout mice of CNP, NPR-C, and cGKII have suggested that the CNP/GC-B/cGKII system is important for normal endochondral ossification and skeletal growth. CNP null mice exhibit dwarfism as a result of impaired endochondral ossification; proliferation and differentiation of chondrocytes in the growth plate of long bones, where CNP, GC-B, NPR-C, and cGKII transcripts are expressed, are markedly reduced. Dwarfism has also been observed in mice lacking cGKII. Targeted disruption

and three independent loss-of-function mutations that naturally occur in the murine NPR-C gene result in skeletal overgrowth; a decrease in endogenous CNP clearance by NPR-C appears to activate the CNP/GC-B pathway.

The CNP/GC-B signaling pathway has been proposed to play an important role in regulating proliferation and migration of vascular SMCs in proliferative vascular lesions such as atherosclerosis and restenosis after coronary angioplasty. It is suggested that both CNP secretion from endothelial cells and GC-B expression in proliferating vascular SMCs are augmented in these lesions.

## GC-C

Disruption of the murine GC-C gene results in resistance to pathogenic bacteria that secrete STa and cause acute secretory diarrhea. In enterocytes of the intestinal villi, intracellular cGMP elevation caused by STa binding to GC-C stimulates cGKII. Cystic fibrosis transmembrane conductance regulator (CFTR) is activated by cGKII and promotes massive intestinal fluid secretion and diarrhea. Mice deficient in cGKII are resistant to STa, and a report indicates that CFTR null mice show lesser responses to STa than do wild-type mice. GC-C null mice display no prominent abnormal phenotypes other than the STa resistance in a disease-free, highly controlled environment, and physiological roles of GC-C still remain unclear.

## GC-E

A series of studies has shown that signaling through GC-E is essential for normal retinal function. In GC-E null mice, a decrease in the number of cone photoreceptors occurs as mice age, whereas rod photoreceptors appear to remain normal. Missense and frame-shift mutations in the human GC-E gene are associated with Leber's congenital amaurosis (LCA), in which early-onset widespread degeneration of rod and cone photoreceptors occurs. A dominant form of cone-rod dystrophy (CORD6) is also associated with GC-E mutations in humans.

## Other pGCs

GC-D may be important to sense a particular subset of odorants or pheromones. A subset of olfactory neurons selectively expresses GC-D, a cGMP-stimulated PDE (PDE2), and a subunit of a cGMP-gated channel that is identified in cone photoreceptors. This subset of neurons also appears to

innervate atypical glomeruli in the olfactory bulb known as necklace glomeruli, which have been suggested to mediate some pheromone responses.

GC-F is thought to play an important role in the visual signal transduction because of its retina-specific expression and homology to GC-E. However, no direct evidence has been uncovered to support this hypothesis. Neither endogenous ligands nor regulatory molecules have been found for GC-G, and the physiological role(s) of GC-G remains unknown.

### Soluble GCs

The NO/sGC pathway appears to play important roles in blood pressure regulation and renal excretion. Disruption of the murine inducible NO synthase gene results in systemic and pulmonary blood pressure elevation. It has been suggested that the NO/sGC system enhances glomerular filtration and inhibits tubular reabsorption in the kidney. In the kidney of the Dahl salt-sensitive rat, gene expression of the sGC  $\beta_2$  isoform is augmented and that of the  $\beta_1$  isoform is repressed compared with the Dahl salt-resistant rat. The dominant-negative effect of the  $\beta_2$  isoform on the  $\alpha_1/\beta_1$  heterodimer might be responsible for blood pressure elevation in the Dahl salt-sensitive rat on high salt intake.

Soluble GC  $\beta_2$  gene expression is detected in human gastric carcinoma tissues but not in the normal gastric tissue where the  $\alpha_1$  and  $\beta_1$  isoforms are expressed. The dominant-negative effect of the  $\beta_2$  isoform may have a role in tumorigenesis.

### Acknowledgment

This work was supported in part by an award to D.L.G. from the Sandler Program for Asthma Research, the Howard Hughes Medical Institute, and the Cecil H. and Ida Green Center for Reproductive Biology Sciences.

### See Also the Following Article

Adenylyl Cyclase

### Further Reading

- Andreopoulos, S., and Papapetropoulos, A. (2000). Molecular aspects of soluble guanylyl cyclase regulation. *Gen. Pharmacol.* **34**, 147–157.
- Dizhoor, A. M. (2000). Regulation of cGMP synthesis in photoreceptors: Role in signal transduction and congenital diseases of the retina. *Cell. Signalling* **12**, 711–719.
- Gibson, A. D., and Garbers, D. L. (2000). Guanylyl cyclases as a family of putative odorant receptors. *Annu. Rev. Neurosci.* **23**, 417–439.
- Koesling, D. (1999). Studying the structure and regulation of soluble guanylyl cyclase. *Methods* **19**, 485–493.
- Lucas, K. A., Pitari, G. M., Kazeronian, S., Ruiz-Stewart, I., Park, J., Schulz, S., Chepenik, K. P., and Waldman, S. A. (2000). Guanylyl cyclases and signaling by cyclic GMP. *Pharmacol. Rev.* **52**, 375–413.
- Potter, L. R., and Hunter, T. (2001). Guanylyl cyclase-linked natriuretic peptide receptors: Structure and regulation. *J. Biol. Chem.* **276**, 6057–6060.
- Tamura, N., Chrisman, T. D., and Garbers, D. L. (2001). The regulation and physiological roles of the guanylyl cyclase receptors. *Endocrine J.* **48**, 611–634.
- van den Akker, F. (2001). Structural insights into the ligand binding domains of membrane bound guanylyl cyclases and natriuretic peptide receptors. *J. Mol. Biol.* **311**, 923–937.
- Wedel, B. J., and Garbers, D. L. (2001). The guanylyl cyclase family at Y2K. *Annu. Rev. Physiol.* **63**, 215–233.

estimate is that approximately half of boys will develop some degree of gynecomastia during puberty. Gynecomastia generally has its onset between 10 and 12 years of age and peaks between 13 and 14 years of age. The gynecomastia usually involutes within 18 months and is completely resolved by 16 to 17 years of age in most male adolescents. The final peak of gynecomastia is during adulthood, when up to two-thirds of males between 50 and 80 years of age are found to have some degree of gynecomastia when carefully examined.

## PATHOLOGICAL CAUSES OF GYNECOMASTIA

[Table I](#) outlines the broad categories of causes of gynecomastia. Gynecomastia may be found either in patients with germ cell tumors of the testicle, which secrete human chorionic gonadotropin (hCG), or in patients with Sertoli cell or Leydig cell tumors, which secrete excessive quantities of estrogens. Feminizing adrenocortical adenomas or carcinomas also either directly produce high quantities of estrogens or cause hyperestrogenemia through excessive production of estrogen precursors, such as androstenedione, that are aromatized to estrogens. Some tumors, such as giant cell carcinomas of lung, hepatoblastomas, and gastric or renal carcinomas, may secrete hCG, which stimulates the testes to produce excessive quantities of estrogen, resulting in gynecomastia.

Primary hypogonadism, which is due to a testicular pathology, results in a lowering of testosterone and, in some instances, an elevation of estrogen concentrations. A mild degree of testicular failure can be found in some men as part of the aging process. This may be accompanied by gynecomastia, which in the past has been termed "involutional gynecomastia." Secondary hypogonadism, which is due to a defect in the

**Table I Pathological Causes of Gynecomastia**

- |                          |
|--------------------------|
| • Tumors                 |
| Testes                   |
| Adrenal                  |
| Other                    |
| • Hypogonadism           |
| • Androgen insensitivity |
| • Hyperthyroidism        |
| • Enhanced aromatization |
| • Drugs                  |

**Table II Drugs Associated with Gynecomastia**

- |                                 |
|---------------------------------|
| • Hormones                      |
| • Antiandrogens                 |
| • Antiulcer drugs               |
| • Cancer chemotherapeutic drugs |
| • Cardiovascular drugs          |
| Antihypertensives               |
| Digitoxin                       |
| Spironolactone                  |
| • Phytoestrogens                |
| • Psychoactive drugs            |

pituitary or hypothalamus and results in a loss of appropriate gonadotropin stimulation of the testes, is a rare cause of gynecomastia.

Androgen insensitivity from a defect in the intracellular androgen receptor leads to an inability of androgens to act at the target tissues. This results in unopposed estrogen activity and is a rare cause of gynecomastia. Approximately a quarter of males with hyperthyroidism from Graves' disease may develop gynecomastia due to an increase in the concentration of SHBG, which binds testosterone more avidly than it binds estrogen, leading to a greater lowering of the free, biologically active testosterone level relative to the free, biologically active estradiol level.

As noted previously, the majority of circulating estrogens in males are derived from peripheral extraglandular aromatization of estrogen precursors. Increased aromatization is found in a variety of clinical situations, including aging, obesity, hyperthyroidism, liver disease, congenital adrenal hyperplasia, and Klinefelter's syndrome, and as a primary defect with excessive aromatase activity from persistence of an unregulated fetal aromatase enzyme.

A variety of drugs are also associated with gynecomastia, and the broad groups are listed in [Table II](#). A strong relationship between the occurrence of gynecomastia and medication use has been shown for androgens and anabolic steroids, estrogens and estrogen agonists, antiandrogens such as flutamide, the antibiotic ketoconazole, cimetidine, spironolactone, and alkalating agents. There are a number of other sources of estrogen that occasionally cause gynecomastia. These include the use of over-the-counter phytoestrogens and androstenedione used by athletes. In addition, there may be occupational exposure and percutaneous absorption from anti-balding creams or a sexual partner's use of an estrogen-containing vaginal cream.

## EVALUATION OF THE PATIENT WITH GYNECOMASTIA

The first question that must be answered is whether the breast enlargement is gynecomastia. The proper way in which to examine the male breast is to have the patient lie down with arms extended over or behind the head and for the examiner to take his or her thumb and forefinger and place them spread apart over the patient's breasts with the nipple centered in between. When the two digits are gently moved toward the nipple, gynecomastia will be detected as a firm or rubbery, mobile, round mound of tissue that arises concentrically from beneath the nipple and areola. The two most important conditions that should be differentiated from gynecomastia are pseudogynecomastia, which represents fatty enlargement of the breast without glandular proliferation, and breast carcinoma. With pseudogynecomastia, there will be no mound of tissue felt as the fingers close in on the nipple. Some of the features of breast carcinoma that distinguish it from gynecomastia include an eccentric location; mass that is hard, firm, and possibly fixed to underlying tissues; skin dimpling; retraction or crusting of the nipple or a frank nipple discharge; and the presence of axillary lymphadenopathy.

Once it is determined that the patient indeed has gynecomastia, the next question is whether medications are involved. A very careful history concerning medications, herb and vitamin intake, illicit drug use, and possible sources of environmental exposure to estrogens should be taken. If the patient is taking a medication known to be associated with gynecomastia, it should be stopped and the patient should be reexamined in 1 month.

The next question is whether the patient is pubertal. Because pubertal gynecomastia occurs frequently and is self-limited, its presence requires only reassurance that gynecomastia is a normal part of the pubertal process as well as a follow-up at 6 months or longer if the gynecomastia persists.

Another important question concerns whether the gynecomastia is of recent onset or is painful or tender. Gynecomastia of any cause undergoes the same pattern of histological evolution. The "florid phase" is found during the first 6 months. The breast ducts proliferate and exhibit epithelial hyperplasia, and there is an increase in the periductal and stromal constructive tissue, increased vascularity, and a substantial degree of periductal edema. Between 6 and 12 months is a transitional phase. The "involutional phase" is seen in patients with long-standing gynecomastia that has been present for 1 year or longer.

There is a marked reduction in the epithelial proliferation and increased stromal hyalinization, dilation of the ducts, and fibrosis. The presence of pain or tenderness indicates that the patient is in the florid phase and that the gynecomastia has been of relatively recent onset. It is also during this phase that medical therapy directed toward the gynecomastia is most likely to be effective. Once the tissue has entered the inactive or fibrotic stage, medical therapy is unlikely to be beneficial.

The patient should undergo a full physical examination, with particular attention given to the breast exam as described previously, the thyroid for signs of hyperthyroidism, an abdominal examination for signs of cirrhosis or adrenocortical masses, and a careful testicular examination, evaluating for the presence of masses and consistency, especially looking for the reduced size or decreased consistency consistent with hypogonadism.

If the patient is not in the pubertal age group and is not ingesting medications known to be associated with gynecomastia, laboratory investigation should include measurements of hCG, luteinizing hormone (LH), estradiol, and testosterone. An elevated hCG should lead to a testicular ultrasound, looking for a mass that would be compatible with a germ cell tumor. A normal ultrasound suggests that the patient has either an extragonadal germ cell tumor or an hCG-secreting nontrophoblastic neoplasm, which should be evaluated through additional imaging studies. An elevated LH and a decreased testosterone are compatible with primary hypogonadism, whereas a low LH and a low testosterone suggest secondary hypogonadism. In that instance, a serum prolactin should be measured to evaluate the patient for a prolactin-secreting pituitary tumor. If the prolactin is normal, the serum concentration of androstenedione and estrone should be determined. A normal or low level suggests secondary hypogonadism, whereas elevation of these hormones is indicative of the rare 17-ketosteroid reductase enzyme deficiency.

An elevated LH and an elevated testosterone are compatible with either hyperthyroidism or androgen resistance due to an androgen receptor defect. These can be discriminated through measurements of thyroid function tests, including free T<sub>4</sub> and thyroid-stimulating hormone (TSH). An elevated thyroxine (T<sub>4</sub>) and a suppressed TSH are diagnostic of hyperthyroidism, whereas normal thyroid function tests indicate that the patient has androgen resistance.

A testicular ultrasound should be carried out in patients who have an elevation of the estradiol concentration and a decreased or normal LH level.

A testicular mass in this setting is indicative of a Leydig or Sertoli cell tumor. If the testicular ultrasound is normal, an adrenal CT scan or MRI should be performed, looking for an adrenal neoplasm. A normal result from that exam suggests that the patient has increased extraglandular aromatase activity. If all results are normal, it is likely that the patient has idiopathic gynecomastia with enhanced sensitivity of the breast tissue to normal amounts of estrogen or that the inciting factor that initiated the gynecomastia has been corrected.

## TREATMENT

The indication for therapy includes anxiety, embarrassment, and pain. It should be noted that approximately 85% of patients who have gynecomastia of recent onset will undergo a spontaneous resolution, whereas patients with long-standing gynecomastia will have an approximately 10% chance of spontaneous resolution. Also, as noted previously, long-standing gynecomastia is less likely to respond to medical therapy than is gynecomastia of recent onset due to the presence of fibrosis in the breast tissue when the gynecomastia has been present for more than 1 year.

There are several therapies that have been tried for gynecomastia. Although there are many anecdotal reports of improvement on various therapies, there are few well-designed studies that have critically examined the different therapies. Many of the studies have been hampered by small numbers and lack of placebo controls. The medical therapies that have been tried include administration of testosterone and its metabolites, danazol (an impeded androgen with

weak androgenic and progestational activity), estrogen receptor antagonists such as clomiphene citrate and tamoxifen, and aromatase inhibitors such as testolactone and anastrozole. The best studied of these therapies is tamoxifen, which does appear to be superior to placebo in regard to reduction of breast pain and tenderness as well as breast size.

In a patient who has long-standing gynecomastia or who fails to have an adequate response to medical therapy, surgical removal of the breast glandular tissue should be considered. There are a number of approaches and methods that include sharp excision, liposuction, or a combination of these methods. The approach that is most commonly used currently involves a periareolar incision, a sharp excision of the breast glandular tissue, and liposuction for contouring the breasts. Complications of the procedures include nipple-areolar numbness, hematomas, seromas, scarring, inverted nipples, nipple necrosis, and depressive deformities of the chest wall as well as inadequate tissue removal along with redevelopment of the gynecomastia.

## See Also the Following Articles

Androgen Insensitivity Syndrome • Breast Disease: Impact of Sex Steroid Replacement • Estrogen and the Male • Graves' Disease • Sexual Function and Androgens

## Further Reading

- Braunstein, G. D. (1999). Aromatase and gynecomastia. *Endocrine-Related Cancer* **6**, 315–324.
- Gruntmanis, U., and Braunstein, G. D. (2001). Treatment of gynecomastia. *Curr. Opin. Invest. Drugs* **2**, 643–649.
- Mathur, R., and Braunstein, G. D. (1997). Gynecomastia: Pathomechanisms and treatment strategies. *Horm. Res.* **48**, 95–102.

- Types Ia and Ib are generally small pedunculated hamartomas that are attached to the tuber cinereum (Ia) or to the mammillary bodies (Ib) and are usually associated with precocious puberty.
- Types IIa and IIb are relatively large lesions that displace the hypothalamus slightly (IIa) or markedly (IIb) and are usually associated with gelastic or other types of seizures. Surgery has been recommended for types Ia, Ib, and IIa, especially when epileptic activity and behavioral abnormalities are severe and difficult to control.

A further classification has been proposed based on magnetic resonance imaging (MRI):

- The parahypothalamic type, attached to the floor of the third ventricle or suspended from the floor by a peduncle, is generally associated with precocious puberty.
- The intrahypothalamic type, enveloped by the hypothalamus to displace the third ventricle, is more likely to be associated with gelastic seizures.

## HISTOLOGY AND PATHOPHYSIOLOGY

Hypothalamic hamartoma is thought to be a developmental aberration, but its origin is unclear. Remnants of brain tissue left along the floor of the third ventricle when the chorda withdraws may account for its origin in the central nervous system (CNS). Histologically, it consists mainly of normal brain elements: neurons, glial cells, and fiber bundles that are frequently myelinated and connected to surrounding hypothalamic structures. Frequently, however, hamartomas do not reproduce the normal architecture of neighboring tissue.

The hypothalamic hamartoma functions as an ectopic luteinizing hormone-releasing hormone (LHRH) pulse generator that escapes the intrinsic CNS inhibitory mechanism. In the mouse, monkey, and human, LHRH neurons originate in the medial olfactory placode of the developing nose, migrate across the nasal septum, and enter the forebrain with the nervus terminalis, arching into the septal-preoptic area and hypothalamus. In hypothalamic hamartoma, a significant number of LHRH neurons migrate beyond the medial basal hypothalamus to the region of the mammillary bodies and tuber cinereum and form one of the constituents of the heterotopic mass of CNS tissue. The defect in migration could be related to an imbalance of diffusible chemotrophic factors that are secreted by restricted cell populations within the brain or to neural cell adhesion

molecules, which play an important role in axonal path-finding.

Since the number of LHRH neurons is limited, it is possible that a majority of the LHRH neurosecretory neurons may migrate into the hamartomas during CNS development. On the other hand, hypothalamic hamartoma may be related to aberrant differentiation among other cell types of neural primordia, including progenitor cells with the capacity to form LHRH neurosecretory neurons.

Histological specimens observed after surgery reveal hamartomas of low cell density containing irregularly structured groups of ganglionoid cells with variably sized unipolar and bipolar neurons interspersed among glial cells with myelinated and unmyelinated fibers connected to surrounding hypothalamic structures. The tissue is highly vascular and many of the vessels have fenestrated endothelium and double basement membranes. Each vessel is almost totally surrounded by axons. There is no or very little tendency to proliferation.

Immunohistochemical studies confirm the neuronal origin of hypothalamic hamartoma, showing positive staining for neuron-specific enolase, synaptophysin, and neurofilament protein.

Different immunohistochemical studies demonstrate the presence of membrane-bound, electron-dense granules (100 nm in diameter), which contain LHRH within the perikarya, the axons, and the axons terminals and which are the elements of an independent neuroendocrine unit.

The examination of two hypothalamic hamartomas associated with sexual precocity revealed that they contained astroglial cells expressing transforming growth factor- $\alpha$  (TGF- $\alpha$ ), but not LHRH neurons. These findings indicate that some hypothalamic hamartomas induce sexual precocious puberty by activating endogenous LHRH secretion via astroglial-derived factors, such as TGF- $\alpha$  and/or TGF- $\beta$ . This activation appears to require a close proximity of the hypothalamic hamartoma to either LHRH neurons or their axonal processes in the median eminence. In some cases, the hamartoma itself does not initiate precocious sexual maturation due to its location, but rather a lesion of the adjacent hypothalamic tissue resulting from surgery may cause the activation of astroglial cells, which may then lead to increased LHRH secretion from hypothalamic LHRH neurons.

A number of findings suggest that hamartomas are themselves epileptogenic. Electroencephalography (EEG) recordings revealed focal spikes arising from the depth contacts within hypothalamic hamartomas,

whereas electrical stimulation studies reliably reproduce gelastic episodes, suggesting a close relationship between hamartomas and the generation of laughing attacks. The most fascinating studies based on ictal single-photon-emission computed tomography demonstrated marked blood flow in the hypothalamus and thalamic structures during gelastic events. Improvement in intractable epilepsy has been reported in some cases after the resection of the hamartoma.

Clinical and experimental evidence suggests that the hypothalamus and adjacent structures, in particular the mammillary bodies and their immediate connections, may constitute an important subcortical pathway for seizure propagation. Sessile hypothalamic hamartomas with displacement of the hypothalamus are associated with seizures.

Evidence shows that, unlike laughing and focal seizures, slow spike-and-wave discharges and associated tonic and atonic seizures do not arise directly from the hamartoma. Indeed, postoperatively, these seizures may progressively “run down” after removal of the hamartoma, suggesting that they are the result of secondary epileptogenesis.

## CLINICAL PRESENTATION

### Endocrine Aspects

Hypothalamic hamartoma is considered the most common organic cause associated with CPP. The incidence of hypothalamic hamartoma in determining CPP varies widely, ranging from 2 to 28%.

The age of onset of the first signs of CPP has been reported as being between 0 and 4 years (frequently before 2 years of age), whereas CNS lesions (gliomas, germinomas, arachnoid cysts) tend to occur in association with CPP later in life.

CPP associated with hypothalamic hamartoma has all of the hormonal hallmarks of puberty, including plasma  $\beta$ -estradiol (E2) levels, a pubertal pattern of pulsatile luteinizing hormone (LH), and a pubertal LH response after gonadotropin-releasing factor administration. Higher levels of peak LH response and a peak LH/follicle-stimulating hormone (FSH) ratio after LHRH administration have been reported in girls with CPP associated with hypothalamic hamartoma compared to idiopathic precocious puberty.

### Neurological and Behavioral Aspects

The classic presentation of hypothalamic hamartoma is represented by gelastic seizures, which begin in infancy, often in the neonatal period, and are later

followed by the development of seizures with focal motor features and autonomic manifestations, such as flushing and cardiorespiratory changes. Partial seizures, suggestive of a temporal or frontal lobe origin, may also occur. In many children, a generalized epileptic encephalopathy develops, with tonic and atonic seizures with slow spike-and-wave discharges on interictal EEG, intellectual deterioration, and sometimes with marked behavioral disturbances including rage attacks.

Several studies reported a characteristic, recognizable epileptic syndrome that shows a typical evolution over time. The syndrome begins with laughing attacks; the resemblance to normal laughter is so close in some patients that delayed diagnosis is possible. The pattern is quite stereotyped, with some variations over time. After a few years, the laughing attacks are associated with momentary loss of awareness, facial myoclonus, and/or abnormal eye movements. Autonomic phenomena are common. In later childhood, typically between the ages of 4 and 10 years, features of secondary generalized epilepsy appear with multiple seizure patterns including tonic, atonic, and tonic-clonic seizures and progressive cognitive impairment.

The occurrence of cognitive deficits is significantly associated with the presence, severity, and frequency of gelastic/complex partial seizures. In a study with a large cohort of patients, it has been demonstrated that patients with hamartomas display a statistically significant increase in comorbid psychiatric conditions, including oppositional defiant disorders, attention-deficit/hyperactivity disorder, high rates of conduct disorders, speech retardation/learning impairment, and anxiety and mood disorders. There are no definitive data to indicate exactly how the epileptic pathology interferes with a patient's development and behavior and why a majority of patients have a cognitive-behavioral deficit, which often worsens with age.

Brief infantile attacks are not associated with any significant change in EEG. However, in later childhood, with the progression of the epileptic syndrome, EEG abnormalities develop with generalized suppression of background rhythm, generalized low-voltage fast activity, or both. Interictally, there are generalized slow spike-and-wave discharges and background activity that is abnormally slow for the child's age as hallmarks of secondary generalized epilepsy.

## MAGNETIC RESONANCE IMAGING

MRI has become the diagnostic method of choice in the detection of hypothalamic hamartoma. In general,

hamartomas are uniformly isointense to gray matter on T1-weighted MRI scans and slightly hyperintense or isointense on T2-weighted images. They are not enhanced by gadolinium (Figs. 1A and 1B). There are some cases where hypothalamic hamartomas appear slightly hyperintense on T1-weighted images and, very rarely, they may appear hypointense in the center of the lesion with normal isointensity at the periphery. The variance in myelinated axons and the presence of gliosis can affect the variability of the T2 signal.

Another characteristic feature of hypothalamic hamartomas is their relative stability in size, shape, and signal in long-term MRI follow-up during and after medical treatment. Signal or cystic changes have occasionally been observed; in these cases, the differential diagnosis should include low-grade astrocytomas and gangliogliomas. The MRI findings that commonly indicate hypothalamic hamartoma are as follows: an isointense mass with a pedunculated or sessile attachment to the hypothalamus without contrast enhancement, stable in size and shape during follow-up, associated with endocrinological and clinical features of CPP, occurring mostly before 2 years of age, and associated with gelastic seizures.

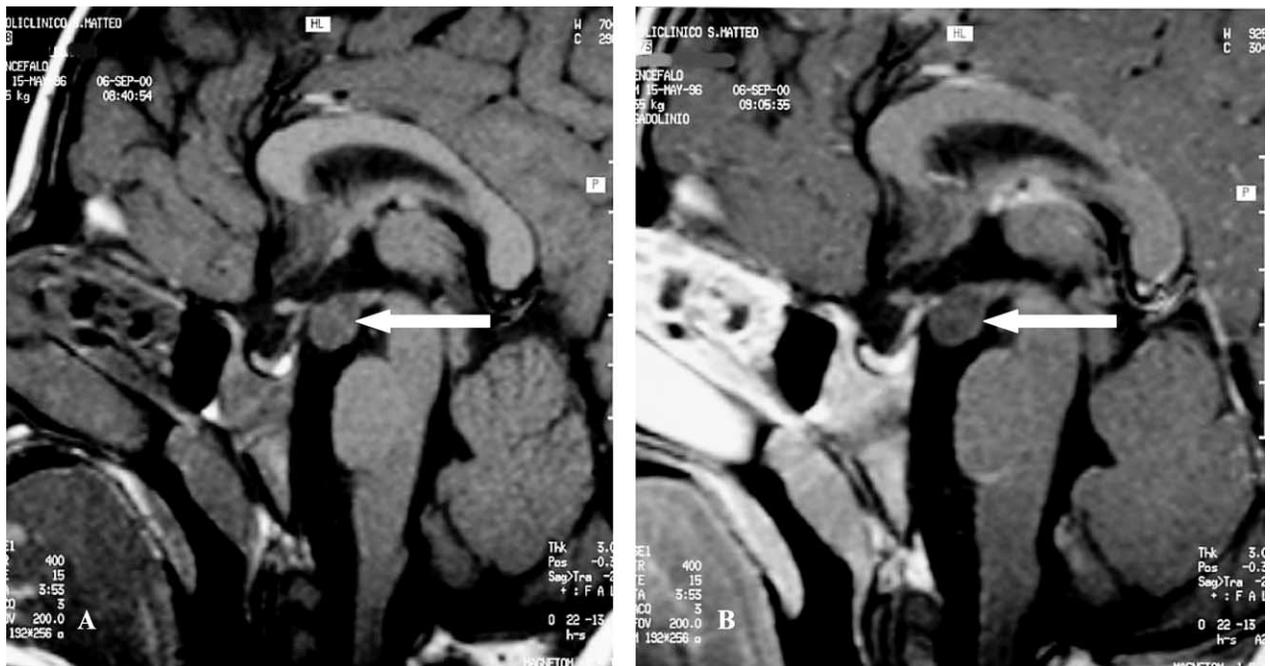
## TREATMENT

### Precocious Puberty

The aims of treatment are to arrest physical maturation, to prevent early menarche or sexual adult psycho-physical maturation, and to improve adult height to within the range of target height combined with normal body proportions.

Successful suppression treatment with gonadotropin-releasing hormone (GnRH) analogue preparations has been reported by several groups for a duration of up to 8 years with a homogenous endocrinological response characterized, after an initial phase of increased serum LH and FSH levels (flare-up), by a reduction to prepubertal ranges of basal LH, FSH, testosterone, and E2, a prepubertal peak LH and FSH response to GnRH stimulation test, and a prepubertal LH/FSH ratio with improvement of adult height toward predicted final height and close to target height; no adverse events have been reported.

A controversial outcome of precocious puberty has been reported after surgical resection of hamartomas. It has been suggested that surgery is safe and can be considered an option when compared to the cost of medical treatment, monthly treatment (or every 3 months) administration until puberty (sometimes



**Figure 1** Sagittal T1-weighted MRI shows sessile hamartoma in the typical location in the third ventricle, isointense with gray matter before (A) and after (B) gadolinium administration (white arrow).

poorly tolerated), painful injections in some cases, the psychological implications, and, in some rare cases, an intolerance to treatment. Total resection, however, cannot be achieved in all hamartomas.

Females with hypothalamic hamartoma have a higher incidence of irregular menses, obesity, and neurological and behavioral problems with a normal reproductive axis within 4–5 years after discontinuation of GnRH analogue therapy; pregnancies resulted in normal, live infants.

### Seizures

It has become clear that epileptic encephalopathy associated with hypothalamic hamartoma is a treatable condition. Strong evidence exists that removal, destruction, or disconnection of the hamartoma leads to remarkable control of the seizures, as well as to an improvement in behavior and probable cessation of cognitive decline.

Gelastic seizures are resistant to all available medications. Rarely, treatment of the focal seizures and of tonic and atonic attacks can be moderately successful with conventional anti-epileptic drugs. However, experience suggests that long-lasting control of these seizures can be effected by the complete removal, destruction, or disconnection of the hamartoma, thanks to important improvements in imaging and surgical techniques. There exist various possible surgical approaches to the lesion and the debate is focused on the best technique to use (i.e., it must be tailored to the specific surgical anatomy of the hamartoma) in order to guarantee, in each single case, the possibility of total removal or disconnection of the lesion. The transcallosal approach (suited to intraventricular lesions) appears to be the technique with the highest rate of success, with 90% of cases free or virtually free of all seizures at 1 year post-surgery. Prolonged follow-up is necessary to evaluate the real effects on cognitive impairment, whereas behavioral abnormalities tend to show marked improvement in many patients. Other surgical choices include the pterional approach (particularly useful for peduncular lesions) and the transventricular endoscopic approach or destruction of the lesion with radiofrequency or gamma knife radiosurgery.

Secondary effects of surgery that have been described are small strokes with good recovery, encephalomalacia, third-nerve palsy, transient memory impairment, and transient diabetes insipidus. Postoperative mortality has also been reported.

## ASSOCIATED CONDITIONS

### Pallister-Hall Syndrome

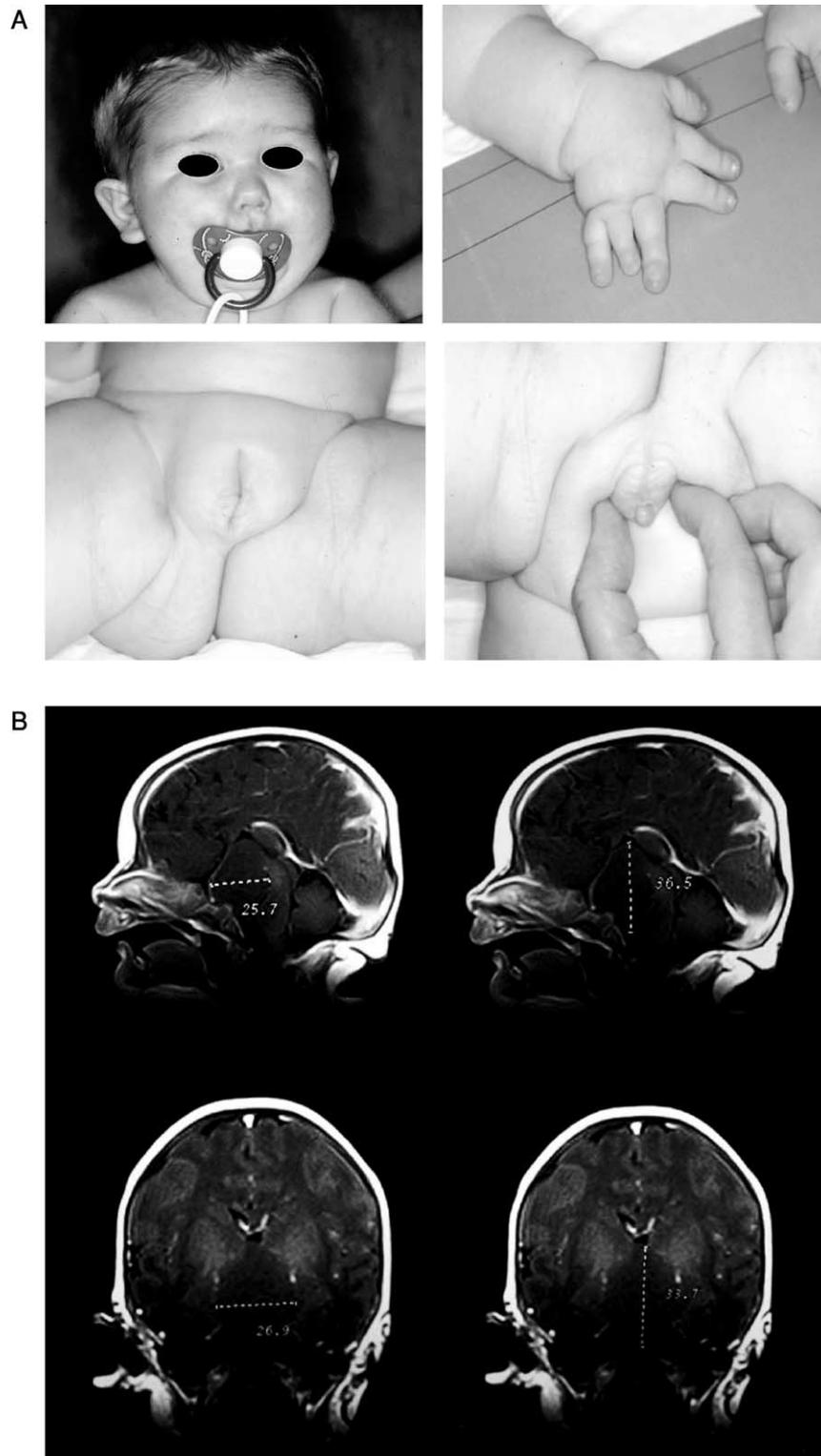
Pallister-Hall syndrome was first described in 1980 as a lethal congenital malformation syndrome associated with hypothalamic hamartoblastomas, postaxial polydactyly, craniofacial malformations, and imperforate anus. Additional abnormalities include developmental and postnatal retardation; holoprosencephaly; pituitary agenesis/dysgenesis with hormone dysfunction leading to micropenis, cryptorchidism, hypopituitarism, or panhypopituitarism; laryngeal clefts; bifid epiglottis; buccal frenula; small nose/anteverted nares; low-set/posteriorly angulated ears and microphthalmia; limb/skeletal malformations such as short fourth metacarpals and nail dysplasia; involvement of other organs with abnormal lung lobulation; renal agenesis and/or dysplasia; and congenital heart defects (Figs. 2A and 2B). The syndrome is considered a clinical, variable “iceberg” disorder with wide phenotypic variability in which adrenal insufficiency is the most common cause of perinatal death.

This disorder is inherited as an autosomal-dominant trait and has been reported in association with an unbalanced chromosome translocation between chromosomes 3p and 7q and it was mapped to chromosome 7p13. Mutations in the transcription regulator gene *GLI3* have been identified in patients with Pallister-Hall syndrome, but also in four other different autosomal-dominant phenotypes: the Greig cephalopolysyndactyly syndrome, preaxial polydactyly type IV, postaxial polydactyly type A, and postaxial polydactyly type A/B.

Consensus guidelines in 1996 developed diagnostic criteria for the delineation of familial Pallister-Hall syndrome, which include the presence of a hypothalamic hamartoma, central polydactyly most commonly affecting the third or fourth digit, and, in the first-degree relatives of an index case, a hypothalamic hamartoma or similar digital abnormalities associated with an autosomal dominant. An association with endocrine dysfunction requires hormone replacement treatment. Adult height above the target height has been reported in a patient treated with growth hormone for up to 7 years. Surgical correction of visceral abnormalities is mandatory.

### Other Conditions

Hamartomas may be found in syndromes of midline defects such as oral–facial–digital syndrome or solitary maxillary incisor or in association with CNS malformations including agenesis of the corpus



**Figure 2** (A) Frontal view showing a phenotype suggestive of congenital hypopituitarism. Note the hands with polydactyly; micropenis and cryptorchidism are associated with panhypopituitarism. (B) Sagittal and coronal T1-weighted MRI demonstrating the mass (dotted lines) compatible with hamartoblastoma and the absence of pituitary tissue.

callosum or holoprosencephaly. They have also occasionally been described in Laurence-Moon-Biedl syndrome.

## Acknowledgments

The authors thank Dr. Luigi Gargantini (Clinica Pediatrica, Ospedale di Treviglio, Treviglio, Italy) for providing the case of Pallister-Hall illustrated in Fig. 2.

## See Also the Following Articles

FSH (Follicle-Stimulating Hormone) • LH (Luteinizing Hormone) • Pituitary Tumors, Clonality • Precocious Puberty, Central (Female) • Precocious Puberty, Central (Male)

## Further Reading

- Arita, K., Ikawa, F., Kurisu, K., Kurisu, K., Sumida, M., Harada, K., Ouzumi, T., Monden, S., Yoshida, J., and Nishi, Y. (1999). The relationship between magnetic resonance imaging findings and clinical manifestations of hypothalamic hamartoma. *J. Neurosurg.* **91**, 212–220.
- Berkovic, S., Arzimanoglou, A., Kuzniecky, R., Harvey, A. S., Palmi, A., and Andermann, F. (2003). Hypothalamic hamartoma and seizures: A treatable epileptic encephalopathy. *Epilepsia* **44**, 969–973.
- Debeneix, C., Bourgeois, M., Trivin, C., Sainte-Rose, C., and Brauner, R. (2001). Hypothalamic hamartoma: Comparison of clinical presentation and magnetic resonance images. *Horm. Res.* **56**, 12–18.
- Feuillan, P. P., Jones, J. V., Barnes, K., Oerter-Klein, K., and Cutler, G. B. (1999). Reproductive axis after discontinuation of gonadotropin-releasing hormone analog treatment of girls with precocious puberty: Long-term follow-up comparing girls with hypothalamic hamartoma to those with idiopathic precocious puberty. *J. Clin. Endocrinol. Metab.* **84**, 44–49.
- Judge, D. M., Kulin, H. E., Page, R., Santen, R., and Trapukdi, S. (1977). Hypothalamic hamartoma: A source of luteinizing-hormone-releasing factor in precocious puberty. *N. Engl. J. Med.* **296**, 7–10.
- Jung, H., and Ojeda, S. R. (2002). Pathogenesis of precocious puberty in hypothalamic hamartoma. *Horm. Res.* **57**, 31–34.
- Kang, S., Graham, J. M., Olney, A. H., and Biesecker, L. G. (1997). GLI3 frameshift mutations cause autosomal dominant Pallister-Hall syndrome. *Nat. Genet.* **15**, 266–268.
- Kuzniecky, R., Guthrie, B., Mountz, J., Bebin, M., Faught, E., Gilliam, F., and Liu, H. G. (1997). Intrinsic epileptogenesis of hypothalamic hamartomas in gelastic epilepsy. *Ann. Neurol.* **42**, 60–67.
- Mahachoklertwattana, P., Kaplan, S. L., and Grumbach, M. M. (1993). The luteinizing hormone-releasing hormone-secreting hypothalamic hamartoma is a congenital malformation: Natural history. *J. Clin. Endocrinol. Metab.* **77**, 118–124.
- Partsch, C. J., Heger, S., and Sippel, W. G. (2002). Management and outcome of central precocious puberty. *Clin. Endocrinol.* **56**, 129–148.
- Valdueza, J. M., Cristante, L., Dammann, O., Bentele, K., Vortmeyer, A., Saeger, W., Padberg, B., Freitag, J., and Herrmann, H. D. (1994). Hypothalamic hamartomas: With special reference to gelastic epilepsy and surgery. *Neurosurgery* **34**, 949–958.
- Weissenberger, A. A., Dell, M. L., Liow, K., Theodore, W., Frattali, C. M., Hernandez, D., and Zimetkin, A. J. (2001). Aggression and psychiatric comorbidity in children with hypothalamic hamartomas and their unaffected siblings. *J. Am. Acad. Child Adolesc. Psychiatry* **40**, 696–703.
- Zúñiga, O. F., Tanner, S. M., Wild, W. O., and Mosier, D. (1983). Hamartoma of CSN associated with precocious puberty. *Am. J. Dis. Child.* **137**, 127–133.

light of a higher prevalence of autoimmune hypothyroidism in iodine-sufficient regions as compared with iodine-deficient regions. Overt hypothyroidism in patients on a very rich iodine diet (e.g., seaweed, kelp) can be reversible on avoiding the source of the iodine excess. Otherwise, lifelong treatment with  $T_4$  is indicated.

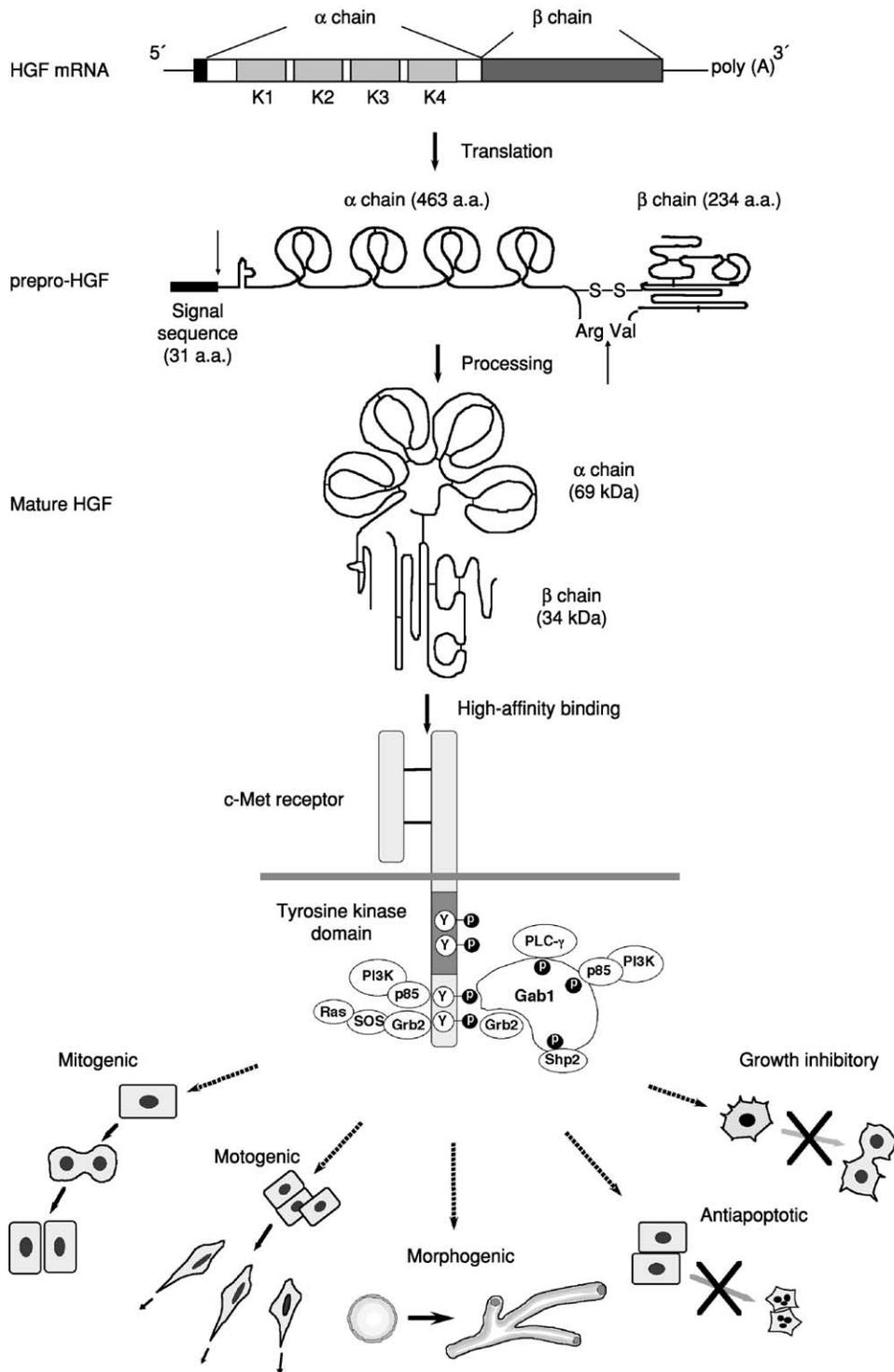
In the atrophic variant of chronic autoimmune thyroiditis (atrophic myxedema), fibrosis is the predominant feature along with lymphocytic infiltration. In the less common goitrous variant originally described by Hashimoto, the histology remains essentially unaltered after 20 years. The goiter is diffuse and has a firm "rubbery" consistence; it does not regress in 43% of patients despite  $T_4$  treatment. Some patients have an initial transient hyperthyroid stage labeled as Hashitoxicosis. It disappears in a few months along with the decline in the causative TSH receptor-stimulating antibodies, which may give way to TSH receptor-blocking antibodies.

## See Also the Following Articles

Autoimmune Polyglandular Syndrome • Graves' Disease • Hyperthyroidism, Childhood and Adolescence • Iodine • Smoking and the Thyroid • Thyroid Disease and Pregnancy

## Further Reading

- Hashimoto, H. (1912). Zur Kenntniss der Lymphomatösen Veränderung der Schilddrüse (struma lymphomatosa). *Arch. Klin. Chir.* **97**, 219.
- Hayashi, Y., Tamai, H., Fukata, S., *et al.* (1985). A long-term clinical, immunological, and histological follow-up study of patients with goitrous chronic lymphocytic thyroiditis. *J. Clin. Endocrinol. Metab.* **61**, 1172–1178.
- Pedersen, I. B., Knudsen, N., Jørgensen, T., Perrild, H., Ovesen, L., and Laurberg, P. (2002). Large differences in incidences of overt hyper- and hypothyroidism associated with a small difference in iodine intake: A prospective comparative register-based population survey. *J. Clin. Endocrinol. Metab.* **87**, 4462–4469.
- Roitt, J. M., Doniach, D., Campbell, P. N., and Hudson, R. V. (1956). Autoantibodies in Hashimoto's disease (lymphocytic thyroiditis). *Lancet* **2**, 820.



**Figure 1** Structure and biological functions of HGF. Schematic structures of HGF mRNA, prepro-HGF, and mature HGF are shown in the upper half. Binding of HGF to the c-Met receptor evokes various biological effects on cellular behavior in a target cell-specific manner. See text for details.

product highly homologous to HGF. HGF-like protein was later shown to be identical to macrophage-stimulating protein. HGF-like protein specifically binds and activates Ron receptor tyrosine kinase, a family of the *c-Met* receptor.

The receptor encoded by *c-met* protooncogene is synthesized as a single polypeptide chain of 1436 amino acids that is cleaved into a 50-kDa  $\alpha$ -chain and a 145-kDa  $\beta$ -chain (Fig. 1). The  $\alpha$ -chain is exposed extracellularly, whereas the  $\beta$ -chain is a transmembrane subunit containing an intracellular tyrosine kinase domain. Binding of HGF to the Met receptor induces activation of tyrosine kinase, resulting in the subsequent phosphorylation of C-terminally clustered tyrosine residues and recruitment of various intracellular signaling molecules (Fig. 1).

Studies on the structure–function relationship in the HGF molecule revealed a functional domain for receptor binding and subsequent biological activities. Both NK1 and NK2, smaller N-terminal variants of HGF, can bind the *c-Met* receptor; thus, NK1 serves a minimum set of domains responsible for binding to the *c-Met* receptor. Interestingly, NK1 and NK2 exhibit antagonistic activity on HGF-induced mitogenesis, whereas they function as agonists in terms of cell motility activity. On the other hand, NK4, composed of N-terminal hairpin and four kringle domains, was characterized as a competitive antagonist for HGF.

## BIOLOGICAL ACTIVITIES

HGF exhibits multiple biological activities for a wide variety of cells (Fig. 1). Target cells of HGF include most epithelial cells (e.g., hepatocytes, renal tubular cells, neurons, keratinocytes), vascular endothelial cells, and several mesenchymal/stromal cell types.

HGF is mitogenic for differentiated epithelial cells (e.g., hepatocytes, renal tubular cells, bronchial and alveolar epithelial cells, gastric epithelial cells, keratinocytes), vascular endothelial cells, mesenchymal/stromal cells (e.g., cardiomyocytes, articular chondrocytes, osteoclast-like cells), and stem cells such as hepatic stem cells, hematopoietic progenitor cells, and skeletal satellite cells, precursor cells that differentiate into myocytes. On the other hand, HGF inhibits growth of several distinct types of cancer cells and cell lines (e.g., HepG2 hepatocellular carcinoma cells, B16F1 melanoma cells). HGF as “motogen” stimulates motility and migration of various cells. Among the multipotent characteristics of HGF, the morphogenic activity of HGF is unique. HGF induces branching tubulogenesis in epithelial

cells, including renal tubular cells, mammary gland epithelial cells, and hepatic bile duct epithelial cells.

HGF has potent antiapoptotic activity in several types of cells, including hepatocytes, renal epithelial cells, vascular endothelial cells, cardiac myocytes, and neurons. HGF increases activity and/or gene expression of proteases involved in the breakdown of extracellular matrix components, including urokinase-type plasminogen activator and matrix metalloproteases. Induction of these proteases plays an important role in branching tubulogenesis, migration, or invasion of cells. It is notable that HGF is involved in a phenotypic transition between epithelial and mesenchymal cell types in a cell type-specific manner; HGF induces epithelial-to-mesenchymal transition of dermomyotome cells in embryos, whereas it inhibits transition of renal tubular epithelial cells to myofibroblasts.

## DEVELOPMENTAL ROLES

Interactions between the epithelium and mesenchyme (i.e., epithelial–mesenchymal interactions) mediate crucial aspects of normal development, affecting tissue induction, morphogenesis, and organogenesis. Growth, differentiation, and morphogenesis of developing epithelia are regulated either inductively or permissively by neighboring mesenchyme in the developing kidney, lung, pancreas, placenta, mammary gland, salivary gland, tooth, and so forth. During rat and murine development, the *c-Met* receptor gene is predominantly expressed in epithelia, whereas the HGF gene is in mesenchymal cells in close vicinity to epithelia in various organs, including the kidney, pancreas, liver, lung, intestine, stomach, salivary gland, limb bud, and tooth. Several distinct approaches indicate that HGF plays roles in the development and morphogenesis of organs as a mediator in epithelial–mesenchymal interactions.

Essential roles of HGF in the development of mammalian fetal tissues were defined by targeted disruption of HGF or the *c-Met* receptor gene. These knockout mice are embryonic lethal due to impaired organogenesis of the placenta and liver. In the placenta, the number of epithelial trophoblasts in the labyrinthine layer is markedly reduced, leading to impaired exchange of oxygen and nutrients between maternal and embryonic bloodstreams. The embryonic liver is reduced in size and shows extensive apoptotic cell death, indicating that hepatoblasts/hepatocytes in the embryonic liver require HGF for proliferation and/or survival. An essential role for HGF in organogenesis was also demonstrated in the amphibian *Xenopus*.

**Table I** Developmental and Physiological Roles of HGF and Potential Clinical Application

Developmental roles	
Growth, antiapoptosis, and/or morphogenesis	
Growth and morphogenesis of the liver	
Growth of the placental labyrinthine layer	
Morphogenesis of the kidney (tubulogenesis)	
Morphogenesis of the lung	
Morphogenesis of the mammary gland	
Morphogenesis of the tooth germ	
Migration	
Migration of myogenic precursor cells from the demomyotome to the limb and diaphragm	
Migration and proliferation of myogenic precursor cells to the tongue	
Chemoattractive and/or neurotropic	
Attraction of axons of motor neurons to the limb	
Guidance or attraction of cranial motor neurons	
Physiological roles	
Liver regeneration following partial hepatectomy, ischemia, or administration of hepatotoxins	
Renal regeneration following unilateral nephrectomy, ischemia, or administration of nephrotoxins	
Lung regeneration following partial pneumonectomy or chemical injury	
Cardiac protection following ischemia	
Epithelia of the stomach and intestine	
Cutaneous wound healing	
Angiogenesis	
Enhancement of insulin secretion and antiapoptosis in pancreatic beta-cells	
Potential therapeutic application	
Hepatic diseases: acute hepatitis, fulminant hepatitis, fatty liver, transplantation, liver cirrhosis	
Renal diseases: acute renal failure, chronic renal failure, transplantation	
Lung diseases: acute lung, lung fibrosis, transplantation	
Gastrointestinal: gastric ulcer, diabetes mellitus	
Cardiovascular diseases: cardiac infarction, cardiomyopathy, arteriosclerosis obliterans	
Nervous system: cerebral ischemia, amyotrophic lateral sclerosis	
Others: graft versus host-disease, cartilage defect	

Involvement of HGF in morphogenesis of organs was also revealed in organ culture experiments. Cultured embryonic organs undergo morphogenic steps that mimic their development *in vivo*. In organ cultures, neutralizing anti-HGF antibody inhibited morphogenesis of developing lung epithelia and branching tubulogenesis of developing epithelia in the kidney and mammary gland. In tooth germ culture, antisense oligonucleotide to HGF induced impaired morphogenesis of tooth epithelium. Thus,

HGF is a mesenchymal-derived paracrine factor that supports morphogenesis of developing epithelia during organogenesis of the lung, kidney, and tooth.

One of the interesting developmental functions of HGF is the decisive role in generation of skeletal muscle that derives from long-range migrating muscle precursor cells. Ablation of HGF or the *c-met* gene in mice results in complete absence of skeletal muscles of the limb and diaphragm, whereas other muscle groups form normally. In these mice, the long-range migration of c-Met-positive myogenic precursor cells from dermomyotome in the somite to limb buds and diaphragm is impaired. Because HGF is strongly expressed in limb bud mesenchyme and septum transversum (which develops into the diaphragm), HGF provides spatially defined chemoattractant-like motogenic signals for migration of myogenic precursor cells. Similar chemoattractant-like function of HGF was noted in migration of myogenic precursor cells into the tongue.

In the nervous system, HGF has chemoattractant-like functions. Developing axons can be guided to their targets by diffusible molecules and factors bound to the cell surface that act as chemoattractants or repellents. Transplanted limb buds attract motor axons in amphibians and birds, and similarly, explanted mouse limb bud mesenchyme attracts axons of motor neurons. HGF was identified to be a chemoattractant factor derived from limb bud mesenchyme. Developmental and morphogenic roles of HGF are summarized in [Table I](#).

## PHYSIOLOGICAL FUNCTIONS

Regeneration of the liver is a most dramatic phenomenon in higher animals. When 70% of the liver is resected, the cells in the remaining liver rapidly proliferate and the original liver mass is restored within a week. Given that HGF was initially implicated to be a hepatotropic factor that enhances liver regeneration, the hepatotropic roles of HGF have been demonstrated extensively. HGF and the c-Met receptor, both expressed in a variety of tissues, play a role in regeneration and/or protection of other tissues and organs besides the liver.

In rats and mice, a range of hepatic injuries (e.g., partial hepatectomy, acute hepatitis, ischemia, physical crush) increase blood HGF levels and HGF mRNA expression in the injured liver and in intact distant organs such as the lung, kidney, and spleen. Similarly, increases in plasma HGF levels and HGF gene expression in injured and noninjured organs occur following various types of injuries (e.g., partial

resection, ischemia, chemical injury) of other organs such as the kidney, lung, heart, muscle, and pancreas. Local injury to the gastric mucosa and skin increases local expression of HGF and the c-Met receptor.

Involvement of endogenous HGF in tissue regeneration and/or protection was demonstrated by approaches using a neutralizing antibody against HGF. Neutralization of endogenous HGF in rats or mice increased tissue damage and cellular apoptosis while it suppressed tissue regeneration in cases of hepatic, renal, and lung injuries. In case of cardiac ischemia reperfusion injury, neutralization of HGF resulted in increased apoptosis of cardiomyocytes and expansion of infarcted area, leading to cardiac dysfunction in rats. Local injection of anti-HGF antibody delayed gastric ulcer healing. Importantly, when endogenous HGF was neutralized in mice with renal injury, mice were susceptible to renal fibrosis, indicating that HGF is antifibrotic.

It is worth noting that increased HGF levels in sera or tissue fluid were noted in patients with distinct types of diseases, including acute hepatitis, fulminant hepatitis, acute renal rejection after transplantation, pneumonia, cardiac infarction, severe acute pancreatitis, and Crohn's, Huntington's, Parkinson's, and Alzheimer's diseases. Increased HGF levels in sera or tissue fluids are likely to reflect physiological responses to tissue and cellular damage.

Vascular endothelial cells from various tissues express the c-Met receptor, and HGF is a potent angiogenic growth factor that stimulates growth and migration of endothelial cells. HGF induces new blood vessel formation *in vivo*. However, because abnormal vascular formation was not seen in knock-out mice of HGF or the *c-met* gene, HGF might not play a major role in neovascularization during embryogenesis. Physiological roles of HGF are listed in [Table I](#).

## INVOLVEMENT IN TUMORIGENESIS AND TUMOR PROGRESSION

Ligand-dependent or-independent constitutive activation of receptor tyrosine kinases is often associated with the tumorigenic transformation of cells. Genetic analysis of papillary renal carcinoma in patients indicated that missense mutations in the *c-met* gene are the causative genetic disorder in inherited and some sporadic papillary renal carcinomas. All mutations that occur in the *c-met* gene are missense, and the c-Met receptor with these mutations had enhanced

tyrosine kinase activity, indicating that these c-Met receptor mutations are likely to be a gain-of-function mutation. In addition to genetic mutation, aberrant activation of the c-Met receptor through establishment of an autocrine loop between HGF and the c-Met receptor, or overexpression of the c-Met receptor, is likely to be involved in tumorigenesis and/or tumor progression toward metastatic tumor. Overexpression of the c-Met receptor was noted in a wide variety of tumor cells and tissues, including carcinoma, lymphomas, and soft tissue tumors.

Although aberrant activation of the c-Met receptor is involved in the tumorigenic transformation of cells, a more definite significance of the c-Met receptor activation in tumor cells is likely to be in progression to invasive and metastatic tumor. Invasion of tumor cells is regulated by distinct cellular functions, including cell-cell adhesion, cell-extracellular matrix association, proteolytic breakdown of the extracellular matrix, and cellular locomotion. Because HGF affects these processes, HGF exhibits profound effects on the invasive behavior of a wide variety of tumor cells. Gene transfer experiments that involve autocrine activation of the c-Met receptor confer invasive and metastatic behavior in cancer cells. Furthermore, although tumor-stromal interaction closely affects invasive and metastatic behavior of carcinoma cells, HGF is a predominant stromal-derived mediator in tumor-stromal interaction. HGF is a critical factor for tissue regeneration, whereas cancer cells use the HGF-c-Met pathway toward tumorigenic and/or metastatic progression. The simile that "cancer is a never-healing wound" seems pertinent here.

## THERAPEUTIC APPROACHES

Based on evidence that activation of the c-Met receptor leads to tissue protection and repair against tissue injury, whereas it also leads to invasive and metastatic progression of cancer cells in tumor tissues, two distinct therapeutic approaches can be considered. One is application of HGF (i.e., c-Met agonist) for treatment of organ injuries, and the other is application of an antagonistic or inhibitory molecule against HGF-c-Met pathway to inhibit cancer invasion and metastasis. Experiments to explore the therapeutic potential of HGF for treatment of diseases have been extensive, and various disease models and experimental animals have been used ([Table I](#)). Several approaches were done using HGF gene therapy.

## Approaches Using HGF or HGF Gene Transfer for Therapeutic Purposes

Administration of human recombinant HGF into mice or rats stimulates proliferation of hepatocytes after liver insult caused by partial hepatectomy or hepatotoxins. Through its antiapoptotic action, HGF abrogated the onset of acute hepatitis and fulminant hepatic failure. Likewise, administration of HGF accelerated tissue regeneration and/or attenuated tissue injury in other organs. HGF suppressed the onset of acute renal failure caused by administration of nephrotoxic drugs or renal ischemia. HGF enhanced the mitogenesis of bronchial and alveolar epithelial cells following lung injury caused by partial pneumonectomy or hydrochloric acid. HGF administered to rats with cardiac ischemia reperfusion injury reduced the extent of apoptosis in cardiac myocytes, thereby reducing the infarcted area. The submucosal application of HGF accelerated gastric ulcer healing in rats. Mice with the transgenic expression of HGF in pancreatic beta cells were resistant to type 1 diabetes. In articular tissues, HGF effectively repairs osteochondral defects in a rabbit model.

Chronic inflammatory diseases are characterized by fibrotic changes in tissues, including liver cirrhosis, chronic renal failure, lung fibrosis, and cardiomyopathy. These fibrotic diseases are progressive and currently incurable. Administration of HGF into rats with liver cirrhosis decreased the accumulation of hepatic extracellular matrix components and abrogated the mortality rate due to hepatic dysfunction. HGF prevented renal fibrosis and concomitant renal dysfunction in models for chronic renal failure or renal fibrosis. Likewise, HGF administration prevented lung fibrosis caused by bleomycin. Although mechanisms responsible for antifibrotic action of HGF are not fully understood, HGF does suppress key events leading to tissue fibrosis. HGF suppresses expression of transforming growth factor- $\beta$ , a key growth factor whose overexpression induces tissue fibrosis. HGF is antiapoptotic and mitogenic for epithelial cells (e.g., hepatocytes, renal tubular cells, alveolar epithelial cells) and endothelial cells, whereas HGF enhances protease activities responsible for the degradation of extracellular matrix components.

Neurotropic actions of HGF were expanded to therapeutic approaches for the treatment of injury or diseases of the nervous system. Infusion of HGF into the brain prevents neuronal death in the hippocampus and in the cerebral cortex. Transgenic expression of HGF in neurons resulted in retardation of disease progression and prolongation of the life span in a

transgenic mouse model for amyotrophic lateral sclerosis, a fatal neurodegenerative disorder characterized by the progressive loss of motor neurons and the degeneration of motor axons.

HGF gene therapy induces angiogenesis in models for hind limb ischemia and cardiac infarction. In 2001, the first clinical trial of HGF gene therapy for treatment of patients with arteriosclerosis obliterans was initiated, and validity of this therapy has been confirmed.

## Approaches Using an HGF Antagonist for Cancer Treatment

The antagonistic molecule designated NK4 was isolated as a competitive inhibitor for binding of HGF to the c-Met receptor. NK4 is composed of N-terminal 447 amino acids of the  $\alpha$ -chain of HGF and so contains the N-terminal domain and subsequent four kringle domains. NK4 has receptor-binding domains (i.e., NK1 or NK2) but has no agonistic activities after binding to the c-Met receptor.

Treatment with NK4 protein or HGF gene therapy inhibited tumor growth, invasion, and/or metastasis in experimental cancer treatment models, including pancreatic, breast, lung, gallbladder, colon, and ovarian cancers. Likewise, NK4 treatment prolonged the survival of mice with pancreatic, ovarian, and colon cancers.

An alternative approach to block c-Met-dependent signaling events was demonstrated using synthetic tripeptide inhibitors that block association of GRB2 with the c-Met receptor. These compounds inhibited migration, invasion, and the morphogenesis induced by HGF. Thus, these compounds are potential candidate molecules to inhibit c-Met-dependent cancer invasion and metastasis.

## PERSPECTIVE

Developmental and morphogenic roles of HGF may facilitate understanding as to how HGF orchestrates organogenesis and morphogenesis, including spatial and timely expression of HGF and the *c-met* gene in distinct cell types as well as specific changes in signal transduction and gene expression toward morphogenic processes. These studies are closely related and recapitulated in tissue regeneration in adult tissues. Further studies on mechanisms for HGF-related tissue regeneration will lead to a fundamental understanding of growth factor-network and cellular interactions involved in tissue regeneration and homeostasis.

Approaches for potential therapeutic use of both agonists and antagonists in the HGF-c-Met system will explore new therapeutic methods for treatment of patients with various diseases and malignant tumors.

### See Also the Following Articles

EGF and Related Growth Factors • Fibroblast Growth Factor (FGF) • GI Hormones as Growth Factors • Growth Factor Receptors • Insulin-like Growth Factors • Peptide Hormones and Growth Factors: Cellular Signaling Mechanisms • Platelet-Derived Growth Factor (PDGF)

### Further Reading

- Balkovets, D. F., and Lipschutz, J. H. (1999). Hepatocyte growth factor and the kidney: It is just not for the liver. *Int. Rev. Cytol.* **186**, 225–260.
- Birchmeier, C., and Gherardi, E. (1998). Developmental roles of HGF/SF and its receptor, the c-Met tyrosine kinase. *Trends Cell Biol.* **8**, 404–411.
- Matsumoto, K., and Nakamura, T. (1996). Emerging multipotent aspects of hepatocyte growth factor. *J. Biochem. (Tokyo)* **119**, 591–600.
- Matsumoto, K., and Nakamura, T. (2001a). Hepatocyte growth factor: Renotropic role and potential therapeutics for renal diseases. *Kidney Int.* **59**, 2023–2038.
- Matsumoto, K., and Nakamura, T. (2001b). HGF-c-Met receptor pathway in tumor invasion–metastasis and potential cancer treatment with NK4. In “Growth Factors and Their Receptors in Cancer Metastasis” (W. G. Jiang, K. Matsumoto, and T. Nakamura, eds.), pp. 241–276. Kluwer Academic, Dordrecht, Netherlands.
- Nakamura, T., Nishizawa, T., Hagiya, M., Seki, T., Shimonishi, M., Sugimura, A., Tashiro, K., and Shimizu, S. (1989). Molecular cloning and expression of human hepatocyte growth factor. *Nature* **342**, 440–443.
- Vande Woude, G. F., Jeffers, M., Cortner, J., Alvord, G., Tsarfaty, I., and Resau, J. (1997). Met-HGF/SF: Tumorigenesis, invasion, and metastasis. In “Plasminogen-Related Growth Factors” (G. R. Bock and J. A. Goode, eds.), pp. 119–132. John Wiley, New York.

**Table I Etiologies of Adult Hypogonadism**

Primary
Testicular trauma
Chemotherapy
Radiation
Klinefelter's syndrome
Mumps orchitis
Malnutrition
Excessive heat exposure
Drugs or medications
Environmental toxins
Surgical manipulation
Autoimmunity
Granulomatous disease
Secondary
Cushing's syndrome
Hyperprolactinemia
Panhypopituitarism
Isolated gonadotropin deficiency
Biologically active adrenal tumors
Obesity
Hypothyroidism
Primary or secondary causes
Medications
Hemochromatosis
Hypercortisolemia
Severe systemic illness
Sickle cell disease
AIDS
Uremia
Cirrhosis

primary gonadal disorders, secondary disorders that reflect dysfunction at the level of the hypothalamic-pituitary axis, and defects with both primary and secondary hypogonadism components (Table I). Among the more common causes are hyperprolactinemia, hypercortisolemia, obesity, and medications. The medications most commonly noted to contribute to hypogonadism are ketoconazole, glucocorticoids, spirinolactone, cimetidine, phenytoin, and alcohol. As pituitary radiation becomes more common as a treatment for various pituitary tumors, its role in secondary hypogonadism and perhaps panhypopituitarism must be considered.

### Testosterone Decline in the Aging Male

There is no male equivalent to female menopause, although some comparison can be made to declining testosterone production and increased gonadotropin levels with increasing male age. However, the loss of

testosterone production in men is gradual and relatively subtle. Both total and free (unbound) testosterone may decrease by approximately 1 to 2% per year from 30 to 40 years of age. Several reports suggest that there is also an alteration in the testosterone/estradiol ratio in that estradiol levels remain the same while testosterone production decreases over time. This results in an increase of SHBG, leading to even further decreases in unbound testosterone. The increases in gonadotropins in this setting are relatively modest. Older men with low to low-normal serum testosterone will have LH and FSH levels slightly above or comparable to the upper range of normal in young adult men.

### DIAGNOSIS

Recognition of the manifestations and physical findings, such as loss of libido, erectile dysfunction, decreased bone density, gynecomastia, and small testes, leads to further evaluation by laboratory testing. Initial testing should include a morning plasma total testosterone performed by radioimmunoassay and gonadotropin levels. In approximately 30% of normal males, levels of testosterone are highest in the morning and are decreased by up to one-third by afternoon and evening, owing to the natural circadian rhythm of testosterone production and release. A low to low-normal morning total testosterone level, usually 200 to 250 ng/dl in most laboratories, warrants further investigation with determination of the free or unbound fraction of testosterone. It should be remembered that testosterone values, even in young men, can vary greatly and can appear to be in the low range of normal despite normal sexual function.

For those with testosterone levels that are clearly in the hypogonadal range, ascertainment of the etiology of hypogonadism is important and the gonadotropin levels are helpful in this respect (for etiologies of hypogonadism, see Table I). In older men with subtle symptoms of testosterone deficiency due to age, the diagnosis may be somewhat difficult. Although levels of total testosterone less than 200 to 250 ng/dl are generally considered hypogonadal, low-normal levels present a dilemma. Partial androgen deficiency in the aging male is an entity of growing interest, owing to a fair proportion of older men with subtle symptoms of hypogonadism and less than clear biochemical evidence of classic hypogonadism. In addition, it is unknown whether the requirements of elderly males are similar to those of younger men. Currently, there is no one parameter that adequately reflects biological

activity. Consequently, one must use good clinical judgment when evaluating hypogonadism, with treatment aims to alleviate symptoms and improve quality of life.

## INDICATIONS FOR HORMONE REPLACEMENT

Patients with elevated gonadotropins and decreased testosterone (hypergonadotropic hypogonadism) have primary hypogonadism and warrant therapy with testosterone. However, most patients with primary hypogonadism have associated failure of spermatogenesis, and so androgen replacement cannot provide fertility.

Patients with low levels of gonadotropins and low testosterone (hypogonadotropic hypogonadism) have secondary hypogonadism. Testosterone is indicated to reverse hypogonadism. However, if fertility is desired, it is possible with the use of human chorionic gonadotropin (hCG) or recombinant LH to raise testosterone as well as human menopausal gonadotropin (hMG) or recombinant FSH to induce spermatogenesis. In patients with hypothalamic disorders resulting in hypogonadotropic hypogonadism, fertility can be induced with either administration of GnRH or the use of gonadotropins. The one caveat in this situation is the possibility of GnRH receptor defect as the etiology, thereby precluding effective treatment with the use of GnRH.

Although the treatment is controversial, older men with normal to borderline elevated levels of gonadotropin and normal to low-normal levels of testosterone, but with suggestive symptoms of hypogonadism, have been administered testosterone. Use of testosterone in this situation has not proved to be beneficial to sexual function, prevention of osteoporosis, or prevention of muscle loss. Although some features of male aging resemble the signs and symptoms of young hypogonadal men, the significance of the partial and gradual declines of testosterone levels in older men remains unclear from observational studies. Without guidance from more rigorous clinical trials, potential benefits of androgen administration must be weighed against potential adverse effects.

Although not strictly in the realm of androgen replacement, androgens are also used to treat adolescents with constitutional delay of puberty, to treat infants with micropenis, and as a preventive agent in cases of hereditary angioedema (C1 esterase inhibitor deficiency).

## TREATMENT EFFECTS

The goals of androgen replacement therapy are alleviation of symptoms and maintenance of secondary sexual characteristics, sexual function, libido, muscle strength, and mass. Changes and improvement in sexual function and libido often begin within days to weeks of initiating therapy, whereas appearance changes evolve over a period of several months to years. Treating frank hypogonadism with androgens generally has significant effects, whereas treatment of the androgen decline of the aging male has effects that are more modest. As mentioned earlier, several studies have noted an increased sense of well-being, only modest changes in muscle mass and lean body mass, and no improvement in strength.

Hypogonadism is a significant risk factor for osteoporosis in men, and androgen replacement can arrest further bone loss and restore bone density. However, it must be pointed out that androgen replacement in osteoporosis is worthwhile only after other secondary causes of bone loss have been ruled out. In addition to probable direct effects, androgen is converted to estradiol, which also acts to decrease bone resorption and increase bone density.

There is evidence that lipid metabolism may be affected by androgens. Men generally have lower high-density lipoprotein (HDL) cholesterol, higher triglycerides, and higher low-density lipoprotein (LDL) cholesterol than do premenopausal women. It is also noted that during adolescence, HDL cholesterol levels decrease with concomitant increases in triglyceride and LDL cholesterol in males. Abuse of exogenous androgens by young men can likewise result in decreased HDL cholesterol levels. On the other hand, this is not necessarily seen with hormone replacement therapy. Reasons for this include the supraphysiological doses of androgens used by abusers and the use of androgens not aromatizable to estrogen. Studies of nonaromatizable androgens in normal men have shown that HDL cholesterol and apolipoprotein A-I levels decreased, whereas LDL cholesterol levels increased. However, the effect of testosterone on lipid profiles is not clearly understood. Most clinical trials of intramuscular testosterone therapy in older men show little, if any, effect on HDL cholesterol. Conversely, numerous cross-sectional studies report a positive association between exogenous testosterone and HDL cholesterol.

There may be other potential effects of androgens based on smaller studies. A tendency toward

decreased arterial pressure and improved insulin sensitivity has been reported. A decrease in abdominal fat, an increase in lipolysis, inhibition of lipoprotein lipase, and mobilization of triglycerides have also been reported in association with androgen administration. Future investigations might help to further define the physiology of testosterone's precise role in lipids and cardiovascular disease.

Although not commonly used in this manner today, androgens can be beneficial for various forms of anemia. In the kidney, androgens stimulate erythropoietin production, whereas synthesis of clotting factors is increased in the liver. Availability of commercially produced erythropoietin makes androgen use for the treatment of anemia rare. Patients with hypogonadism can often see increases in hematocrit with androgen therapy.

## PRECAUTIONS

For individuals over 45 years of age, prostate-specific antigen and digital rectal exam are recommended. Untreated hypogonadal men generally have smaller prostates than do age-matched eugonadal men. With androgen replacement, prostate size can be expected to increase to match their eugonadal counterparts. However, caution should be exercised when considering treatment in men with known benign prostatic hyperplasia. Men with symptoms of urinary obstruction should not be treated with androgen given that benign prostatic hyperplasia can be hastened with the use of this therapy. Prostatic carcinoma is an absolute contraindication to the use of androgen replacement.

Several other precautions must be noted due to the effects of testosterone on various tissues. In older men, testosterone can have a profound effect on hematocrit and can cause polycythemia in as many as 25% of patients under treatment, often resulting in termination of therapy. Sleep apnea can also be precipitated or worsened. Fluid retention, gynecomastia, and cardiovascular disease have been noted to some degree, especially with higher doses and with nonaromatizable androgens. Although it is commonly believed that androgen administration may lead to behavioral disturbances, this has not been borne out in studies, especially when maintaining eugonadal levels of testosterone. Aggressive behavior can be monitored with patient self-reports with the aid of spouses or partners. In summary, evaluation prior to treatment should include a history for potential sleep apnea, prostate-related disorders, behavioral disturbances, and measurement of hematocrit.

## ANDROGEN PREPARATIONS

Multiple forms of testosterone therapy for male hormone replacement have been produced through biochemical modification due to the bioavailability characteristics of naturally occurring testosterone. When administered by mouth, testosterone is promptly degraded by the liver once it enters the portal circulation, and relatively little reaches the systemic circulation. Parenterally, testosterone is both absorbed and degraded rapidly, making it unfavorable for treatment. As a consequence, a variety of chemical modifications have resulted in the currently available forms of androgen replacement (Table II). Most of

Table II Androgen Preparations

Form	Agent	Usual dose
Oral	Fluoxymesterone	5–20 mg daily
	Methyltestosterone	10–50 mg daily
	Testosterone undecanoate	80–160 mg daily
Intramuscular	Testosterone propionate	25–50 mg three times per week
	Testosterone cypionate	200 mg every 10–14 days
	Testosterone enanthate	200 mg every 10–14 days
	Human chorionic gonadotropin (can also be given subcutaneously)	2500–5000 IU two or three times per week
	Human menopausal gonadotropin	75–100 IU three times per week
Transdermal	Testosterone patch	1 patch (4 or 6 mg) daily
	Testosterone gel	5–10 g daily
	Dihydrotestosterone gel	5–10 g daily
Miscellaneous	GnRH pump	5–20 µg every 120 min

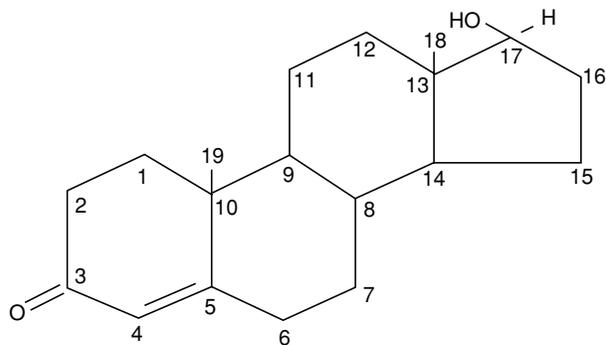


Figure 1 Testosterone.

the modifications involve esterification of the  $17\beta$ -hydroxyl group; alkylation of the  $17\beta$ -position; alteration of the ring structure at the 1, 2, 9, or 11 carbons; or a combination of these changes (Fig. 1).

Esterification of the  $17\beta$ -hydroxyl group allows formation of fat-soluble compounds prepared as injectable testosterone esters. Commonly available preparations include testosterone propionate, testosterone cypionate, and testosterone enanthate. The main difference among these compounds is the length of the carbon chain in the ester. Longer chain modifications make the compounds more fat soluble, prolonging the action of the ester through slow release. Testosterone propionate has the shortest half-life and requires administration every 2 to 3 days. Testosterone cypionate and enanthate have a longer half-life, are used more often, and have a dosing frequency of 7 to 21 days. In hypogonadal men 20 to 40 years of age, 200 mg of testosterone cypionate or enanthate leads to supraphysiological levels of total testosterone at 1 to 5 days, with subnormal levels typically by 14 days. Anecdotally, this is supported by patients who feel a pronounced effect of therapy during the first week after testosterone injection and who feel occasional diminished libido during the last few days before an injection. Testosterone ester injections are relatively low in cost and are known to be safe and effective through years of clinical experience. However, this form of therapy also entails the need for regular intramuscular injections, a non-physiological pattern of testosterone levels over the dosing interval, and potentially wide variations in mood and libido over the dosing interval.

The testosterone patch and testosterone gel are two forms of the more recent transdermal testosterone delivery systems. Testosterone patches are applied to either scrotal or non-scrotal skin. Patches are applied daily to clean intact skin without hair and provide physiological testosterone levels. However,

application to scrotal skin results in higher levels of DHT, owing to the higher levels of  $5\alpha$ -reductase enzyme in the scrotum, and the long-term effects on the prostate are not known. Difficulties encountered with the testosterone patch include local skin reactions, cosmetic unacceptability, and relatively high cost, all leading to discontinuation in 30 to 50% of men who try the patch.

Testosterone gel has become particularly popular since its introduction. Packaged as a once-daily application, it consists of a hydroalcoholic gel that dries quickly when applied to the skin surface. Absorption occurs quickly, and increased serum testosterone levels can be seen within 30 min of application. Similarly to the testosterone patch, steady physiological levels are achieved. However, testosterone gel is potentially transferable via prolonged skin contact. Clinical studies have shown that as little as 15 min of skin contact within up to 12 h of application can result in dermal transfer. Thus, patients must be warned to apply the gel to clean intact skin that is not easily in contact with others. Although most men on transdermal testosterone achieve eugonadal levels of testosterone, the peak level achieved is significantly less than that achieved with intramuscular injections. Consequently, there are occasional patients who favor the intramuscular administration if they are accustomed to the effects of a supraphysiological dose that is often felt during the first week after an injection. Local skin reactions can occur, although not as commonly as with the testosterone patch.

Although not currently available in the United States, testosterone undecanoate is an orally active ester that is absorbed via intestinal lymphatics. Shortcomings of this preparation include short duration of action, the need to administer three times per day, and variable absorption, with consequent variations in levels of measured testosterone within the same patient and among different patients.

Other oral preparations for male hypogonadism include fluoxymesterone and methyltestosterone. Formed from  $17\alpha$  substitution, these agents should be avoided in the treatment of hypogonadism. They are known to cause cholestatic jaundice, abnormal liver function tests, and (rarely) tumors of the liver and peliosis hepatitis.

In addition to testosterone, its metabolites (e.g., DHT) have been developed for use in hypogonadal men. DHT is a selective androgen produced from testosterone by the  $5\alpha$ -reductase enzymes 1 and 2 that is not aromatized to estrogen. In Europe, DHT is marketed as a 2.5% hydroalcoholic gel that is applied transdermally. Because DHT is the major androgen

within the prostate, there is very little role for its use in men as hormonal replacement therapy.

For patients desiring fertility who have hypothalamic or pituitary lesions or any congenital or acquired defect involving GnRH, spermatogenesis can be achieved and maintained assuming that the gonads are capable of functioning. This requires treatment with LH and FSH to drive Leydig cell function and Sertoli cell function, respectively. Conventional therapy includes the use of hCG for its LH activity, followed by the use of hMG, a purified extraction from the urine of postmenopausal women providing FSH activity. Therapy is applied in two phases. The induction phase begins with hCG alone at doses of 2500 to 5000 IU two or three times per week, either subcutaneously or intramuscularly, for a period of 4 to 6 months. The dose may need titration, with a goal of testosterone in the normal range. Once this phase has been completed successfully, hMG (75–150 IU) is administered three times per week intramuscularly. hCG doses need to be monitored and adjusted further at this point because dose reductions are often necessary. Treatment should be carried out for at least 4 months, owing to the lengthy cycle of spermatogenesis. Potential complications that require consideration during therapy include local skin reactions, hypersensitivity and anaphylactic reactions, and thromboembolism.

Pulsatile GnRH therapy is also available to those with hypothalamic dysfunction resulting in hypogonadotropic hypogonadism. Treatment requires a portable pump attached to the skin via a subcutaneous butterfly needle. Doses range from 5 to 20  $\mu$ g every 120 min. Efficacy is noted by induction of sperm in the ejaculate and an increase in testicular size.

## MONITORING

So long as patients can be monitored for potential adverse effects, therapy should be lifelong. As with any treatment, goals, benefits, and risks must be reassessed continually. In general, the continuing assessment of androgen replacement therapy relies on the clinical response. In addition, levels of testosterone can be obtained with initiation of treatment and the timing of testing based on the anticipated peak or trough, depending upon the androgen preparation. Digital rectal exam and prostate-specific antigen should be checked at initiation of treatment and at frequent intervals thereafter. Hematocrit, lipid profile, and liver function tests (if indicated) should also be

examined prior to therapy, after 1 to 2 months, and at yearly intervals thereafter.

## CONCLUSION

Hormone replacement therapy for the hypogonadal man is fairly straightforward. Once an etiology is ascertained, treatment with testosterone for those who are clearly hypogonadal has proven benefits regarding libido, sexual function, muscle mass and body habitus, and (for those with osteoporosis) bone mass. There are also a number of preparations available, with promising results shown in the newer transdermal gel preparations so far as ease of use and physiological effects are concerned. However, what is unclear is the effect of hormone replacement therapy in those with low-normal testosterone levels or the so-called partial androgen deficiency of aging males. The signs and symptoms of aging often seem similar to those of hypogonadal young men. Unfortunately, although testosterone and its metabolites can be measured, there is no clinically useful measure of the biological activity of androgens. Data on aging males reveal only modest, if any, positive effects. However, caution must be exercised when considering androgen replacement therapy for the aging male. Experience with hormone replacement therapy in postmenopausal women has shown that data from initial small series can be misleading or, at the very least, do not provide sufficient information until larger trials are performed. This was borne out recently by the data from the Women's Health Initiative Study. It is hoped that the future will offer the promise of further study in the matter of hormone replacement in the aging male and will delineate more clearly the risks and benefits of androgen supplementation.

## See Also the Following Articles

Aging and the Male Reproductive System • Constitutional Delay of Growth and Puberty (CDGP) • Delayed Puberty, Male • Estrogen and the Male • Gynecomastia • Hypergonadotropic Hypogonadism • Hypopituitarism, Hormonal Therapy for • Sexual Function and Androgens

## Further Reading

- Bagatell, C. J., and Bremner, W. J. (1996). Androgens in men: Uses and abuses. *N. Engl. J. Med.* **334**, 707–714.
- Bhasin, S. (1992). Clinical review 34: Androgen treatment of hypogonadal men. *J. Clin. Endocrinol. Metab.* **74**, 1221–1225.
- Griffin, J. E., and Wilson, J. D. (1998). Disorders of the testes and the male reproductive tract. *In* "Williams Textbook

- of Endocrinology" (J. D. Wilson, D. W. Foster, H. M. Kronenberg, and P. R. Larsen, eds.), 9th ed., pp. 819–876. W. B. Saunders, Philadelphia.
- Snyder, P. J. (2001). Effects of age on testicular function and consequences of testosterone treatment. *J. Clin. Endocrinol. Metab.* **86**, 2369–2372.
- Tenover, J. L. (1998). Male hormone replacement therapy including "andropause." *Endocrinol. Metab. Clinics North America* **27**, 969–987.
- Vermuelen, A. (2001). Androgen replacement therapy in the aging male: A critical evaluation. *J. Clin. Endocrinol. Metab.* **86**, 2380–2390.
- Wang, C., and Swerdloff, R. S. (2002). Should the nonaromatizable androgen dihydrotestosterone be considered as an alternative to testosterone in the treatment of the andropause? *J. Clin. Endocrinol. Metab.* **87**, 1462–1466.
- Zitzmann, M., and Nieschlag, E. (2000). Hormone substitution in male hypogonadism. *Mol. Cell. Endocrinol.* **161**(1–2), 73–88.

Table I Transdermal Hormone Patches

Name	Active hormone	Doses available
Alora	Estradiol 17- $\beta$	.05, .075, and 1 mg/day
Climara	Estradiol 17- $\beta$	.025, .05, .075, and .1 mg/day
Vivelle	Estradiol 17- $\beta$	.0375, .05, .075, and .1 mg/day
Vivelle-Dot	Estradiol 17- $\beta$	.0375, .05, .075, and .1 mg/day
CombiPatch	Estradiol 17- $\beta$ ; norethindrone acetate	50 $\mu$ g/140 $\mu$ g and 50 $\mu$ g/250 $\mu$ g

(e.g., Alora, Climara, Vivelle, Esclim), and each allows for delivery of varying doses of estradiol (0.025–0.1 mg/day) (Table I). Twice-weekly application of most of these patches is recommended due to the pharmacokinetics of these preparations. A gel containing estradiol is in development for marketing in the United States.

Estrogens can also be prescribed for transvaginal absorption with the indication of vaginal atrophy. Creams containing estradiol (e.g., Estrace cream) and conjugated equine estrogens (e.g., Premarin cream) are designed for application into the vagina. Estrogen-containing rings (e.g., Estring) and estradiol vaginal tablets (e.g., Vagifem) have also been developed for treatment of vaginal atrophy.

### TRANSDERMAL PROGESTIN THERAPY

There is no approved product for transdermal administration of progestin only. Progesterone-containing creams are marketed, but they have not been documented to prevent the increased risk of endometrial cancer that may result from unopposed estrogen treatment. Although the preparation of vaginal progesterone (e.g., Crinone) has not been traditionally used for hormone replacement, it remains a viable option that is under study for prevention of endometrial hyperplasia associated with estrogen replacement. The vehicle in this vaginal gel is an emulsion of water, oil, and polycarbophil polymer. This preparation has been known to cause some vaginal irritation in patients who use this product. However, vaginal administration of progestins may allow for an alternative treatment for those women who do not tolerate oral progestins.

### COMBINED THERAPIES ADMINISTERED TRANSDERMALLY

A patch that delivers both estradiol and a progestin is convenient for women with uteri; unopposed estrogen is not desirable in these women. One system (CombiPatch) delivers estradiol and a progestin, norethindrone (Table I). This patch is available in two

different dosing combinations. The hormones are contained within a matrix system similar to that of the newer estrogen patches. Twice-weekly application is required.

### TRANSDERMAL TESTOSTERONE THERAPY

Transdermal preparations for testosterone replacement have been marketed for men who experience hypogonadism. A transdermal testosterone patch that delivers lower and more physiological androgen levels for women is in development.

Testosterone has been thought to enhance libido and the sense of well-being, but the adverse effects of oral testosterone on circulating lipids has led to concerns about its use in women. Transdermal therapy may be optimal because it will alleviate a first-pass androgen effect on the liver; moreover, in one study, it led to no significant changes in circulating lipoprotein levels.

If studies support a benefit for testosterone replacement in women, it will be important to determine which groups of women will benefit from this therapy. It will also be essential to better delineate the risks of androgen replacement in women.

### CONCLUSIONS

Transdermal hormone replacement offers another avenue for hormone replacement for postmenopausal women and their physicians. The available preparations are outlined in Table I. This route of administration avoids a first-pass liver effect. In addition, new preparations that allow for the administration of a progestin along with the estrogen therapy are now available. Advances in this area of hormone replacement are likely to be greatly beneficial to this growing population.

### See Also the Following Articles

Breast Disease: Impact of Sex Steroid Replacement • Estrogen Replacement, Oral • Estrogen Replacement, Vaginal •

Hormone Replacement Therapy, Male • Hot Flash: Impact of Sex Steroid Replacement • Osteoporosis in Reproductive Age Women: Impact of Sex Steroid Replacement

### Further Reading

Archer, D. F., Furst, K., Tipping, D., Dain, M. P., and Vandepol, C. (1999). A randomized comparison of continuous combined transdermal delivery of estradiol-norethindrone acetate and

estradiol alone for menopause: CombiPatch Study Group. *Obstet. Gynecol.* **94**, 498–503.

Shifren, J. L., Braunstein, G. D., Simon, J. A., *et al.* (2000). Transdermal testosterone treatment in women with impaired sexual function after oophorectomy. *N. Engl. J. Med.* **343**, 682–688.

Slater, C. C., Hodis, H. N., Mack, W. J., Shoupe, D., Paulson, R. J., and Stanczyk, F. Z. (2001). Markedly elevated levels of estrone sulfate after long-term oral, but not transdermal, administration of estradiol in postmenopausal women. *Menopause* **8**, 200–203.

Table 1 The Clinical Continuum of a Hot Flash

Symptom/Sign	Description
Core temperature	A small (0.03°C) temperature elevation preceding the hot flash; temperature then drops (0.1–0.7°C) several minutes after the hot flash starts; felt as a chill
Sensation	Sudden feeling of intense heat sometimes accompanied by anxiety and feeling of suffocation
Heart rate	Increases by 5–35 beats per minute; sense of palpitations
Cutaneous blood flow	Increases due to vasodilatation; observed as flushing
Sweating	Rapid onset and profuse over head, face, and chest
Sleep disturbances	Night sweats

menopausal, whether naturally or as a result of interventions such as ovariectomy, chemotherapy, and radiation or following administration of gonadotropin-releasing hormone (GnRH) agonists and antagonists. Rarely, premenopausal women report experiencing hot flushes during the postpartum period. Hot flushes have also been reported by men on abrupt decline of androgens such as occurs following orchiectomy or treatment with GnRH agonists and antagonists.

Most epidemiological studies in North America and Europe have found that the majority of women had at least some hot flashes with the onset of menopause. The incidence of hot flashes is highest during the first postmenopausal years, ranging from 58 to 93%. The incidence of hot flashes in perimenopausal women ranges from 28 to 65%. Surgically induced menopause is characterized by a somewhat higher prevalence of hot flashes as compared with natural menopause. Hot flashes typically last from 6 months to 5 years after natural menopause, with an average duration of 2 years. There is a direct correlation between severity of symptoms and duration of symptoms; thus, moderate to severe vasomotor flashes tend to persist for a longer time.

There is a large cross-cultural variability in the reported prevalence of vasomotor symptoms among menopausal women. Hot flashes are reported to be much less common among Mayan women in Mexico, Indonesian women (10–20%), and Chinese women (10–25%). The reasons for the differences of prevalence between Eastern and Western countries might be attributed to different cultural perspectives of the events or to a phytoestrogens-rich diet.

No significant association has been found between the occurrence of hot flashes and sociodemographic variables such as employment status, social class, income rate, and marital status or between the occurrence of hot flashes and gynecological factors such as age of menarche and number of pregnancies. However, mean body weight seems to relate to hot

flashes, as asymptomatic women usually have higher mean body weight and higher blood level of total circulating estrogens than do women with hot flashes.

## PHYSIOLOGY OF HOT FLASHES

The underlying physiological mechanisms of the hot flash are not completely understood. Estrogen withdrawal at menopause is clearly the precipitating factor for hot flashes, but the relationship that might exist between estrogen and hot flashes is still illusive. Postmenopausal women suffering from severe vasomotor symptoms were found to have lower levels of circulating estrogens, and a smaller fraction of free estradiol (unbound to sex hormone-binding globulin), as compared with asymptomatic women. However, the presence of low blood estrogen level does not always lead to hot flashes. Other hypoestrogenic states, such as prepuberty, anorexia nervosa, and gonadal dysgenesis, are not characterized by hot flashes. It seems that the rapid decline in blood estrogen level is responsible for the occurrence of vasomotor symptoms. The onset of hot flashes after oophorectomy or after GnRH analogue administration supports this contention. So does the observation that although estrogen therapy usually improves hot flashes, they often return when estrogen therapy is discontinued. The influence of estrogen on temperature regulation is still not fully understood; however, it was demonstrated that estrogen can modulate the firing rate of thermosensitive neurons in the preoptic area of the hypothalamus in response to thermal stimulation in the rat. The responsiveness of vascular smooth muscle to catecholamines is modulated by estrogen, and it has been shown to be greater in women with hot flashes. Estrogen appears to enhance  $\alpha_2$ -adrenergic activity, so estrogen withdrawal may result in reduced  $\alpha_2$ -adrenergic activity and increased vasomotor function. Thus, estrogen may play both peripheral and central roles in the physiology of hot flashes. It was once believed that hot flashes were triggered by a sudden

downward resetting of the hypothalamic set point because there was no evidence of increased core body temperature. Recently, however, a central noradrenergic mechanism was asserted to be responsible for small core body temperature elevations that precede menopausal hot flashes. The thermoneutral zone within which sweating, peripheral vasodilatation, and shivering do not occur was virtually nonexistent in women with hot flashes. It was suggested that a small increase in core body temperature, acting within a narrowed thermoneutral zone, triggers a hot flash.

## DIFFERENTIAL DIAGNOSIS OF HOT FLASHES

Most perimenopausal women presenting with hot flashes do not require further investigation. Menstrual cycle irregularity usually corresponds with fluctuating hot flashes, but regular menses should not defer the diagnosis. In a few cases where the diagnosis of menopause is unclear, measurements of follicle-stimulating hormone (FSH) and luteinizing hormone (LH) may be of value. The levels of the gonadotropic hormones tend to fluctuate, especially in the premenopause; thus, repeated measurements may be needed. Conditions such as hyperthyroidism, panic attacks, carcinoid syndrome, pheochromocytoma, and niacin flush may mimic menopausal hot flashes and should be considered in the differential diagnosis of hot flashes that do not resolve following estrogen therapy.

## SEX STEROID THERAPY

The available therapies do not “cure” hot flashes; rather, they provide symptomatic relief. Sometimes, the hot flashes are totally eliminated, at least for the duration of the treatment; other times, they are significantly relieved and enable women to maintain an acceptable quality of life.

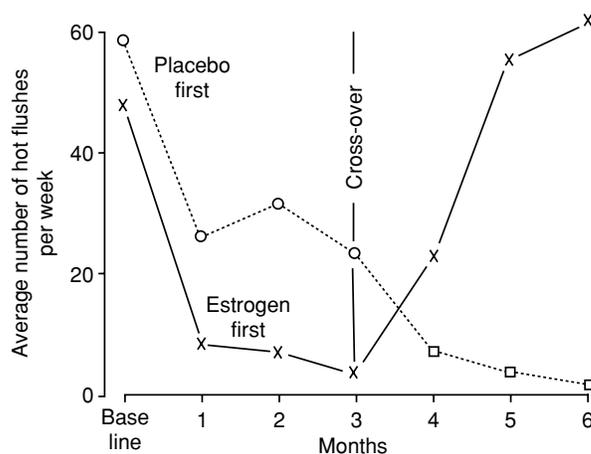
Comparison of various therapies for hot flashes with placebo demonstrates considerable effectiveness of the latter; therefore, all data regarding effectiveness of various drug regimens for hot flashes should originate from randomized, double-blind, placebo-controlled crossover studies.

Hormonal replacement therapy (HRT) is by far the most effective therapy for treating estrogen-deficient hot flashes. For women who have absolute contraindications to HRT, hot flashes can be reduced with other medications and alternative approaches.

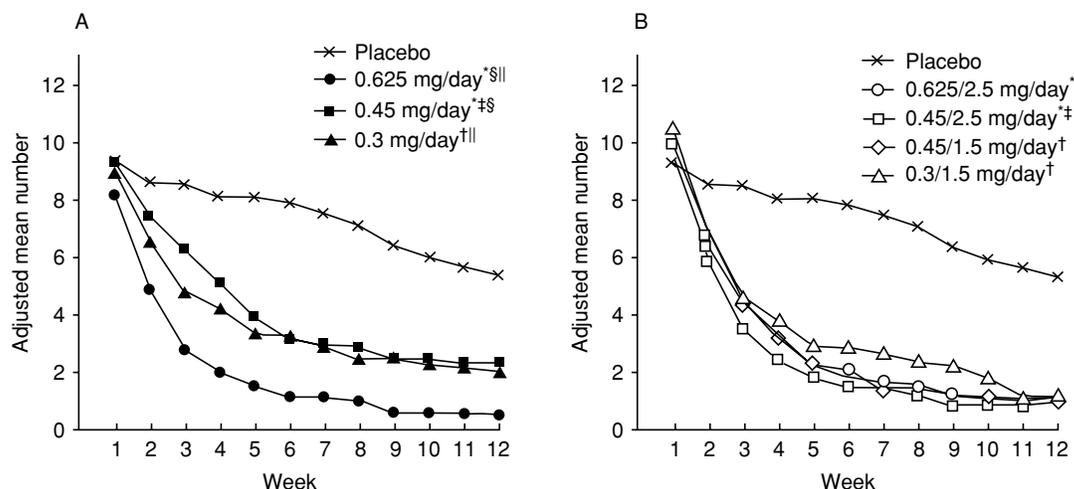
## Estrogen

Estrogen treatment has been used in many forms, including crude extracts, for approximately 100 years. It was suggested as a treatment for hot flashes due to the co-appearance of these symptoms with declining blood estrogen levels and not due to understanding the physiology and pathogenesis of hot flashes. The effect of estrogens on climacteric symptoms is not always noted immediately and might be fully realized only after several weeks. Women on a cyclic estrogen regimen might experience recurrence of hot flashes during the intervening days between two cycles of therapy. Sometimes, the hot flashes recur following complete discontinuation of the treatment, and it is not known whether the treatment only postpones the climacteric symptoms or whether the women would have had the hot flashes for that duration regardless of the treatment. However, the effect on hot flashes may persist for several weeks after discontinuation of therapy because some estrogen formulations, such as conjugated equine estrogen, are stored in fatty tissue that serves as a reservoir for the drug.

The “traditional” estrogen therapy for relief of hot flashes, which remains the most commonly used one in the United States, is oral conjugated equine estrogen (CEE) in doses ranging from 0.375 to 1.25 mg. Other oral preparations are available, including 1 to 2 mg of 17- $\beta$  estradiol. Transdermal preparations are also available and liberate 25 to 200  $\mu$ g/day of estradiol. Estrogen is also available as subcutaneous implants, injectables, and vaginal creams; most are effective at relieving climacteric symptoms. We



**Figure 1** Hot flush count during the 6-month trial. Reproduced by permission from the BMJ Publishing Group (Coope, J., Thomson, J. M., and Poller, L. (1975). Effects of “natural oestrogen” replacement therapy on menopausal symptoms and blood clotting. *Br. Med. J.* 4, 141).



**Figure 2** Mean daily numbers of hot flashes by week for the eight treatment groups. (A) The placebo and the CEE alone groups. (B) The placebo and the CEE/MPA groups. Data are adjusted for baseline. All values are for mean numbers of hot flashes, except CEE 0.3/MPA 1.5 at week 1. \*Difference from placebo was significant ( $P < 0.5$ ) at weeks 2 to 12. †, Difference from placebo was significant ( $P < 0.5$ ) at weeks 3 to 12. ‡, Difference between CEE 0.45 and CEE 0.45/MPA 2.5 was significant ( $P < 0.5$ ) at weeks 3, 4, 5, and 9. §, Difference between CEE 0.625 and CEE 0.45 was significant ( $P < 0.5$ ) at weeks 2 to 12. ||, Difference between CEE 0.625 and CEE 0.3 was significant ( $P < 0.5$ ) at weeks 4, 5, 6, 9, 10, and 12. Reproduced by permission from the American Society for Reproductive Medicine (Utian, W. H., *et al.* (2001). Lower HRT doses and vasomotor symptoms. *Fertil. Steril.* 75, 1070).

recently examined the efficacy of a vaginal ring delivering estradiol and progesterone in relieving climacteric symptoms and found that this modality is indeed effective in relieving hot flashes and might serve as a promising method for long-term HRT.

However, the most studied therapeutic regimen for hot flashes is the oral route, and it has been shown repeatedly that oral estrogen relieves approximately 90% of hot flashes within 3 months as compared with approximately 60% relief achieved with placebo (Fig. 1). Furthermore, estrogens were also efficient in relieving other climacteric symptoms such as vaginal dryness, sleep disturbances, anxiety and irritability, and memory loss. The first studies used doses as high as 1.25 mg of CEE (not commonly used today), whereas more recent studies demonstrated effectiveness in relieving hot flashes using doses as low as 0.375 mg of CEE or other estrogens in equivalent doses (Fig. 2). Long-term HRT is associated with adverse side effects such as breast tenderness and irregular vaginal bleeding, which are often given as reasons for discontinuation of therapy. To overcome this obstacle, a tendency toward lowering the dose used has been studied and encouraged recently. The hypothesis behind this attitude is based on the intuitive feeling that lowering the dose of HRT would be associated with reduced risk. There are indications from oral contraceptive experience that lowering the dose significantly decreased severe side effects, such as

thromboembolic phenomena and arterial occlusion events, without interfering with contraceptive effectiveness. For the time being, healthy, nonobese, postmenopausal women experiencing severe vasomotor symptoms should be treated with the lowest dose of estrogen shown to be effective.

Another route for delivering estrogens to relieve hot flashes is the transdermal approach. Avoiding the first pass in the liver is associated with significantly less effect on the angiotensin–renin system as well as with significantly less induction of triglyceride production. Thus, the transdermal approach may be the “route of choice” for hypertensive women, for those with hypertriglyceridemia, or for patients at increased risk for thromboembolism. Transdermal estrogen patches that provide continuous delivery of estrogens have been shown to be effective at relieving hot flashes in a dose–response manner. Transdermal patches delivering 25 to 200  $\mu\text{g}/\text{day}$  of estrogens were studied in regard to relieving hot flashes. All doses of estradiol reduced hot flashes; however, the highest dose (200  $\mu\text{g}/\text{day}$ ) was found to be the most efficient and resulted in 91% reduction in the number of hot flashes. The objectively monitored reduction in hot flashes has been reported to be higher than the subjectively reported reduction (85 vs 74%), and this is explained by the fact that estrogens relieve mainly hot flashes and, to a lesser degree, other climacteric symptoms that have a significant adverse effect on women’s well-being.

## Progesterone

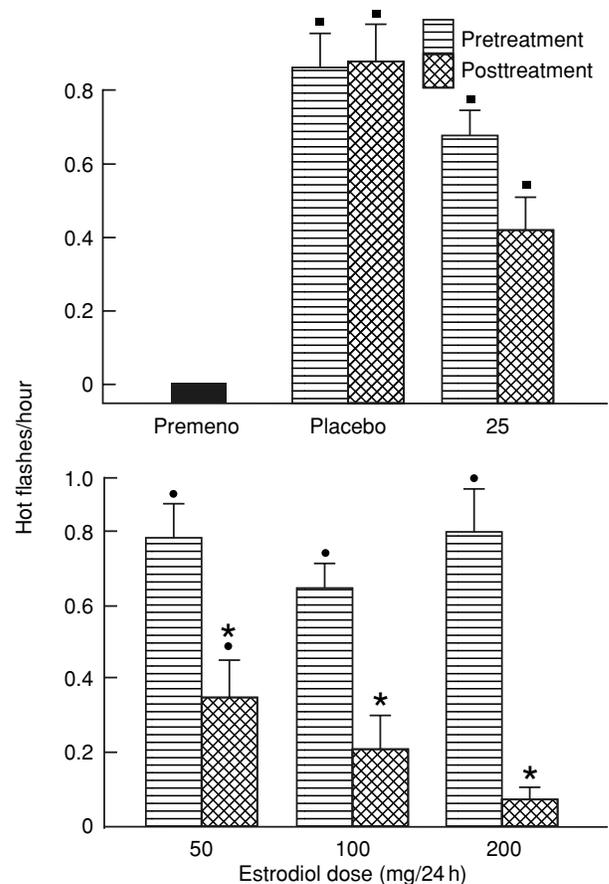
The combination of estrogen and progesterone is recognized as necessary treatment for postmenopausal women with intact uteri. The addition of progestin is aimed at providing protection against endometrial hyperplasia and adenocarcinoma. The inclusion of progesterone often raises concerns as to whether the beneficial effect of estrogen will be compromised. However, in regard to symptoms reduction, it has been suggested that progestin not only will not interfere with the estrogenic effect but might have independent benefits (Fig. 3). Several double-blind, placebo-controlled studies have shown that medroxyprogesterone (MPA), a C-21 nonestrogenic steroid, decreased the frequency of hot flashes. It has been shown that 150 mg/month of MPA injected intramuscularly reduced climacteric symptoms. A rate of 75% reduction effect was achieved with a 50 mg dose, rising to 90% following 150 mg of MPA by week 4 of treatment. Another study demonstrated a 74% decline in the number of hot flashes following 20 mg/day of MPA taken orally by the third month of treatment, as compared with a 26% reduction achieved with placebo. The same effect was achieved by megestrol acetate (MA), a synthetic derivation of natural progesterone, taken orally at a dose of 20 mg twice daily for 4 weeks. Norethindrone was also shown to significantly reduce hot flashes in a double-blind, crossover design study.

## Androgens

Cotreatment of androgens with estrogen is specifically intended to increase low libido. The use of this combination therapy may also be an option for patients whose vasomotor symptoms are not adequately controlled with estrogen monotherapy. The additive effect of methyltestosterone in estrogen replacement therapy was demonstrated in a double-blind placebo-controlled trial. The lower dose of estrified estrogens plus methyltestosterone provided greater relief of menopausal symptoms than did corresponding estrified estrogens, and the extent of relief was similar to that observed with a higher dose of estrified estrogens alone.

## CONCLUSIONS

Hot flashes are commonly experienced during menopause. For many women, vasomotor symptoms are severe enough to significantly compromise the overall quality of life and the general sense of well-being. Standard HRT is effective and results in alleviation of symptoms within days to a few weeks. Lower doses of estrogen treatment have also been shown to be



**Figure 3** Rates of occurrence (means  $\pm$  SEs) of hot flashes in the study groups and premenopausal women (Premeno) before and during transdermal estradiol administration. Reproduced by permission from The Endocrine Society (Steingold, K. A., Laufer, Y., Chetkowski, R. J., et al. (1985). Treatment of hot flashes with transdermal estradiol administration. *J. Clin. Endocrinol. Metab.* **61**, 630).

effective. Progesterone probably has an independent effect on alleviation of hot flashes. For some women, an estrogen–androgen combination therapy might provide sustained relief. Only modest and delayed improvement of symptoms could be expected by alternative nonhormonal preparations.

## See Also the Following Articles

Breast Disease: Impact of Sex Steroid Replacement • Estrogen Replacement, Oral • Estrogen Replacement, Vaginal • Hormone Replacement Therapy, Male • Hormone Replacement, Transdermal • Osteoporosis in Reproductive Age Women: Impact of Sex Steroid Replacement

## Further Reading

Bachmann, G. A. (1999). Vasomotor flushes in menopausal women. *Am. J. Obstetr. Gynecol.* **180**, S312–S318.

- Freedman, R. R. (2001). Physiology of hot flashes. *Am. J. Hum. Biol.* **13**, 453–464.
- Hochner-Celnikier, D. (2002). "Vaginal Ring Delivery of E+P." Society for Gynecological Investigation, Los Angeles.
- Kronenberg, F. (1999). Hot flashes. In "Treatment of the Postmenopausal Woman: Basic and Clinical Aspects" (R. A. Lobo, ed.), pp. 157–177. Lippincott Williams & Wilkins, Philadelphia.
- Lobo, R. A., and Whitehead, M. I. (2001). Is low-dose hormone replacement therapy for postmenopausal women efficacious and desirable? *Climacteric* **4**, 110–119.
- Mattox, J. H., and Shulman, L. P. (2001). Combined oral hormone replacement therapy formulations. *Am. J. Obstetr. Gynecol.* **185**, S38–S46.
- Stevenson, J. C., Crook, D., Godsland, I. F., Lees, B., and Whitehead, M. I. (1993). Oral versus transdermal hormone replacement therapy. *Int. J. Fertil.* **38**, 30–35.
- Utian, W. H., Shoupe, D., Bachmann, G., Pinkerton, J. V., and Pickar, J. H. (2001). Relief of vasomotor symptoms and vaginal atrophy with lower doses of conjugated equine estrogens and medroxyprogesterone acetate. *Fertil. Steril.* **75**, 1065–1079.

**Table 1 Methods for the Identification of Endocrine Controls of Eating**

If a particular control of eating ("the effect"; for example, meal initiation or satiation) is under the control of a hormone, then:

- Intravenous infusion of the hormone, or agonists or secretagogues for it, should produce the effect (for example, for satiation, perception of fullness and a reduction in meal size). Furthermore, infusion of doses that produce levels of the hormone that occur in the blood when the effect normally occurs (for example, for a satiation effect, at meal end) should be sufficient.
- Intravenous infusion of a selective and potent antagonist to the hormone at the time the effect normally occurs should reverse the effect (for example, if the hormone produces satiation, administration of the antagonist during the meal should increase meal size).
- Local infusion of the hormone and of its antagonist at the site of the hormone's action should produce the effect and reverse it, respectively.
- Removal of the hormone or of the receptors mediating the effect should prevent the effect (for example, for a satiation effect, increase meal size). Classically, this is done by removal of the gland producing the hormone; it is now possible to engineer transgenic animals lacking the gene for the hormone or its receptor.
- When the hormone but not its receptor, has been removed, replacement of the hormone in physiological patterns should normalize the effect.

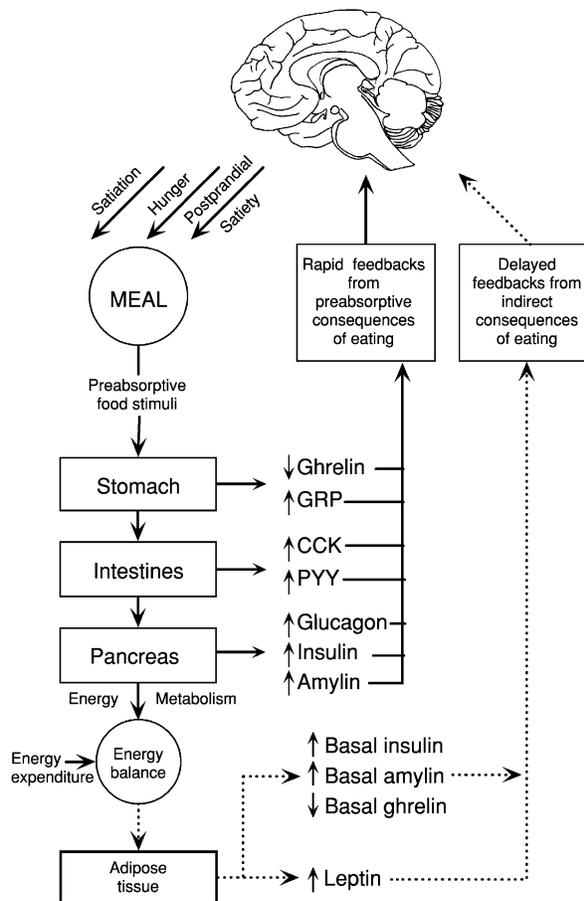
whether the organism gains or loses stored energy. In the adult, this translates into increases or decreases in body adiposity. This loop is closed by negative feedback signals related to body adiposity that affect eating behavior. Again, endocrine signals are the best understood of such feedback signals (see Fig. 1).

Finally, eating is also subject to influences, such as stress and reproduction, that are independent of feedback signals from ingested food. Endocrine signals are also important mediators of these influences.

## SATIATION

### Cholecystokinin

The gut peptide hormone cholecystokinin (CCK) is synthesized by endocrine cells dispersed along the small intestine and is released by preabsorptive food stimuli. CCK, which was discovered and named by the pioneering gut endocrinologist Andrew Ivy in 1928, has long been accepted as the major physiological control of gall bladder contraction and enzyme secretion from the exocrine pancreas. It now appears that CCK is also responsible for a substantial fraction of the satiating effect of ingested food.



**Figure 1** Schematic overview of endocrine feedback signals controlling eating. The diagram begins with eating (i.e., a meal) under way (upper left). The best-understood endocrine controls of eating are CCK, GRP, glucagon, and amylin, which are released from the gut during the meal by preabsorptive actions of ingested food signals and provide immediate direct negative feedbacks to the brain that contribute to meal termination (i.e., produce satiation) (arrows in solid lines). Absorbed nutrients affect energy balance and body adiposity, and several hormones are secreted in proportion to body adiposity. These contribute delayed negative feedbacks inhibiting eating, again apparently mainly by acting on brain processes of satiation (arrows in dashed lines). Some endocrine signals, such as ghrelin, PYY, and basal amylin, may affect hunger or postprandial satiety, but these contributions are not yet well understood.

### Exogenous CCK and Satiation

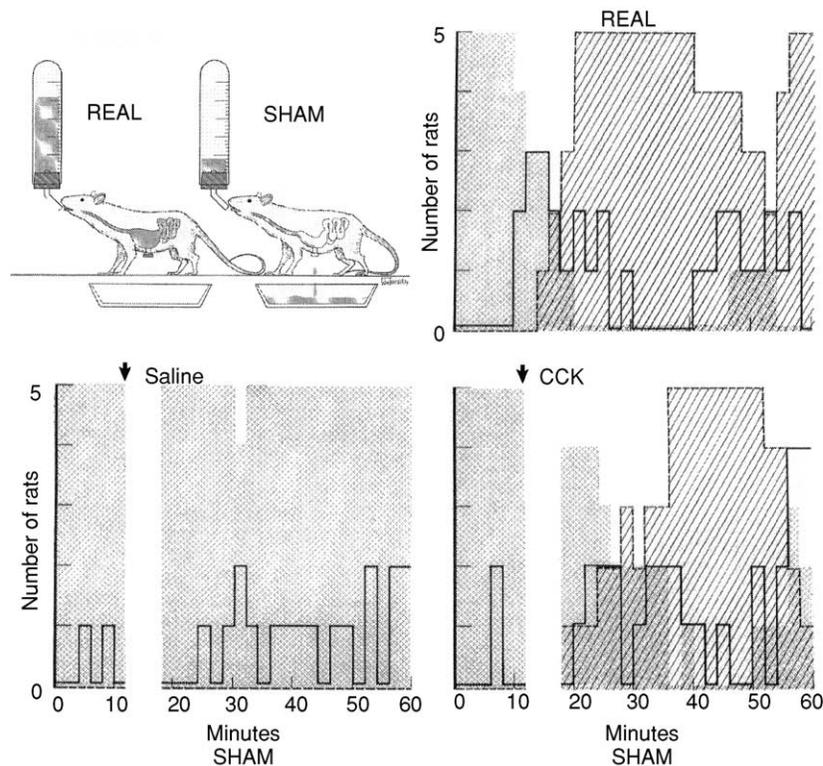
In 1973, Gerard Smith, James Gibbs, and their colleagues at Cornell University Medical College in New York began publication of a series of classic studies that both advanced the operational sophistication of the study of eating and identified CCK as a putative endocrine control of satiation. CCK is one of a number of gut hormones that is released during meals. These investigators intraperitoneally injected a CCK-rich porcine gut extract that had been prepared by Viktor Mutt of the Karolinska Institute

in Stockholm, Sweden. The injections elicited a dose-related decrease of liquid food intake in food-deprived rats, but the injections did not decrease water intake in water-deprived rats and did not elicit a conditioned taste aversion. Synthetic CCK-8, the carboxyl-terminal octapeptide of CCK, produced a similar behaviorally selective inhibition of eating. Neither desulfated CCK-8 nor gastrin, which is structurally similar to CCK, affected eating, in agreement with the structure-activity profiles of CCK's visceral effects. Finally, the investigators modified Pavlov's classical sham feeding paradigm to demonstrate that food-deprived rats with open gastric cannulas sham feed essentially indefinitely and that CCK is sufficient to elicit satiation in these animals (Fig. 2). Note that the sham feeding results can be considered analogous to the traditional gland removal technique because the ingested food does not reach the small

intestine to stimulate CCK release. Taken together, these data indicated that exogenous CCK could selectively elicit satiation. This has been confirmed in numerous animal species and in normal-weight and obese humans. In humans, intravenous infusions mimicking normal prandial increases are sufficient.

### Endogenous CCK and Satiation

The development of potent and selective CCK receptor antagonists during the 1980s made it possible to test whether endogenous CCK is normally part of the mechanisms by which food ingestion terminates meals. It is. In both animals and humans, premeal treatment with antagonists of the CCK-1 receptor (formerly called the CCK-A receptor) blocks the satiating action of exogenous CCK and, when injected alone, increases meal size. For example, intravenous



**Figure 2** CCK eliciting the behavioral sequence of satiation in sham-feeding rats. Meals of liquid diet were observed in 5 rats with surgically implanted gastric cannulas. The rats either “real” fed, with the cannulas closed, or “sham” fed, with the cannulas open so that the food drained without accumulating in the stomach or entering the intestines. Data are the numbers of rats during successive 2-min blocks that were feeding (gray bars), locomoting or grooming (solid lines), or resting (diagonally lined bars). In the real condition, rats fed for 10 to 20 min, locomoted and groomed for a few minutes, and finally rested. This is the behavioral sequence of satiation. In the sham condition, either saline or CCK was injected (arrows; observations were interrupted for 6 min following injections). Saline-injected sham-feeding rats did not satiate but fed almost continuously during the 60-min test. CCK-injected rats stopped feeding, groomed and locomoted, and rested just as did real-feeding rats. Modified with permission from Antin, J., Gibbs, J., Holt, J., Young, R. C., and Smith, G. P. (1975). Cholecystokinin elicits the complete behavioral sequence of satiety in rats. *J. Comp. Physiol. Psychol.* 89, 784-790.

infusion of a CCK-1 receptor antagonist that was begun just before volunteers began a lunch buffet consisting of a wide variety of appetizing foods significantly increased meal size (expressed as total energy content of the various foods) without affecting the volunteers' enjoyment of their meals or their subjective sense of normal satiation. These data establish CCK as a part of the normal physiological process of satiation in humans as well as in animals.

### Mechanism of CCK Satiation

Local infusion experiments reveal that the receptors initiating CCK's satiating action are located in the proximal small intestine. Their activation increases neural activity in the vagal afferent terminals at this site. This neural message is relayed to the nucleus tractus solitarius (NTS) in the caudal brainstem. From the NTS, it is relayed into local brainstem neural networks and into widespread forebrain neural networks. The CCK-1 receptor has both low- and high-affinity states; receptors in the low-affinity state mediate CCK's satiating action. Interestingly, in rats, intraperitoneal injections of CCK are more satiating than intravenous infusions of CCK, indicating that CCK's satiating action in the rat is via a paracrine mode of action rather than an endocrine one. CCK probably has an endocrine mode of action in humans and monkeys given that plasma CCK does increase during meals and intravenous infusions within the range of postprandial increases are effective. Another species difference is that CCK's actions on exocrine pancreas function are mediated by CCK-1 receptors in rats but are mediated by CCK-2 receptors in humans. This may facilitate development of a CCK agonist for therapeutic use in overeating.

### Disorders of CCK Satiation

Spontaneous mutations have been identified in the CCK-1 receptor. Rats without CCK-1 receptors overeat at every meal, gain weight, and become diabetic (Fig. 3). Humans without CCK-1 receptors apparently do the same. This syndrome indicates not only that CCK is an important part of the natural process of satiation but also that the physiological system controlling food intake, however complicated, is not completely redundant. Rather, the lack of a single basic control of meal size can produce uncompensated hyperphagia and obesity.

Disorders in CCK satiation occur in eating disorders. Patients with bulimia nervosa display reduced prandial CCK secretion and reduced subjective experience of satiation during meals. Although these

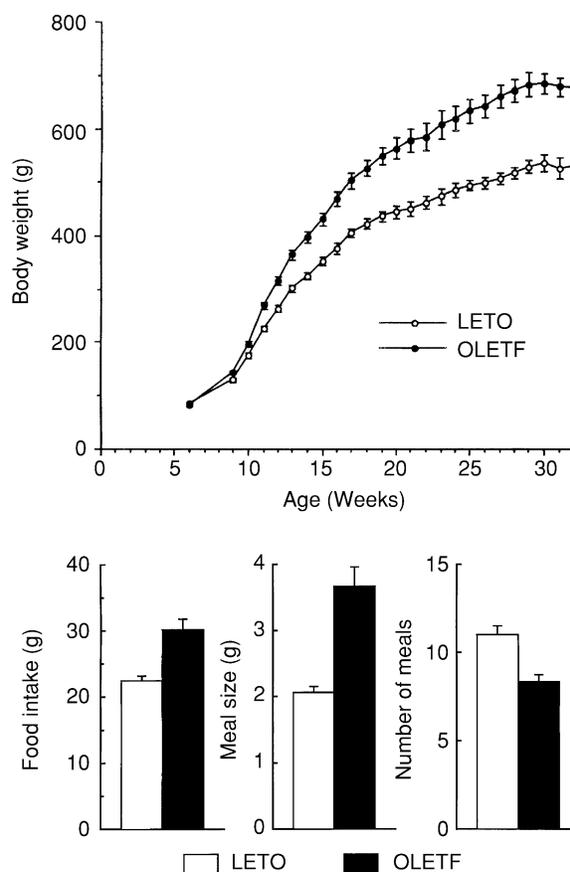
abnormalities improve together with improvement in binge eating, it is possible that they both facilitate the onset of the disorder and retard recovery from it.

### Other Gut Peptides in Satiation and Postprandial Satiety

CCK research has been paradigmatic for analyzing the effects of several other peptides that are released by the action of food in the gut and, at least under some circumstances, may control eating.

#### Pancreatic Glucagon

Glucagon is released from the pancreas during meals, apparently via a cephalic phase reflex, which is a



**Figure 3** Null mutation of the CCK-1 receptor leading to increased meal size and obesity. Top graph shows weights of normal male rats (LETO strain) and the same rats lacking the CCK-1 receptor (OLETF). Lower graphs show daily food intakes and spontaneous meal sizes and meal frequencies (means  $\pm$  SEMs) in the same rats at approximately 30 weeks of age. Modified with permission from Moran, T. H., Katz, L. F., Plata-Salaman, C. R., and Schwartz, G. J. (1998). Disordered food intake and obesity in rats lacking cholecystokinin A receptors. *Am. J. Physiol.* 274, R618–R625.

neuroendocrine reflex whose afferent receptors are olfactory, gustatory, or other receptors in the head. Glucagon administration elicits a behaviorally specific, dose-related reduction in meal size under several conditions, including tests of spontaneous feeding in rats. Glucagon's satiating action appears to be a physiological function, at least in rats, because antagonism of endogenous glucagon by prandial administration of specific antibodies increases meal size. Infusion of glucagon during meals also produces a selective satiating effect in humans. Glucagon satiation originates in the liver because hepatic portal vein glucagon infusions are more effective than vena cava infusions. Vagal afferents transmit a neural satiation signal from the liver to the brain. Whether glucagon's potent hepatic metabolic effects are the cause of this neural signal remains unclear.

### **Insulin**

Insulin secretion increases rapidly with eating as a result of cephalic and intestinal phase reflexes (the latter arise from excitation of intestinal receptors by preabsorptive food stimuli) and of direct substrate actions on the pancreatic beta cells. Acute insulin administration has been demonstrated to decrease meal size under some test conditions in rats, but it has been difficult to establish a reliable dose-response relationship and no satiating effect has been found in humans. Nevertheless, prandial antagonism of endogenous insulin by infusion of specific antibodies increases meal size in rats, suggesting that insulin has at least a permissive effect in normal satiation.

### **Amylin**

Amylin, or islet amyloid polypeptide, is synthesized by the pancreatic beta cells and is sequestered and released together with insulin. Amylin fulfills the criteria for a satiation signal in rats because prandial amylin administration selectively decreases meal size, whereas prandial antagonism of endogenous amylin with receptor antagonists increases meal size. Amylin's satiating action has not been tested in humans. Amylin's satiating action appears to originate through activation of receptors in the area postrema (AP), a structure just dorsal to the NTS with a permeable blood-brain barrier.

### **Gastrin-Releasing Peptide**

Gastrin-releasing peptide (GRP) is synthesized, stored, and secreted by enteric neurons in the stomach and intestines and acts locally via a neurocrine mode. Administration of GRP, or of GRP agonists such as bombesin, selectively reduces meal size in animals and

humans. Local administration studies suggest that the critical site of secretion and action is the stomach. Thus, GRP may mediate some of the satiating effect of the mechanical stimuli produced when ingested food fills the stomach. This peripheral action of GRP is transmitted to the brain via both vagal and spinal visceral afferents (which project indirectly to the NTS).

### **Peptide YY(3-36)**

Peptide YY(3-36) is a gut peptide released mainly from the ileum and colon. It has recently been reported that PYY(3-36) administration produces a dose-related inhibition of eating in rodents and humans. PYY(3-36)'s relatively slow postprandial release and the long-lasting inhibition of eating produced by PYY(3-36) administration suggest that this hormone may signal postprandial satiety rather than satiation. The physiological status of this effect has not yet been proven with antagonist experiments. PYY(3-36) may act centrally in the arcuate nucleus of the hypothalamus, where it is an agonist for the NPY Y2 receptor, an inhibitory presynaptic receptor on NPY neurons that stimulate eating. Thus, it may enter the same neural network as do leptin and insulin.

## **HUNGER**

Ghrelin, discovered in 1999, is the first candidate endocrine signal for hunger. It is the only hormone known whose infusion increases eating in animals or humans. Ghrelin is synthesized and released primarily by endocrine cells in the stomach and small intestine. Although the specific stimuli for ghrelin secretion are unknown, the pattern of plasma ghrelin levels is consistent with a hunger-inducing action. Ghrelin levels increase before meals and decrease after eating (in humans, ghrelin levels also decrease during the hours after midnight but increase again before breakfast). Food deprivation also increases ghrelin levels. It is not yet known whether infusion doses of ghrelin that mimic the premeal level are sufficient to increase eating or whether antagonism of endogenous ghrelin can decrease meal size.

## **ADIPOSIITY SIGNALS**

Body weight, which is influenced by food intake in an important way, feeds back onto the control of meal size via several hormones. One of these, leptin, is a secretory product of the adipose tissue. Three others,

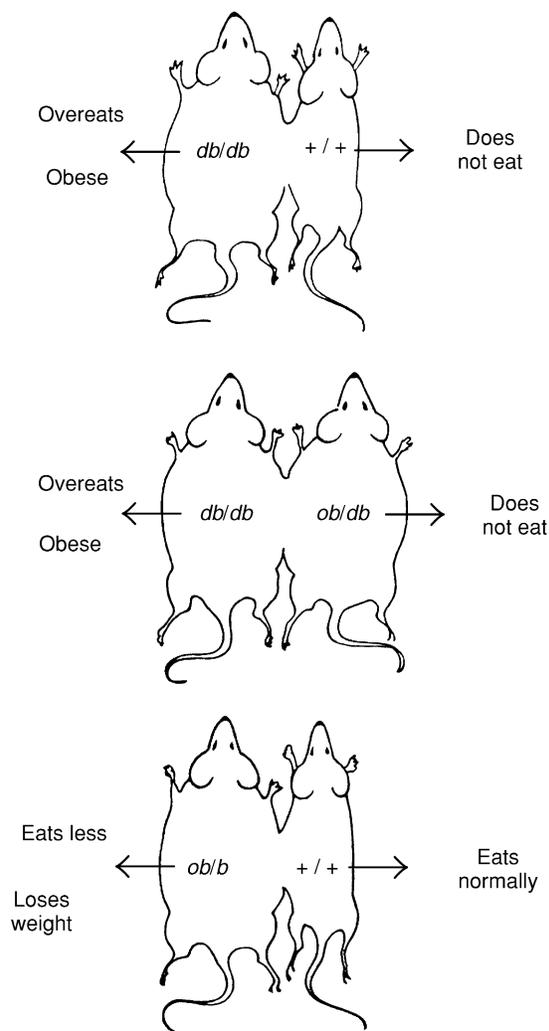
insulin, amylin, and ghrelin, are gut peptides whose secretion is linked to adipose tissue mass as well as to preabsorptive, meal-to-meal changes in the gut.

## Leptin

The search for a humoral signal that linked eating and adiposity was stimulated by early analyses of several spontaneous single-gene mutations that produced dramatic phenotypes in mice, especially the obese (*ob*) and diabetes (*db*) genes (Fig. 4). Obese (*ob/ob*) and diabetic (*db/db*) mice display identical syndromes of increased meal size, hyperphagia, obesity, and diabetes. These syndromes are now known to arise from disruptions of the leptin signaling pathway. The wild-type *ob* gene, *lep*, encodes the peptide leptin that is secreted by adipose (and other) cells, and the wild-type *db* gene, *lep-r*, encodes the leptin receptor. The leptin system appears to operate as a negative-feedback control of body fat because plasma leptin levels are closely correlated with body fat mass and because leptin administration decreases meal size, presumably by increasing satiation (Fig. 5). Leptin acts in the brain to reduce meal size; there are leptin receptors in several brain loci, leptin is transported across the blood–brain barrier by a specific transport mechanism, and leptin administration directly into the cerebral ventricles, the hypothalamus, or the NTS area of the caudal brainstem inhibits eating. Furthermore, at least in the short term, central administration of leptin antagonists stimulates feeding. Leptin also appears to have other functions related to nutritional status, including reducing the neuroendocrine responses to food deprivation and regulating the onset of puberty.

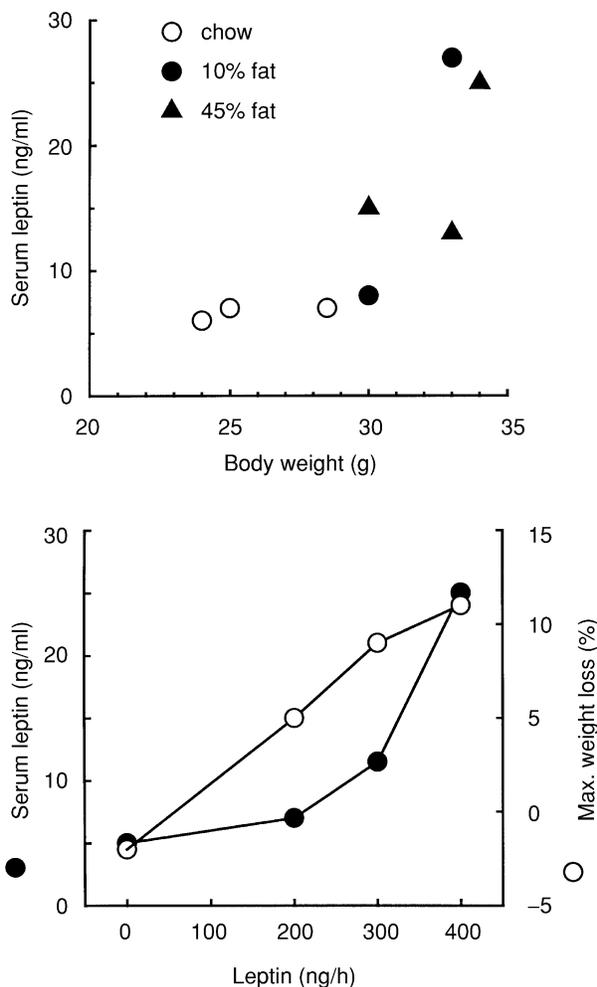
Leptin's actions in the hypothalamus have received particular attention. Leptin receptors in the arcuate nucleus of the hypothalamus are on neurons that express various interneuronal signaling molecules that have been linked to the control of feeding, including NPY,  $\alpha$ -MSH, and agouti-related peptide, and a great deal of neuroanatomical, neuropharmacological, and neurophysiological information has been amassed about the operation of the neural networks in which these neurons function. This information has not yet shown exactly how leptin reduces meal size. Leptin may do so by influencing the neural networks that determine the satiating potencies of CCK and perhaps of other gut peptides discussed previously.

A challenge for the hypothesis that leptin operates as a negative-feedback regulator of body weight is that animals and humans with apparently normal leptin systems can easily be induced to become obese, for example, by the easy availability of palatable,



**Figure 4** Effects of parabiosis of normal mice and mice with mutations of the leptin signaling system. Parabiotic mice are surgically joined so that blood contents are shared. Each mouse continues to eat independently, and parabiosis of like-genotype mice has no effect on feeding and weight gain. Hyperphagic, obese *db/db* mice continue to overeat when joined with wild type (*+/+*) mice or *ob/ob* mice, but their *+/+* and *ob/ob* partners stop eating and starve. In contrast, when *ob/ob* mice are joined with *+/+* mice, their feeding normalizes and they lose weight. These results suggest that *ob/ob* mice lack a hormone controlling eating and that *db/db* mice lack the hormone's receptor. This hormone was subsequently shown to be leptin. Modified with permission from Coleman, D. L. (1978). Obese and diabetes: Two mutant genes causing diabetes–obesity syndromes in mice. *Diabetologia* **14**, 141–148.

energy-rich foods. Furthermore, obese hyperleptinemic animals are less responsive to exogenous leptin, perhaps because leptin transport into the brain is reduced. Thus, leptin (as well as other adiposity signals controlling eating) may be a relatively ineffective



**Figure 5** The physiology of the leptin system. As shown in the upper graph, circulating leptin level increases with increased body weight, brought about by feeding mice regular chow or chow enriched with 10 or 45% fat, which they overeat. As shown in the lower graph, continuous subcutaneous leptin infusion in doses that produce leptin levels in the physiological range decrease body weight, which in similar tests was shown to result mainly from decreased food intake. Such data indicate that endogenous leptin is a physiological control of eating. However, note that this control is insufficient to reverse the increased adiposity produced by high-fat feeding. Modified with permission from Van Heek, M., Compton, D. S., France, C. F., Tedesco, R. P., Fawzi, A. B., Graziano, M. P., Sybertz, E. J., Strader, C. D., and Davis, H. J., Jr. (1997). Diet-induced obese mice develop peripheral, but not central, resistance to leptin. *J. Clin. Invest.* 99, 385–390 (upper graph); and Halaas, J. L., Boozer, C., Blair-West, J., Fidathusein, N., Denton, D. A., and Friedman, J. M. (1997). Physiological response to long-term peripheral and central leptin infusion in lean and obese mice. *Proc. Natl. Acad. Sci. USA* 94, 8878–8883 (lower graph).

controller of body weight whose potency can be overwhelmed by stimuli that increase feeding. A few humans with homozygous mutations of the leptin or leptin receptor gene have been identified; not

surprisingly, these individuals display dramatic syndromes such as *ob/ob* and *db/db* mice. Whether less complete defects in the leptin system occur commonly in human obesity and whether manipulation of the leptin system might be therapeutically useful have not yet been determined.

## Insulin

Insulin appears to be another endocrine signal linking eating and adiposity. Basal plasma insulin levels vary in proportion to body adiposity, an active transport mechanism carries peripheral insulin through the blood–brain barrier, there are insulin receptors in the arcuate nucleus of the hypothalamus and other brain areas, local administration of insulin into the hypothalamus decreases eating and body weight in several species (including nonhuman primates), and intracerebral administration of insulin antagonists increases food intake in rats and sheep. It is also noteworthy that centrally administered insulin can inhibit feeding without eliciting the hormone's potent anabolic effects in peripheral tissues. Whether selective defects in the brain insulin-signaling pathway can lead to overeating and obesity is unknown. Because adipocytes require insulin to deposit fat, weight gain cannot occur during peripheral insulin insufficiency, even if food intake increases, as occurs in uncontrolled diabetes mellitus.

In general, basal insulin appears to act like leptin in the control of appetite. Like leptin, its main effect is on meal size and it appears to increase CCK's satiating potency. Also similar to leptin, it appears that insulin is a less potent inhibitory influence on appetite than the stimulatory influences that so easily induce obesity and that obesity-related hyperinsulinemia produces insulin tolerance. One important functional difference between the two hormones' actions on feeding has emerged, at least in rodents. Insulin apparently acts more potently than leptin on eating in male animals, whereas leptin acts more potently than insulin on eating in female animals.

## Amylin

Amylin is cosecreted with insulin, and basal amylin levels also increase when body adiposity increases. Recent evidence indicates that basal amylin provides another negative-feedback adiposity signal to the brain. Chronic intracerebral infusion of amylin decreased food intake and body adiposity in rats, and similar infusion of an amylin antagonist had the opposite effects. Interestingly, this action of amylin

may be related to decreases in meal number rather than to decreases in meal size, as produced by leptin and insulin.

## HYPOTHALAMIC–PITUITARY–GONADAL AXIS

### Estradiol and Increased Satiation

Hormones of the hypothalamic–pituitary–gonadal (HPG) axis, in addition to controlling the physiological operation of the reproductive system, act on the brain to affect behavior. For example, female rats and mice are sexually receptive and mate only during the brief periovulatory period (estrus) of the cycle, and this responsiveness is controlled by estradiol and progesterone. Estradiol, but not progesterone, also controls eating. Cycling female rats' food intake is less during estrus than during other days of the cycle, food intake following ovariectomy is tonically increased (producing increases in body adiposity), and a physiological pattern of estradiol replacement normalizes both eating effects and body weight. Estradiol's effects on eating are expressed as changes in meal size; meal frequency changes tend to compensate for the meal size effect. Furthermore, at least in rats, estradiol's effects are not due to decreases in the palatability of sweet or fat foods. Thus, estradiol increases satiation.

Food intake is also decreased during the periovulatory phase of nonhuman primates and women. In women, this amounts to a decrease of a few hundred calories per day. In some studies, this decrease appears to occur throughout the follicular phase of the cycle (i.e., 10 days or more) and so could have body weight implications. Because estradiol increases throughout the follicular phase until ovulation, it is likely that the follicular and periovulatory decreases in eating are due to estradiol, but this has not been proven.

The ovarian cycles of rats and mice end within hours of ovulation, without the prolonged luteal phase that occurs in the ovarian (menstrual) cycles of primates. Some data indicate that food intake may increase during the luteal phase in nonhuman primates and women. It has been suggested that this increase is due to the luteal increase in progesterone. Pharmacological doses of progesterone have some stimulatory effect on eating in rodents, but no feeding effect of a physiological dose of this hormone has ever been documented.

It is unclear whether there is any human parallel to the dramatic increase in eating produced by ovariectomy (oophorectomy) in animals. Women do

increase body fat content by 1 to 2 kg (and lose nearly as much lean body tissue) across the menopausal transition, but changes in eating have not been documented. Estradiol and other HPG hormones have metabolic effects that contribute to the weight gain effects of ovariectomy in animals and may account for the changes in body composition in postmenopausal women.

Food selection may also be influenced by ovarian hormones in women, especially in women with the syndrome of premenstrual (luteal) dysphoria. Such effects appear to represent individuated brain responses to normal steroidal hormone signals that are unlike the normal responses to varying hormone signals that produce estradiol's effects on meal size.

### Mechanism of Estradiol's Satiating Effect

In rats, estradiol increases the satiating potency of CCK and glucagon. Thus, like leptin and insulin, estradiol may affect feeding by an effect on the central neural processing of gut hormone satiation signals, perhaps in the NTS.

Estradiol interacts with the classic estradiol receptor, ER $\alpha$ , to inhibit feeding. ER $\alpha$  is widespread in the brain, and which population(s) initiates its feeding effects is unknown, although it is clear that it is not the ventromedial hypothalamus, where estradiol acts to stimulate sexual receptivity.

### Pathophysiology

Eating disorders, including anorexia nervosa, bulimia nervosa, and binge eating disorder, occur much more often in women than in men. They are also associated with disorders in the subjective experiences of hunger and satiation as well as disordered eating during meals. In addition, they are associated with disorders in endocrine processes underlying satiation such as CCK secretion. Because these same endocrine processes mediate some or all of estradiol's satiating effects, it is possible that disorders of these mechanisms may be part of the pathophysiology of eating disorders and part of the reason for women's increased vulnerability to eating disorders.

## HYPOTHALAMIC–PITUITARY–ADRENAL AXIS

### Stress Anorexia

The classical hypothalamic–pituitary–adrenal (HPA) axis stress response consists of release of corticotropin-releasing hormone (CRH) from the

hypothalamus into the hypophyseal portal system, leading to adrenocorticotrophic hormone (ACTH) release from the anterior pituitary, and then glucocorticoid release from the adrenal cortex (the primary glucocorticoid in humans is cortisol and in most rodents is corticosterone). Stress also stimulates the sympathetic nervous system, leading to release of epinephrine from the adrenal medulla. The possibility that the decreases in eating associated with stress (including normal stresses such as exercise) are caused by these hormones has been investigated extensively. Although administration of each of these hormones has inhibited eating in some animal models, a clear picture of the endocrine control of “stress anorexia” has yet to emerge. Undoubtedly, both the farrago of experimental treatments and natural situations that are subsumed under the rubric of “stress” and the highly interactive actions of the HPA axis with the autonomic nervous system, metabolic regulation, and immune system contribute to this situation.

## CRH

CRH is secreted into the hypophyseal portal system by neurons with cell bodies in the parvocellular paraventricular nucleus (PVN). CRH-containing neurons, however, also project to other brain sites, including other hypothalamic areas and the NTS, and local administration of CRH into the brain has been shown to inhibit eating. Such inhibition can occur independent of HPA axis activation, suggesting that CRH may control eating in situations not involving stress. These actions of CRH apparently represent neurotransmitter functions rather than endocrine functions.

The recent discovery of another endogenous ligand for CRH receptors, urocortin (UCN), has led to the hypothesis that UCN mediates some or all of these neural actions of CRH. In some studies, UCN has affected eating more potently than has CRH, whereas CRH has affected other stress-related responses more potently than has UCN. Such functional selectivity may occur at the receptor level. CRH-2 receptors have a higher affinity for UCN than does CRH, and CRH-1 receptors have a higher affinity for CRH than does UCN.

## Glucagon-like Peptide-1

Glucagon-like peptide-1 (GLP-1) is another molecule with both peripheral hormonal and central neurotransmitter effects. Hormonal GLP-1 is synthesized and released from gut endocrine cells and appears to

act as an incretin (a hormone acting to increase insulin secretion). Neuronal GLP-1 is synthesized by a small cluster of neurons in the caudal brainstem that project to the hypothalamus. This GLP-1 appears to be a key neurotransmitter in the neural network, mediating the anorexia produced by many “stressful” treatments.

## Oxytocin and Arginine Vasopressin

The posterior pituitary hormones arginine vasopressin and oxytocin are increased by various stress treatments, and both can produce anorexia. Because both hormones also affect ACTH secretion, their anorectic actions may be related to HPA axis function. But both molecules also occur in neurons. Oxytocinergic neurons project from the PVN to the brainstem, where they appear to produce anorexia by activating the neuronal GLP-1 system described previously.

## Glucocorticoids and Obesity

### **Cushing’s Syndrome**

Glucocorticoids play a crucial role in a number of obesity syndromes. The hyperphagia and obesity of all gene mutation models, including the *ob/ob* and *db/db* mice discussed previously, are reversed by adrenalectomy and are reinstated by corticosteroid treatment. Cushing’s syndrome, which includes insulin resistance, decreased glucose tolerance, hyperlipidemia, and central (visceral) obesity, results from increases in glucocorticoid secretion. Such increases are usually secondary to increased CRH or ACTH secretion, for example, as can be caused by pituitary tumors. Cushing’s syndrome can also result from an aberrant glucocorticoid response to gastric inhibitory peptide secreted during meals, suggesting that overeating alone may also contribute to it.

### **Metabolic Syndrome**

Central obesity, insulin resistance, decreased glucose tolerance, and hyperlipidemia most often occur in the absence of elevated plasma glucocorticoids. This has been called the metabolic syndrome. Recent work indicates that obesity and metabolic syndrome can be produced by disordered intracellular metabolism of glucocorticoids. The enzyme 11 $\beta$ -hydroxysteroid dehydrogenase type 1 regulates the amount of glucocorticoid reaching its intracellular receptors by controlling the interconversion of active glucocorticoids and their inactivated forms. Transgenic mice that overexpress this enzyme only in white adipose tissue developed central obesity, metabolic syndrome, and increased plasma leptin and insulin but maintained

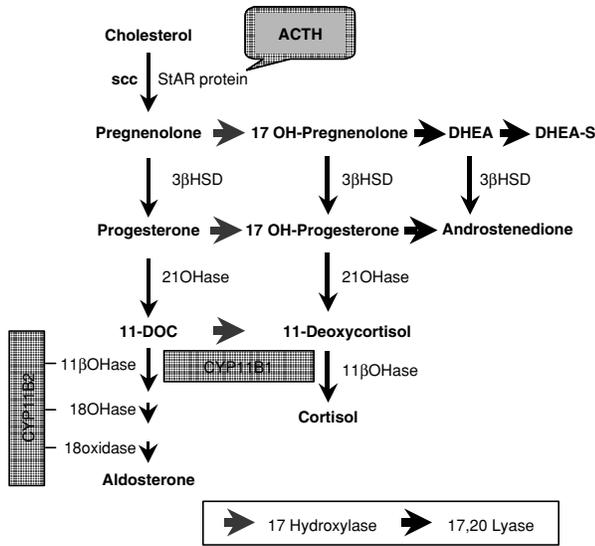
normal plasma glucocorticoid levels. Thus, hepatic glucocorticoid metabolism may be a factor in human obesity in the absence of any sign of endocrine dysfunction. The transgenic mice also overate, perhaps in response to decreased circulating metabolic fuels secondary to increased lipogenesis.

### See Also the Following Articles

Anorexia Nervosa • Caloric Restriction, Aging and Oxidative Stress • Eating Disorders and the Reproductive Axis • Obesity, Childhood and Adolescence • Obesity Regulation

### Further Reading

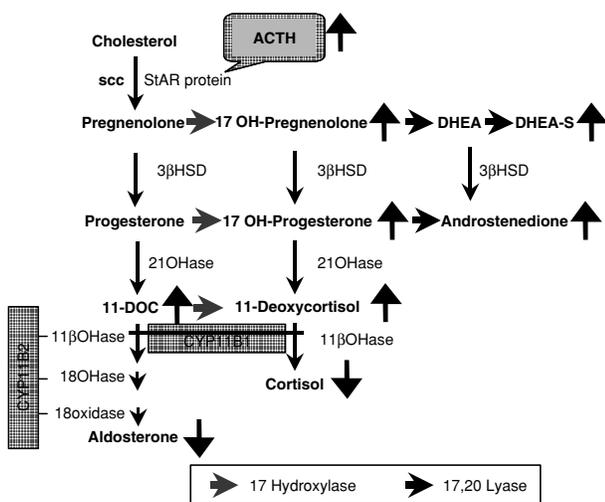
- Asarian, L., and Geary, N. (2002). Cyclic estradiol treatment normalizes body weight and restores physiological patterns of spontaneous feeding and sexual receptivity in ovariectomized rats. *Horm. Behav.*, **42**, 461–472.
- Batterham, R. L., Cowley, M. A., Small, C. J., Herzog, H., Cohen, M. A., Dakin, C. L., Wren, A. M., Brynes, A. E., Low, M. J., Ghatei, M. A., Cone, R. D., and Bloom, S. R. (2002). Gut hormone PYY(3–36) physiologically inhibits food intake. *Nature* **418**, 650–654.
- Beglinger, C., Degen, L., Matzinger, D., D’Amato, M., and Drewe, J. (2001). Loxiglumide, a CCK-A receptor antagonist, stimulates calorie intake and hunger feelings in humans. *Am. J. Physiol.* **280**, R1149–R1154.
- Cummings, D. E., Weigle, D. S., Frayo, R. S., Breen, P. A., Ma, M. K., Dellinger, E. P., and Purnell, J. Q. (2002). Plasma ghrelin levels after diet-induced weight loss or gastric bypass surgery. *N. Engl. J. Med.* **346**, 1623–1630.
- Geary, N. (1997). Appetite. In “Encyclopedia of Human Biology” (R. Dulbecco, ed.), Vol. 1, pp. 433–443. Academic Press, San Diego.
- Geary, N. (1999). Effects of glucagon, insulin, amylin, and CGRP on feeding. *Neuropeptides* **33**, 400–405.
- Geary, N. (2001). Estradiol, CCK, and satiation. *Peptides* **22**, 1251–1263.
- Geary, N., and Smith, G. P. (2000). Appetite. In “Kaplan and Sadock’s Comprehensive Textbook of Psychiatry” (B. J. Sadock and V. A. Sadock, eds.), 7th ed., Vol. 1, pp. 209–218. Lippincott Williams & Wilkins, Philadelphia.
- Grill, H. J., and Kaplan, J. M. (2002). The neuroanatomical axis for the control of energy balance. *Front. Neuroendocrinol.* **23**, 2–40.
- Masuzaki, H., Paterson, J., Shinyama, H., Morton, N. H., Mullins, J. J., Seckl, J. R., and Flier, J. S. (2001). A transgenic model of visceral obesity and the metabolic syndrome. *Science* **294**, 2166–2170.
- Moran, T. H. (2000). Cholecystokinin and satiety: Current perspectives. *Nutrition* **16**, 858–865.
- Rinaman, L., and Rothe, E. E. (2002). GLP-1 receptor signaling contributes to anorexigenic effect of centrally-administered oxytocin in rats. *Am. J. Physiol.* **283**, R99–R106.
- Rubinow, D. R., Schmidt, P. J., Roca, C. A., and Daly, R. C. (2002). Gonadal hormones and behavior in women: Concentration versus context. In “Hormones, Brain, and Behavior” (D. Pfaff ed.), vol. 5, pp. 37–73. Academic Press, San Diego.
- Smith, G. P. (ed.) (1998). “Satiation: From Gut to Brain.” Oxford University Press, New York.
- Woods, S. C., and Stricker, E. M. (1999). Food intake and metabolism. In “Fundamental Neuroscience” (M. J. Zigmond, F. E. Bloom, S. C. Landis, J. L. Roberts, and L. R. Squire, eds.), pp. 1091–1109. Academic Press, San Diego.



**Figure 1** CYP11B1 and CYP11B2 deficiencies and their relations to steroid biosynthesis. SCC, side chain cleavage enzyme; StAR, steroidogenic acute regulatory protein; 3 $\beta$ HSD, 3 $\beta$ -hydroxysteroid dehydrogenase; 21-OHase, 21-hydroxylase; 11 $\beta$ OHase, 11 $\beta$ -hydroxylase; 18-OHase, 18-hydroxylase.

nonclassical form. The renin-angiotensin pathway is the main route by which CYP11B2 is activated. Therefore, since the excess DOC suppresses plasma renin activity, serum aldosterone levels are low, despite completely normal CYP11B2 enzyme activity.

As mentioned previously, CYP11B2 catalyzes three reactions in the zona glomerulosa of the adrenal gland: 11 $\beta$ -hydroxylase, corticosterone methyl-oxidase I (CMO-I) (i.e., 18-hydroxylase), and corticosterone methyl-oxidase II (CMO-II) (i.e., 18-oxidase).



**Figure 2** Alterations in steroid biosynthesis caused by CYP11B1 deficiency. Vertical bold arrows indicate increase or decrease in serum concentration. Abbreviations are as in Fig. 1.

Deficiencies of both CMO-I and CMO-II cause salt wasting without virilization. Since the genetics and clinical syndromes are entirely different for CYP11B1 and CYP11B2 deficiencies, these aspects are discussed separately.

### CYP11B1 DEFICIENCY: MOLECULAR GENETICS

The CYP11B1 and CYP11B2 enzymes are encoded for by genes on chromosome 8, with each containing nine exons encompassing approximately seven kilobase pairs of DNA. They are relatively close to each other (40 kb apart) and are 95% homologous. CYP11B1 is regulated by the action of ACTH on the surface of cells in the adrenal cortex. The ACTH receptor is a typical seven-transmembrane G protein-coupled receptor. G stimulatory protein activation increases the activity of adenylate cyclase, with a subsequent increase in cyclic AMP. Cyclic AMP then binds to the appropriate response elements in the CYP11B1 promoter, thus activating mRNA expression and protein translation.

The mutations in CYP11B1 mainly cluster in exons 2 and 6–8 (Table I). The majority cause classical CAH due to a completely nonfunctional 11 $\beta$ -hydroxylase protein. A large group of patients of Moroccan Jewish descent bare a unique common point mutation that causes a change in amino acid 448 from arginine to histidine (R448H). In offspring of Moroccan Jews living in Israel, the incidence of this disorder is 1 in 5000–7000, with an allele frequency of 1 in 70–84 and a carrier frequency of 1 in 35–42. Considering this ethnic group’s isolated existence in the Atlas Mountains of Morocco, with a high rate of inbreeding for approximately two millennia, this common mutation is probably caused by a founder effect. The same mutation has also been identified in other populations, and a different mutation at the same amino acid residue has been identified (R448C), thus making amino acid 448 a hot spot. Other described defects are mostly in regions of the gene important for enzyme function—substrate binding, heme iron binding, binding of oxygen molecules, and proton transfer to the bound oxygen. Like other cytochrome P450 adrenal enzymes, both forms of the 11 $\beta$ -hydroxylase enzymes depend on electrons transported from NADPH for their hydroxylation activity. These electrons are transferred to adrenodoxin reductase and then to adrenodoxin and finally to the enzyme. Gene mutations in regions important for adrenodoxin binding result in total inactivation of the enzyme.

Table 1 Mutations in CYP11B1 and Their Effects<sup>a</sup>

Exon	Mutations causing classic CAH	Nonclassic mutations
1	P32del1nt	P42S
2	L105del28nt, W116X, Q121ins5nt, V129M	None reported
3	K174X	N133H
4	W247X	None reported
5	T318M	None reported
6	A331V, Q338X, Q356X, E371G, R374Q	T319M
7	R384Q, R384G, N394ins2nt	None reported
8	Y423X, L464ins3nt, V441G, R448H (Moroccan Jewish), R448C	None reported
9	None reported	None reported
Multiple	Hybrid gene; promoter and exons 1–6 of CYP11B2 and exons 7–9 of CYP11B1	

<sup>a</sup>Abbreviations used: del, deletion; ins, insertion; nt, nucleotides. Amino acid mutations: A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine; X, stop codon.

In a particularly interesting mutation, a hybrid gene is formed that includes the promoter and exons 1–6 of CYP11B2 and exons 7–9 of CYP11B1. The resulting enzyme has *in vitro* activity of both CYP11B1 and CYP11B2 but is not expressed in the zona fasciculata since the promoter and regulatory elements of CYP11B1 are replaced by those of CYP11B2.

An interesting phenomenon reported by our group is secondary 11 $\beta$ -hydroxylase deficiency. This occurs together with 21-hydroxylase deficiency and is probably caused by a local inhibitory effect of the excess androgen caused by the primary 21-hydroxylase defect.

## CLINICAL PRESENTATION: CLASSICAL 11 $\beta$ -HYDROXYLASE CAH

### Diagnosis

Most females with the severe or classical form of CAH are diagnosed in the neonatal period after presenting with ambiguous genitalia at birth or in early childhood. Occasionally, females are so severely virilized at birth that their external genitalia are male looking, with a penile urethra and fused labioscrotal folds. This leads to errors in gender assignment. Males, on the other hand, have normal genitalia at birth, but

marked virilization becomes evident (with or without associated hypertension) in childhood. The evaluation of ambiguous genitalia includes verification of the karyotype. A rapid fluorescent *in situ* hybridization for Y chromosome sequences can be obtained in many clinical laboratories within 24–48 h. Negative findings on this test suggest that the patient's genome lacks the Y chromosome, and this must be confirmed by a complete karyotype assessment. At the same time that blood is drawn for genetic studies, biochemical testing is also done. One should determine cortisol, 11-deoxycortisol, 17-hydroxyprogesterone, DOC, aldosterone, DHEA-S, androstenedione, and testosterone concentrations as well as plasma renin and aldosterone, which are particularly useful when hypertension is suspected. Biochemical diagnosis of 11-hydroxylase deficiency is made when the relative values of 11-deoxycortisol/cortisol and DOC/aldosterone are more than three standard deviations higher than the average before or 1 h after intravenous stimulation with 250  $\mu\text{g}/\text{m}^2$  aqueous ACTH.

As previously noted, there is a good association between baseline and stimulated steroid levels. In our experience, baseline levels of the appropriate steroid metabolites are usually sufficiently elevated to allow for a correct diagnosis to be made in classical CAH.

Male patients are not usually diagnosed early in childhood unless either a previous index female case has occurred in the family or hypertension is noted. Rarely, a neonatal salt-wasting crisis can occur. In most cases, male children present with rapid somatic growth, penile but not testicular enlargement, and growth of pubic hair. X-ray of the left hand and wrist will confirm advanced bone age. Biochemical testing in such cases should be performed as for female patients.

### Treatment

After biochemical testing has been performed, treatment should be initiated, even prior to receipt of the results, with 15–25  $\text{mg}/\text{m}^2$  of hydrocortisone in order to affect replacement of glucocorticoids and to suppress excess androgen formation. We often recommend starting treatment in the neonate with 0.2  $\text{mg}/\text{day}$  of 9-fludrocortisone in order to prevent salt wasting and because 21-hydroxylase deficiency often cannot be ruled out initially. If the diagnosis of 11 $\beta$ -hydroxylase deficiency is confirmed, mineralocorticoid treatment can be gradually withdrawn over the first few months of life. In addition to the medical treatment described, surgical correction of the genitalia (i.e., clitoral reduction and vaginoplasty) is

performed preferably during the first year of life. Therefore, a pediatric surgeon or urologist experienced in this type of delicate surgery should be part of the team. It is also recommended that the parents be involved in the early stages of the diagnostic workup so that they become as comfortable as possible with the diagnosis and understand the treatment in which they will have to take part.

Some children are born with such severe virilization that the genitalia appear completely male and the lack of testes in the "scrotum" is overlooked. Indeed, in our series, approximately half of the female children were mistakenly reared as boys, and other series have also included many such cases. Since these "males" have ovaries and do not have testes, the Müllerian system, including the fallopian tubes, uterus, and the upper two-thirds of the vagina, develops normally. In these cases, it is a dilemma whether to change the gender because of possible female fertility or to remove the female organs and continue to rear the child as a male. The decision depends on the age at diagnosis and the cultural context.

Follow-up evaluation and maintenance therapy are ideally performed in a coordinated manner by a team including endocrine, urological, psychological, and genetics expertise. The endocrinologist should monitor compliance with hydrocortisone by regularly measuring ACTH, plasma renin activity, and 11-deoxycortisol and androgen levels and by following growth, bone age, and pubertal development. One should be aware that excess androgens may trigger the onset of early central puberty and should appropriately monitor the patient for signs of such occurrence by regular physical examination and, if necessary, testing for elevated gonadotropin levels. The urologist or surgeon performing vaginoplasty and clitoral reduction will also be required to follow the female patient to ensure adequate genitalia for intercourse. This may require dilatation of the vaginal opening during adolescence and, occasionally, additional surgery. Psychological expertise is required early in life to help the parents come to terms with gender assignment and later to help the adolescent with concerns associated with sexual activity. Genetic counseling should be considered part of the initial therapy, and the possibility of early diagnosis and prenatal treatment should be discussed with the endocrinologist.

### Treatment during Acute Illness

Patients with 11 $\beta$ -hydroxylase deficiency cannot mount a sufficient stress response and should receive

appropriate stress doses of glucocorticoids as for other patients with adrenal insufficiency.

### PROGNOSIS

If treated appropriately, patients presenting with the classical form of 11 $\beta$ -hydroxylase deficiency should theoretically achieve a final height near their genetic potential. Several studies of 21-hydroxylase-deficient patients have documented that many patients do not achieve this final height, probably as a result of undertreatment or overtreatment at various stages of life. A significant loss of final height has been attributed to delayed diagnosis; therefore, the final height of patients with salt-wasting 21-hydroxylase deficiency (who are diagnosed at an early age because of the dramatic presentation) has been reported to be better than that of the simple virilizing form. Similarly, it is understandable that 11 $\beta$ -hydroxylase-deficient patients tend to be more severely height compromised, although few studies have dealt with this issue.

### Prenatal Diagnosis

Although population screening has not been set up for this form of CAH, prenatal diagnosis may be done when a previous sibling has been affected by searching for one of the known mutations. If the mutation is unknown, levels of tetrahydro-11-deoxycortisol and 11-deoxycortisol can be determined in the amniotic fluid and maternal urine, as reported by our group; however, to facilitate this, prenatal treatment must be temporarily halted.

### Prenatal Treatment

*In utero* virilization of female genitalia occurs in CAH because of the accumulation prior to the blocked enzyme of steroid precursors, which are then converted to potent androgens. Early prenatal treatment of the mother with dexamethasone can prevent virilization of female genitalia. Dexamethasone crosses the placenta and suppresses the fetal hypothalamic-pituitary-adrenal axis. Although this treatment is successful if initiated early, it must be initiated prior to diagnosis and can be terminated only after chorionic villous sample (CVS) confirms that the fetus is unaffected, heterozygous, or male. A prerequisite for this is knowledge of the mutation in an index case in the same family so that the CVS can be amplified by polymerase chain reaction and examined rapidly for the mutation. Treatment is controversial because in a

large number of pregnancies the mother is inevitably treated unnecessarily with a drug that may cause her harm (at least seven of eight pregnancies; i.e., all males, four of eight pregnancies; female carriers, two of eight pregnancies; and females without the mutation, one of eight pregnancies). Also, the long-term effects of dexamethasone therapy on the fetus throughout gestation are unknown. Most of the reported experience regards 21-hydroxylase deficiency, and the major side effects are those of transient iatrogenic Cushing's syndrome in the mother. Only two female children homozygotes for 11 $\beta$ -hydroxylase deficiency have been reported as treated prenatally: One was born virilized and the other had normal female genitalia. Before treatment is contemplated, the parents must be made aware of the risks and benefits involved.

### NONCLASSICAL 11 $\beta$ -HYDROXYLASE DEFICIENCY CAH

Patients with this form of the disease present with clinical signs of hyperandrogenism (i.e., hirsutism, acne, male-type balding, and menstrual irregularities). Several mutations causing nonclassical CAH have been described (Table I), including missense mutations (N133H and T319M) and a nonsense mutation (Y423X). Significantly elevated stimulated levels of the appropriate metabolites define biochemical enzyme deficiency. However, this does not correlate with mutations in the CYPB1 gene. Indeed, in one report, only two of five women with ACTH-stimulated 11-deoxycortisol levels more than three times higher than the 95th percentile had mutations in the gene. Secondary deficiency of 11 $\beta$ -hydroxylase activity has been reported in patients with late-onset 21-hydroxylase deficiency, probably due to the effects of elevated circulating androgens on the CYP11B1 enzyme.

Patients with nonclassical CAH are diagnosed during evaluation of hyperandrogenism presenting during adolescence or late childhood, and treatment is not dependent on the genetic status. Suppressive therapy with hydrocortisone in childhood or dexamethasone after growth is complete will prevent hyperandrogenism and its effects and therefore is indicated.

### Special Considerations

11 $\beta$ -Hydroxylase deficiency was initially defined as the hypertensive form of CAH. However, our experience, and that of a Saudi Arabian group, includes several

patients in whom neonatal salt wasting occurred. The renal distal tubule is comparatively mineralocorticoid resistant in the neonatal period, and this gives rise to normally elevated levels of renin and aldosterone. It appears that the mineralocorticoid effect of the accumulated precursors is, in some infants, insufficient to prevent renal salt wasting. Furthermore, some precursors (e.g., 17-hydroxyprogesterone) have an antiminerocorticoid effect. When one suppresses ACTH with hydrocortisone, previously accumulated weakly mineralocorticoid precursors (e.g., 11-deoxycorticosterone) are reduced, thus increasing the likelihood of a salt-wasting crisis. Therefore, it is prudent to initiate therapy with mineralocorticoids and gradually reduce the dose over the first year of life.

### Incidentalomas

Adrenal incidentalomas are tumors discovered when abdominal imaging is performed. In up to 70% of these patients, elevated serum 17-hydroxyprogesterone levels have been noted. Evidence of 11-hydroxylase deficiency has been reported for a group of patients with incidentalomas based on ACTH-stimulated ratios of 11-deoxycortisol:cortisol and 11-deoxycorticosterone:corticosterone compared with controls. It is thought that the preexisting undiagnosed deficiency caused adrenal overstimulation, thus producing these small adrenal tumors.

### Heterozygotes for CYP11B1 Mutations

CYP11B1-deficiency CAH is an autosomal recessive condition. Several studies have evaluated whether heterozygotes have any detectable abnormality. This is a controversial issue. Our group and others have presented evidence that they do not, although other groups have contested this.

### CYP11B2 DEFICIENCY

P450c11aldo, encoded by the CYP11B2 gene, catalyzes the final steps of aldosterone synthesis in the mineralocorticoid pathway. This enzyme is expressed only in the zona glomerulosa of the adrenal gland. CYP11B2 is regulated by angiotensin II via a G protein-coupled receptor that activates the phospholipase C pathway, culminating in increased intracellular calcium. In turn, calcium is thought to activate a calmodulin-dependent protein kinase that phosphorylates various transcription factors. These factors then increase expression of CYP11B2.

Potassium also has a direct effect on the glomerulosa cells by membrane depolarization. The change in transmembrane voltage causes increased calcium influx via voltage-dependent calcium channels. The rest of the activation pathway is the same as the angiotensin II pathway. ACTH is also thought to increase expression of CYP11B2 but appears to have only a short-term effect.

Deficiency of the aldosterone synthase enzyme complex presents with congenital salt wasting due to isolated mineralocorticoid deficiency. Such patients are not virilized and have normal levels of cortisol and sex steroids. The reduced aldosterone production is due to insufficient enzyme activity in the CYP11B2 complex that includes 11 $\beta$ -hydroxylase, CMO-I, and CMO-II activities, all of which are encoded by the CYP11B2 gene. As previously mentioned, the largest number of cases (with CMO-II deficiency) have been reported among Jews of Iranian descent. These Jews originated in a highly inbred community from Ispahan, Iran, that may be traced back to approximately 500 B.C. The incidence of this condition in Iranian Jews is estimated to be at least 1:4000. All these patients are homozygous for two mutations in CYP11B2: R181W (on exon 3) and V386A (on exon 7). The R181W mutation completely abolishes both CMO-I and CMO-II activity *in vitro*. The second mutation on exon 7, codon 386, produces only a minimal reduction in CMO-I activity. Interestingly, individuals homozygous for either one of these mutations alone are completely asymptomatic. A similar requirement for double homozygosity has been described for other combinations [i.e., T318M and V386A (exons 5 and 7) and also R181W (exon 3) and C372del1nt (exon 6)]. Double homozygosity for E198D (exon 3) and V386A (exon 7) is associated with CMO-I deficiency. Clinical CMO-I deficiency has been reported with double homozygosity for E198D and V386A. Homozygosity for a single V35del2nt and R384P mutation also causes clinical CMO-I deficiency.

The diagnosis of the two types of CYP11B2 deficiency is made based on dramatically increased levels of 18-hydroxycorticosterone in CMO-II, with a marked increase in the 18-hydroxycorticosterone:aldosterone ratio compared with mildly decreased levels in CMO-I deficiency. Serum aldosterone is undetectable in type I deficiency but usually within normal limits in type II.

Both forms of aldosterone synthase deficiency present clinically in the same way. A salt-wasting syndrome with hypovolemia occurs at a few days to a few weeks of age (usually after the first week of life due to the protective effect of maternal steroids). Such

patients are initially stabilized with intravenous saline infusion. Once stable, oral salt supplementation at a dose of 1 or 2 g/day and oral 9-fluorocortisone at a dose of 0.1–0.3 mg/day usually suffice. During childhood, patients can be weaned from therapy, and adults usually do not require treatment. This phenomenon is explained by the ability of these patients to form 11-deoxycorticosterone, a weak mineralocorticoid that is sufficient to prevent salt loss in the mature distal nephron but insufficient in the infant. As a result, although the 18-hydroxycorticosterone:aldosterone ratio is persistently elevated into adulthood, the plasma renin normalizes with age.

Some patients present in a milder fashion and are not diagnosed in infancy but are diagnosed when severe hyponatremia and hyperkalemia occur during gastroenteritis. Otherwise, they may be diagnosed asymptotically upon screening of a family after another infant sibling presents with salt wasting. These mild cases probably do not require treatment except in conditions of severe sodium loss, when they might not be able to mount a sufficient mineralocorticoid response.

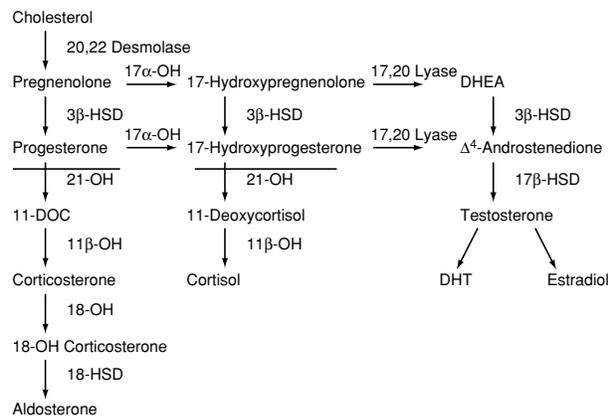
### See Also the Following Articles

Congenital Adrenal Hyperplasia, Prenatal Diagnosis and Therapy • 3 $\beta$ -Hydroxysteroid Dehydrogenase Deficiency • 17 $\alpha$ -Hydroxylase/17,20-Lyase Deficiency • 21-Hydroxylase Deficiency, Classical • 21-Hydroxylase Deficiency, Genetics of

### Further Reading

- Al-Jurayan, N. A. M. (1995). Congenital adrenal hyperplasia due to 11-beta hydroxylase deficiency in Saudi Arabia: Clinical and biochemical characteristics. *Acta Paediatr.* **84**, 651–654.
- Cerame, B. I., Newfield, R. S., Pascoe, L., *et al.* (1999). Prenatal diagnosis and treatment of 11 beta-hydroxylase deficiency congenital adrenal hyperplasia resulting in normal female genitalia. *J. Clin. Endocrinol. Metab.* **84**, 3129–3134.
- Eldar-Geva, T., Hurvitz, A., Vecsel, P., Palti, Z., Milwidsky, A., and Rosler, A. (1990). Secondary biosynthetic defects in women with late-onset congenital adrenal hyperplasia. *N. Engl. J. Med.* **323**, 855–863.
- Geley, S., Kapelari, K., Johrer, K., *et al.* (1996). CYP11B1 mutations causing congenital adrenal hyperplasia due to 11 beta-hydroxylase deficiency. *J. Clin. Endocrinol. Metab.* **81**, 2896–2901.
- Gillis, D., Speiser, P., Zhou, Z., and Rosler, A. (2000). Combined 21-hydroxylase and 11-beta hydroxylase deficiency: Patient report and molecular basis. *J. Pediatr. Endocrinol. Metab.* **13**, 945–949.
- Hochberg, Z., Shechter, J., Benderly, A., Leiberman, E., and Rosler, A. (1985). Growth and pubertal development in patients with congenital adrenal hyperplasia due to 11-beta hydroxylase deficiency. *Am. J. Dis. Child.* **139**, 771–776.
- Joehrer, K., Geley, S., Strasser-Wozak, E. M. C., Azziz, R., Wollmann, H. A., Schmitt, K., *et al.* (1997). CYP11B1 mutations

- causing non-classical adrenal hyperplasia due to 11 beta-hydroxylase deficiency. *Hum. Mol. Genet.* **6**, 1829–1834.
- Manoli, I., Kanak-Gantenbein, Ch., Voutetakis, A., Maniati-Christidi, M., and Dacou-Voutetakis, C. (2002). Early growth, pubertal development, body mass index and final height of patients with congenital adrenal hyperplasia: Factors influencing the outcome. *Clin. Endocrinol. (Oxford)* **57**, 669–676.
- Peter, M., and Sippell, W. G. (1997). Evidence for endocrinological abnormalities in heterozygotes for adrenal 11 beta-hydroxylase deficiency of a family with the R448h mutation in the CYP11B1 gene. *J. Clin. Endocrinol. Metab.* **82**, 3506–3508.
- Peter, M., Dubuis, J. M., and Sippell, W. G. (1999) Disorders of the aldosterone synthase and steroid 11 beta-hydroxylase deficiencies. *Horm. Res.* **51**, 211–222.
- Portrat, S., Mulatero, P., Curnow, K. M., Chaussain, J. L., Morel, Y., and Pascoe, L. (2001). Deletion hybrid genes due to unequal crossing-over between CYP11B1 (11beta-hydroxylase) and CYP11B2 (aldosterone synthase) cause steroid 11beta-hydroxylase deficiency and congenital adrenal hyperplasia. *J. Clin. Endocrinol. Metab.* **86**, 3197–3201.
- Reincke, M., Peter, M., Sippell, W. G., and Allolio, B. (1997). Impairment of 11 beta-hydroxylase but not 21-hydroxylase in adrenal incidentalomas. *Eur. J. Endocrinol.* **136**, 196–200.
- Rosler, A., and Cohen, H. (1995). Absence of steroid biosynthetic defects in heterozygote individuals for classic 11 beta-hydroxylase deficiency due to a R448H mutation in the CYP11B1 gene. *J. Clin. Endocrinol. Metab.* **80**, 3771–3773.
- Rosler, A., and White, P. C. (1993). Mutations in human 11-beta hydroxylase genes; 11-Beta hydroxylase deficiency in Jews of Morocco and corticosterone methyl-oxidase II deficiency in Jews of Iran. *J. Steroid Biochem. Mol. Biol.* **45**, 99–106.
- Rosler, A., Weshler, N., Leiberman, E., Hochberg, Z., Weidenfeld, J., Sack, J., and Chemke, J. (1988). 11-Beta hydroxylase deficiency congenital adrenal hyperplasia: Update of prenatal diagnosis. *J. Clin. Endocrinol. Metab.* **66**, 830–838.
- Singer-Granick, C., Hill, I. D., and Rolston, R. K. (1995) Does primary salt-wasting occur in 11-beta hydroxylase deficiency? *J. Pediatr. Endocrinol. Metab.* **8**, 275–280.
- Stocco, D. M., and Clark, B. J. (1996). Regulation of the acute production of steroids in steroidogenic cells. *Endocr. Rev.* **17**, 221–244.
- White, P. C. (2001). Steroid 11 beta-hydroxylase deficiency and related disorders. *Endocrinol. Metab. Clin. North Am.* **30**, 61–79.
- White, P. C., New, M. I., and Dupont, B. (1987). Congenital adrenal hyperplasia. *N. Engl. J. Med.* **316**, 1519–1524.



**Figure 1** Schematic representation of steroid hormone synthesis with 21-hydroxylase enzyme block resulting in accumulation of precursors and synthesis of excess sex hormones. Enzymes: 17 $\alpha$ -OH, 17 $\alpha$ -hydroxylase; 3 $\beta$ -HSD, 3 $\beta$ -hydroxysteroid dehydrogenase; 21-OH, 21-hydroxylase; 11 $\beta$ -OH, 11 $\beta$ -hydroxylase; 18-OH, 18-hydroxylase; 18-HSD, 18-hydroxysteroid dehydrogenase; 17 $\beta$ -HSD, 17 $\beta$ -hydroxysteroid dehydrogenase. Intermediate products: 11-DOC, 11-deoxycorticosterone; DHEA, dehydroepiandrosterone.

## CLINICAL FEATURES

The phenotypic spectrum of classical 21-hydroxylase deficiency depends mainly on the genetic sex and residual enzyme activity, if any. Two distinct phenotypes, the salt-wasting and simple virilizing forms, are commonly recognized, although intermediate forms are also known to occur. Patients with nonclassical congenital adrenal hyperplasia do not present at birth, and often female patients present around puberty or later with hyperandrogenic features such as premature pubarche, hirsutism, acne, and menstrual disturbances.

### Salt-Wasting Form

Approximately 75% of patients with classical 21-hydroxylase deficiency have severely impaired hydroxylation at C21 and so cannot synthesize adequate amounts of either cortisol or aldosterone. Inadequate aldosterone reserve leads to excessive sodium loss through the kidney, colon, and sweat glands. Concomitant severe cortisol deficiency exacerbates the systemic effects of aldosterone deficiency and results in severe hyponatremic dehydration with vomiting, hypotension, and shock (adrenal crisis). This is often the presenting symptom and occurs 1 to 4 weeks after birth in undiagnosed patients. These adrenal crises are fatal if untreated, and they remain

an important cause of mortality, especially in untreated patients and those under 5 years of age. In addition to salt wasting, these patients have virilization because accumulated precursor 17-hydroxyprogesterone is diverted to produce excess adrenal androgens, which are further converted peripherally to more potent androgens such as testosterone. Genetically female infants with classical 21-hydroxylase deficiency are virilized to a variable degree *in utero* and are born with ambiguous genitalia. The skin of the labia majora is thickened and wrinkled like scrotal skin and is fused to a variable extent, concealing the vaginal orifice. The urethra and vagina open into a common urogenital sinus, and there is invariably a degree of clitoromegaly. In severe cases, the clitoris may resemble a penis and the labia may be partially fused, so that a mistaken diagnosis of hypospadias with cryptorchidism can be made. In extreme cases, complete labial fusion, a phallic urethra, and an external meatus at the penile tip can occur, mimicking a complete male external phenotype with bilateral cryptorchidism. However, these children have normal female organs internally (fallopian tubes, uterus, and upper vagina), and the ovaries are normally situated. In genetic males, sexual differentiation is normal, and there is usually only minimal enlargement, if any, of the penis.

### Simple Virilizing Form

Some 25% patients with classical 21-hydroxylase deficiency have 1 to 2% residual enzyme activity, such that the adrenal cortex secretes adequate aldosterone to maintain normal serum electrolytes but cortisol production is low enough to stimulate ACTH and, hence, androgen hypersecretion (Fig. 2). These patients do not have salt wasting but have virilization. In



**Figure 2** Gingival hyperpigmentation due to increased secretion of ACTH.

genetic females, this results in ambiguous genitalia. Genetic male newborns escape detection at birth and are often not diagnosed until a few years later when they present with premature sexual development.

### Untreated or Late Presentations

Often, classical 21-hydroxylase deficiency might not be diagnosed at birth for a variety of reasons and so might not be treated. The severe salt-wasting form is fatal if unrecognized and untreated during the first few weeks of life. An affected male infant who dies after several weeks of persistent vomiting may be thought to have died of other causes (e.g., pyloric stenosis, infection, gastroenteritis). However, patients with simple virilizing forms survive with progressive postnatal virilization, rapid somatic growth, and precocious puberty. These children will have compromised adult height due to premature fusion of epiphyses. Some 46,XX patients in developing countries may be raised as males.

### DIAGNOSIS

Classical 21-hydroxylase deficiency should be clinically considered as a possible diagnosis in all children with genital ambiguity (see Fig. 5). Salt loss might not commence at birth, and electrolytes usually remain normal for the first week of life. Hyponatremic dehydration with shock during the first few weeks of life, especially in boys or in infants with ambiguous genitalia, is a strong clinical indicator of the severe salt-wasting form of congenital adrenal hyperplasia. Hypoglycemia may also be a feature in the very young. Less severe forms can present with lethargy, poor feeding, and poor weight gain. Precocious puberty and virilization in older boys and girls can be a clinical indicator for the untreated simple virilizing form of congenital adrenal hyperplasia. Very high levels of 17-hydroxyprogesterone, the main substrate for the enzyme 21-hydroxylase, are found and are diagnostic. Salt wasters tend to have higher levels than do non-salt wasters. However, some 10% of severely affected infants may have falsely low initial values on day 1 of life. In addition, most sick premature babies have elevated levels without an enzyme defect. This is partly because the fetal zone of the adrenal cortex is still active and secretes a range of steroids that can cross-react with 17-hydroxyprogesterone in an immunoassay. False-positive 17-hydroxyprogesterone results in samples from newborn infants may occur if the laboratory does not include an

additional step to reduce the amount of cross-reacting steroids. The measurement of urinary pregnanetriol (a metabolite of 17-hydroxyprogesterone) can also be of diagnostic value. An ACTH stimulation test can accurately differentiate among various enzyme defects in steroidogenesis by comparing precursor-to-product ratios after stimulation, but this usually is not necessary for the diagnosis of classical 21-hydroxylase deficiency because basal values are diagnostic.

CYP21 genotyping and mutation analysis does not help in making treatment decisions but is helpful in prenatal diagnosis when there is a known proband in the family. There is good correlation between genotype and phenotype, allowing DNA analysis to predict, with certain reliability, the residual enzyme activity and consequent clinical expression.

### PATIENT MANAGEMENT

#### Conventional Hormone Replacement Therapy

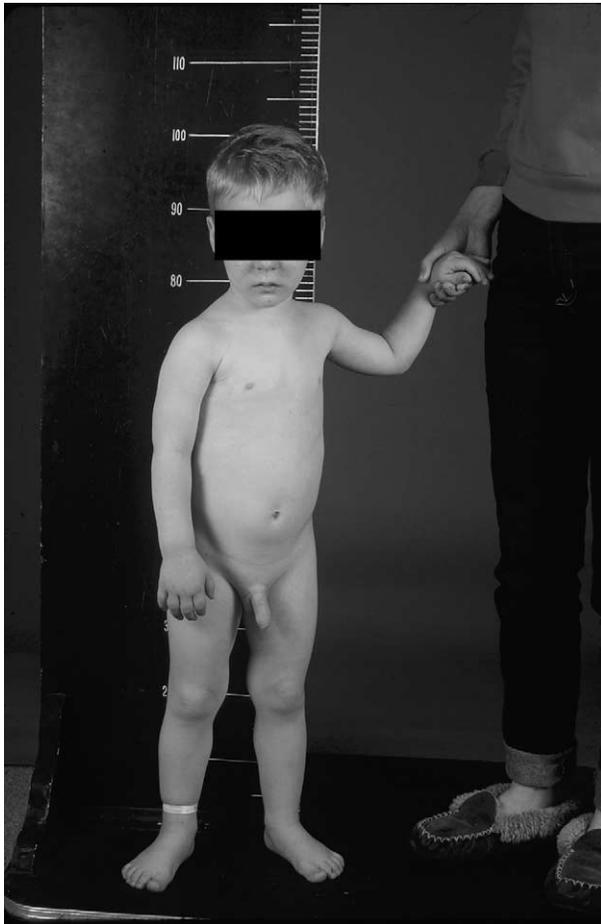
The aims of treatment are to replace the missing hormones while at the same time suppressing the excess production and accumulation of precursors. It appears that supraphysiological doses of hydrocortisone are required to inhibit ACTH secretion adequately and to suppress excess androgen production. However, finding the appropriate dose for each patient is a fine balancing act because inadequate treatment will not suppress ACTH adequately and overtreatment results in unwanted side effects of steroids, principally growth retardation and Cushing's syndrome.

#### Salt-Wasting Form

Hydrocortisone (10–20 mg/m<sup>2</sup>/day in two or three divided doses) and fludrocortisone (0.1–0.2 mg/day) are usually preferred for replacing cortisol and aldosterone, respectively, in these patients, and lifelong treatment is required. Salt supplements and higher doses of fludrocortisone may be necessary during infancy. Salt wasting generally seems to get better with age, and older children and adults often have lower fludrocortisone dose requirements without salt supplementation.

#### Simple Virilizing Form

Adequate hydrocortisone replacement alone is enough to normalize growth and stop progressive virilization in these patients, but hydrocortisone treatment is often unable to reverse the virilization that has already occurred (Fig. 3). In some patients who are



**Figure 3** 2½-year-old boy showing signs of hyperandrogenism with facial acne, pubic hair, and increased penile size. Also note height above 98th centile.

diagnosed very late, the elevated adrenal androgen levels suppress gonadotropin secretion. When treatment with cortisol suppresses these androgens, follicle-stimulating hormone (FSH) and luteinizing hormone (LH) are secreted, resulting in true or central precocious puberty. This should be treated with long-acting luteinizing hormone-releasing hormone (LHRH) analogues because there is a strong risk of reduced final adult height due to early and premature fusion of epiphyses. Compliance is another major problem in adolescent boys because they have few clinical symptoms on a daily basis.

#### **Adrenal Crisis**

Patients with classical 21-hydroxylase deficiency are unable to mount an adequate cortisol response during periods of stress, such as infection and surgery, and require additional pharmacological doses of

hydrocortisone. Steroid cover for stress should provide 30 to 60 mg of hydrocortisone/m<sup>2</sup>/day in divided doses. If children are vomiting or have diarrhea, the hydrocortisone should be injected subcutaneously or intramuscularly. Care should also be taken to recognize and treat hypoglycemia in young children with an adrenal crisis. Untreated and unrecognized stress can be fatal and remains one of the potential causes of mortality in young children.

### **Ambiguous Genitalia**

#### **Gender Assignment at Birth**

The established practice is to minimize the period of gender uncertainty by definitive diagnosis and treatment as early as possible. The female sex of rearing is appropriate for most 46,XX patients because it maximizes the chance of fertility. Feminizing genital surgery (clitoral reduction and vaginoplasty) helps psychological adjustment of the parents and reduces the risk of the patient becoming stigmatized. However, the practice of early genital surgery has been questioned during recent years by certain patient advocacy groups, which have urged physicians to assist patients' families, and later patients themselves, in participating in informed decision making regarding the genital surgery. In our opinion, this does not apply to girls with congenital adrenal hyperplasia. A medical benefit from early vaginoplasty is a reduced risk of recurrent urinary tract infection.

#### **Gender Reassignment during Childhood**

Some 46,XX patients with markedly masculinized genitalia at birth are inadvertently assigned to the male gender (Figs. 4 and 5). In developing countries, some untreated patients survive and postnatal virilization is progressive. When the correct diagnosis is made later on but still during infancy or early childhood, many physicians recommend sex reassignment to female, taking into account the behavior (masculine or feminine) of the child. Again, female gender rearing is usually considered best for most 46,XX patients. Some 46,XX patients have been reared in the male gender until late childhood or early adolescence, and these are best to remain as boys unless they themselves feel otherwise. A small proportion of patients may self-initiate gender reassignment (female to male) around puberty or later. Such patients have usually been exposed to excessive androgen levels during childhood due to poor compliance with medical therapy.



**Figure 4** Phenotypically male genitalia with complete labial fusion, labial scrotalization, clitoromegaly resembling a penis with urethral opening at the tip, and bilateral cryptorchidism (Prader stage 5) in a 46,XX infant with 21-hydroxylase deficiency.

### Genital Surgery

Genital reconstructive surgery is performed to make the appearance of the external genitalia concordant with the sex of rearing. The mainstays of surgery in a virilized 46,XX patient assigned a female sex are clitoral reduction and vaginoplasty. With increasing masculinization of the infant, the repair becomes technically more demanding. Some surgeons believe that the repair is more difficult with increasing age because the urogenital sinus lengthens with age and greater degrees of vaginal mobilization are then required. The surgical treatment for 46,XX patients raised as males or those self-initiating gender reassignment include removal of ovaries and uterus, staged hypospadias repair, and insertion of testicular prostheses. These patients would also require testosterone replacement therapy for life.



**Figure 5** Ambiguous genitalia in a 46,XX newborn showing labial scrotalization and hyperpigmentation, labial fusion, clitoromegaly mimicking a penis, and bilateral cryptorchidism.

### Adult Issues

#### Short Stature

Most patients with classical 21-hydroxylase deficiency, both men and women and regardless of the age at diagnosis or the quality of endocrine treatment, achieve a final height that is shorter than that of the population in general and the mid-parental height. This loss of height potential can be due to iatrogenic hypercortisolism, adrenal hyperandrogenism, or a combination of both. In general, if the hydrocortisone dose is sufficient to suppress 17-hydroxyprogesterone to less than 20 nmol/L, adrenal androgens are more completely suppressed. Therefore, women with classical 21-hydroxylase deficiency may be androgen deficient. Inadequate dose results in hyperandrogenism, which accelerates linear growth and epiphyseal maturation. Unfortunately an excessive hydrocortisone dose also suppresses growth, so correct treatment involves careful clinical and biochemical monitoring to find the fine line between inadequate and excessive doses of hydrocortisone.

#### Continuing Hyperandrogenism

Adult females with untreated or undertreated classical 21-hydroxylase deficiency continue to have problems of hyperandrogenism such as hirsutism, menstrual disturbances, and decreased fertility. Increased body hair and acne appear during adolescence and persist into adulthood. Noncompliance with treatment or failure to monitor therapy closely is associated with a higher proportion of these problems. Girls may have delayed menarche and ovarian dysfunction, similar to polycystic ovarian syndrome, which can result in later reduced fertility.

### **Fertility Issues in Females**

Female patients with classical 21-hydroxylase deficiency, particularly the salt wasters, have reduced fertility. They tend to have anovulatory cycles in addition to menstrual disturbances during times of hyperandrogenism. Increased levels of progesterone exert an independent contraceptive effect (“mini-pill effect”) with prevention of conception even in the presence of ovulation. Masculinization of the external genitalia, an inadequate introitus, and other structural factors related to genital reconstructive surgery (e.g., vaginal stenosis, clitoral dysfunction, poor surgical repair) play an important part in reducing heterosexual activity.

Despite these factors, studies show an apparent improvement in fertility due to good compliance, surgical advances in genital reconstruction, careful monitoring before and during pregnancy, and use of ovulation induction agents and other assisted reproductive techniques such as *in vitro* fertilization. In general, most pregnancies have been carried to term successfully with healthy outcomes. Excessive maternal androgens could virilize the genitalia of an unaffected female fetus; therefore, a mother with classical 21-hydroxylase deficiency needs preconception counseling and careful monitoring during pregnancy.

### **Fertility Issues in Males**

Men have impaired gonadal function less frequently than do women. Testicular integrity is normal even during periods of noncompliance. Reduced sperm counts found in men with poorly controlled classical 21-hydroxylase deficiency do not usually preclude fertility. However, a more potent cause of infertility is obstruction of the hilum of the testis by hyperplastic adrenal rests. In most cases, these cause severe oligospermia or even complete azoospermia.

### **Adrenal Hyperplasia and Adrenal Rest Tumors**

The incidence of adrenal “incidentalomas” increases with age and is higher in patients with classical 21-hydroxylase deficiency and heterozygotes than in the general population. The incidence is highest in untreated or inadequately treated classical 21-hydroxylase deficiency patients, especially those with the salt-wasting variant. These adrenal masses are nearly always benign. Occasionally, the adrenal “tumor” is the presenting sign in a previously undiagnosed adult patient. Ectopic adrenal masses or rest tumors are common in the testes of inadequately treated or noncompliant patients, but they may also develop at other sites along the line of testicular descent. These benign masses often have a rock-hard

consistency mimicking testicular malignancies, and the patient is at risk for having one or both testes removed if the correct diagnosis is not made in time. The ultrasound appearance of adrenal rests in the testes is said to be diagnostic.

### **Gender Identity**

Pre- and perinatal exposure of the female to high levels of adrenal androgens may influence gender identity and gender-related behavior during childhood and adult life. Most 46,XX patients with varying degrees of genital ambiguity who have been assigned or reassigned to the female gender during early childhood have remained females lifelong, although their behavior may be somewhat masculine. Nearly all 46,XX patients who have been reared as males until late childhood or adolescence elect to remain boys. A small number of patients have requested surgical gender reassignment around puberty or later. In general, 46,XX patients with classical 21-hydroxylase deficiency have a stable female gender identity. As such, they differ from patients with mixed gonadal dysgenesis and partial androgen insensitivity, who are more likely to identify as males (Figs. 5 and 6).

### **Sexual Orientation and Sexual Activity**

Women with classical 21-hydroxylase deficiency have been found in some (but not all) studies to have an increased rate of bisexual or homosexual orientation, which again is attributed to varying degrees of brain androgenization during the pre- and perinatal periods. This bisexual or homosexual orientation is demonstrated in sexual imagery, fantasies, and attractions, but only few of these women consider themselves to be lesbians or have overt homosexual involvement. More strikingly, women with classical 21-hydroxylase deficiency show greater avoidance of all forms of sexual activity than do other women. Whether this is due to psychological differences or to physical problems, such as an insensitive (or absent) clitoris, vaginismus, failure to lubricate, or a tight introitus, is unclear. Reduced sexual activity undoubtedly contributes greatly to infertility.

## **FAMILY MANAGEMENT**

### **Genetic Counseling for the Family**

The identification of a genetic abnormality in any member has wide-ranging implications for the whole family. It is appropriate to consider all family members when planning diagnosis and treatment. Parents should be counseled about risks of the problem in



**Figure 6** Urogenital sinogram showing urethra, vagina, and cervical indentation (same patient as in Fig. 4).

their other children and future pregnancies. All siblings of the index case have a 25% risk of having 21-hydroxylase deficiency and, hence, should be screened.

### Prenatal Diagnosis and Treatment

Preconception counseling should be provided to carrier parents given that there is a 25% risk of an affected fetus in all of their pregnancies. Both parents and the index case should be genotyped if prenatal diagnosis is desired. There are two alternative approaches to prenatal diagnosis. One is to establish the HLA haplotypes for the index case and both parents and then compare them with the haplotype of the fetus. The other is to identify the CYP21 mutation in the index case and screen the fetus for the same mutation. Prenatal treatment with dexamethasone is started as early as the fifth week in the pregnancy. The sex of the fetus is established first by

chorionic villus sampling and a fluorescence *in situ* hybridization (FISH) test using Y-sequence probes. If the fetus is female, further testing is carried out to determine whether classical 21-hydroxylase deficiency is present. The dexamethasone treatment is continued to term only if the fetus is an affected female and is discontinued in other pregnancies with a male or an unaffected female fetus. There are no significant side effects of dexamethasone treatment prior to diagnosis at 9 to 11 weeks in the unaffected fetus or pregnant mother. However, the sudden cessation of dexamethasone may cause steroid withdrawal symptoms (e.g., peeling skin, malaise, hypotension). When dexamethasone treatment is continued to term, the mother tends to gain more weight and is also at risk for high blood pressure, emotional disturbances, loss of bone mineral density, and glucose intolerance. Prenatal treatment has been shown to be effective in reducing genital virilization, and in two-thirds of cases, the genitalia of the infant are either normal or so minimally virilized that surgery is not required. Follow-up studies are not yet available to show whether the treated children are less masculine in their behavior or less prone to confusion about their gender identity.

### PUBLIC HEALTH ISSUES

Classical 21-hydroxylase deficiency is a condition that fulfills all of the recommended criteria for newborn screening. The incidence of the disorder in the general population is relatively high, the condition can result in high morbidity and mortality if left undetected, effective treatment is available, and (most important) the means exist for reliable and efficient newborn screening. In general, a blood sample collected and dried on filter paper on day 2 or 3 of life is analyzed for 17-hydroxyprogesterone. Neonatal screening detects cases of CAH that are not suspected clinically. Lives are saved and incorrect gender assignments are prevented. In addition, the early diagnosis of non-salt-losing classical 21-hydroxylase deficiency permits early treatment, and this prevents progressive virilization and loss of adult potential.

### FUTURE RESEARCH ISSUES

Most of the adverse outcomes in children treated for classical 21-hydroxylase deficiency are attributable to the side effects of the supraphysiological doses of glucocorticoids that are prescribed in an attempt to lower elevated levels of adrenal androgens. Some alternative therapies are being evaluated.

## Peripheral Blockade of Androgen Action

The basis of this alternative approach lies in using a reduced dose of steroids in combination with drugs that block androgen action. An androgen antagonist such as flutamide is used, and an aromatase inhibitor is used to block the conversion of androgen to estrogen. Estrogen blockade is essential because estrogens can advance skeletal maturity and induce premature epiphyseal fusion. Preliminary results from this trial have shown promising results in the short term, with normalized growth rates and bone maturation. Treatment of a pregnant woman with androgen blockers can interfere with normal sex differentiation in a male fetus, and sexually active women should be cautioned to use contraception.

## Intra-adrenal Blockade of Androgen Production

This approach takes an additional step in preventing the excess androgen production at the adrenal level without resorting to high doses of hydrocortisone. Ketoconazole blocks adrenal steroid production at several enzymatic steps and can potentially reduce androgen production at the adrenal level. In the future, ongoing pharmaceutical research may provide a specific 17,20-lyase inhibitor capable of blocking the conversion of 17-hydroxyprogesterone to androstenedione.

## CRH/ACTH Antagonists

The tropic hormone ACTH primarily drives adrenal hypersecretion of androgens, and if ACTH antagonists were to become available, they would represent a potential new approach to the treatment of classical 21-hydroxylase deficiency. The genes controlling the melanocortin receptor superfamily (of which the ACTH receptor is one member) have been cloned.

Corticotropin-releasing hormone (CRH) antagonists can also be used to inhibit ACTH release and break the negative feedback loop. The administration of a CRH antagonist, antalarmin, to rats decreases ACTH and cortisol secretion without causing overt adrenal insufficiency.

## Adrenalectomy

Adrenalectomy is advocated as a treatment for classical 21-hydroxylase deficiency by some endocrinologists. Although still a controversial matter, it is

gaining credibility and acceptance. The benefit of this approach is that lower steroid doses can be used, with fewer side effects, because there is no longer the need to suppress adrenal androgens. Adrenalectomy can be done using a laparoscopic approach, which is safer and more convenient than open surgery. Suitable candidates for adrenalectomy are patients with a null mutation in CYP21 because their adrenal glands are performing no useful function. Lifelong glucocorticoid and mineralocorticoid replacement therapy is essential following adrenalectomy.

## Gene Therapy

The 21-hydroxylase gene is exclusively localized in the adrenal cortex, making it a suitable candidate for tissue-targeted gene therapy. Precise regulation is not absolutely essential given that modestly increased gene expression is not harmful and, moreover, the cortisol pathway is tightly regulated by a negative feedback mechanism. An animal model of classical 21-hydroxylase deficiency, the 21-hydroxylase-deficient mouse, has been used to test gene therapy. The experiments involved adenoviral transfection of the adrenals with the normal human CYP21 gene. The results were promising. Provided that metabolic regulation can be obtained, gene therapy may become an ideal treatment for the future.

## LHRH Agonist/GH Therapy

LHRH analogues are useful to control rapid progression of central precocious puberty and to optimize growth in a child. Growth hormone (GH) administration is shown to increase adult height in children with idiopathic short stature or Turner syndrome. LHRH analogue treatment for precocious puberty in children with classical 21-hydroxylase deficiency has been shown to improve final adult height, with or without GH therapy.

## Deferment of Genital Surgery

The ethics of carrying out feminizing genital surgery in an infant or a child too young to give consent has been questioned. Certain patient advocacy groups recommend that cosmetic genital surgery be deferred until the patient is old enough to give informed consent and to indicate his or her preferences. In our opinion, this is not an approach applicable to genetic females with classical 21-hydroxylase deficiency

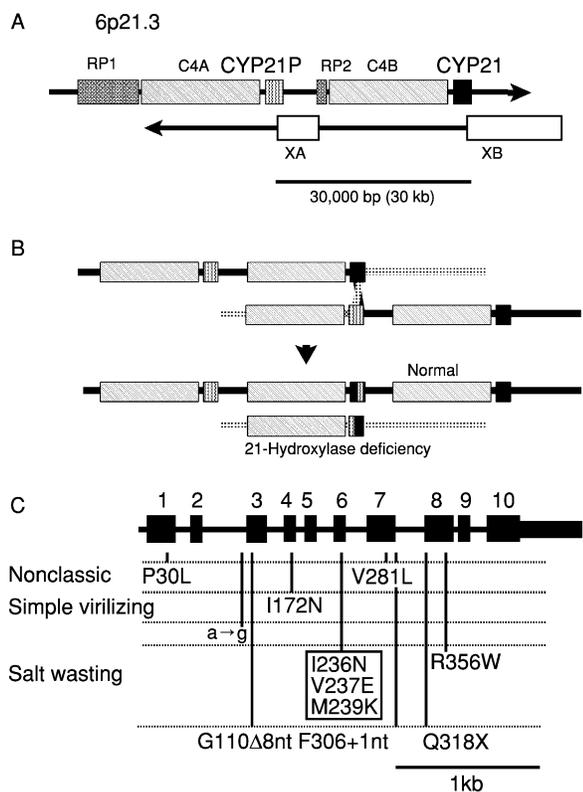
because they are fertile and because nearly all of them grow up identifying as female.

### See Also the Following Articles

Androgens, Gender and Brain Differentiation • Endocrine Disrupters and Male Sexual Differentiation • 3 $\beta$ -Hydroxysteroid Dehydrogenase Deficiency • 11 $\beta$ -Hydroxylase Deficiency • 17 $\alpha$ -Hydroxylase/17,20-Lyase Deficiency • 21-Hydroxylase Deficiency, Genetics of • Pseudohermaphroditism, Male, Due to 5 $\alpha$ -Reductase-2 Deficiency

### Further Reading

- Cabrera, M. S., Vogiatzi, M. G., and New, M. I. (2001). Long term outcome in adult males with classic congenital adrenal hyperplasia. *J. Clin. Endocrinol. Metab.* **86**, 3070–3078.
- Collett-Solberg, P. (2001). Congenital adrenal hyperplasia: From genetics and biochemistry to clinical practice (part 1). *Clin. Paediatr.* **40**, 1–16.
- Collett-Solberg, P. (2001). Congenital adrenal hyperplasia: From genetics and biochemistry to clinical practice (part 2). *Clin. Paediatr.* **40**, 125–132.
- Jaaskelainen, J., and Voutilainen, R. (2000). Long term outcome of classical 21-hydroxylase deficiency: Diagnosis, complications, and quality of life. *Acta Paediatr.* **89**, 183–187.
- Migeon, C. J., and Wisniewski, A. B. (2001). Congenital adrenal hyperplasia owing to 21-hydroxylase deficiency: Growth, development, and therapeutic considerations. *Endocrinol. Metab. Clinics North America* **30**, 193–206.
- New, M. I. (1998). Diagnosis and management of congenital adrenal hyperplasia. *Annu. Rev. Med.* **49**, 311–328.
- Speiser, P. W. (2001). Congenital adrenal hyperplasia owing to 21-hydroxylase deficiency. *Endocrinol. Metab. Clinics North America* **30**, 31–59.



**Figure 1** Genetics of steroid 21-hydroxylase deficiency. (A) Map of the genetic region around the 21-hydroxylase (*CYP21*) gene. Arrows denote the direction of transcription. *CYP21P*, 21-hydroxylase pseudogene; *C4A* and *C4B*, genes encoding the fourth component of serum complement; *RP1*, gene encoding a putative nuclear protein of unknown function; *RP2*, truncated copy of this gene. *XB*, tenascin-X gene (not shown full length). and *XA*, a truncated copy of this gene, are on the opposite chromosomal strand. (B) An unequal crossover generating a *CYP21* deletion. For clarity, only the two *C4* and two *CYP21* genes are shown on each chromosome. When these are misaligned during meiosis, a crossover can generate two daughter chromosomes—one with three copies of the *C4*-*CYP21* tandem and the other with one copy. The latter is often a 21-hydroxylase deficiency allele. (C) Diagram of a *CYP21P* gene. Exons are numbered. *CYP21P* has nine mutations that may be transferred into *CYP21* by gene conversion, causing 21-hydroxylase deficiency. These are arranged in the diagram so that those causing increasing enzymatic compromise are arrayed from top to bottom. Dotted lines divide mutants into groups with similar activities. The mutations at the bottom are nonsense mutations or frameshifts that are predicted to completely prevent protein synthesis. The mutations are associated with different forms of 21-hydroxylase deficiency as marked. As an example of mutation terminology, P30L is proline-30 to leucine. a→g is a mutation at nucleotide (nt) 656 in intron 2. The three mutations in the box are invariably inherited together. There are several dozen additional rare mutations that collectively account for 5% of all 21-hydroxylase deficiency alleles (not shown). Δ, deletion; +, insertion. Other single-letter amino acid codes: A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine.

overlaps a truncated copy of this gene (*TNXA*) that does not encode a functional protein.

*CYP21* is located approximately 600 kb centromeric of HLA-B and 400 kb telomeric of HLA-DR. It is transcribed in the telomeric to centromeric direction.

*CYP21* and *CYP21P* each contain 10 exons spaced over 3.1 kb. Their nucleotide sequences are 98% identical in exons and approximately 96% identical in introns. *CYP21P* contains several mutations preventing synthesis of a functional protein, including an A→G substitution 13 nucleotides (nt) before the end of intron 2 that results in aberrant splicing of pre-mRNA; an 8-nt deletion in exon 3 and a 1-nt insertion in exon 7, each of which shifts the reading frame of translation; and a nonsense mutation in codon 318 of exon 8. There are also eight missense mutations.

## HLA LINKAGE

Congenital adrenal hyperplasia (CAH) due to 21-hydroxylase deficiency is inherited as a monogenic autosomal recessive trait closely linked to the HLA complex, meaning that siblings who have 21-hydroxylase deficiency are almost invariably HLA identical. Before cloning of *CYP21*, HLA typing was the main way to perform prenatal diagnosis. In addition, particular forms of 21-hydroxylase deficiency are associated with particular combinations of HLA antigens or haplotypes; this phenomenon is referred to as genetic linkage disequilibrium. An association between the salt-wasting form of the disease and HLA-A3;Bw47;DR7 is most characteristically seen in northern European populations. In addition to 21-hydroxylase deficiency, this haplotype usually carries a null allele at one of the two *C4* loci encoding the fourth component of serum complement. This represents a contiguous gene syndrome due to a single deletion of *C4B* and *CYP21* genes. The nonclassic form of 21-hydroxylase deficiency is often associated with HLA-B14;DR1, particularly in eastern European Jewish populations. This haplotype is associated with the V281L mutation in *CYP21* and with a duplication of complement *C4A* and the *CYP21P* pseudogene. Finally, HLA-A1;B8;DR3 is negatively associated with 21-hydroxylase deficiency. This haplotype has a *C4A* null allele and is associated with deletion of the *C4A* and *CYP21P* genes. Thus, even before the genes were sequenced, comparison of a very few individuals homozygous for HLA-A3;Bw47;DR7 or A1;B8;DR3 strongly suggested that the *CYP21* gene was an active gene, whereas *CYP21P* was a pseudogene.

## MUTATIONS CAUSING 21-HYDROXYLASE DEFICIENCY

Approximately 95% of mutations causing 21-hydroxylase deficiency are apparently the result of either of two types of recombinations between *CYP21*, the normally active gene, and the *CYP21P* pseudogene. These two mechanisms are unequal crossing-over during meiosis, resulting in a complete deletion of *C4B* and a net deletion of *CYP21*, or apparent gene conversion events that transfer deleterious mutations normally present in *CYP21P* to *CYP21*.

Because particular mutations occur in many unrelated kindreds, each mutation, and the degree of enzymatic compromise it causes, may be correlated with the different clinical forms of 21-hydroxylase deficiency (i.e., salt wasting, simple virilizing, and nonclassic disease). In addition, the functional effects of missense mutations have been assessed *in vitro* by recreating them in *CYP21* cDNA and expressing the mutant cDNA in cultured mammalian cells or yeast.

### Deletions and Large Gene Conversions

Large deletions involving *C4B* and *CYP21* comprise approximately 20% of alleles in patients with classic 21-hydroxylase deficiency in most populations but are rarer in some Latin American countries. Deletions usually extend approximately 30 kb from somewhere between exons 3 and 8 of *CYP21P* through *C4B* to the corresponding point in *CYP21*, yielding a single remaining *CYP21* gene in which the 5'-end corresponds to *CYP21P* and the 3'-end to *CYP21*. Deleterious mutations within the *CYP21P* portion render such a gene incapable of encoding an active enzyme. All patients who carry homozygous deletions suffer from the salt-wasting form of the disorder.

Unequal crossovers may occur anywhere within the duplicated 30-kb region, including the *RP*, *C4*, and *TNX* genes, but only crossover breakpoints within or 3' of the *CYP21* genes cause 21-hydroxylase deficiency. There is apparently no strong selection against chromosomes with one or three copies of the 30-kb region as long as *CYP21* remains intact, and such rearrangements are seen on 16 and 12% of chromosomes 6, respectively.

One kindred has been described that carries an unusual deletion extending into the *TNXB* gene. The patients in this kindred have a contiguous gene syndrome including 21-hydroxylase deficiency and a form of Ehlers-Danlos syndrome due to loss of function of tenascin-X.

In most studies, deletions have been detected by genomic blot hybridization as an absence (or diminished intensity in heterozygotes) of gene-specific fragments produced by digestion with several restriction enzymes. Large gene conversions, in which multiple mutations are transferred from *CYP21P* to *CYP21*, are also detected by this approach when gene-specific restriction endonuclease sites are affected. Large conversions account for approximately 10% of alleles in classic 21-hydroxylase deficiency.

Reliable differentiation of deletions and large gene conversions requires analysis of several different restriction digests in which the sites used to distinguish *CYP21P* and *CYP21* are widely spaced. This is required because the remaining *CYP21*-like gene on a chromosome with a deletion consists of the 3'-end of *CYP21* "spliced" onto the 5'-end of *CYP21P* so that the missing fragment in each restriction digest does not necessarily correspond in size to the *CYP21*-specific fragment in normal chromosomes. In practice, genomic blot hybridizations are no longer routinely used for molecular diagnosis because they are more laborious and less informative than polymerase chain reaction (PCR)-based techniques. However, PCR cannot distinguish deletions from large gene conversions.

### Nonsense and Frameshift Mutations

Two other mutations normally found in *CYP21P* completely prevent synthesis of an intact enzyme and cause salt-wasting 21-hydroxylase deficiency if they occur in *CYP21*: the nonsense mutation in codon 318 (Q318X) and the 8-nt deletion in exon 3. The 1-nt insertion in exon 7 of *CYP21P* has generally not been identified as an independent mutation in patients with 21-hydroxylase deficiency.

### Mutations Affecting Pre-mRNA Splicing

The nucleotide 13 bp before the end of intron 2 (nt656) is A or C in normal individuals. Mutation to G constitutes the single most frequent allele causing classic 21-hydroxylase deficiency.

This mutation causes aberrant splicing of intron 2 with retention of 19 nucleotides normally spliced out of mRNA, resulting in a shift in the translational reading frame. Almost all of the mRNA is aberrantly spliced, but in cultured cells a small amount of normally spliced mRNA is detected. If no other mutations were present, a small amount of normal enzyme might thus be synthesized.

It is not known what proportion of mRNA is normally spliced in the adrenal glands of patients with this mutation. Most (but not all) patients who are homozygous or hemizygous for this mutation have the salt-wasting form of the disorder, indicating that they have insufficient enzymatic activity to permit adequate aldosterone synthesis. Occasionally, presentation of salt-wasting signs is delayed until several months of age in patients carrying this mutation. Putative asymptomatic nt656g homozygotes have been reported but in fact represent PCR typing artifacts.

### Missense Mutations

The missense mutations found in *CYP21P* have varying effects on enzymatic activity when transferred into *CYP21* as gene conversions; they are discussed in ascending order of severity.

Two mutations with relatively mild effects on activity are found most frequently in patients with nonclassic 21-hydroxylase deficiency. V281L occurs in all or nearly all such patients who carry the HLA haplotype B14;DR1, an association that is presumably due to a founder effect. In certain populations (such as Jews of eastern European origin), this is a very common genetic polymorphism, with a gene frequency of more than 10%. In contrast, direct molecular screening of normal newborns in New Zealand yielded a carrier frequency of 2%. Overall, approximately 70% of all nonclassic alleles carry the V281L mutation. This mutation results in an enzyme with 50% of normal activity when 17-hydroxyprogesterone is the substrate but only 20% of normal activity for progesterone. One study suggested that the mutant enzyme is not normally localized in the endoplasmic reticulum, whereas another proposed that heme binding was affected.

The P30L mutation also yields an enzyme with 30–50% of normal activity when expressed in cultured cells. However, enzymatic activity is rapidly lost when the cells are lysed, suggesting that the enzyme is relatively unstable. Patients carrying this mutation tend to have more severe signs of androgen excess than patients carrying V281L. P30L is found in approximately one-sixth of alleles in patients with nonclassic disease, but it may comprise a higher percentage of such alleles in Japan.

As is the case with other microsomal P450 enzymes, *CYP21* is targeted and anchored to the membrane of the endoplasmic reticulum mainly by a hydrophobic “tail” at the amino terminus; this tail is required for enzymatic activity and stability. Most

P450 enzymes have one or more proline residues separating this tail from the remainder of the polypeptide. These residues are predicted to create a turn in the polypeptide chain, and P30L may abolish this turn. Based on studies of other P450 enzymes, this may lead to improper folding of the polypeptide and may interfere with localization in the endoplasmic reticulum. Indeed, the P30L mutant of *CYP21* is poorly localized to the endoplasmic reticulum in some but not all studies.

The I172N mutation is the only one specifically associated with the simple virilizing form of the disease. It results in an enzyme with approximately 1% of normal activity. Improper localization to the endoplasmic reticulum has been demonstrated in some but not all studies. The mutation may disrupt an intramolecular hydrophobic interaction, stabilizing the secondary structure of the enzyme. The mutant enzyme is abnormally sensitive to protease digestion and does not incorporate heme properly.

Because aldosterone is normally secreted at a rate 100–1000 times less than that of cortisol, it is apparent that 21-hydroxylase activity would have to decrease to very low levels before it became rate limiting for aldosterone biosynthesis. Apparently, as little as 1% of normal activity allows sufficient aldosterone synthesis to prevent significant salt wasting in most cases. R356W and a cluster of three missense mutations (I235N/V236E/M238K) both abolish enzymatic activity.

### Other Mutations

Mutations that are apparently not gene conversions (i.e., they are not usually found in *CYP21P*) account for 5–10% of 21-hydroxylase deficiency alleles in most populations. The most frequent of these is P453S, which occurs in a number of different populations. This suggests that *CYP21P* may carry P453S as an occasional polymorphism and that this mutation is transferred to *CYP21* in the same way as the other mutations frequently causing 21-hydroxylase deficiency.

Novel mutations are easy to detect using automated sequencing technologies in centers with well-developed prenatal or neonatal screening programs and thus have been reported at an increased rate during the past few years; most are unique. However, several codons, including W22, P30, G291, R356, and R483, have undergone several independent mutations; thus, these areas may be hot spots for such events.

## Normal Polymorphisms

Several normal polymorphisms have been detected in *CYP21* in the course of initial sequencing of cloned genes by several groups. An extra leucine near the N terminus (this has confused numbering of other mutations in some reports) and D183E also occur in *CYP21P* and presumably represent gene conversions that do not affect activity. K102R, S268T, and N493S do not represent gene conversions.

## DE NOVO RECOMBINATIONS

CAH is unusual among genetic diseases in the high proportion of mutant alleles generated by intergenic recombination. Both *de novo* deletions and *de novo* gene conversions have been documented; the latter usually involve the intron 2 nt656g mutation and comprise approximately 1% of 21-hydroxylase deficiency alleles.

*De novo* recombinations involving *CYP21* have also been documented by PCR in normal sperm and leukocytes. Unequal crossing over is detected only in sperm, confirming that this process takes place only during meiosis. Gene conversions, however, take place at equal frequencies in somatic cells and gametes, suggesting that gene conversions occur mainly in mitosis and that meiotic recombination (i.e., double crossing over) contributes little, if at all, to this process.

The high rate of recombinations involving the *CYP21* genes may reflect their location in the major histocompatibility complex, in which a high recombination rate between genes encoding transplantation antigens may increase the diversity of the immune response and be evolutionarily favored. The mechanism by which recombination rates might be increased is not known.

In addition, both *CYP21* and *CYP21P* have high rates of single nucleotide polymorphisms, particularly in intron 2. The significance of this is uncertain, but it may mean that additional mechanisms besides intergenic recombination generate sequence diversity within the major histocompatibility complex.

## CORRELATIONS BETWEEN GENOTYPE AND PHENOTYPE

The classification of 21-hydroxylase deficiency into salt wasting, simple virilizing, and nonclassic types is a useful way to roughly grade the severity of the disease and to predict the therapeutic interventions

that will likely be required. If molecular diagnosis could predict this classification, it would increase the utility of prenatal diagnosis and neonatal screening and it might serve as a useful diagnostic adjunct to ACTH stimulation tests.

The simplest way to correlate genotype and phenotype is to determine which mutations are characteristically found in each type of 21-hydroxylase deficiency. This is most informative for frequently occurring mutations. As mentioned previously, deletions and large conversions are most often found in salt-wasting patients, the intron 2 nt656g mutation is found in both salt-wasting and simple virilizing patients, I172N is characteristically seen in simple virilizing patients, and V281L and P30L are found in nonclassic patients. This distribution is consistent with the compromise in enzymatic activity conferred by each mutation.

However, patients are usually compound heterozygotes for different mutations, and so this approach has little predictive value in itself. A useful analytic strategy is to consider that 21-hydroxylase deficiency is a recessive disease, and thus the phenotype of each patient is likely to reflect his or her less severely impaired allele. If mutations are provisionally classified by the degree of enzymatic compromise—severe (also termed type A), moderate (type B), or mild (type C)—then one might hypothesize that salt-wasting patients would have severe/severe genotypes; simple virilizing patients would have severe/moderate or moderate/moderate genotypes; and nonclassic patients would have severe/mild, moderate/mild, or mild/mild genotypes. Approximately 90% of patients can be correctly classified in this manner.

The salt wasting, simple virilizing, and nonclassic categories are qualitative in nature, and the distinction between simple virilizing and nonclassic disease is necessarily difficult in males in whom signs of androgen excess cannot be detected at birth. Therefore, attempts have been made to correlate genotype with quantitative measures of disease severity, such as basal and ACTH-stimulated 17-hydroxyprogesterone levels, plasma renin:urinary aldosterone ratios, and Prader genital virilization scores. In general, these are no better correlated with genotype than are the broader clinical categories. There is excellent discrimination between severe and mild genotypes but a high degree of overlap between moderate genotypes and those either more and less affected.

Several explanations for the less than complete correspondence between genotype and phenotype are possible. The most obvious is that the severity of the disease falls on a continuum and patients with

disease severity near the “borders” of the various classifications may easily fall on either side of these borders. Several mutations and genotypes seem to be particularly associated with this problem. First, although the intron 2 nt656g mutation is classified as severe, it is clearly “leaky” and may yield enough normally spliced mRNA to ameliorate the enzymatic deficiency in some patients. Second, the I172N mutant has marginal enzymatic function (1% of normal), and this apparently is not always sufficient to prevent salt wasting. Third, many patients who are “discordant” for genotype and phenotype are compound heterozygotes for mutations with different effects on enzymatic activity; thus, some of these patients may have *in vivo* enzymatic activities intermediate between those seen in patients who are homozygous for each mutation. Consistent with this idea, presumed compound heterozygotes for a classic and nonclassic allele as a group have higher stimulated 17-OHP levels than presumed homozygotes for nonclassic alleles. Fourth, in studies relying on detection of known mutations, additional novel mutations within *CYP21* might not be detected and might adversely affect activity.

Finally, genetic or environmental factors other than 21-hydroxylase activity may influence phenotype. For example, genetically based variations in androgen biosynthesis or sensitivity to androgens would be expected to influence the expression of signs of androgen excess.

## MOLECULAR DIAGNOSTICS

One of the most troubling signs of classic 21-hydroxylase deficiency is genital ambiguity in affected females caused by excessive exposure to adrenal androgens *in utero*. This may be at least partially prevented by administration of dexamethasone to the mother starting as early as possible in each pregnancy. Unnecessary treatment of males and unaffected females may be minimized by prenatal diagnosis using chorionic villus sampling and direct detection of *CYP21* mutations after amplification of the gene using PCR. This approach detects mutations on 95% of chromosomes. The same technology can be used as

an adjunct to hormonally based newborn screening for 21-hydroxylase deficiency, although it has not been widely implemented.

A number of sources of confusion should be kept in mind in carrying out molecular diagnosis of this disorder. First, gene conversions may transfer several adjacent mutations to *CYP21* so that more than one mutation may be observed on a single chromosome. Second, if PCR is used to specifically amplify *CYP21*, gene conversions involving the sites used for the primers may prevent one or both chromosomes from being amplified, leading to inaccurate genotyping. This problem is detected by performing PCRs in several overlapping segments. Third, a small percentage of alleles are *de novo* mutations. Fourth, particular mutant alleles may be preferentially amplified in PCR, causing heterozygotes to be typed as homozygous affected.

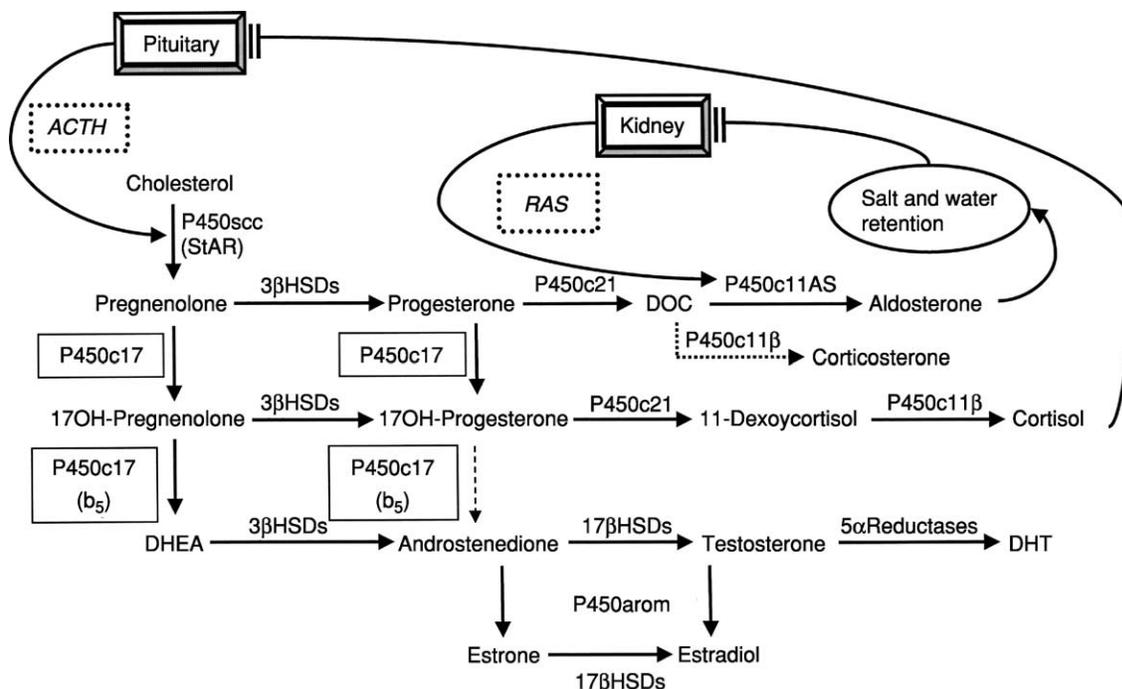
These problems are addressed by examining parents to determine segregation of each mutation and thus ensure that mutations have been identified on both chromosomes in the proband. Use of highly polymorphic microsatellite markers linked to *CYP21* can help identify affected chromosomes when a proband is available.

## See Also the Following Articles

Congenital Adrenal Hyperplasia, Prenatal Diagnosis and Therapy • 3 $\beta$ -Hydroxysteroid Dehydrogenase Deficiency • 11 $\beta$ -Hydroxylase Deficiency • 17 $\alpha$ -Hydroxylase/17,20-Lyase Deficiency • 21-Hydroxylase Deficiency, Classical

## Further Reading

- Merke, D. P., Bornstein, S. R., Avila, N. A., and Chrousos, G. P. (2002). NIH conference. Future directions in the study and management of congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Ann. Intern. Med.* **136**(4), 320–334.
- New, M. I., Carlson, A., Obeid, J., Marshall, I., Cabrera, M. S., Goseco, A., Lin-Su, K., Putnam, A. S., Wei, J. Q., and Wilson, R. C. (2001). Prenatal diagnosis for congenital adrenal hyperplasia in 532 pregnancies. *J. Clin. Endocrinol. Metab.* **86**, 5651–5657.
- White, P. C., and Speiser, P. W. (2000). Congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Endocr. Rev.* **21**, 245–291.



**Figure 1** Major human steroidogenic pathways, highlighting the roles of P450c17. Aldosterone synthesis does not use P450c17; in contrast, cortisol production requires 17 $\alpha$ -hydroxylase activity, and sex steroid biosynthesis requires 17,20-lyase activity as well. The synthesis of cortisol is regulated by ACTH, and cortisol feeds back to attenuate ACTH secretion. Aldosterone production is mainly controlled by the renin-angiotensin system (RAS).

to the appearance of a “17,20-lyase enzyme” in the adrenal glands, but when P450c17, which is present in the adrenal glands from birth, was shown to possess both 17 $\alpha$ -hydroxylase and 17,20-lyase activities, alternate explanations for the acquisition of adrenal DHEA synthesis were sought. Thus, 17,20-lyase activity is not only controlled in a tissue-specific manner but also, in the adrenal glands, it is regionally and temporally or developmentally regulated. These phenomena are partially explained by the biochemistry of P450c17.

### Biochemistry

P450c17 is a 57-kDa “microsomal” cytochrome P450 located in the smooth endoplasmic reticulum. P450c17 receives two electrons from NADPH via the flavoprotein P450 oxidoreductase (CPR) in two discrete one-electron steps, with substrate and molecular oxygen binding between these transfers of electrons. Although the catalytic cycle of P450c17 uses the same initial steps whether pregnenolone, progesterone, or their 17 $\alpha$ -hydroxylated derivatives are substrates, P450c17 catalyzes at least three fundamentally distinct chemical transformations. First, the

17 $\alpha$ -hydroxylase reaction is a typical P450-mediated hydroxylation, in which an oxygen atom derived from molecular oxygen is inserted into a C–H bond. Human P450c17 also 16 $\alpha$ -hydroxylates at least progesterone and DHEA, presumably by the same mechanism. Second, human P450c17 converts 17 $\alpha$ -hydroxysteroids to C<sub>19</sub> steroids via the 17,20-lyase reaction, in which a carbon–carbon bond is oxidatively cleaved to yield the 17-ketosteroid and acetic acid. Unlike the 17 $\alpha$ -hydroxylase reaction, which is approximately equally efficient for both  $\Delta^5$  and  $\Delta^4$  steroids, the human 17,20-lyase reaction is approximately two orders of magnitude more efficient with the  $\Delta^5$  steroid 17 $\alpha$ -hydroxypregnenolone than with its  $\Delta^4$  analog, 17 $\alpha$ -hydroxyprogesterone. This discrepancy results from the combination of a 10-fold higher  $K_m$  and a 10-fold slower  $V_{max}$  for 17 $\alpha$ -hydroxyprogesterone compared to 17 $\alpha$ -hydroxypregnenolone. Kinetic constants obtained for the four principal reactions using his-tagged modified P450c17 purified from *Escherichia coli* and native microsomal P450c17 expressed in yeast are remarkably consistent, confirming differences in both affinity and steady-state turnover rates for the 17,20-lyase substrates.

Third, in the presence of cytochrome b<sub>5</sub>, an alternate pathway of pregnenolone metabolism forms androst-5,16-diene-3 $\beta$ -ol via a different carbon-carbon bond cleavage reaction. In pigs, this pathway forms pheromones, but the roles of these  $\Delta^{5,16}$ -dienes in humans are not known.

The electron transfer proteins CPR and cytochrome b<sub>5</sub>, collectively referred to as redox partners, profoundly influence both the rate of P450c17 catalysis and the ratio of the 17 $\alpha$ -hydroxylase and 17,20-lyase activities. Increased abundance of CPR favors the 17,20-lyase reaction, which ordinarily occurs at a rate only 1–10% as fast as the 17 $\alpha$ -hydroxylations. More dramatically, the addition of cytochrome b<sub>5</sub> to either purified, modified human P450c17 in reconstituted assay systems or yeast microsomes containing P450c17 and CPR increases 17,20-lyase activity more than 10-fold. At least equimolar amounts of cytochrome b<sub>5</sub> relative to P450c17 are required to augment 17,20-lyase activity, but very high molar ratios inhibit catalytic activity, probably by competing with P450c17 for electrons from CPR. Furthermore, apob<sub>5</sub>, which lacks a heme and therefore cannot participate in electron transfer reactions, stimulates 17,20-lyase activity just as well as does holob<sub>5</sub> but does not inhibit it at higher molar ratios. These data indicate that cytochrome b<sub>5</sub> does not act as an electron donor to P450c17 but acts as an allosteric facilitator of the P450c17–CPR catalytic complex. Although the mechanisms whereby CPR and cytochrome b<sub>5</sub> regulate P450c17 activities are not fully elucidated, it is clear that these proteins can profoundly modulate the 17,20-lyase activity of P450c17.

Also influencing 17,20-lyase activity is serine phosphorylation of P450c17. When P450c17 is dephosphorylated *in vitro*, it loses nearly all of its 17,20-lyase activity, but 17 $\alpha$ -hydroxylase activity is not affected. The specific serine residues involved and the responsible kinase have not been identified, so the mechanism by which phosphorylation increases 17,20-lyase activity is unknown but may also be involved in fostering the interaction of P450c17 with its redox partners.

## CLINICAL MANIFESTATIONS

### Combined 17 $\alpha$ -Hydroxylase/ 17,20-Lyase Deficiency

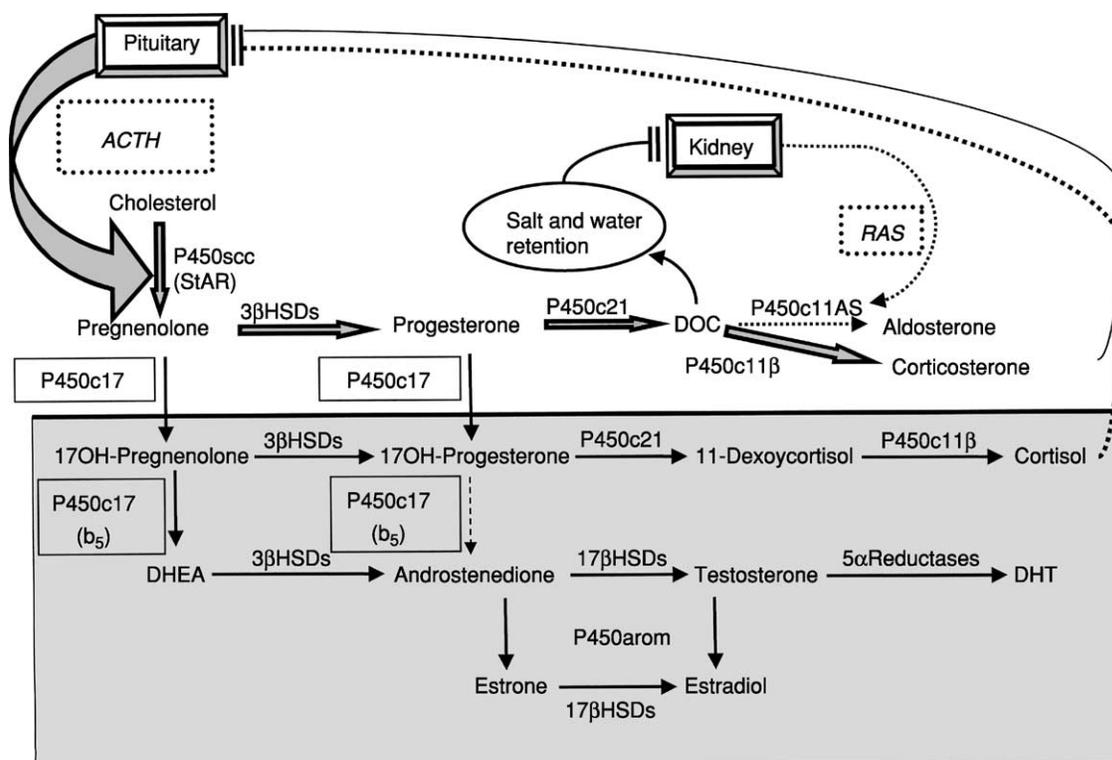
Although rodents survive admirably without adrenal P450c17, mutations in the *CYP17* gene do cause human disease. Patients with combined 17 $\alpha$ -hydroxylase/

17,20-lyase deficiency (17-OHD) cannot make cortisol, but they do make corticosterone. The lower glucocorticoid potency of corticosterone compared to cortisol results in increased production of cortisol precursors when adrenal 17 $\alpha$ -hydroxylase activity is absent. Consequently, the zona fasciculata produces increased amounts of corticosterone and 11-deoxycorticosterone (DOC) in 17-OHD, and the potent mineralocorticoid activity of DOC causes hypertension and hypokalemia while suppressing plasma renin activity and the secretion of aldosterone from the zona glomerulosa (Fig. 2). Hypertension can begin at any age from infancy to late adulthood and can become fixed if not treated for several years.

Mutations that are completely devoid of enzyme activity cause sexual infantilism and failure to progress into puberty (and adrenarche). Genetically female patients can have substantial breast development and cyclical menses if only a small percentage of activity is retained. Genetically male patients with complete deficiencies also have female external genitalia and remain prepubertal. In contrast, partial 17-OHD leads to undervirilization and sexual ambiguity in 46,XY individuals. The potential for further virilization at puberty with appropriate hormonal replacement therapy must be considered when assigning a sex of rearing and developing a management strategy for these patients.

### Isolated 17,20-Lyase Deficiency

Cases of isolated 17,20-lyase deficiency (ILD) were described in the early 1970s, when it was believed that 17 $\alpha$ -hydroxylase and 17,20-lyase were distinct enzymes. Subsequent clinical studies and genetic analyses of the *CYP17* genes from some of these early cases proved that they were actually examples of partial 17-OHD, demonstrating that ILD is not only difficult to diagnose, especially in the newborn, but also either extremely rare or did not exist. By definition, P450c17 mutations in patients with ILD must retain most 17 $\alpha$ -hydroxylase activity; therefore, these patients do not suffer any consequences attributable to adrenal 17 $\alpha$ -hydroxylase deficiency. However, ILD patients have varying degrees of impaired sexual development and intersex features depending on the genetic sex and the severity of the mutation. The few cases studied are 46,XY individuals with sexual ambiguity, but this male predominance probably represents an ascertainment bias because genetic females with ILD might not come to medical attention unless the deficiency was severe enough to cause amenorrhea or pubertal failure.



**Figure 2** Deranged pathways in 17-OHD. The absence of P450c17 activities abrogates cortisol and sex steroid production (shaded region). Lack of cortisol relieves the negative feedback at the pituitary, and ACTH production increases. Abundant production of corticosterone, a weak glucocorticoid, compensates for the lack of cortisol, but the potent mineralocorticoid 11-deoxycorticosterone (DOC) accumulates and causes salt and water retention, hypertension, and suppression of renin and aldosterone.

## MOLECULAR GENETICS

### Combined 17 $\alpha$ -Hydroxylase/ 17,20-Lyase Deficiency

#### **Deletions, Truncations, Frameshifts, and Splicing Errors**

Although no complete deletions of the *CYP17* gene have been reported, the substitution of 518 base pairs (bps) (comprising most of exon 2 and part of exon 3) with 469 bps of unknown DNA has been found to cause 17-OHD (Table I). The first reported P450c17 mutation was the duplication of a CATC following Ile479. This 4-bp duplication was originally observed in Canadian Menonites but has subsequently been found in at least six Dutch Frieslander families. This mutation leaves 95% of the protein unaffected and creates a mutant enzyme in which only the last 25 residues are altered, but the encoded protein is wholly devoid of enzymatic activity. The crucial nature of the carboxy terminus of P450c17 is also indicated by the complete absence of activity when 9 bps are deleted in-frame, removing residues Asp487, Ser488, and Phe489 within the last 17 residues of the protein.

Computer modeling suggests that the P450c17 polypeptide chain forms two sets of strands for two  $\beta$ -sheet units before terminating at the protein surface. In forming these  $\beta$ -sheets, the last 45 residues fold down from the protein surface into the protein core to a point above the heme, forming the top of the substrate-binding pocket and probably contributing structural elements required for catalysis. Thus, the mutation altering the reading frame after Ile479, the mutation deleting residues 487–489, and the mutation Gln461→stop all disrupt or lack this critical stretch of residues and hence are inactive despite retaining the heme binding site.

Several other truncations, frameshifts, and splicing mutations have been described as early as codons 1 and 17 and as late as codon 438. All these errors delete Cys442, which forms the axial sulfhydryl ligand to the heme iron and is thus essential for enzyme activity.

#### **Missense Mutations/Single Amino Acid Changes**

Amino acid substitution mutations in P450c17 can be grouped into mutations in or near the

Table I Mutations Causing 17-OHD: Insertions, Deletions, Frameshifts, and Truncations

Mutation	Exon	Activities <sup>a</sup>		Comment
		17-OHase	Lyase	
Met 1 Ile	1	ND <sup>b</sup>	ND	Possible start codon at Met49
Trp17Stop	1	0	0	Early truncation
7bp dup 118–120	2	0	0	Frameshift, early truncation
G→T +5 intron 2	12	0	0	Frameshift, early truncation
518-bp substitution	2–3	0	0	Frameshift, early truncation
Glu194Stop	3	0	0	Early truncation
Arg239Stop	4	0	0	Early truncation
del TG300–301	5	0	0	Frameshift, early truncation
G→A +5 intron 7	17	0	0	Frameshift, early truncation
del G codon 438	8	0	0	Truncation prior to heme
Gln461Stop	8	0	0	Altered C terminus, ?steroid binding
4-bp dup 478–479	8	0	0	Altered C terminus, ?steroid binding

<sup>a</sup>Expressed as percentage of wild-type activity.

<sup>b</sup>ND, not determined.

substrate-binding pocket, mutations in or near the heme-binding region, and mutations elsewhere in the protein. Mutations Gly90Asp, Arg96Trp, Ser106Pro, and insIle112 map to the  $\beta$ -sheet-rich amino terminus, in or close to regions predicted to form the substrate-binding pocket, and all of these mutations are devoid of activity. The sensitivity of P450c17 to minor changes in this region of the protein is further demonstrated by the finding that the conservative replacement of Ser106 with Thr, the corresponding residue found in rainbow trout P450c17, also abolishes most enzymatic activity.

The best example of a mutation that disrupts heme binding is Arg440His, located two residues away from the heme-liganding Cys442. By comparison with bacterial P450 enzymes of known structure, the positive guanidinium group on Arg440 is predicted to pair with the negative charge on one of the two propionic acid side chains of the heme protoporphyrin ring. Analogous mutations of other P450 enzymes, such as Arg435Cys in P450arom, also cause complete loss of enzyme activity. In the “meander” region, located just prior to the heme-binding domain in the linear sequence of all P450s, two P450c17 mutations have been identified. Both Phe417Cys and the adjacent mutation, Pro409Arg, completely destroy enzyme activity, probably due to impaired heme incorporation (Table II).

Mutation His373Leu yields a protein that fails to incorporate heme, but the large distance of His373 from the heme, both in linear sequence and in

predicted three-dimensional structure, suggests that this mutation causes more global structural alterations that secondarily impair heme binding. Mutation Pro342Thr retains approximately 20% residual activity, and mutations  $\Delta$ Phe53/54 and Tyr64Ser retain 5–15% of wild-type 17 $\alpha$ -hydroxylase and 17,20-lyase activities. These proteins must incorporate heme to have activity, so the effects of these mutations, which are not directly located in crucial regions for enzyme function, must be more subtle than in the completely inactive mutations.

### Isolated 17,20-Lyase Deficiency

The first published cases of ILD that were proven by molecular genetics and biochemical analyses are two 46,XY patients from rural Bahia, Brazil. These patients, from two unrelated but consanguineous families, had ambiguous genitalia and normal 17-hydroxysteroid excretion rates but markedly reduced C<sub>19</sub> steroid production. Each patient was homozygous for an arginine substitution, either Arg347His or Arg358Gln (Table II). These mutations retained the capacity to hydroxylate pregnenolone and progesterone when expressed in transfected cells or yeast microsomes. However, these mutations converted only trace amounts of 17 $\alpha$ -hydroxypregnenolone to DHEA when conditions were optimized by using an excess of both CPR and cytochrome b<sub>5</sub>. Surprisingly, competition experiments using [<sup>3</sup>H]pregnenolone demonstrated that 17 $\alpha$ -hydroxypregnenolone binds

Table II Mutations Causing 17-OHD: In-Frame Insertions and Deletions; Missense

Mutation	Exon	Activities <sup>a</sup>		Comment
		17-OHase	Lyase	
Arg35Leu	1	35	35	Membrane anchor
del Phe53/54	1	<25	<5	Substrate binding region
Tyr64Ser	1	15	ND <sup>b</sup>	Substrate and membrane binding
Gly90Asp	2	ND	ND	Substrate binding region
Phe93Cys	2	10	10	Substrate binding region
Arg96Trp	2	<5	<5	Substrate binding region
Ser106Pro	2	0	0	Substrate binding region
Ile112 duplication	2	0	0	Substrate binding region
Phe114Val	2	<5	0	Substrate binding region
Asp116Val	2	35	10	Substrate binding region
Asn177Asp	3	10	10	Substrate binding region
Pro342Thr	6	20	20	Near redox partner binding region
His373Leu	6	0	0	No heme binding
Arg347His <sup>c</sup>	6	65	≪5	Redox partner binding site
Arg347Cys	6	15	0	Redox partner binding site
Arg358Gln <sup>c</sup>	6	65	≪5	Redox partner binding site
Pro409Arg	7	0	0	Meander region
Arg416Cys	8	10	10	Meander region
Phe417Cys	8	0	0	Meander region, no heme binding
Arg440His	8	0	0	No heme binding
del Asp487–Phe489	8	0	0	Altered C terminus, ?steroid binding
Arg496Cys	8	<10	<5	Altered C terminus, ?steroid binding
Arg496His	8	35	35	Altered C terminus, ?steroid binding

<sup>a</sup>Expressed as percentage of wild-type activity.

<sup>b</sup>ND, not determined.

<sup>c</sup>Isolated 17,20-lyase deficiency.

to the mutant enzymes with an affinity equivalent to that of the wild-type enzyme, proving that these mutations did not affect substrate access or binding.

To understand the biochemistry of these extraordinary mutations, we turned to computational chemistry for insight. Arginines 347 and 358 are located in the J' and K helices, beneath the heme ring on the distal surface of P450c17 that interacts with redox partners and on the opposite side of the heme ring from the substrate-binding pocket. The mutations Arg347His and Arg358Gln neutralize some of the positive charges predicted to lie on this surface and to interact with the negative charges in the FMN-binding domain of CPR. The neutralization of two other positive charges on this surface (lysine 89 and arginine 449) by site-directed mutagenesis also resulted in preferential loss of 17,20-lyase activity. Furthermore, the neutralization of the corresponding arginines of rat P450c17 by mutagenesis (mutation Arg346Ala in particular) also preferentially disrupted 17,20-lyase activity. The consistent loss of 17,20-lyase

activity in these mutations, despite their wide distribution in the linear sequence of P450c17, makes a compelling argument that residues 89, 347, 358, and 449 congregate in the redox partner binding site and are critical for 17,20-lyase activity. This observation is similar to those of earlier biochemical studies that highlighted the importance of CPR and cytochrome b<sub>5</sub> abundance for optimal 17,20-lyase reaction. Both the Arg347His mutation and a closely related Arg347Cys mutation have recently been found in Dutch patients with well-characterized ILD.

The mutation Phe417Cys was formerly implicated as a cause of ILD. Although the residue Phe417 is not predicted to lie within the redox partner binding surface, the aromatic side chain may participate in forming one edge of this surface. However, more careful study demonstrated that this mutation was completely devoid of both 17 $\alpha$ -hydroxylase and 17,20-lyase activities. This example illustrates the complexity of distinguishing ILD from partial 17-OHD and emphasizes the need for detailed

biochemical analyses to prove that a P450c17 mutation causes ILD. It is possible that mutations elsewhere in P450c17, such as in the active site, may cause ILD, but as of mid-2003, the only mutations proven to cause ILD all neutralize positive charges in the redox partner binding site.

Finally, one male pseudohermaphrodite with congenital methemoglobinemia due to a mutation in the gene for cytochrome b<sub>5</sub> has been reported. It is possible that undervirilization of this patient resulted from fetal testosterone deficiency due not to a mutation in the *CYP17* gene but, rather, the loss of the stimulatory influence of cytochrome b<sub>5</sub> on 17,20-lyase activity. Unfortunately, a detailed endocrine evaluation of this patient was not included with the report, but this case indicates the physiologic importance of b<sub>5</sub> in P450c17 physiology and androgen biosynthesis.

## CONCLUSION

Human deficiencies in P450c17 can manifest clinically in a variety of forms depending on the specific enzymatic properties of the mutation. The variable clinical expression is due in part to the multiple activities and functions of P450c17 in human physiology. ILD can exist when mutations in the redox partner binding site selectively impair 17,20-lyase activity, but this disease is difficult to distinguish from partial 17-OHD. Careful clinical, genetic, and biochemical studies are necessary to precisely define diseases other than complete, classical 17-OHD.

## See Also the Following Articles

3 $\beta$ -Hydroxysteroid Dehydrogenase Deficiency • 11 $\beta$ -Hydroxylase Deficiency • 21-Hydroxylase Deficiency, Classical • 21-Hydroxylase Deficiency, Genetics of

## Further Reading

- Auchus, R. J. (2001). The genetics, pathophysiology, and management of human deficiencies of P450c17. *Endocrinol. Metab. Clin. North Am.* **30**, 101–119.
- Auchus, R. J., and Miller, W. L. (1999). Molecular modeling of human P450c17 (17 $\alpha$ -hydroxylase/17,20-lyase): Insights into reaction mechanisms and effects of mutations. *Mol. Endocrinol.* **13**, 1169–1182.
- Auchus, R. J., Lee, T. C., and Miller, W. L. (1998). Cytochrome b<sub>5</sub> augments the 17,20 lyase activity of human P450c17 without direct electron transfer. *J. Biol. Chem.* **273**, 3158–3165.
- Biason, A., Mantero, F., Scaroni, C., Simpson, E. R., and Waterman, M. R. (1991). Deletion within CYP17 gene together with insertion of foreign DNA is the cause of combined complete 17 $\alpha$ -hydroxylase/17,20-lyase deficiency in an Italian patient. *Mol. Endocrinol.* **5**, 2037–2045.
- Chung, B., Picado-Leonard, J., Haniu, M., Bienkowski, M., Hall, P. F., Shivley, J. E., and Miller, W. L. (1987). Cytochrome P450c17 (steroid 17 $\alpha$ -hydroxylase/17,20 lyase): Cloning of human adrenal and testis cDNAs indicates the same gene is expressed in both tissues. *Proc. Natl. Acad. Sci. USA* **84**, 407–411.
- Geller, D. H., Auchus, R. J., Mendonça, B. B., and Miller, W. L. (1997). The genetic and functional basis of isolated 17,20 lyase deficiency. *Nature Genet.* **17**, 201–205.
- Gupta, M. K., Geller, D. H., and Auchus, R. J. (2001). Pitfalls in characterizing P450c17 mutations associated with isolated 17,20-lyase deficiency. *J. Clin. Endocrinol. Metab.* **86**, 4416–4423.
- Katagiri, M., Kagawa, N., and Waterman, M. R. (1995). The role of cytochrome b<sub>5</sub> in the biosynthesis of androgens by human P450c17. *Arch. Biochem. Biophys.* **317**, 343–347.
- Kater, C. E., and Biglieri, E. G. (1994). Disorders of steroid 17 $\alpha$ -hydroxylase deficiency. *Endocrinol. Metab. Clin. North Am.* **23**, 341–357.
- Lee-Robichaud, P., Wright, J. N., Akhtar, M. E., and Akhtar, M. (1995). Modulation of the activity of human 17 $\alpha$ -hydroxylase-17,20-lyase (CYP17) by cytochrome b<sub>5</sub>: Endocrinological and mechanistic implications. *Biochem. J.* **308**, 901–908.
- Nakajin, S., and Hall, P. F. (1981). Microsomal cytochrome P450 from neonatal pig testis. Purification and properties of a C<sub>21</sub> steroid side-chain cleavage system (17 $\alpha$ -hydroxylase-C<sub>17,20</sub> lyase). *J. Biol. Chem.* **256**, 3871–3876.
- Onoda, M., and Hall, P. F. (1982). Cytochrome b<sub>5</sub> stimulates purified testicular microsomal cytochrome P450 (C<sub>21</sub> side-chain cleavage). *Biochem. Biophys. Res. Commun.* **108**, 454–460.
- Picado-Leonard, J., and Miller, W. L. (1987). Cloning and sequence of the human gene encoding P450c17 (steroid 17 $\alpha$ -hydroxylase/17,20 lyase): Similarity to the gene for P450c21. *DNA* **6**, 439–448.
- van der Akker, E. L. T., Koper, J. W., Boehmer, A. L. M., Themmen, A. P. N., Verhoef-Post, M., Timmerman, M. A., Otten, B. J., Drop, S. L. S., and de Jong, F. H. (2002). Differential inhibition of 17 $\alpha$ -hydroxylase and 17,20 lyase activities by three novel missense CYP17 mutations identified in patients with P450c17 deficiency. *J. Clin. Endocrinol. Metab.* **87**, 5714–5721.
- Zhang, L. H., Rodriguez, H., Ohno, S., and Miller, W. L. (1995). Serine phosphorylation of human P450c17 increases 17,20 lyase activity: Implications for adrenarche and the polycystic ovary syndrome. *Proc. Natl. Acad. Sci. USA* **92**, 10619–10623.

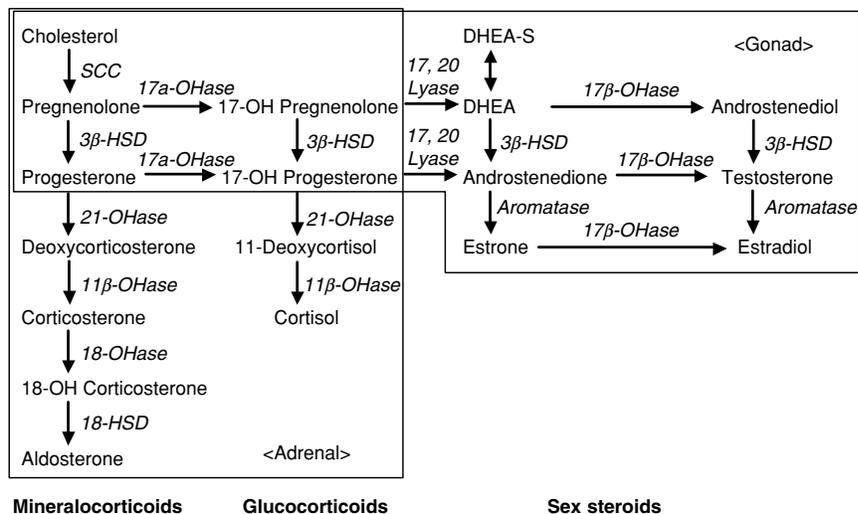


Figure 1 Simplified scheme for steroidogenesis.

## LABORATORY FINDINGS OF 3 $\beta$ -HSD DEFICIENCY

An elevated ratio of  $\Delta^5$  to  $\Delta^4$  steroids is considered the best biological finding for the diagnosis of 3 $\beta$ -HSD deficiency. However, it is well-known that levels of 17-hydroxyprogesterone and  $\Delta^4$ -androstenedione and other  $\Delta^4$  steroids in plasma are frequently elevated in 3 $\beta$ -HSD-deficient patients. Such observations suggest functional 3 $\beta$ -HSD that is expressed in peripheral tissues and responsible for extra-adrenal and extragonadal conversion of 5-ene-hydroxysteroid precursors into corresponding  $\Delta^4$ -3-ketosteroids.

The ACTH-stimulated hormonal profiles in reported non-salt-losing 3 $\beta$ -HSD-deficiency children with *HSD3B2* gene mutation and in children with precocious puberty without *HSD3B2* gene mutation were unequivocally distinguished by the hormonal parameter in the C<sub>21</sub> glucocorticoid pathway. Hormonal criteria to identify inherited 3 $\beta$ -HSD deficiency have been proposed based on hormonal findings in patients with genotype-proven 3 $\beta$ -HSD deficiency and in those with a normal genotype (Table I).

## MOLECULAR BASIS OF 3 $\beta$ -HSD DEFICIENCY

The structure of the 3 $\beta$ -HSD gene family has been characterized in humans and several other vertebrate species. The *HSD3B1* gene encodes a 372-amino acid protein and is mainly expressed in the placenta and peripheral tissues, such as skin and mammary glands.

In contrast, the *HSD3B2* gene encodes a 371-amino acid protein that shares 93.5% identity with *HSD3B1*, and it is expressed in the adrenal and gonads. The structures of *HSD3B1* and *HSD3B2* genes consist of four exons within a DNA fragment of 7.8 kb and have been assigned to the chromosome 1p13.1 region at 1 or 2 cM of the centromeric marker D1Z5.

## Salt-Losing Form of Classic 3 $\beta$ -HSD Deficiency

Nineteen different mutations have been detected in the *HSD3B2* gene in 21 patients suffering from the salt-losing form. No mutations have been detected in the *HSD3B1* gene (Table II; Fig. 2).

It is believed that no functional type II 3 $\beta$ -HSD isoenzyme is expressed in adrenals and gonads of patients bearing mutations leading to a putative truncated protein (W171X, 558/insC/559, n687delC, R249X, n867delG, and Y308X), which is consistent with the severity of disease. Furthermore, A10E, E142K, Y253N, A10E, L205P, P222Q, P222T, T259M, T259R, and 687del27 mutants show no detectable activity. Therefore, in compound heterozygous patients with E142K:W171X or 186/ins/187:Y253N, it is believed that no functional type II 3 $\beta$ -HSD isoenzyme is expressed. G15D mutation not only decreases the apparent affinity for both substrate and cofactor but also alters the proper intracellular localization or its association with intact membranes *in vivo*, which may exert some strain, preventing adoption of its final maximally efficient conformation. Interestingly, in the case of L108W and P186L

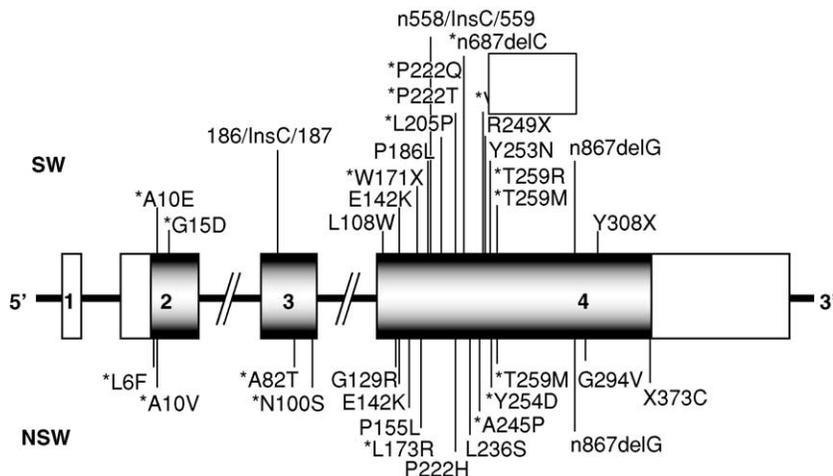
Table I Proposed Hormonal Criteria for Genotype-Proven 3 $\beta$ -HSD Deficiency<sup>a</sup>

	Neonates (age < 42 d)		Children			Adolescents/Adults	
			Genotype-proven patients		Genotype-normal patients w/PP		
	Genotype-proven patients w/AG or SW	Genotype-normal patients w/AG	AG	PP $\pm$ AG	T II-III	Genotype-proven males w/AG	Genotype-normal females w/hirsutism and menstrual disorder
			T I	T II-III			
ACTH-stimulated $\Delta$ 5-17P (nmol/L)	$\geq 378$ ( $\geq 5.3$ SD)	ND	$\geq 165$ ( $\geq 35$ SD)	$\geq 294$ ( $\geq 54$ SD)	$\leq 72$ ( $\leq 11$ SD)	$\geq 289$ ( $\geq 21$ SD)	$\leq 150$ ( $\leq 12$ SD)
Baseline $\Delta$ 5-17P (nmol/L)	$\geq 84$ ( $\geq 6.8$ SD)	ND	$\geq 26.4$ ( $\geq 12$ SD)	$\geq 29$ ( $\geq 10$ SD)	$\leq 35$ ( $\leq 12$ SD)	$\geq 159$ ( $\geq 74$ SD)	$\leq 45$ ( $\leq 16$ SD)
ACTH-stimulated $\Delta$ 5-17P/F	$\geq 434$ ( $\geq 6.4$ SD)	ND	$\geq 216$ ( $\geq 23$ SD)	$\geq 363$ ( $\geq 38$ SD)	$\leq 67$ ( $\leq 5$ SD)	$\geq 4010$ ( $\geq 221$ SD)	$\leq 151$ ( $\leq 10$ SD)
Baseline $\Delta$ 5-17P/F	$\geq 464$ ( $\geq 4.6$ SD)	ND	$\geq 94$ ( $\geq 15$ SD)	$\geq 103$ ( $\geq 16$ SD)	$\leq 59$ ( $\leq 8$ SD)	$\geq 1943$ ( $\geq 193$ SD)	$\leq 43$ ( $\leq 5$ SD)

<sup>a</sup>The genotype-proven patients' lowest hormonal value and the value expressed as the SD above mean value of the appropriate control/normal subjects and the hormonal criteria for exclusion of 3 $\beta$ -HSD deficiency by the genotype-normal patients' highest hormonal value and the value expressed as the SD above mean value of the control/normal subjects are shown. The hormonal criteria for infants aged 2.5–6 and 7–20 months are not defined due to insufficient data. AG, Ambiguous genitalia; ND, no sufficient data available; PP, premature pubarche; pts, patients; SW, salt wasting; T I, Tanner I public hair stage; T II-III, Tanner II-III public hair stage. Modified from Lutfallah *et al.*, 2002.

Table II Mutations in the Type II 3 $\beta$ -HSD Gene in Patients with Classical 3 $\beta$ -HSD Deficiency

Family	Origin	Patient	Karyotype	Salt-wasting	Mutant alleles	Reference
1	French-Canadian	1	46, XX	Yes	A 10E/A10E	Alos <i>et al.</i> , 2000
2	French-Canadian	2	46, XY	Yes	A 10E/A10E	Alos <i>et al.</i> , 2000
3	Algerian	3	46, XY	Yes	G15D/G15D	Rheume <i>et al.</i> , 1995
4	Spanish/Portugese	4	46, XY	Yes	L108W/P186L	Sanchez <i>et al.</i> , 1994
5	American	5	46, XY	Yes	E142K/W171X	Simard <i>et al.</i> , 1993
6	Japanese	6	46, XY	Yes	L205P/L205P	Katsumata <i>et al.</i> , 1995
		7	46, XX	Yes		
7	Algerian	8	46, XY	Yes	P222Q/P222Q	Moisan <i>et al.</i> , 1999
		9	46, XX	Yes		
8	Dutch	10	46, XY	Yes	Y253N/558/insC/559	Simard <i>et al.</i> , 1993
9	French	11	46, XY	Yes	T259M/867delG	Moisan <i>et al.</i> , 1999
		12	46, XX	Yes		
10	Japanese	13	46, XY	Yes	T259R/T259R	Tajima <i>et al.</i> , 1995
		14	46, XX	Yes		
11	Sri-Lanka	15	46, XY	Yes	687del27/687del27	Moisan <i>et al.</i> , 1999
12	Sri-Lanka	16	46, XY	Yes	687del27/687del27	Moisan <i>et al.</i> , 1999
13	Mexican-Hispanic	17	46, XY	Yes	V248N/V248N	Chang <i>et al.</i> , 1993
14	Eastern European Jewish	18	46, XY	Yes	P222T/P222T	Pang <i>et al.</i> , 2002
15	French-Canadian	19	46, XY	Yes	V248N/R249X	Chang <i>et al.</i> , 1993
16	Pakistani	20	46, XY	Yes	318delC/273delAA	Zhang <i>et al.</i> , 1996
17	Taiwanese	21	46, XY	Yes	T259M/T259M	Zhang <i>et al.</i> , 2000
18	French	22	46, XY	No	N100S/N100S	Mebarki <i>et al.</i> , 1995
19	French	23	46, XY	No	P155L/G294V	Moisan <i>et al.</i> , 1999
		24	46, XY	No		
20	Brazilian	25	46, XX	No	P222H/G129R	Moisan <i>et al.</i> , 1999
		26	46, XX	No		
21	Egypt	27	46, XY	No	A10V/A10V	Moisan <i>et al.</i> , 1999
		28	46, XY	No		
22	Brazilian	29	46, XX	No	A82T/A82T	Moisan <i>et al.</i> , 1999
		30	46, XY	No		
		31	46, XY	No		
23	Brazilian	32	46, XX	No	A82T/A82T	Moisan <i>et al.</i> , 1999
24	American	33	46, XX	No	G129R/6651(G to T)	Rheume <i>et al.</i> , 1995
		34	46, XY	No		
25	Scottish	35	46, XY	No	L173R/L173R	Moisan <i>et al.</i> , 1999
		36	46, XX	No		
26	French	37	46, XY	No	L236S/867delG	Moisan <i>et al.</i> , 1999
27	Turkish	38	46, XY	No	A245P/A245P	Simard <i>et al.</i> , 1993
28	American	39	46, XX	No	Y254D/Y254D	Sanchez <i>et al.</i> , 1994
29	Brazilian	40	46, XX	No	T259M/T259M	Moisan <i>et al.</i> , 1999
30	Brazilian	41	46, XX	No	T259M/T259M	Moisan <i>et al.</i> , 1999
31	Pakistani	42	46, XY	No	L6F/L6F	Zhang <i>et al.</i> , 2000
32	?	43	46, XY	No	N100S/266delA	McCartin <i>et al.</i> , 2000
33	?	44	?	No	A82T/A82T	McCartin <i>et al.</i> , 2000
34	?	45	?	No	T259M/T259M	Marui <i>et al.</i> , 2000
35	?	46	46, XY	No	G128R/P222M	Marui <i>et al.</i> , 2000
36	Caucasian	47	46, XX	No	X373C/E142K	Pang <i>et al.</i> , 2000



**Figure 2** Mutations in the type II 3 $\beta$ -HSD gene in patients with 3 $\beta$ -HSD deficiency. Asterisks indicate homozygous mutations.

mutations, low but detectable isoenzyme activities have been detected in intact transfected cells. The L108W and P186L mutant proteins possess a marked reduced affinity for substrate and the cofactor. The *in vitro* overall efficiency was ~0.1–0.3%. Such a low level of activity is clearly inadequate for the formation of a sufficient number of mineralocorticoids to avoid severe salt losing. These findings suggest that the absence of functional type II 3 $\beta$ HSD due to *HSD3B2* mutation causes the severe salt-losing form of 3 $\beta$ -HSD deficiency.

### Non-Salt-Losing Form of Classic 3 $\beta$ -HSD Deficiency

Sixteen different point mutations have been detected in the *HSD3B2* gene in 19 families. In general, the presence of mutant 3 $\beta$ -HSD in patients with non-salt-wasting 3 $\beta$ -HSD deficiency corresponds closely with the severity of the disease state. L6F, A245P, L173R, G129R, A82T, A10V, and G294V mutations found as homozygous mutations showed decreased but detectable enzyme activity in transfection studies. In contrast, P155L, P222H, and Y254D mutant proteins found as compound heterozygous mutations showed no enzyme activity. Patients with N100S mutation demonstrated a poor male differentiation without severe salt losing. The overall efficiency of the N100S protein is similar to that of L108W and P186L. These findings suggest that the weak residual activity of N100S is sufficient to prevent salt loss at high renin levels, although it is insufficient for preserving normal male differentiation. Homozygous mutation T259M detected in two patients with non-salt-losing

3 $\beta$ -HSD deficiency showed no 3 $\beta$ -HSD activity and the instability of the mutant protein.

Interestingly, amino acids Gly15, Ala82, Asn100, Leu108, Gly129, Glu142, Leu173, Tyr253, Tyr254, and Thr259, which are the sites of the missense mutations, are conserved among all members of the vertebrate 3 $\beta$ -HSD family. Therefore, characterization of the molecular basis of classical 3 $\beta$ -HSD deficiency provides an understanding of the structure–function relationship of the 3 $\beta$ -HSD family.

### Nonclassic 3 $\beta$ -HSD Deficiency

3 $\beta$ -HSD deficiency in older females is called nonclassic or late-onset 3 $\beta$ -HSD deficiency. However, no mutations have been detected in the coding regions, exon–intron junctions, or the 3' noncoding region, including the polyadenylation site of both *HSD3B1* and *HSD3B2* genes. The fact that no mutations have been detected in patients suffering from nonclassic 3 $\beta$ -HSD deficiency suggests that this disorder does not result from a mutant type II 3 $\beta$ -HSD isoenzyme.

### See Also the Following Articles

Congenital Adrenal Hyperplasia, Prenatal Diagnosis and Therapy • 11 $\beta$ -Hydroxylase Deficiency • 17 $\alpha$ -Hydroxylase/17,20-Lyase Deficiency • 21-Hydroxylase Deficiency, Classical • 21-Hydroxylase Deficiency, Genetics of

### Further Reading

Alos, N., Moisan, A.-M., Ward, L., Desrochers, M., Legault, L., Leboeuf, G., Van Vliet, G., and Simard, J. (2000). A novel A10E homozygous mutation in the *HSD3B2* gene causing severe

- salt-wasting 3 $\beta$ -hydroxysteroid dehydrogenase deficiency in 46,XX and 46,XY French-Canadians: Evaluation of gonadal function after puberty. *J. Clin. Endocrinol. Metab.* **85**, 1968–1974.
- Bongiovanni, A. M. (1962). The adrenogenital syndrome with deficiency of  $\beta$ -hydroxysteroid dehydrogenase. *J. Clin. Invest.* **41**, 2086–2092.
- Lutfallah, C., Wang, W., Mason, J. I., Chang, Y. T., Haider, A., Rich, B., Castro-Magana, M., Copeland, K. C., David, R., and Pang, S. (2002). Newly proposed hormonal criteria via genotype proof for type II 3 $\beta$ -hydroxysteroid dehydrogenase deficiency. *J. Clin. Endocrinol. Metab.* **87**, 2611–2622.
- Morel, Y., Mebarki, F., Rheaume, E., Sanchez, R., Forest, M. G., and Simard, J. (1997). Structure–function relationships of 3 $\beta$ -hydroxysteroid dehydrogenase: Contribution made by the molecular genetics of 3-beta-hydroxysteroid dehydrogenase deficiency. *Steroids* **62**, 176–184.
- Morisan, A. M., Ricketts, M. L., Tardy, V., Desrochers, M., Mebarki, F., Chaussain, J. L., Cabrol, S., Raux-Demay, M. C., Forest, M. G., and Sippell, W. G. (1999). New insight into the molecular basis of 3 $\beta$ -hydroxysteroid dehydrogenase deficiency: Identification of eight mutations in the HSD3B2 gene in eleven patients from seven new families and comparison of the functional properties of twenty-five mutant enzymes. *J. Clin. Endocrinol. Metab.* **84**, 4410–4425.
- Nishi, T., and Tezuka, T. (1992). Mild adrenal 3 $\beta$ -hydroxysteroid dehydrogenase deficiency in children with accelerated growth, premature puberche and/or hirsutism. *Eur. J. Pediatr.* **151**, 19–23.
- Rheaume, E., Simard, J., Morel, Y., Mebarki, F., Zachamann, M., Forest, M. G., New, M. I., and Labrie, F. (1992). Congenital adrenal hyperplasia due to point mutations in the type II 3 $\beta$ -hydroxysteroid dehydrogenase gene. *Nature Genet.* **1**, 239–245.



## Hyperadrenalism

see Hypercortisolism and Cushing's Syndrome

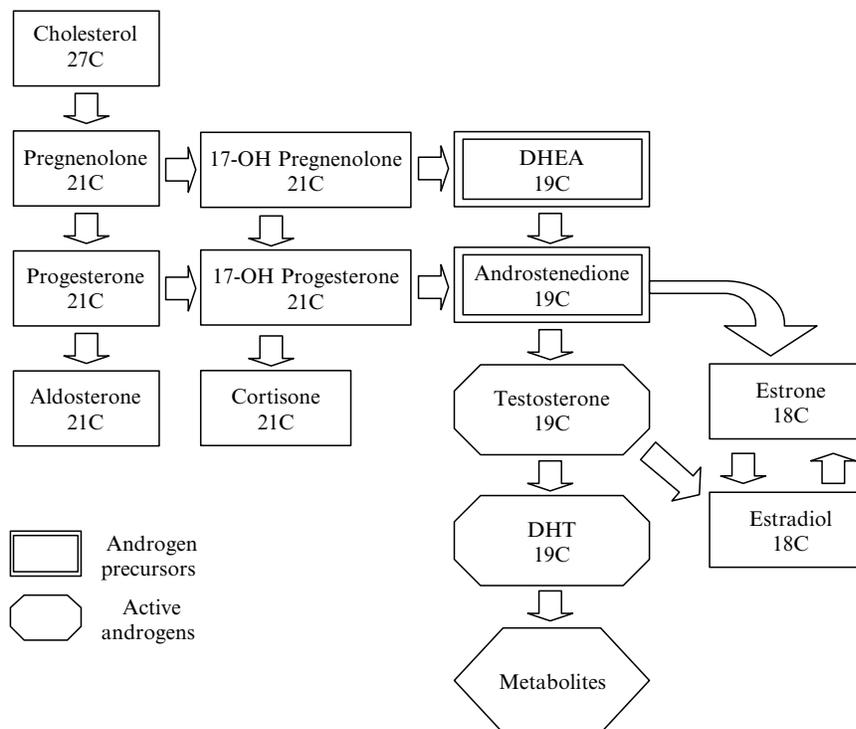


Figure 1 Steroid biosynthesis pathway and classification of the main androgens.

at the level of hair follicles and derivatives of the urogenital sinus and urogenital tubercle.

The human ovary produces all three classes of steroid sex hormones: progestins (C-21 compounds), androgens (C-19 compounds), and estrogens (C-18 compounds). It differs from the adrenal gland in that it is deficient in 21-hydroxylase and 11 $\beta$ -hydroxylase reactions; therefore, glucocorticoids and mineralcorticoids are not produced in ovarian tissue. The adrenal sex steroids are intermediate compounds in the production of glucocorticoids and mineralcorticoids; thus, 21-hydroxylase deficiency at the adrenal level or androgen-producing tumor cells will lead to androgen accumulation and hyperandrogenemia.

### Blood Transport

Androgens are present in the bloodstream in bound and unbound forms. The androgenic activity is mediated by free unbound androgens, hence the primary importance of the free portion of androgens. One or 2% of free circulating testosterone is accepted as normal for females. The rest is bound to albumin and sex hormone-binding globulin (SHBG). This globulin is produced in the liver, and its serum levels are a sensitive marker of free testosterone and, hence, androgenic

activity. Normal-high levels of SHBG are maintained during pregnancy, hyperthyroidism, or estrogen administration. SHBG is reduced in overweight patients, those with hyperandrogenemia, or during treatment with corticoids, progestins, and growth hormones. Hyperinsulinemia, related to insulin resistance that characterizes PCOS, is marked by low SHBG levels.

### ANDROGEN SOURCES

Research suggests that the main androgen production occurs in the ovaries, adrenals, and peripheral conversion sites, such as the liver, skin, and adipose tissue. In normal women, the production rate of androstenedione is 3 mg/day and that of testosterone is 0.2–0.3 mg/day, of which approximately 50% is derived from peripheral conversion of androstenedione. The remaining 50% of testosterone is produced in approximately equal amounts by the ovaries and the adrenals. In women with signs of hyperandrogenism, such as hirsutism and acne, ovarian and adrenal testosterone production may increase, accounting for up to 75% of total testosterone. Although androstenedione has the highest production rate, testosterone is the major circulating androgen and its free portion is correlated with hyperandrogenism. The testosterone

target organs are the Wolffian duct structures, such as the epididymis, the vas deferens, and the seminal vesicle. At the androgen receptor level, DHT is most important in many sensitive tissues, including hair follicles, male external genitalia, urethra, and prostate. In males, DHT is a product of intracellular testosterone conversion mediated by 5 $\alpha$ -reductase, whereas in females DHT is in part derived from androstenedione and DHEA. Other important androgens, particularly DHEAS, are products of the adrenals.

## HYPOTHESIS UNDERLYING FUNCTIONAL HYPERANDROGENISM

Functional hyperandrogenism is not a very well-defined entity. It comes to the clinician's attention when it causes functional or esthetic inconvenience. Many authors have reported normal androgen levels in hirsute women, suggesting the presence of an idiopathic form as the most frequent cause of this problem. On the other hand, a study by Rossi *et al.* on this group of patients using two stimulation tests, gonadotropin-releasing hormone analogue triptoreline and adrenocorticotropin hormone, found mild to moderate ovarian and adrenal abnormalities in the steroidogenesis of 35.4% of these patients.

As shown in Table I, a variety of causes of functional hyperandrogenism have been noted in the literature based on ovarian, adrenal, or skin level abnormalities. Increased cutaneous 5 $\alpha$ -reductase with overproduction of DHT in hirsute women compared with nonhirsute causes hyperactivity of existing hair follicles and pilosebaceous glands. PCOS patients with insulin resistance and hyperinsulinemia have a

**Table I** Possible Causes of Functional Hyperandrogenism<sup>a</sup>

Skin-level abnormalities	
Increased cutaneous 5 $\alpha$ -reductase	→ overproduction of DHT
Ovarian-level abnormalities	
PCOS (thickened ovarian stroma)	→ overproduction of testosterone
Reduced aromatization of androgens into estrogens	→ accumulation of testosterone
Dysregulation of the 17 $\alpha$ -hydroxylase/17,20-lyase enzyme complex	→ increase in testosterone
Adrenal-level abnormalities	
Dysregulation of the 17 $\alpha$ -hydroxylase/17,20-lyase enzyme complex	→ increase in testosterone

<sup>a</sup>Abnormalities at various tissue levels will result in an increase in androgen precursors/active androgens.

thickened ovarian stroma that is capable of excessive androgen production. Reduced aromatization of androgens into estrogens at the level of granulosa cells in the ovaries of hyperandrogenic-anovulatory women results in androgen accumulation. Dysregulation of the 17 $\alpha$ -hydroxylase/17,20-lyase enzyme complex causes both ovarian and adrenal hyperandrogenism through accumulation of androgens due to overactivity of this enzyme complex.

## SIGNS, SYMPTOMS, AND DIAGNOSIS OF HYPERANDROGENISM

### Signs and Symptoms

As described in Table II, functional hyperandrogenism is usually discovered as a result of undesirable signs and symptoms. Hirsutism, defined as excess facial and body hair in women, and acne are probably the main complaints of women with hyperandrogenism. Other physical signs and symptoms include menstrual disturbances, irregular ovulation, muscle development, obesity, clitoral hypertrophy, male pattern baldness, voice deepening, and changes in libido.

### Diagnosis

Diagnosis of hyperandrogenism is based on medical history, physical examination, and laboratory testing. The main purpose of these is to rule out serious underlying disorders, such as virilizing ovarian or adrenal tumors, congenital adrenal hyperplasia (CAH), or Cushing's syndrome. The most frequent causes of hyperandrogenism are PCOS and severe insulin resistance.

Medical history should take into account regular drug intake (diazoxide, glucocorticoids, cyclosporine, and phenytoin) as well as complaints, such as the

**Table II** Symptoms and Signs of Hyperandrogenism

Hirsutism
Acne
Menstrual disturbances
Irregular ovulation
Muscle development
Obesity
Clitoral hypertrophy
Male pattern baldness
Voice deepening
Changes in libido

presence and progression of hirsutism and acne, baldness, and deepening in the voice. Hyperandrogenic women often suffer menstrual irregularities with oligomenorrhea and breakthrough bleeding because of anovulation and unopposed estrogen action on the endometrium.

Physical examination will determine the amount and distribution of facial and body hair. It can be scored on a scale of 1 to 4 for 11 different body areas according to the Ferriman and Gallwey method.

Laboratory findings can be based on total and free testosterone levels, SHBG, and androstenedione as markers of ovarian androgen overproduction. On laboratory examination performed during the first days of the menstrual cycle, PCOS patients may be characterized by high luteinizing hormone levels in presence of normal follicle-stimulating hormone levels. High 17-hydroxyprogesterone levels are used to detect CAH, while high DHEAS levels may be a marker of an adrenal androgen-producing tumor.

## TREATMENT

Treatment of hyperandrogenism should be directed toward the individual problem. A noteworthy feature of hirsutism is the slow response to treatment because the change in the hair growth cycle is time-consuming.

Hirsute anovulatory women wishing to conceive may be treated by ovulation induction using clomiphene citrate or gonadotropins. Azziz *et al.* performed a multicenter, double-blind, placebo-controlled trial of PCOS women treated with troglitazone, an insulin-sensitizing agent of the thiazolidinedione class. Troglitazone administration improved the ovulatory dysfunction, hirsutism, hyperandrogenemia, and insulin resistance of PCOS patients in a dose-related fashion.

If a patient is interested in a contraceptive method, an oral contraceptive is a good treatment option, particularly one that contains an antiandrogenic progestin such as cyproterone acetate, which is capable of binding to the androgen receptor.

Ovarian androgen secretion may be decreased by gonadotropin-releasing hormone analogs, which suppress pituitary follicle-stimulating hormone and

luteinizing hormone secretion. This blocks the pituitary-ovarian axis and prevents cyclic ovarian follicle development and hormonal production.

Other treatment options include spironolactone, an aldosterone antagonist diuretic, and flutamide, a nonsteroidal antiandrogen. Both act at the androgen receptor level and compete with testosterone and DHT.

The 5 $\alpha$ -reductase activity inhibitor finasteride blocks conversion of testosterone into DHT. It is used to treat prostate cancer and can be used to treat hirsutism. Patients treated with finasteride should have an effective method of contraception because DHT is essential for the development of the urogenital sinus and urogenital tubercle into the male external genitalia, urethra, and prostate.

## See Also the Following Articles

Hyperandrogenism, Gestational • Hyperandrogenism, Hyperinsulinemic • Ovarian Androgen-Producing Tumors • Polycystic Ovary Syndrome (PCOS)

## Further Reading

- Azziz, R., Ehrmann, D., Legro, R. S., Whitcomb, R. W., Hanley, R., Fereshetian, A. G., O'Keefe, M., and Ghazzi, M. N. (2001). PCOS/Troglitazone Study Group. *J. Clin. Endocrinol. Metab.* **4**, 1626-1632.
- Kahsar-Miller, M. D., Nixon, C., Boots, L. R., Go, R. C., and Azziz, R. (2001). Prevalence of polycystic ovary syndrome (PCOS) in first-degree relatives of patients with PCOS. *Fertil. Steril.* **75**(1), 53-58.
- Penning, T. M. (1997). Molecular endocrinology of hydroxysteroid dehydrogenases. *Endocr. Rev.* **18**(3), 281-305.
- Rittmaster, R. S. (1996). Functional hyperandrogenism. In "Reproductive Endocrinology, Surgery, and Technology" (E. Y. Adashi, J. A. Rock, and Z. Rosenwaks, eds.), pp. 1501-1520. Lippincott-Raven, Philadelphia.
- Rosenfield, R. L. (2001). Polycystic ovary syndrome and insulin-resistant hyperinsulinemia. *J. Am. Acad. Dermatol.* **45**(3 Suppl.), S95-S104.
- Rossi, R., Tauchmanova, L., Luciano, A., Valentino, R., Savastano, S., Battista, C., Di Martino, M., and Lombardi, G. (2001). Functional hyperandrogenism detected by corticotropin and GnRH-analogue stimulation tests in women affected by apparently idiopathic hirsutism. *J. Endocrinol. Invest.* **24**(7), 491-498.
- Speroff, L., Glass, R. H., and Kase, N. G. (1999). "Clinical Gynecologic Endocrinology and Infertility," 6th ed. Lippincott Williams & Wilkins, Baltimore.

testosterone concentrations are approximately three or four times higher than umbilical cord concentrations (100–140 compared to 33.5 ng/dl).

### **Protective Mechanisms against Maternal and Fetal Virilization**

Most pregnant women and their infants are not virilized despite increases in serum total and free androgen concentrations during pregnancy. Mechanisms postulated to account for this lack of virilization are discussed in the following sections.

#### **Increased Serum SHBG Concentrations**

Much of the increase in serum androgen concentrations consists of SHBG-bound androgens, with relatively little increase in free androgens. Therefore, much of the increase in circulating androgens may not be available for action at the target tissues.

#### **Progesterone Competition for Androgen Receptor Binding**

Progesterone production increases dramatically (approximately 10-fold) during pregnancy and approaches 250 mg per day at term. Despite the low affinity of androgen receptors for progesterone, the increased progesterone production is so great that there may be inhibition of testosterone binding to the androgen receptor.

#### **Progesterone Competition for Androgen Activation at Target Tissues**

Testosterone is converted in target tissues with the aid of 5 $\alpha$ -reductase to the more biologically active dihydrotestosterone. Progesterone also has weak affinity for 5 $\alpha$ -reductase, but due to its excess amount, it may inhibit conversion of testosterone to dihydrotestosterone.

#### **Placental Androgen Aromatization**

The placenta has a massive capacity to convert androgens to estrogens. Therefore, it is highly likely that most of the increased androgen production occurring during pregnancy is metabolized by the placenta, thereby affording protection from virilization to both the mother and the female fetus. In one report, a pregnant woman had a serum testosterone concentration of 15,000 ng/dl but the cord serum concentration was only 252 ng/dl. Interestingly, the cord serum concentration of estradiol was elevated compared to that of the mother, suggesting conversion of testosterone to estradiol and indicating a protective mechanism against the passage of testosterone from mother to fetus. In addition, female infants in pregnancies

complicated with placenta aromatase deficiency are virilized. These findings suggest that fetal androgen exposure may be diminished by placenta aromatization of androgen, but aromatization does not fully explain the lack of fetal virilization. In the case mentioned previously, the cord serum concentration of 252 ng/dl would normally be sufficient for virilization, but the female infant was not virilized. This finding and others suggest that decreased androgen action may also protect the female fetus against virilization in normal pregnancies.

### **CAUSES OF GESTATIONAL HYPERANDROGENISM**

Hyperandrogenism in pregnancy is nearly always the result of conditions arising during pregnancy; hyperandrogenism in a nonpregnant woman often results in anovulation and infertility. The incidence of hyperandrogenism during pregnancy is low, with the two most common causes being gestational luteomas and hyperreactio luteinalis (theca lutein cysts of the ovary). It is important for the clinician to appreciate that an increase in androgen production in a pregnant woman can cause hirsutism and virilization in the mother and virilization of the female fetus. The extent of fetal virilization or risk to the fetus are variable, depending on the time of onset of the increased maternal androgen production, its severity, and other unknown factors.

#### **Luteomas**

Luteomas are benign, solid ovarian tumors composed of hyperplastic masses of large lutein cells. Luteomas regress and disappear after delivery; therefore, the incidence of these lesions is probably greater than reported because most are unnoticed because they secrete little androgen or have minimal hormonal effects. These subclinical luteomas may be discovered incidentally at the time of cesarean section or other surgery for unrelated reasons. There is evidence that the incidence is increased in the African American population. Pregnant women with luteomas may have increased serum concentrations of testosterone, dihydrotestosterone, and androstenedione. Maternal hirsutism or virilization are seen in approximately 30–35% of women with pregnancy-related luteomas, suggesting that either serum androgen concentrations are often not elevated significantly or androgen action is blunted. Such blunting could occur because of the increased availability of SHBG during pregnancy so that less testosterone is available to peripheral tissues

or because the tissue effects of testosterone are blunted by the actions of estrogens. There are cases suggesting resistance to androgen in which women with luteomas of pregnancy had marked elevations in serum testosterone concentrations ( $>2500$  ng/dl) and were not virilized even though normal gestational concentrations are 50–120 ng/dl.

There is an approximately 80% chance that a female infant born to a virilized mother with a gestational luteoma will be virilized. If the mother is not virilized, neither is the infant. In order to estimate the risk of virilization to a fetus *in utero*, one must consider the sex of the fetus, the duration of exposure to the androgens, and, most importantly, the stage of pregnancy during which the exposure occurred. Cord serum androgen concentrations have been measured in only a few infants born to mothers with marked hyperandrogenism due to luteomas of pregnancy. Two cases had normal cord androgen concentrations and the infants were a normal male and a nonvirilized female. Two other infants had a two- to eightfold elevation in cord serum testosterone concentrations: One was a premature nonvirilized male and the other a virilized female. Based on these very limited data, it appears that the risk of virilization to a female infant may be related to cord serum testosterone concentrations but not maternal concentrations. Maternal hyperandrogenism is therefore necessary but not sufficient to cause virilization of the female infant, implying that some of the previously mentioned protective mechanisms may be in play.

### Hyperreactio Luteinalis

The incidence of hyperreactio luteinalis or theca lutein cysts is increased in women with multiple or isoimmunized gestations, underlying diabetes, molar pregnancies, or choriocarcinomas (trophoblastic disease), where the incidence is highest at 10–22%. The size of ovaries containing theca lutein cysts in pregnant women ranges from normal to 10–15 cm each. Approximately 30% of pregnant women with hyperreactio luteinalis are virilized and many, if not all, virilized women have high serum testosterone and androstenedione concentrations. Cord serum testosterone concentrations in infants of virilized mothers are either normal or increased. Despite the presence of maternal virilization and, in some cases, elevated cord testosterone concentrations, none of the reported female infants have been virilized. Therefore, maternal virilization and even fetal hyperandrogenemia seem to have no relationship to fetal virilization in this disorder.

### Administration of Progestins and Androgens

The administration of progestin, androgen, and even diethylstilbestrol to pregnant women has been linked to signs of androgen excess in women and masculinization of female fetuses. A similar effect can be seen with placental aromatase deficiency in which the conversion of androgen to estrogen is decreased, leading to androgen accumulation. Female external genital development occurs between 7 and 12 weeks of gestation; therefore, exposure to androgens at this time may result in partial or complete labial fusion and clitoral hypertrophy. After week 12 of gestation, labial fusion does not occur but clitoral hypertrophy remains a risk. Male fetuses do not appear to be affected.

### Other Ovarian Tumors

Numerous other ovarian tumors have been associated with hyperandrogenism in pregnancy, including Sertoli–Leydig cell tumors, Krukenberg tumors, lipoid cell tumors of the ovary, Brenner cell tumors, and sclerosing stromal cell tumors. Sertoli–Leydig cell tumors (arrhenoblastomas) have been reported in 18 pregnant women; the rarity is a reflection of the effects of hyperandrogenism and its association with anovulation. However, there have been a few cases in which the tumor was thought to be present at the time of conception. The risk of malignancy in Sertoli–Leydig cell tumors is higher in the pregnant state (40–50 vs 12–22% in the nonpregnant state). Pregnant women are virilized at a rate similar to that of nonpregnant women (approximately 69–87%). Krukenberg tumors, which are ovarian metastasis of primary tumors of the gastrointestinal tract, are solid and bilateral in approximately 80% of cases. In a few cases reported during pregnancy, most of the women had hirsutism or virilization and half delivered virilized female infants. Polycystic ovary syndrome (PCOS) with no associated ovarian tumor has been associated with increased virilization in pregnancy, although very few cases have been described despite the relatively common occurrence of PCOS. One case reported associated fetal virilization but the PCOS was not confirmed histologically.

### Role of Human Chorionic Gonadotropin

It has been postulated that human chorionic gonadotropin (hCG) can contribute to the pathogenesis of and hormone secretion by luteomas and theca lutein cysts. In the case of luteoma, this concept is supported by the regression of the tumor mass and prompt

decline in androgen levels after delivery. However, most luteomas are diagnosed late in pregnancy, well after the time of peak hCG production. Also, theca lutein cysts, not luteomas, are the characteristic ovarian lesions among women who produce excess amounts of hCG, such as those with trophoblastic disease. Therefore, hCG may be necessary for hormone production by luteomas, but there may be other factors causing growth and androgen production in late gestation.

Women with theca lutein cysts tend to have increased hCG production. However, administration of hCG alone does not cause these cysts, and not all women with massive elevations in serum hCG concentrations have them. Furthermore, the cysts may persist for several months after the evacuation of molar pregnancies, which is followed by a rapid decline in serum hCG concentrations. Therefore, as with luteoma, hCG is not the only factor involved in the pathogenesis of this condition.

## DIAGNOSIS AND MANAGEMENT

Gestational-related hyperandrogenism may cause hirsutism and virilization of the mother and virilization of the female fetus (Fig. 1). The major causes of

pregnancy-related hyperandrogenism are shown in Table I.

## Signs and Symptoms

Consideration of androgen excess in pregnancy usually occurs when a pregnant woman presents with a rapid onset of masculinization. Signs and symptoms include hirsutism, temporal balding, acne, clitoromegaly, and deepening of the voice. There may be palpable pelvic or abdominal masses due to tumors or cysts, which could include luteomas, hyperreactio lutealis, other ovarian tumors, and adrenal tumors. Exogenous hormones or placental aromatase deficiency are not expected to be associated with masses. Sometimes, the first suspicion of gestational hyperandrogenism occurs at the time of delivery of a virilized female infant. It is important to remember that luteomas and theca lutein cysts undergo spontaneous regression in the postpartum period.

## Diagnostic Approach

A thorough evaluation is warranted when evaluating a woman with gestational hyperandrogenism because of the possibility that the masculinizing lesion may be malignant. Sertoli–Leydig cell tumors, granulosa

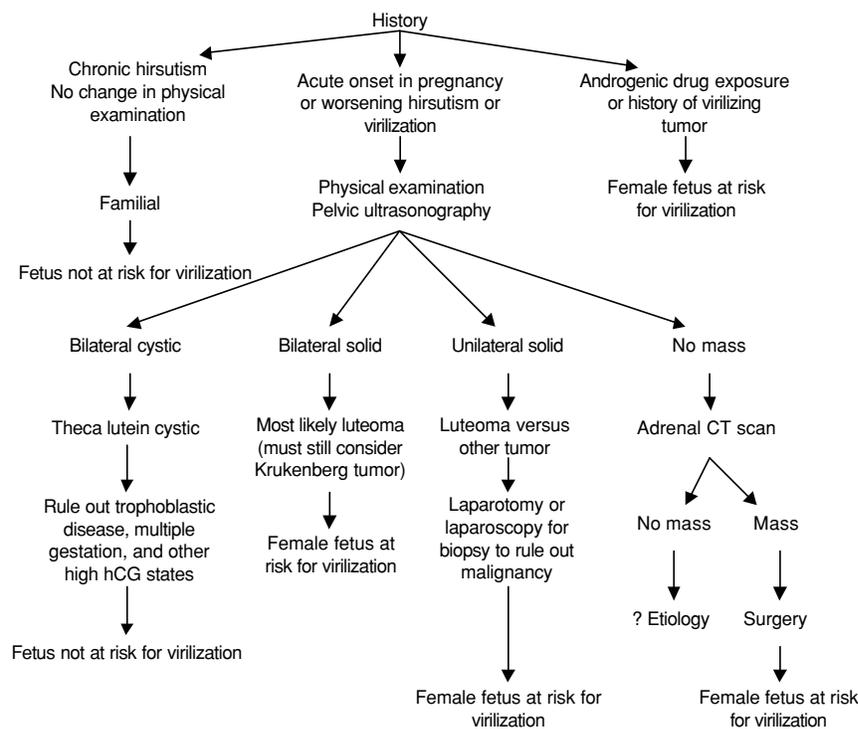


Figure 1 Diagnosis and management of hirsutism or virilization in pregnancy.

**Table I Causes of Gestational Hyperandrogenism<sup>a</sup>**

Luteoma
Theca-lutein cysts
Exogenous progestin or androgen administration
Placental aromatase deficiency
Sertoli-Leydig cell tumor (arrhenoblastoma)
Miscellaneous ovarian tumors
Krukenberg tumors (gastrointestinal cancer metastatic to ovaries)
Polycystic ovary disease
Adrenal tumors

<sup>a</sup>Adapted from McClamrock and Adashi (1996).

theca cell tumors, ovarian mucinous cystadenocarcinomas, Krukenberg tumors, and adrenocortical carcinomas are examples of malignant tumors. Cushing's syndrome, although rare during pregnancy, should also be considered.

Information should be gathered regarding the time of onset of hirsutism and virilization. Exposure to elevated levels of androgens during early pregnancy may cause labiascrotal fusion and clitoromegaly, which may be irreversible, whereas exposure after 12 weeks of gestation causes only clitoral hypertrophy. If the mother has a history of chronic hirsutism, theca lutein cysts are more likely, although recurrence of virilizing tumors has been reported. In early pregnancy, one should consider trophoblastic disease.

Ultrasonography may be helpful in distinguishing solid and cystic ovarian tumors, uteralateral from bilateral ovarian disease, and ovarian from adrenal tumors. Theca lutein cysts are almost always bilateral, whereas most malignant tumors are unilateral, with the exception of Krukenberg tumors.

**Table II** lists characteristics of the major causes of gestational hyperandrogenism. Ultrasonography cannot reliably predict fetal sex until midgestation. As a result, its use for evaluation of fetal virilization is limited. There is very little information on the use of laparoscopy in diagnosis, but it could potentially be helpful in the first trimester.

Maternal serum hormone determinations are of questionable benefit because there are elevations of androgens in normal pregnancy, and even extremely elevated levels are not always associated with fetal virilization.

Percutaneous umbilical blood sampling for karyotyping and measurement of circulating fetal androgen levels could potentially provide helpful prognostic information if it could be done early enough in gestation. However, the procedure cannot be performed prior to differentiation of the fetal gonads and,

**Table II Characteristics of the Major Causes of Gestational Hyperandrogenism<sup>a</sup>**

Characteristic	Cause			
	Luteoma	Theca-lutein cyst	Sertoli-Leydig tumor	Krukenberg tumor
Incidence	?	?	Very rare	Very rare
Bilaterally	~47%	~96%	~5%	>80%
Maternal androgen excess	~35%	~30%	69–87%	80–100%
Fetal virilization	High if mother virilized	0%	High if mother virilized	80–100%
Malignant	0%	0%	~45%	100%

<sup>a</sup>Adapted from McClamrock and Adashi (1992).

therefore, cannot be performed early enough to allow preventative intervention.

### Prognosis and Therapy

Luteomas should be suspected in women with bilateral solid lesions, but Krukenberg tumors are still possible. Luteomas should regress and disappear after delivery. The risk of maternal hirsutism or virilization is approximately 35% in women with pregnancy-related luteoma and approximately 80% of female fetuses of effected mothers have been virilized. If the mother has no clinical manifestations of androgen excess, fetal virilization is rare. The fetus does not appear to be at risk for virilization in the woman with hyperreactio luteinalis even when maternal serum androgen concentrations are elevated. These cysts also regress after delivery and should not require surgical therapy.

Approximately 7% of ovarian tumors in pregnancy are malignant, but if the pregnant woman is virilized and has a unilateral solid ovarian mass, the risk of malignancy is approximately 50%. The risk of maternal and fetal virilization due to exogenous hormone exposure is poorly studied. The effects are expected to vary according to when the fetus was exposed. There is no evidence that cord blood measurements of androgens or other hormones are helpful in determining the risk of fetal virilization in these cases.

### Management

It is important to remember that male infants are not thought to be at risk and that exposure of a female fetus to high androgen concentrations in early pregnancy may be associated with irreversible virilization.

Elevated cord serum concentrations of testosterone and androstenedione (based on four cases) appear to be necessary for virilization of a female infant. Therefore, if cord levels are low, virilization is not expected. Surgery is not generally indicated for women with luteomas or theca lutein cysts because both lesions regress in the postpartum period.

Antepartum treatment remains generally unavailable despite our ability to recognize the clinical manifestations and the ability to evaluate fetal androgen exposure.

## FUTURE DIRECTIONS

Given the limitations in antepartum therapy, future efforts may be aimed at determining which patients are at risk and at developing screening tests to identify gestational hyperandrogenism earlier in pregnancy. With earlier diagnosis, it may be possible to implement therapies that decrease androgen exposure, thereby limiting fetal exposure to androgens during the critical period of sex determination.

## See Also the Following Articles

Androgens, Gender and Brain Differentiation • Hyperandrogenism, Functional • Hyperandrogenism, Hyperinsulinemic • Pregnancy Endocrinology

## Further Reading

- Bammann, B. L., Coulam, C. B., and Jiang, N. S. (1980). Total and free testosterone during pregnancy. *Am. J. Obstet. Gynecol.* **137**, 293.
- Baxi, L., Holub, D., and Hembree, W. (1988). Bilateral luteomas of pregnancy in a patient with diabetes. *Am. J. Obstet. Gynecol.* **159**, 454.
- Hensleigh, P. A., Carter, R. P., and Grotjan, H. E., Jr. (1975). Fetal protection against masculinization with hyperreactio luteinalis and virilization. *J. Clin. Endocrinol. Metab.* **40**, 816.
- McClamrock, H. D., and Adashi, E. Y. (1992). Gestational hyperandrogenism. *Fertil. Steril.* **57**, 257.
- McClamrock, H. D., and Adashi, E. Y. (1996). Gestational hyperandrogenism. In "Reproductive Endocrinology, Surgery, and Technology" (E. Y. Adashi, J. A. Rock, and Z. Rosenwaks, eds.). Lippincott-Raven, Philadelphia.
- Nagamani, M., Gomez, L. G., and Garza, J. (1982). In vivo steroid studies in luteoma of pregnancy. *Obstet. Gynecol.* **59**, 105S.
- Rivarola, M. A., Forest, M. G., and Migeon, C. J. (1968). Testosterone, androstenedione and dehydroepiandrosterone in plasma during pregnancy and at delivery: Concentration and protein binding. *J. Clin. Endocrinol. Metab.* **28**, 34.
- Saez, J. M., Forest, M. G., Morera, A. M., and Bertrand, J. (1972). Metabolic clearance rate and blood production rate of testosterone and dihydrotestosterone in normal subjects during pregnancy and in hyperthyroidism. *J. Clin. Invest.* **51**, 1226.
- Verkauf, B. S., Reiter, E. O., Hernandez, L., and Burns, S. A. (1977). Virilization of mother and fetus associated with luteoma of pregnancy: A case report with endocrinologic studies. *Am. J. Obstet. Gynecol.* **129**, 274.
- Wolff, E., Glasser, M., Gordon, G. G., et al. (1973). Virilization luteoma of pregnancy. Report of a case with measurements of testosterone and testosterone binding in plasma. *Am. J. Med.* **54**, 229.

active hormone. Several end organs have androgen receptors whose sensitivity to the hormone varies greatly, including total insensitivity seen in feminizing syndromes. In the skin, conversion to potent DHT by  $5\alpha$  reductase enzymes determines the response of the pilosebaceous unit, which controls hair growth and sebaceous gland secretion.

## ETIOLOGY AND CLINICAL MANIFESTATIONS OF HYPERANDROGENISM

Hyperandrogenism etiology varies; it is principally ovarian, mainly involving the polycystic ovary, and rarely due to an enzymatic defect or neoplasia. In the adrenal, it is caused by an enzymatic defect in congenital adrenal hyperplasia, usually partial  $21$ -hydroxylase deficiency and rarely Cushing's disease and neoplasia. The clinical response to excess androgens varies and is genetically and race dependent. Caucasians are most sensitive and Asians are least sensitive. Oily skin, acne, and excess coarse facial, chest/breast, back, and genital area hair may be found. In more severe cases, signs of virilization (clitoromegaly and voice changes) may be present. The ovary frequently contains small cysts and there is thecal thickening and stromal hypertrophy. The endometrial lining may cause hyperplasia. Menstrual dysfunction ranges from minimal to anovulation and amenorrhea, frequently leading to infertility.

## POLYCYSTIC OVARY SYNDROME

Polycystic ovary syndrome (PCOS) is a multifaceted syndrome with a range of clinicopathological manifestations associated with altered androgen metabolism and frequently combined with altered glucose and lipid metabolism. The initiating defect is unclear, but central to it is increased ovarian androgen production due to LH-induced up-regulation of theca cells. Local inhibin B may be involved by stimulating androgen production. In the ovary, insulin acts through the insulin-like growth factor-1 receptor, but at high levels it may act through its own receptor. Insulin abnormalities are variable, and circulating levels in PCOS are modestly elevated. Any insulin abnormality, including altered receptor function, may lead to PCOS. Consequently, in addition to reproductive dysfunction, patients are at risk for a number of serious medical conditions, including diabetes and cardiovascular disease.

PCOS generally begins at menarche and is progressive. By the fourth decade, 40–50% of

women with PCOS develop diabetes mellitus. Insulin abnormalities are not fully characterized since the increase in liver glucose production is beyond the control of insulin. Measurable levels of testosterone, preferably the free form, and DHEAS are the hallmark for laboratory diagnosis. Very high levels are due to tumors, whereas lower levels may be due to PCOS. The androgen source is determined by administering oral contraceptives (estrogen/progestin) that suppress mostly androgens of ovarian origin. Dexamethasone administration (which blocks ACTH), on the other hand, will lower adrenal production of androgens. Management is cosmetic and dermatologic for skin disorders, and oral contraceptives with a weak androgenic component (norgestimate) are effective in lowering both ovarian and adrenal sources of androgen production. If the adrenal component is significant, low doses of prednisone may be administered. Less common is the use of spironolactone or ketoconazole. The addition of oral hypoglycemics, such as troglitazone, has been shown to significantly affect the insulin response to glucose challenge and to improve the response to clomiphene citrate in ovulation-induction regimens.

## IMPLICATIONS OF SYNDROME X

The central defect is insulin resistance, which is compensated by increased insulin secretion. However, the precise definition of what constitutes resistance is vague; based on standard criteria, patients are not diabetic. The associated defects are variable but may be severe, and the etiology may be partially genetic and partially environmental (obesity and lack of physical exercise). Insulin resistance is manifested in the muscle and fat cells. Defects include abnormalities in glucose secretion and a mild increase in uric acid levels due to decreased renal clearance. The lipid abnormalities are increased triglycerides, secretion of very low-density lipoprotein, and decreased high-density lipoprotein secretion by the liver due to chronic hyperinsulinemia. Syndrome X is associated with increased heart rate, sodium retention, and hypertension. Changes in the coagulation profile as reflected by increased plasminogen activator inhibitor-1 levels are present. The result of these abnormalities is an increased risk for coronary heart disease.

## See Also the Following Articles

Hyperandrogenism, Functional • Hyperandrogenism, Gestational • Hypertension and Diabetes • Insulin-Resistant

States, Role of Free Fatty Acids (FFA) • Polycystic Ovary Syndrome (PCOS) • Steroid Metabolism and the Metabolic Syndrome

### Further Reading

- American College of Obstetrics and Gynecology (1995). Hyperandrogenic chronic anovulation, technical bulletin No. 202. ACOG.
- Ehrmann, D. A., Barnes, R. B., and Rosenfield, R. L. (2000). Hyperandrogenism, hirsutism and the polycystic ovary syndrome. In "Textbook of Endocrinology," pp. 2122–2137. Saunders, Philadelphia.
- Legro, R. S., Driscoll, D., Strauss, J. F., *et al.* (1998). Evidence for a genetic basis for hyperandrogenemia in polycystic ovary syndrome. *Proc. Natl. Acad. Sci. USA* **95**, 14956–14960.
- Miller, W. L. (1999). Congenital adrenal hyperplasia in the adult patient. *Adv. Intern. Med.* **44**, 155–173.
- Reaven, G. M. (2000). Syndrome X. In "Textbook of Endocrinology," pp. 954–958. Saunders, Philadelphia.
- White, P. C., and Speiser, P. W. (2000). Congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Endocr. Rev.* **21**, 245–291.

infants and increasing to 1500 mg/day during adolescence and adulthood. At supplementary doses of 1500 mg/day, there appear to be no hazards to calcium administration. Increasing calcium intake to more than 4 g/day can overwhelm the homeostatic mechanisms designed to protect against hypercalcemia, particularly in patients with increased  $1,25(\text{OH})_2\text{D}_3$  levels such as hyperparathyroidism (HPT), sarcoidosis, and exogenous vitamin D administration.

## CALCIUM HOMEOSTASIS

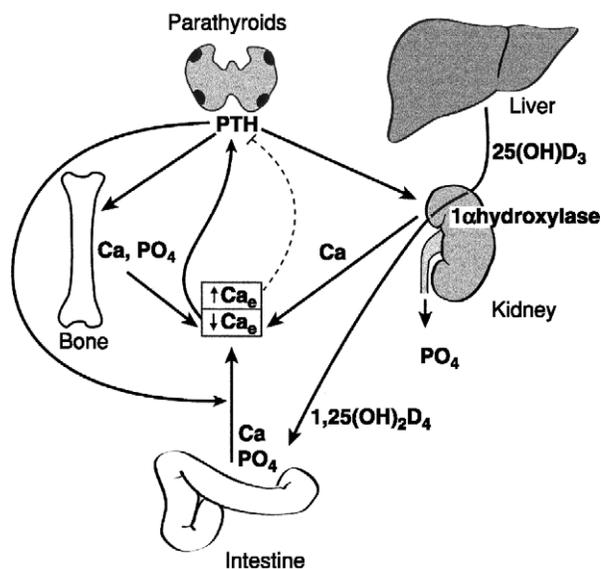
### Regulation of Extracellular Calcium Concentration

Compared with the large changes in cytosolic calcium concentration, variation in extracellular calcium concentration, as measured by the ionized calcium levels in blood, is quite small. Normal extracellular calcium concentration is regulated by three calciotropic hormones: parathyroid hormone (PTH), calcitonin, and vitamin D. These hormones act on three target organs—kidney, bone, and intestine—to maintain serum calcium within a very narrow range. The most critical of these hormones on a minute-to-minute basis is PTH. Acutely, slight changes (1–2% decrease) in extracellular calcium concentration result in a prompt increase in PTH secretion, restoration of extracellular fluid (ECF) calcium to normal, and inhibition of further PTH secretion (Fig. 1). This stability is maintained by exquisitely sensitive calcium ion receptors (CaR) that influence the distribution of calcium in different physiological compartments. The CaR, first identified on parathyroid cells, belongs to family C of the superfamily of G protein-coupled receptors (GPCRs).

Increases in  $[\text{Ca}]_e$  within the millimolar range result in CaR-initiated G protein-dependent activation of diverse intracellular signaling pathways that ultimately inhibit PTH secretion. Figure 1 depicts the major sites of PTH action in calcium homeostasis.

In the kidney, PTH stimulates the renal tubular reabsorption of calcium and excretion of phosphorus and increases synthesis of the active form of vitamin D,  $1,25(\text{OH})_2\text{D}_3$ , from its precursor,  $25(\text{OH})\text{D}_3$ . The kidney filters approximately 10,000 mg/day of calcium, reabsorbs 98%, and excretes 2% of the filtered load. A variety of factors influence renal calcium handling. Total urinary calcium excretion increases with increased dietary calcium. Calcium excretion is facilitated by increased sodium chloride intake due to inhibition of proximal tubular calcium reabsorption as

a result of volume expansion. This relationship is exploited clinically in the use of sodium chloride loading in the treatment of severe hypercalcemia. Calcium excretion is also increased by administration of loop diuretics such as furosemide. PTH blocks sodium-dependent phosphate cotransport in the proximal and distal tubules, resulting in increased phosphate excretion. PTH decreases calcium reabsorption in the proximal tubule but stimulates calcium absorption in the distal tubule, with the overall effect being net calcium absorption. In proximal tubular cells, PTH activates the  $25(\text{OH})\text{D}_3$   $1\alpha$  hydroxylase, which catalyzes the synthesis of the most active metabolite of vitamin D,  $1,25(\text{OH})_2\text{D}_3$ . Increased levels of  $1,25(\text{OH})_2\text{D}_3$  directly stimulate intestinal absorption of calcium and phosphate as well as mobilization of calcium and phosphate from bone. In a negative-feedback loop,  $1,25(\text{OH})_2\text{D}_3$  acts via the vitamin D receptor (VDR) to inhibit PTH gene transcription, PTH secretion, and parathyroid cell proliferation. Conversely, reduced levels of  $1,25(\text{OH})_2\text{D}_3$  are seen in patients with uremia and are associated with parathyroid hyperplasia.

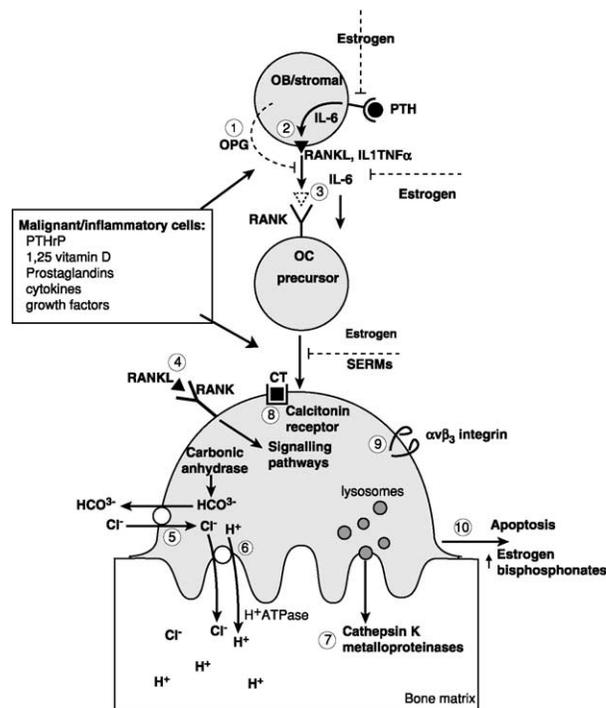


**Figure 1** The calcium homeostatic system. A small decrease in serum-free calcium concentration results in prompt release of PTH by the parathyroid cells. PTH acts on the kidney to increase calcium reabsorption and phosphate excretion. PTH also stimulates the synthesis of  $1,25(\text{OH})_2\text{D}_3$ , which facilitates active absorption of calcium and phosphate from the intestine. In addition, PTH acts on bone to mobilize calcium and phosphate. All of these effects result in a return of extracellular calcium to the normal range. Calcium activation of the calcium sensor inhibits PTH secretion in a classic negative feedback loop.

## The Skeleton as a Calcium Reservoir: Osteoclastic Activity and Hypercalcemia

During development and growth, the skeleton is sculpted or “modeled” into its adult form by removal of bone at one site and deposition at another. At maturity, periodic replacement of old bone by new bone at the same site (remodeling) is carried out such that the entire adult skeleton is regenerated every 10 years or so. Osteoclastic bone resorption and osteoblastic new bone formation are tightly linked under normal conditions by local (autocrine and paracrine) and systemic endocrine factors. Excessive osteoclastic activity is an important mechanism in the development of hypercalcemia, and many treatments are based on modifying osteoclastic regulation. Osteoclasts, the agents of bone resorption, are multinucleated cells of the monocyte/macrophage family (Fig. 2). Their development requires the presence of osteoblast precursors, which express macrophage colony-stimulating factor (M-CSF) and receptor for activation of nuclear factor kappa B (NF- $\kappa$ B) (RANK) ligand (RANKL). PTH and other osteoclastogenic agents act on osteoblasts or stromal precursors to increase the production of RANKL and other local factors, such as interleukin-1 (IL-1) and TNF $\alpha$ , which stimulate the release of soluble factors from osteoblasts and lead to further recruitment, differentiation, and activation of osteoclasts. Malignant and inflammatory cells can overexpress a number of products that stimulate osteoclastogenesis and the osteolytic cascade, including PTHrP, 1,25-vitamin D, prostaglandins, interferons, growth factors, and RANKL. Several factors are responsible for the inhibition of osteoclastic bone resorption. Osteoblasts also secrete osteoprotegerin (OPG), which competes with RANKL for RANK and inhibits osteoclastogenesis. Estrogen directly or indirectly suppresses or regulates the production of many stimuli of osteoclastogenesis. Interaction of calcitonin with its receptors on the mature osteoclast results in potent inhibition of bone resorption.

Osteoclasts secrete protons, collagenase, hyaluronidase, and other matrix-degrading enzymes into the space between their ruffled border and the bone surface (Fig. 2). This leads to the resorption and removal of both the mineral and matrix phases of bone. Subsequently, increased local concentration of a variety of factors, including calcium, leads to decreased osteoclast numbers (apoptosis or programmed cell death) and activity and also sets the stage for osteoblast recruitment, osteoblast activation, and new bone formation.



**Figure 2** Therapeutic targets for modulation of osteoclast activity. Current therapies include estrogen and selective estrogen receptor modulators (SERMs) for suppressing osteoclast formation and include bisphosphonates and calcitonin for inhibiting osteoclast activity. Some potential new targets include blocking the action and signaling of the OPG/RANKL/RANK pathway (1–4), blocking acidification of the resorption lacuna by inhibiting the activity of carbonic anhydrase or H<sup>+</sup>-ATPase (5,6), blocking enzymatic degradation of collagen and matrix proteins (7), developing more robust calcitonin receptor ligands (8), blocking  $\alpha$ v $\beta$ <sub>3</sub> integrin-mediated cytoskeletal signaling (9), and selectively stimulating osteoclast apoptosis (10). OPG, osteoprotegerin; RANK, receptor for activator of nuclear factor kappa B; RANKL, ligand for RANK; OB, osteoblast; OC, osteoclast. Redrawn with permission from Rodan, G., and Martin, J. T. (2000). Therapeutic approaches to bone diseases. *Science* 289, 1508–1514.

## DIFFERENTIAL DIAGNOSIS OF HYPERCALCEMIA

As is evident in Table I, the etiology of hypercalcemia is multifactorial. The symptoms of hypercalcemia (Table II) reflect the critical physiological role of calcium in all tissues. The construction of a working differential diagnosis of hypercalcemia in an individual patient begins by obtaining a good clinical history.

### Exogenous Factors

A variety of common medications influence calcium metabolism. Thiazide diuretics inhibit urinary calcium excretion, resulting in a mild rise in serum

**Table I Differential Diagnosis of Hypercalcemia in General Order of Frequency of Main Categories**

Hyperparathyroidism
Primary hyperparathyroidism
Secondary/Tertiary hyperparathyroidism
Parathyroid carcinoma
Malignancy-associated hypercalcemia
PTHrP mediated
1,25(OH) <sub>2</sub> D <sub>3</sub> mediated
Lytic bone metastases
Exogenous
Drugs: vitamins D and A, lithium, thiazides, calcium
Endocrine disorders
Thyrotoxicosis
Addison's disease
Miscellaneous
Granulomatous diseases
Immobilization

calcium concentration sufficient to render patients with previously unrecognized HPT frankly hypercalcemic.

There appears to be a duration-dependent relationship between lithium use and the development of HPT. In addition, some HPT patients on chronic

lithium therapy might not revert to a normocalcemic state with drug withdrawal and tend to have multiple gland disease at surgery. Short-term therapy may also unmask mild preexisting HPT, as do the thiazide diuretics.

The milk-alkali syndrome, defined as hypercalcemia and metabolic alkalosis occurring with a history of consumption of large quantities of calcium and absorbable alkali in the absence of other explanations for the metabolic abnormalities, was initially described as a complication of treatment for peptic ulcer disease. Such patients can be identified by an appropriately suppressed PTH level in the face of hypercalcemia and can be easily treated by limiting calcium intake.

Vitamin D and its metabolites are used to treat a variety of metabolic bone diseases and to retard the development of secondary HPT in patients with chronic renal failure. However, vitamin D can cause hypercalcemia when given in high doses, and severe illness, including renal impairment and death, have been reported. Hypervitaminosis is diagnosed by elevated levels of serum 25(OH)D<sub>3</sub> and hypercalcemia or hypercalciuria (or both) and by a history of vitamin D intake. Although the half-life of vitamin D metabolites is short, the half-life of vitamin D is weeks to months. Treatment of vitamin D intoxication includes, in addition to drug withdrawal, hydration, calciuresis, and perhaps glucocorticoids.

Hypercalcemia can also occur as a consequence of administration of vitamin A and its analogues (*cis*-retinoic acid), probably due to increased osteoclast-mediated bone resorption.

**Table II Symptoms and Signs of Hypercalcemia**

General
Weakness
Lethargy
Dehydration
Renal
Polyuria
Nephrolithiasis/Nephrocalcinosis
Decreased renal function
Gastrointestinal tract
Anorexia
Nausea
Constipation
Abdominal pain
Central nervous system
Depression
Impaired cognition
Altered mental status
Psychosis
Cardiovascular system
Hypertension
Shortened QT interval
Digitalis sensitivity

### Malignancy-Associated Hypercalcemia

Hypercalcemia resulting from cancer is the most common form of hypercalcemia among hospitalized patients and is second in frequency only to primary HPT as a cause of hypercalcemia among outpatients. Approximately 10 to 25% of patients with squamous carcinoma, ovarian carcinoma, multiple myeloma or lymphoma, renal adenocarcinoma, or carcinoma of the breast will develop hypercalcemia during the course of their disease, and in a substantial portion of these, hypercalcemia will produce clinical signs and symptoms (e.g., polyuria, polydipsia, confusion, obtundation, renal failure) that directly contribute to the demise of the affected patient. Hypercalcemia generally occurs late in the course of the patient's disease. In a minority of patients, malignancy-associated hypercalcemia is due to direct tumor involvement with the

skeleton, a syndrome that is called “local osteolytic hypercalcemia” (LOH). This syndrome is most commonly encountered in patients with breast cancer, multiple myeloma, lymphoma, or leukemia. In these cases, hypercalcemia is due to extensive malignant skeletal involvement. Bone scans, bone biopsies, and autopsies reveal large skeletal tumor burdens. In addition, malignant cells within the marrow space secrete cytokines, which lead to the recruitment and activation of osteoclasts, as shown in Fig. 2. The local activation of osteoclastic bone resorption leads to hypercalcemia, suppression of PTH secretion, hyperphosphatemia, hypercalciuria, suppression of nephrogenous cyclic AMP (cAMP) excretion, and suppression of circulating  $1,25(\text{OH})_2\text{D}_3$ . These findings are summarized in Table III. The diagnosis of LOH is usually straightforward, requiring only the clinical history; extensive biochemical testing is rarely necessary.

A more common type of malignancy-associated hypercalcemia occurs in patients with squamous carcinomas (head and neck, lung, cervix, vulva, esophagus, skin, and other sites), renal carcinomas, ovarian carcinomas, and transitional carcinomas of the bladder. Recently, it has been reported to occur in patients with HTLV-1 lymphoma/leukemia and is being observed increasingly in women with breast cancer in the absence of bone metastases. This “humoral hypercalcemia of malignancy” (HHM) accounts for 80% of patients with malignancy-associated hypercalcemia. Affected patients have minimal skeletal metastatic disease, and hypercalcemia subsides with tumor removal.

The agent responsible for hypercalcemia in these cases, PTHrP, was discovered in 1987. PTHrP, a product of many normal cells, has significant homology with PTH at the amino terminal end of the

molecule and interacts with the PTH/PTHrP shared receptor. Thus, when PTHrP is overexpressed systemically by tumor cells, it can produce most of the physiological effects of PTH, including hypercalcemia, activation of bone resorption, inhibition of renal phosphorus reabsorption, and hypophosphatemia (Table III). PTHrP may also stimulate renal reabsorption of filtered calcium, mimicking the anticalciuric effect of PTH and further exacerbating hypercalcemia. Hypercalcemia leads to suppression of PTH, but nephrogenous cAMP excretion is elevated as a result of the effects of PTHrP on the renal proximal tubular PTH receptor. For reasons that are still unclear, plasma  $1,25(\text{OH})_2\text{D}_3$  concentrations are reduced in HHM. Bone biopsy shows striking activation of osteoclastic bone resorption in the absence of osteoblastic bone formation. The reasons for this uncoupling of bone formation from bone resorption also remain unclear despite having been described two decades ago. These findings are contrasted with those observed in patients with LOH and with primary HPT in Table III.

Rarely, the ectopic secretion of authentic PTH may be responsible for humoral hypercalcemia of malignancy. These patients are usually identified following an unsuccessful attempt at parathyroidectomy.

Another unusual form of malignancy-associated hypercalcemia occurs in patients with lymphomas, which overproduce  $1,25(\text{OH})_2\text{D}_3$ . In this setting,  $1,25(\text{OH})_2\text{D}_3$  stimulation of intestinal calcium absorption and osteoclastic bone resorption results in the development of hypercalcemia. Such patients display hypercalcemia, suppressed PTH values in the setting of high-normal or elevated plasma  $1,25(\text{OH})_2\text{D}_3$  concentrations, and suppressed nephrogenous cAMP excretion.

It is important to recognize that the differential diagnosis of hypercalcemia occurring in a patient with cancer includes the entire differential diagnosis of hypercalcemia. For example, both breast cancer and primary HPT are common conditions and may coexist in a given patient. Because the clinical outcome of a patient whose hypercalcemia is due to cancer is generally poor, identification of a nonmalignant cause of the hypercalcemia may permit its effective treatment and may dramatically change the patient’s prognostic category.

Besides the tumors mentioned previously, nearly any type of cancer can cause hypercalcemia. In addition, benign neoplasms (pheochromocytomas and giant mammary fibro-adenomas) have occasionally been associated with hypercalcemia due to overproduction of PTHrP.

Table III Biochemical Comparison of Local Osteolytic Hypercalcemia, Humoral Hypercalcemia of Malignancy, and Hyperparathyroidism

	Hyperparathyroidism	Humoral hypercalcemia of malignancy	Local osteolytic hypercalcemia
Serum calcium	Increased	Increased	Increased
Renal phosphorus threshold	Decreased	Decreased	—
Parathyroid hormone	Increased	Decreased	Decreased
PTHrP	Decreased	Increased	Decreased
NcAMP	Increased	Increased	Decreased
$1,25(\text{OH})_2\text{D}_3$	Increased	Decreased	Decreased

In summary, the diagnosis of malignancy-associated hypercalcemia is generally straightforward because the tumor burden in such patients is virtually always large. Thus, a thorough history, a careful physical exam, a chest X ray, urinalysis, routine chemistries, and a complete blood count will almost always point to the responsible tumor or to an alternate cause of hypercalcemia. Extensive biochemical testing (e.g., PTH, PTHrP, plasma  $1,25(\text{OH})_2\text{D}_3$ , nephrogenous cAMP) is most appropriate in more difficult cases.

## Other Endocrinopathies

### **Addison's Disease**

The etiology of hypercalcemia in adrenal insufficiency is probably multifactorial. Hypovolemia and decreased glomerular filtration rate lead to a reduction in filtered calcium and an increase in proximal tubular reabsorption of sodium and calcium. Some of the proximal tubular calcium abnormalities may be related to glucocorticoid insufficiency given that they are not corrected by volume resuscitation alone but return to normal with hormone replacement. In Addison's disease, there is also an increase in calcium mobilization from bone, perhaps due to physiological effects of thyroid hormone unopposed by glucocorticoids or to other effects on osteoclast-mediated bone resorption. In any case, hypercalcemia can be an important sign of an impending Addisonian crisis. The treatment is prompt fluid resuscitation with normal saline and glucocorticoid replacement.

### **Thyrotoxicosis and Pheochromocytoma**

Mild hypercalcemia frequently occurs in thyrotoxicosis, probably due to thyroid hormone-induced osteoclastic bone resorption. Hypercalcemia resolves after treatment of the hyperthyroid state. Significant hypercalcemia can be seen in some patients with pheochromocytoma. In some cases, this is caused by concomitant HPT, as in MEN2A syndrome. In other instances, the pheochromocytoma may be producing a substance that increases bone resorption such as PTHrP.

## Miscellaneous

### **Granulomatous Diseases**

Hypercalcemia is a common occurrence in sarcoidosis, with probably 10% of patients developing hypercalcemia and 20% developing hypercalciuria at some time during the course of their disease. The hypercalcemia, hypercalciuria, and accelerated bone loss have

been shown to be due to increased production of  $1,25(\text{OH})_2\text{D}_3$  by disease-associated macrophages. Unlike the hydroxylation reaction in the renal tubular cell, the macrophage reaction is not stimulated by PTH or inhibited by  $1,25(\text{OH})_2\text{D}_3$ . Instead, it is inhibited by glucocorticoids, which form the basis for treatment. In addition, patients with hypercalcemia due to sarcoidosis should have limited dietary intake of calcium and vitamin D and should avoid sunlight exposure. Other granulomatous diseases, such as tuberculosis, silicon-induced granulomatosis, disseminated candidiasis, and leprosy, have also been associated with the development of hypercalcemia. In some cases, hypercalcemia may be due to macrophage production of PTHrP, which stimulates vitamin D activation, and RANKL production, with resultant osteoclast activation.

### **Immobilization**

Prolonged weightlessness or immobilization results in increased bone resorption and hypercalcemia, especially in those patients whose underlying rates of bone turnover are high. Thus, children and young adults, as well as patients with HPT, Paget's disease, or malignancy-associated hypercalcemia, are at special risk. Hypercalcemia begins within days to weeks of bed rest and can result in nephrolithiasis as well as osteopenia in the long term. The actual cause of the observed increase in osteoclastic activity and the observed decrease in osteoblastic function is unknown. PTH and  $1,25(\text{OH})_2\text{D}_3$  levels are appropriately suppressed in these hypercalcemic patients. Although administration of bisphosphonates may have some benefit, the most effective treatment is weight-bearing exercise.

## Primary Hyperparathyroidism

Primary HPT is the most common cause of hypercalcemia in the outpatient population. The diagnosis of HPT is made by demonstrating an inappropriately elevated PTH level for the simultaneously measured serum calcium.

Primary HPT is caused by inappropriate (for the ambient calcium concentration) PTH secretion by a parathyroid adenoma(s), hyperplasia, or carcinoma. In HPT, there is both increased cell proliferative activity (enlarged glands) and decreased sensitivity of the cells to secretory inhibition by calcium (altered set point). Sporadic primary HPT is the most common cause of hypercalcemia in an outpatient setting, with a prevalence of 1 in 500 women and 1 in 2000 men over 40 years of age.

### Genetic Aspects of Hyperparathyroidism

Investigation into the genetics of familial HPT, as seen in MEN1 and MEN2, familial hypocalciuric hypercalcemia (FHH), hyperparathyroidism–jaw–tumor syndrome (HPT-JT/HRPT2), and familial isolated hyperparathyroidism (FIHP) syndromes, has illuminated many aspects of normal and abnormal parathyroid cell function. The genetic abnormalities described in HPT are shown in Table IV.

#### MEN1

Primary HPT is the most common feature of the MEN1 syndrome, present in 87 to 97% of patients. The syndrome was mapped to chromosome 11q13 and the MEN1 gene encoding a 610-amino acid protein, MENIN, whose function is consistent with that of a tumor suppressor gene. Interestingly, allelic loss on chromosome 11 has also been seen in patients with tertiary HPT. The clinical features of HPT in the MEN1 syndrome recapitulate those seen in sporadic

HPT. However, the average age of onset of HPT in MEN1 is 25 years, substantially younger than that of primary HPT due to sporadic adenoma, and instances of hypercalcemia in individuals as young as 4 years of age have been reported. By 50 years of age, essentially 100% of affected family members will demonstrate HPT. Also unlike sporadic HPT, in MEN1 there is a similar prevalence between males and females. At surgery, multiple-gland involvement is the rule, although this may be quite asymmetric. Single adenomas and parathyroid cancer have also been described in the setting of MEN1.

#### MEN2

HPT also occurs as part of the MEN2A syndrome, although its incidence is low (5–20%) compared with MEN1. MEN2A, MEN2B, and familial medullary carcinoma of the thyroid result from a mutation in the RET protooncogene. RET encodes a putative tyrosine kinase receptor, and mutations appear to result in activation of the RET oncogene rather than in loss or inactivation of a tumor suppressor. Unlike the MEN1 gene, mutations of the RET protooncogene apparently are not found in sporadic parathyroid adenomas. The occurrence of HPT, as well as the age at onset, appears to be specific to the kindred in which it occurs. The clinical spectrum of HPT in patients with MEN2A resembles that seen in sporadic disease, with hypercalcemia and nephrolithiasis being the most common features. At surgery, multiple-gland enlargement or an adenoma may be found. Interestingly, HPT is not a feature of MEN2B.

#### Hyperparathyroidism–Jaw–Tumor Syndrome

In the hereditary HPT–JT syndrome, affected patients present with aggressive HPT (adenoma or carcinoma), and fibro-osseous tumors of the mandible or maxilla. Associated lesions include a variety of renal lesions (both benign and malignant). The HPT–JT gene, denoted HRPT/2, maps to chromosome 1q21–31 in the region of the susceptibility gene for hereditary prostate cancer (HPC1).

#### Familial Isolated Hyperparathyroidism

A number of families have been reported with FIHP without other associated clinical manifestations of the MEN or HPT–JT syndromes. In many of these cases, clinical follow-up will disclose the development of syndromic disease. Undoubtedly, linkage to other known syndromes or isolated genetic causes of HPT will be discovered. Clinical and genetic screening should be undertaken in patients whose family history suggests familial disease so as to facilitate treatment

Table IV Genetic Findings in Hyperparathyroidism

Gene	Chromosome	Product	Clinical phenotype
MEN1	11q13	MENIN	MEN1 20–40% sporadic adenomas Familial isolated HPT
RET	10q21	RET	MEN2A
CASR	3q13–21	CaR	FHH (heterozygotes) NSHPT (homozygotes) ADMH
HPC1/ HRPT2	1q24–31 1q21–31	Unknown	Hereditary prostate CA HPT–JT syndrome Familial isolated HPT
RB	13q	Tumor suppressor	Parathyroid CA
p53	17p13	Tumor suppressor	Parathyroid CA
PRAD1	11q13 11p15	Cyclin D1	Sporadic adenoma
VDR	12q	Vitamin D receptor	Primary HPT
Tumor suppressors	1p, 6q, 9p, 15q, 22q	Tumor suppressors	Sporadic HPT
Oncogenes	16p, 19p		Sporadic HPT

planning and follow-up for both the patients and their family members. Patients with FIHP without the MEN1 syndrome have multiple-gland disease and more profound hypercalcemia and higher rates of persistent and recurrent HPT than do those with either MEN1 disease or sporadic HPT. Patients with this condition may be at increased risk for parathyroid carcinoma.

#### **Familial Hypocalciuric Hypercalcemia**

FHH is a usually benign and asymptomatic disorder inherited as an autosomal dominant trait with essentially 100% penetrance. Patients with FHH have hypercalcemia with inappropriately elevated PTH levels. Despite their hypercalcemia, affected individuals exhibit none of the morbidity seen in hypercalcemic patients with HPT, and surgical treatment should not be undertaken. Neonatal severe HPT is a rare disease characterized by extreme hypercalcemia (commonly more than 20 mg/dl), skeletal demineralization, respiratory distress, parathyroid hyperplasia, and hypotonia. Without urgent parathyroidectomy, the disease is usually fatal. FHH and neonatal severe HPT have been shown to be the heterozygous and homozygous, respectively, phenotypic consequences of mutations in the calcium receptor (CaR) gene.

#### **Molecular Mechanisms in Sporadic Disease**

Using molecular tools developed to study familial disease, various workers have investigated sporadic parathyroid tumors, which in most cases have been shown to be clonal in origin. Evidence for both oncogene activation and suppressor gene inactivation (Table IV) has been reported.

#### **Vitamin D Receptor Polymorphisms**

Genetic variation in vitamin D metabolism has been found to be associated with some HPT states, particularly in the postmenopausal setting.

#### **Radiation-Associated Hyperparathyroidism**

HPT also occurs with increased frequency after radiation to the head and neck. HPT has also been described to have increased frequency after <sup>131</sup>I treatment for Graves' disease, although not after radioiodine treatment for malignant thyroid disease. Genetic analyses of radiation-associated parathyroid tumors have disclosed multiple alterations per tumor, most commonly involving losses in 11q and 1p. Such profiles are similar to those seen in adenomas exhibiting MEN1 gene abnormalities and suggest that this gene may be vulnerable to irradiation. These data argue for long-term monitoring of patients for both

parathyroid and thyroid disease after radiation exposure as well as for careful intraoperative assessment of both thyroid and parathyroid glands during surgery for either condition.

#### **Secondary Hyperparathyroidism**

The most common cause of secondary HPT (and of tertiary or refractory HPT) is chronic renal failure. Hyperplasia of the parathyroid glands and increased serum PTH levels occur early in the development of renal dysfunction. The primary factors responsible for the development of HPT in renal failure are phosphorus retention, decreased levels of calcitriol, hypocalcemia, and skeletal resistance to the calcemic action of PTH.

The increase in serum phosphorus seen as a result of decreasing renal function results in a reduction in the level of ionized calcium in the blood and stimulates PTH secretion. PTH reduces the tubular reabsorption of phosphorus, promoting phosphate excretion and leading to a return of both calcium and phosphorus toward normal levels at the expense of the "trade-off" of a higher circulating PTH level. In uncomplicated secondary HPT, patients are usually normocalcemic.

In advanced secondary HPT, autonomous function of hyperplastic parathyroid tissue may lead to hypercalcemia. Similarly, persistence of hyperfunction after successful renal transplantation is termed "tertiary hyperparathyroidism." The incidence and clinical significance of persistent posttransplant abnormalities in parathyroid function are still a matter of debate. As many as one-third of patients may develop hypercalcemia after transplantation, and the majority will resolve over 1 to 3 years following surgery. A minority of patients will continue to exhibit abnormalities of calcium metabolism, and perhaps 5% will require a parathyroidectomy. Indications for surgery are usually a decrease in renal function and evidence of ongoing bone disease.

#### **Parathyroid Carcinoma**

Parathyroid carcinoma is an uncommon cause of hypercalcemia, present in less than 1% of patients with primary HPT. Parathyroid carcinoma has been reported to arise within the context of adenomatous or hyperplastic parathyroid disorders, but no clear evidence exists to suggest that malignant transformation of what was originally benign parathyroid tissue occurs. Parathyroid carcinoma has also been reported after head and neck irradiation and in familial HPT. Clinically, parathyroid carcinoma is typically characterized by severe manifestations of HPT. Patients with parathyroid carcinoma usually have markedly elevated

serum calcium levels, on the order of 14 mg/dl or higher, and concomitant PTH levels that are 3 to 10 times the upper level of normal. Compared with typical patients with HPT, those with carcinoma are almost invariably symptomatic. Patients may exhibit features of bone disease and nephrolithiasis, a combination that is very unusual in patients with primary HPT of benign etiology. Physical signs in parathyroid carcinoma may include a neck mass or hoarseness due to recurrent laryngeal nerve paralysis, both of which are unlikely to occur in patients with benign disease.

In addition to clinical parameters, several intraoperative findings may suggest a diagnosis of parathyroid carcinoma. Parathyroid carcinoma is usually a fibrotic, poorly circumscribed mass that is often adherent to surrounding structures, especially the thyroid gland. Regional lymph nodes may be sites of metastatic spread. The presence of metastatic deposits in regional nodes, clear invasion of contiguous structures, and/or evidence of distant metastases to lung, liver, or bone are the most reliable diagnostic criteria for parathyroid carcinoma. Histological criteria of malignancy are nonspecific but include the presence of fibrous trabeculae traversing the tumor, capsular or vascular invasion, and frequent mitoses.

## TREATMENT OF HYPERCALCEMIA

Treatment of hypercalcemia ideally is directed toward correcting its cause. Successful therapy of malignancy-associated hypercalcemia requires an effective long-term surgical, radiotherapeutic, or chemotherapeutic treatment strategy. Without such a strategy, the overall prognosis is very poor. As a corollary, antihypercalcemic therapy should best be viewed as a temporizing measure that is used to buy time while an effective therapeutic plan aimed at treating the tumor is undertaken. If no effective antitumor treatment is available in a given case, it may be most rational to withhold antihypercalcemic therapy given that death due to hypercalcemia may be preferable to other options.

### Medical Treatment of Hypercalcemia

Regardless of etiology, hypercalcemic crisis or severe hypercalcemia is a life-threatening emergency, and although it is more commonly seen in malignancy, it can complicate primary HPT. Although strict diagnostic criteria are lacking, hypercalcemic crisis can be defined as a calcium level greater than 14 mg/dl associated with acute signs and symptoms attributable

to the elevated calcium such as volume depletion, neurological findings, and cardiac arrhythmias. Treatment of hypercalcemic crisis, whether due to HPT or to malignancy, is two-pronged: promotion of calcium excretion by the kidney and inhibition of bone resorption.

Initially, treatment is directed toward volume restoration with isotonic or half-normal saline. This also provides an obligatory calcium diuresis due to the sodium excretion. Because the volume required may be in the range of 2 to 4 L/day, patients must be carefully observed for signs of fluid overload. Thiazide diuretics must not be used because they increase renal tubular calcium reabsorption. Loop diuretics facilitate calcium excretion but may also result in dehydration and exacerbate the hypercalcemia. They may be required, however, to prevent fluid overload in older patients. Resultant hypokalemia and hypomagnesemia must be corrected. Volume expansion with improvement in glomerular filtration rate and calcium excretion may be expected to lower the serum calcium level by 1.5 to 2.0 mg/dl in 24 to 48 h.

As volume expansion measures are initiated, anti-resorptive therapy should be undertaken. Calcitonin, a potent inhibitor of osteoclastic bone resorption, should be given early and at maximal doses (up to 8 IU/kg by subcutaneous or intramuscular injection every 6 h). A decrease in serum calcium concentration of 2 to 3 mg/dl may be seen within a few hours, with a nadir expected at about 24 h. However, resistance to the hypocalcemic effect of calcitonin is commonly seen within several days of treatment, so it is most useful as part of initial therapy.

Because bone resorption is influenced by the calcium-phosphate ion product, phosphorus has been used both orally and parenterally in the past to lower serum calcium and inhibit osteoclastic activity. Because of the risk of extraskelatal calcification and the availability of safer agents, phosphate administration has little role in the management of hypercalcemia today. Oral phosphorus supplementation, in the form of neutral phosphate (250 mg orally four times per day), should be given only to patients with normal renal function whose phosphorus concentrations are below 2 mg/dl, that is, where complications of hypophosphatemia are of concern. Bisphosphonates ameliorate hypercalcemia by blocking osteoclasts and acting as bone crystal substitutes. They are highly effective at lowering serum calcium within several days and are essentially nontoxic. The preferred drug in the United States is pamidronate, generally given in a dose of 60 to

90 mg intravenously in 500 ml of saline over 4 to 24 h. Response begins within 1 to 2 days and is maximal after 4 to 5 days, with the majority of patients so treated returning to normal serum calcium values. Other effective bisphosphonates include risendronate, zoledronate, and ibandronate.

Glucocorticoids, once widely used in treating hypercalcemia, today have most utility in those cases where a glucocorticoid responsive etiology is the cause of the hypercalcemia. Patients with vitamin D intoxication or granulomatous diseases would be candidates for glucocorticoid therapy. Other malignancies such as myelomas, lymphomas, lymphocytic leukemias, and some breast cancers may be responsive to glucocorticoids.

Recently, several "calcimimetic" organic compounds of the phenylalkylamine family have been shown to allosterically modify the CaR and increase its sensitivity to activation by extracellular calcium. In short-term, placebo-controlled clinical trials, one of these agents, NPS R-568, has been shown to lower PTH and serum calcium levels in patients with both primary and secondary HPT. In animal models of secondary HPT, both continuous and cyclic administration of calcimimetics have decreased parathyroid cell growth.

Finally, dialysis may be indicated in patients with impaired renal function or in extreme emergencies. Weight-bearing mobilization of bedridden patients is important when possible.

## Surgical Treatment of Hypercalcemia Due to Hyperparathyroidism

### Indications for Surgery

In HPT, surgical therapy has a success rate of close to 100% with minimal morbidity and negligible mortality. Patients with symptoms directly attributable to HPT, such as nephrolithiasis and nephrocalcinosis, and those with overt bone disease should be offered surgical treatment.

Currently, many patients with HPT identified as hypercalcemic by automated multiphasic serum chemistry screening do not exhibit the classic signs and symptoms of HPT. Such patients are often referred to as "asymptomatic," although many have subtle findings that are often attributed to other conditions and that are alleviated by parathyroidectomy. Nonetheless, several studies have shown that most patients with asymptomatic HPT can be followed for long periods of time without significant clinical progression of their disease. The National Institutes of

Health recently reported on the results of a second Consensus Conference on indications for surgical intervention in primary HPT. The recommendations are summarized as follows:

1. Serum calcium greater than 1 mg/dl above the upper limits of normal
2. Creatinine clearance reduced by more than 30% compared with age-matched controls
3. 24-h urinary calcium excretion greater than 400 mg/dl
4. Bone density at the lumbar spine, hip, or distal radius that is more than 2.5 *SD* below peak bone mass (T score < 2.5)
5. Patients for whom medical surveillance is either not desirable or not possible
6. Young patients (under 50 years of age)

The role of parathyroidectomy in pregnancy is a matter of some debate. HPT increases the risk of spontaneous abortion, stillbirth, and fetal demise, but some patients with very mild disease may be appropriate for nonoperative management. After delivery, neonates must be watched for signs of hypocalcemia until their suppressed parathyroids regain function. In any case, the development of unexpected complications in the pregnant patient, such as renal calculi, pancreatitis, and peptic ulcer disease, should raise the question of the presence of HPT. Patients with severe or complicated HPT should be referred for surgery.

### Aspects of Surgical Treatment

Successful parathyroid surgery requires an appreciation for the normal anatomical relationships of the thyroid and parathyroid glands, the recurrent and superior laryngeal nerves, and the inferior thyroid arteries as well as their common embryological variations. Although the availability of better imaging techniques and the use of intraoperative PTH determination may permit a less invasive surgical approach in some patients with HPT, the standard of care is still the traditional bilateral dissection with visual assessment of all four glands.

### Role of Imaging Studies

Until recently, the contribution of imaging studies to the management of patients with HPT has been limited by the accuracy of the available modalities. Currently, nuclear imaging with technetium sestamibi is the method of choice for preoperative localization of parathyroid adenomas. Technetium sestamibi exhibits preferential uptake and retention in abnormal parathyroid glands. The addition of single positron

emission computerized tomography (SPECT) analysis to the sestamibi scan provides three-dimensional information about the location of the imaged gland. Ultrasound can also provide three-dimensional information and provides assessment of coexisting thyroid disease that may require evaluation or treatment at the time of parathyroid surgery. In recent reports, the combined application of sestamibi and ultrasound resulted in successful visualization of an adenoma in 96% of cases.

In reoperative situations, all imaging modalities are used, beginning with the least to the most invasive procedures.

If sestamibi and U.S. studies are negative, other anatomical studies, such as computerized tomography (CT) and magnetic resonance imaging (MRI), may be useful. Angiography and selective venous sampling may be required in some cases. Angiography has been reported to localize parathyroid tumors in 60% of cases and may permit angiographic ablation of mediastinal adenomas. Selective venous sampling may provide regional localization data but requires a high level of expertise that is not available in most institutions. Where it is available, analysis of venous samples can be carried out in "real time" in the radiology suite using the "quick" intact PTH assay.

Imaging modalities may contribute to intraoperative management as well. The combination of sestamibi SPECT imaging with three-dimensional reconstruction and probe-directed intraoperative radionuclide detection may allow for a more limited dissection, as may the use of high-resolution intraoperative ultrasound.

In summary, current data do not support the use of routine preoperative imaging in the treatment of HPT unless the intent is to perform a unilateral or minimally invasive exploration, although it appears that accurate localization does shorten the procedure. Preoperative identification of an apparently solitary enlarged parathyroid gland also does not guarantee that it is the only abnormal gland. The role of imaging is, however, critical in the reoperative circumstance.

#### *Approaches to Parathyroidectomy*

Critical to the success of the operation is the maintenance of meticulous homeostasis throughout. In the majority of cases, an adenoma will be present and the initial goal is the identification of an enlarged and a normal-appearing gland on the affected side. If imaging data are available, the suspected side is explored first. If an adenoma is identified grossly and three other apparently normal glands are visualized, one may be biopsied to confirm the presence of

parathyroid tissue. In general, there are no histological criteria that can reliably differentiate among normal, hyperplastic, and adenomatous parathyroid tissue. The adenoma is resected and the procedure is completed. If it is available, demonstration by the quick intraoperative assay that PTH levels have dropped by 50 to 65% of baseline after resection of a presumed adenoma has reliably confirmed that all abnormal tissue has been resected. The most common cause of unsuccessful parathyroid surgery is failure to find an adenoma, and in most cases the adenoma is ultimately found in one of the common sites of embryological derivation.

Given the preponderance of a single adenoma in patients with primary HPT, a unilateral approach in patients whose disease has been localized has long been attractive. Early limited experience suggested that good results could be obtained with a unilateral exploration when an enlarged gland and a normal gland are found on one side. More recently, the availability of convincing preoperative localization studies and quick intraoperative PTH assays has permitted focused dissection with either general or local anesthesia in many patients. It should be recognized that the intent to perform a unilateral exploration is just that. In 50% of cases, a bilateral dissection is ultimately required, either because the adenoma is not found on the initial side or because the patient is found to have more than one abnormal gland or other (usually thyroid) pathology requiring treatment.

The laparoscopic era has produced several endoscopic approaches to diseases of the thyroid and parathyroid glands. These include both totally endoscopic and video-assisted techniques. In a recent multi-institutional study, video-assisted parathyroidectomy could be carried out in 89% of 123 qualified patients. Patients were excluded as candidates for an endoscopic approach if they had negative imaging studies, suspected multiglandular or concomitant thyroid disease, and/or prior neck surgery or irradiation. Disadvantages of this procedure over a conventional approach included the need for general anesthesia and longer operating times. These approaches may find the greatest utility in patients with deeply situated, typically superior adenomas for which a conventional approach would require a more extensive dissection.

#### *Intraoperative Decision Making: Multiple-Gland Disease*

If multiple abnormal glands are encountered, the decision must be made as to how much tissue is to be removed. Double adenomas are estimated to occur in 2 to 5% of the general population and in perhaps 9% of older patients. The presence of more than one

enlarged gland more often implies a diagnosis of hyperplasia; this may be quite asymmetric in nature, and a complete evaluation of all glands is mandatory. Clarification of the status and location of all parathyroids should be carried out before any gland is resected. The role of the quick intraoperative PTH assay in directing the extent of resection is undergoing investigation in several centers.

If all four glands are enlarged, subtotal parathyroidectomy—the resection of three to three and one-half glands—is a reliable treatment. Intraoperative PTH determination may help to guide the extent of resection in many cases. In patients with secondary hyperparathyroidism and MEN1 syndrome, even when four glands have been identified, the cervical thymus should be resected because the incidence of supernumerary glands is significant in these conditions.

A second approach to four-gland hyperplasia is total parathyroidectomy and autotransplantation. After total parathyroidectomy, minced fragments of parathyroid tissue are placed in separate sites of the brachioradialis muscle in the forearm and are secured with nonabsorbable suture material. The theoretical advantage of this approach is that if recurrent (graft-dependent) HPT occurs, treatment requires only removal of some of the grafts under local anesthesia rather than a neck reexploration. Of course, successful application of the technique depends on identification and removal of all parathyroid tissue from the neck (including supernumerary glands).

Besides patients with secondary HPT, other individuals at risk for recurrent HPT who may be candidates for these procedures include patients with familial hyperplasia, particularly those with MEN1.

In summary, the choice of whether to perform a total parathyroidectomy and autotransplant or to perform a subtotal resection in the case of secondary HPT or MEN1 is a matter of personal preference; both are reliable techniques when performed by experienced surgeons.

#### **Parathyroid Carcinoma**

The most effective management of parathyroid carcinoma is complete resection of the tumor at the time of initial exploration. Half of patients so treated may be cured. Where possible, an *en bloc* resection should be carried out, removing the mass with the contiguous thyroid lobe and nodes in the central (paratracheal) compartment. The other parathyroids must be identified because carcinoma and hyperplasia may exist concurrently. Parathyroid carcinoma is usually an indolent malignancy, with local recurrence and distant

metastases to lung and bone occurring over many years. Because the major physiological effects of parathyroid carcinoma are related to unregulated PTH secretion and hypercalcemia rather than to tumor burden, resection of metastatic disease plays an important role in patient management. For similar reasons, reoperation for locally recurrent parathyroid carcinoma is appropriate.

#### **Outcome of Surgical Treatment for Hypercalcemia Due to HPT**

In experienced hands, parathyroidectomy is effective at returning the serum calcium and PTH levels to normal in 95 to 99% of cases of primary HPT. The ability of serum calcium to suppress PTH secretion also promptly returns to normal after resection of parathyroid adenomas, reflecting the normal “set point” of the remaining glands. There are also abundant data documenting the effects of successful parathyroidectomy on associated conditions such as renal function, urolithiasis, and bone disease.

The effect of surgery on the clinical course of other conditions associated with HPT is less clear. Hypertension is more prevalent in HPT patients than in control populations, and some studies have shown an improvement in blood pressure after parathyroidectomy. Likewise, some data indicate improvement in left ventricular hypertrophy and endothelial function after parathyroidectomy. Because correction of HPT results in lowering of IL-6 levels and other acute phase reactants related to cardiovascular risk, it is possible that longitudinal follow-up studies currently in progress may show a beneficial effect on cardiovascular parameters in surgically treated patients.

Several health assessment studies have documented improvement after successful parathyroid surgery in a variety of the neuropsychiatric and musculoskeletal complaints common to the hypercalcemic state. A controlled clinical trial to compare the outcomes of surgical and medical treatment of HPT (including use of calcimimetics as they become available) with respect to these issues will be important in defining the appropriate criteria for surgical treatment of mild or asymptomatic primary HPT.

## **CONCLUSIONS**

Hypercalcemia is a common and, when severe, potentially life-threatening condition. Understanding the basic mechanisms of calcium homeostasis and the abnormalities that occur in various diseases is critical

to safe, successful, and cost-effective patient management. Current research has begun to define many of the molecular mechanisms that underlie the common causes of hypercalcemia. Malignancy-associated hypercalcemia and HPT together account for the vast majority of cases of hypercalcemia. New diagnostic and medical and surgical approaches to these diseases continue to refine treatment and improve patient outcome.

### See Also the Following Articles

Calcitonin, Overview • G Protein-Coupled Receptors • Hypercalciuria • Hyperparathyroidism, Primary • Parathyroid Hormone (PTH) • Skeletal Development • Vitamin D

### Further Reading

Bilezikian, J. P., Potts, J. T., Fuleihan, G. E., Kleerekoper, M., *et al.* (2002). Summary statement from a workshop on asymptom-

atic primary hyperparathyroidism: A perspective for the 21st century. *J. Clin. Endocrinol. Metab.* **87**, 5353–5361.

Brown, E. M., and MacLeod, R. J. (2001). Extracellular calcium sensing and extracellular calcium signaling. *Phys. Rev.* **81**, 239–297.

Kinder, B. K., and Stewart, A. F. (2002). Hypercalcemia. *Curr. Problems Surg.* **39**, 354–447.

Manolagas, S. C. (2000). Birth and death of bone cells: Basic regulatory mechanisms and implications for the pathogenesis and treatment of osteoporosis. *Endocrine Rev.* **21**, 115–137.

Pasieka, J. L., and Parsons, L. L. (1998). A prospective surgical outcome study of relief symptoms following surgery in patients with primary hyperparathyroidism. *World J. Surg.* **22**, 513–519.

Rodan, G., and Martin, J. T. (2000). Therapeutic approaches to bone diseases. *Science* **289**, 1508–1514.

Rose, B., and Shoback, D. (2001). Update of the calcium-sensing receptor and calcimimetics. *Curr. Opin. Endocrinol. Diab.* **8**, 29–33.

Stewart, A. F. (2002). Hyperparathyroidism, humoral hypercalcemia of malignancy, and the anabolic actions of parathyroid hormone and parathyroid hormone-related protein on the skeleton. *J. Bone Min. Res.* **17**, 758–762.

Udelsman, R., Donovan, P. I., and Sokoll, L. J. (2000). One hundred consecutive minimally invasive parathyroid explorations. *Ann. Surg.* **232**, 331–339.

Factors that influence renal tubular calcium reabsorption can be divided into two categories: (1) those that decrease calcium reabsorption, increasing calcium excretion such as hypercalcemia, loop diuretics, extracellular fluid volume expansion, NaCl, protein or glucose intake, phosphate or potassium depletion, and metabolic acidosis; and (2) those that increase calcium reabsorption, thereby reducing urinary calcium, such as PTH, PTH-related protein, thiazide diuretics, metabolic alkalosis, phosphate or potassium administration, hypocalcemia, and extracellular fluid volume depletion.

Urinary calcium should not exceed 300 mg (7.5 mmol)/day, considering the adequate intake of 1000 mg/day as recommended for 30- to 50-year-old adults as part of the Dietary Reference Intake (DRI) or even if the intake goes up to 2000 mg/day. Among normal individuals, urinary calcium increases only about 8% as dietary calcium intake increases. Conversely, patients with IH may exhibit a twofold higher calcium excretion across a similar range of calcium intake. Therefore, compared with people without stones, IH patients show a marked increase in the fraction of diet calcium absorbed and lost in urine as well as in bone mineral lability. Hypercalciuria is consequent to a dysregulation of calcium transport in the gut, kidney, or bone, thereby representing a systemic abnormality in calcium homeostasis. The continued excessive urine calcium excretion may result in bone demineralization.

To establish the diagnosis of normocalcemic IH in stone formers, the serum calcium level should be measured to rule out hypercalcemic disorders. Among hypercalcemic hypercalciuric conditions that may alter urinary calcium in CSF patients, primary hyperparathyroidism (PHP), characterized by the presence of elevated serum calcium with nonsuppressed serum PTH values, is the most common, accounting for more than 50% of cases. On the other hand, PHP affects only 0.5 to 5% of all CSF patients. Other causes of hypercalcemic hypercalciuric states include granulomatous diseases such as sarcoidosis and tuberculosis due to the extrarenal conversion of  $25(\text{OH})\text{D}_3$  to  $1,25(\text{OH})_2\text{D}_3$  in the granulomatous tissue, leading to increased intestinal calcium absorption. Prolonged immobilization, hyperthyroidism, and acromegaly may also be associated with hypercalcemic hypercalciuria. In previous evaluations by our group, the incidence of stones in hyperthyroidism patients had been only 1.6% despite the presence of hypercalciuria in 9% of the patients, whereas the frequency of renal stones and hypercalciuria in acromegalic patients were 9.5 and 36%, respectively. Most likely, those

findings may be ascribed to the earlier control of the hyperthyroid state, resulting in a shorter period of hypercalciuria, in contrast with acromegaly, in which late diagnosis may be associated with long-lasting hypercalciuria. Hypercalcemia associated with malignancy also does not ordinarily result in stone formation because patients do not live long enough to form stones. Milk-alkali syndrome as a cause of hypercalcemic hypercalciuria is unusual since the advent of  $\text{H}_2$  acid-blocking drugs has replaced the need for calcium-containing nonabsorbable antiacids. Familial hypocalciuric hypercalcemia patients are not at increased risk for making stones because they do not present with hypercalciuria.

## PATHOGENESIS OF HYPERCALCIURIA

### Mechanisms

IH may result from disordered regulation of calcium fluxes in the intestine, kidney, and/or bone. The first underlying mechanism seems to be an increased intestinal calcium absorption either occurring directly or mediated by the excess of  $1,25(\text{OH})_2\text{D}_3$ , which would result in hypercalcemia and suppression of PTH, leading to an increased filtered load of calcium. Hypercalciuria would then be ascribed to the combination of increased filtered load of calcium and suppression of PTH. The increased calcium excretion then corrects the hypercalcemia, and serum calcium returns to normal levels. The second alternative mechanism is the reduction on renal tubular reabsorption of calcium, leading to a subsequent decrease in serum calcium level, which in turn stimulates PTH secretion, with subsequent increased  $1,25(\text{OH})_2\text{D}_3$  levels resulting in a secondary increased intestinal calcium absorption. The increased urinary calcium excretion in this condition should persist after an overnight fast. In some patients, the cause of the hyperabsorption would be a defect in renal tubular phosphorus reabsorption, causing hypophosphatemia, high  $1,25(\text{OH})_2\text{D}_3$  levels, and secondary intestinal hyperabsorption of calcium, resulting in hypercalciuria. Finally, the third proposed mechanism is an enhanced bone demineralization as a primary process (not caused by hyperparathyroidism), leading to an increased serum calcium concentration, a decreased level of PTH, and an increased filtered load of calcium, resulting in hypercalciuria. Those patients should also present with higher calcium excretion after an overnight fast, similarly to those with defective renal tubular calcium reabsorption. In summary, in pure intestinal calcium hyperabsorption, namely

“absorptive hypercalciuria,” the levels of serum PTH or urinary cAMP should be low or normal and the overnight fasting should abolish the hypercalciuria. Conversely, the other subtype of hypercalciuria caused by a defect in calcium tubular reabsorption or “renal leak,” defined initially as renal hypercalciuria, would present with an opposite profile, one with high serum PTH and urine cAMP together with high fasting urine calcium.

This pathophysiological classification seemed to be very appealing because of its simplicity, disclosed by Pak and associates in 1975 through the development of a simple acute oral calcium load test that could clearly distinguish between absorptive and renal hypercalciuria subtypes. One of the parameters that helped to distinguish between renal and absorptive hypercalciuric patients using this test was the measurement of the overnight fasting urinary calcium corrected for creatinine excretion (urinary calcium/creatinine ratio), with a cutoff value of 0.11. Absorptive hypercalciuria subtype exhibited a urinary calcium/creatinine ratio lower than 0.11 and an exaggerated response of urinary calcium/creatinine after the oral calcium load, whereas the renal subtype presented a fasting urinary calcium/creatinine higher than 0.11 in addition to high levels of serum PTH. Practical therapeutic implications derived from this test included the recommendation of a low-calcium diet for absorptive-type hypercalciuria and that of thiazide agents for renal or fasting hypercalciuria. In a previous study by our group in 1996, we collected a 24-h urine sample for calcium determination, under conditions of a mean usual calcium intake of 540 mg/day, from CSF patients who previously presented with an “absorptive” or “renal” type of response to the oral calcium load test. We observed that the majority of these patients (63 and 78% of each group, respectively) presented with normocalciuria rather than hypercalciuria. Because this apparently normal calcium excretion might have resulted from a combination of high calcium absorption and low calcium intake, these patients were then challenged with a higher calcium intake of 1500 mg/day given as a supplement for 1 week. Regardless of whether there was a former absorptive or renal type of response to the acute calcium load test, the higher calcium intake disclosed the presence of subpopulations sensitive to calcium intake in previously normocalciuric patients (dietary calcium dependent). Conversely, most of the hypercalciuric patients, when challenged with a higher calcium intake, did not present a further increase in their urinary calcium. This showed that

under conditions of low calcium intake, as was the case for these patients as part of their dietary habit, they were already excreting calcium in excess of their intake, hence being considered as dietary calcium independent. In addition, because the morning urinary fasting calcium/creatinine ratio seemed to be the single parameter that would distinguish between renal and absorptive hypercalciuric patients with a cutoff value of 0.11, this determination was repeated in 31 patients. Fully 87% of them changed their results from values higher than 0.11 to values lower than 0.11. Other investigators also detected renal and absorptive hypercalciuria in the same patients when examined at different points of time. This discrepancy between different determinations of urinary fasting calcium/creatinine may also be ascribed to an incomplete intestinal calcium clearance, a previous higher sodium or protein intake inducing mild acidosis, a higher bone resorption, or even a state of chronic suppression of PTH. Therefore, elevated fasting hypercalciuria might not always reflect the presence of a renal leak, and the single determination of this urinary parameter might not be a specific criterion for patient classification.

Neither cAMP nor serum vitamin D helped to distinguish between absorptive and renal hypercalciuria. Taken together, these data suggested that absorptive and renal hypercalciuria should be considered the same rather than two distinct entities, a hypothesis raised previously by Coe and co-workers in 1982. These investigators observed that feeding IH patients a very low-calcium diet, approximately 140 mg/day for 9 days, all but a few lost more calcium in the urine than they ate, and their urine losses formed a smooth continuum from overlap with the top of the normal values to much higher, without evidence of a break or division point along the obtained curve. The values of PTH were generally low despite calcium losses and did not rise when the normal diet was replaced with the low-calcium diet. Serum calcitriol values ranged from normal to high. On the other hand, some authors have recommended that, for a clear-cut distinction between these subtypes, patients must be evaluated under a more prolonged dietary calcium restriction for at least 4 weeks to disclose some difference.

It seems that IH patients most likely would fit better into two extremes of a continuum and broad behavior, representing a single process for pathogenesis, that varies in severity (some tend to hyperabsorb calcium more than do others) and probably results from abnormal regulation of  $1,25(\text{OH})_2\text{D}_3$ .

## The Role of PTH

Theoretically, the negative calcium balance found among patients with renal hypercalciuria or renal leak could conceivably lead to secondary hyperparathyroidism. Early studies by Coe in 1973 and Pak in 1975 demonstrated that patients with renal hypercalciuria exhibited a higher serum PTH or fasting urinary cAMP. Shortly thereafter, it was shown that high PTH levels were scarce, so that the term “renal hypercalciuria” should instead be “fasting hypercalciuria” whenever one found elevated fasting urinary calcium in the absence of elevated serum PTH. In 1982, Coe observed normal or low values of PTH despite calcium losses in another series of patients. Subsequently, several investigators reported normal parathyroid function in IH either through urinary cAMP determination or by using newer serum PTH assays. In summary, there is no agreement on the existence of secondary hyperparathyroidism in hypercalciuria. However, considering the lack of secondary hyperparathyroidism in most of the recent series, and considering the evidence for high levels or disordered control of  $1,25(\text{OH})_2\text{D}_3$ , it seems reasonable to assume that in this condition, calcitriol may suppress the secretion of PTH, explaining why most IH patients do not exhibit high PTH levels.

## The Role of $1,25(\text{OH})_2\text{D}_3$

Early studies by Broadus, Insogna, and colleagues in 1984 found plasma levels of  $1,25(\text{OH})_2\text{D}_3$  to be in the upper normal range or elevated in the majority of patients with absorptive hypercalciuria when compared with nonhypercalciuric stone formers, probably due to a higher renal synthesis of this hormone. However, normal serum calcitriol levels have also been detected among patients with high rates of calcium absorption. In 1991, Lemann and colleagues speculated that even small elevations of serum  $1,25(\text{OH})_2\text{D}_3$  seem to be sufficient to increase bone resorption when dietary calcium intake is low, indicating that  $1,25(\text{OH})_2\text{D}_3$  may up-regulate its own receptor and so amplify its effect. Therefore, rates of calcium transport by enterocytes that are high in relation to the effects of calcitriol, or possibly abnormal levels of the vitamin D receptor (VDR) amplifying the effect of circulating  $1,25(\text{OH})_2\text{D}_3$ , could explain the occurrence of calcium hyperabsorption and also the mobilization of bone mineral induced by a low-calcium diet despite the presence of normal calcitriol levels among patients with IH. Approximately two decades ago, David A. Bushinsky and his group started

to develop a model of hypercalciuria in rats that were successively inbred for more than 50 generations until they produced a strain of a genetic hypercalciuric (GHS) rat. In this experimental model, Bushinsky was able to show that the increased calcium absorption is mediated by an increase in the number of VDR in the intestine (duodenum), kidney cortex, and bone. Not only are there more receptors, but they are hyperresponsive to  $1,25(\text{OH})_2\text{D}_3$ , as evidenced by a higher calcium efflux from GHS rats' calvariae when exposed to increasing amounts of  $1,25(\text{OH})_2\text{D}_3$  relative to control animals. The increased number of intestinal VDR, even if levels of circulating  $1,25(\text{OH})_2\text{D}_3$  are normal, may result in increased functional VDR- $1,25(\text{OH})_2\text{D}_3$  complexes that exert biological actions in enterocytes to increase intestinal calcium transport. The defective tubular calcium reabsorption in this model is further evidenced by a threefold calcium excretion when compared with normal rats at similar glomerular filtration rates and ultrafiltrable calcium concentration, resulting in similar filtered loads of calcium. Because the renal tubular defect would necessitate bone resorption to maintain normal serum calcium levels, the renal and bone defects have been shown to be independent of each other. The use of a bisphosphonate, a bone resorption blocker agent (alendronate), induced a decrease in urinary calcium excretion in GHS rats, confirming the hypothesis that bone contributes to hypercalciuria.

## The Role of Bone

Decreased bone mineral density (BMD) has been consistently reported in several series of hypercalciuric CSF patients, owing to the important contribution of hypercalciuria to bone demineralization. A histomorphometric study undertaken by our group in 1994 confirmed a low bone volume, a tendency of low bone formation coupled with increased bone resorption and delayed bone mineralization, in male hypercalciuric CSF patients. Further studies showed conflicting data regarding bone resorption, but all suggested a severe mineralization defect among hypercalciuric patients. The reason for such a defect remains unknown considering the adequate levels of serum calcium, phosphorus, and vitamin D. A population-based cohort study by L. Joseph Melton, III, and colleagues in 1998 showed a fourfold increase in vertebral fracture risk among urolithiasis patients. In a large cross-sectional study conducted by Diane S. Lauderdale, Murray J. Favus, and other investigators in 2001

using the Third National Health and Nutrition Examination Survey (NHANES III), a history of kidney stones was found to be associated with lower femoral neck BMD and more prevalent wrist and spine fractures in men after adjustments for age and body mass index. For the men (but not the women) who formed kidney stones, there was a correlation between low milk consumption and low neck BMD. It is interesting that most of the reports on low BMD among hypercalciuric CSF patients in the literature show a higher prevalence in men than in premenopausal women for some unknown reason. However, postmenopausal women with a previous history of renal stones still present a higher frequency of osteopenia than do normal postmenopausal women. Experimental studies in GHS rats highlight the contribution of bone demineralization, as noted previously.

In 1991, Pacifici and co-workers demonstrated that monocytes from patients with fasting hypercalciuria overproduced interleukin-1 (IL-1), a potent *in vitro* and *in vivo* stimulator of bone resorption. Subsequently, in 1996, Weisinger and his associates found increased IL-1 $\alpha$  mRNA in unstimulated monocytes from IH patients as well as higher lipopolysaccharide-induced production of IL-6 and TNF $\alpha$ . Based on these findings, which were corroborated in several other reports, it has been suggested that bone resorption induced by elevated cytokines could represent the primary mechanism leading to hypercalciuria. It still remains to be clarified whether the hypercalciuria is a "bone disease," resulting from elevated bone resorption (primary mechanism), or a "renal tubular disease," in which the renal tubular defect leads to hypercalciuria and the resultant negative calcium balance then stimulates bone resorption (secondary mechanism). The reduction of hypercalciuria by a bone resorption blocker agent, such as alendronate, favors the first mechanism, whereas the improvement of bone mass by a drug that reduces calcium losses by acting on the renal tubules favors the second mechanism.

### The Role of Diet

In the past, calcium restriction became a very popular recommendation, based on the contribution of calcium intake and intestinal calcium hyperabsorption to hypercalciuria, augmenting the risk of stone formation. However, there has been no evidence that a reduction in dietary calcium intake prevents stone recurrence. In addition, in a large prospective epidemiological study conducted by Curhan and his group in 1983, healthy men with different levels of calcium

intake were followed up for 8 years; surprisingly, it was observed that the lower the calcium intake, the higher the incidence of stone formation. A hypothesis to explain such an unexpected effect was a secondary increase in urinary oxalate due to a decreased binding of oxalate to calcium in the gastrointestinal tract. Thus, dietary calcium binds intestinal oxalate, preventing its absorption and subsequent excretion. Nevertheless, one has to consider that the benefits of a high-calcium supply do not apply to calcium supplements, which usually are not taken with meals, thereby losing their oxalate-chelating properties. It is very well established that calcium excretion is not solely affected by the intake of calcium but that other nutrients, such as animal protein, sodium, oxalate, and potassium intake, might influence calcium excretion as well. Finally, focusing on the bone issue, many investigators have addressed the loss of bone mass in hypercalciuric patients, not only suggesting the contribution of a low-calcium diet in such loss but also stressing the deleterious effects of high animal protein and sodium intake on bone. In summary, there are many reasons why calcium restriction should be avoided in hypercalciuric patients. First, there is no clear distinction between absorptive and renal hypercalciuria. Second, no prospective studies have supported the belief that calcium restriction leads to a reduction in stone recurrence. Third, calcium restriction induces secondary hyperoxaluria. Fourth, calcium restriction predisposes to bone loss due to a negative calcium balance. Fifth, chronic calcium restriction might up-regulate VDRs, allowing 1,25-vitamin D to stimulate both intestinal calcium absorption and bone resorption. Sixth, other nutrients, such as protein, sodium, oxalate, and potassium, may affect calcium excretion as well. Animal protein-induced hypercalciuria may be caused by a higher bone resorption and lower tubular calcium reabsorption to buffer the acid load, and it may also be caused by the elevated filtered load of calcium and the presence of non-reabsorbable calcium sulfate in the tubular lumen. The effect of NaCl intake on increasing calcium excretion is well established. Every 100-mmol increase in dietary sodium results in an approximately 25- to 40-mg rise in urinary calcium per day. The adverse effects of a high-NaCl intake and the resultant higher calcium excretion have been reported extensively by many investigators. In a previous study by our group in 2000, Martini and colleagues showed that a high-NaCl consumption (higher than 16 g/day) is associated with low BMD in hypercalciuric stone formers. A carefully controlled long-term study by Loris Borghi in 2002 showed that for men with IH, a diet with a normal amount of calcium (1200 mg/day) but

reduced amounts of animal protein (52 g/day) and salt (50 mmol/day) for 5 years is superior to a low-calcium diet (400 mg/day) in preventing recurrent stones.

## DRUG TREATMENT

Thiazide diuretics are the first-choice drugs for hypercalciuria. Indapamide, a thiazide-like agent, has also been employed. Thiazides increase renal tubular calcium reabsorption, lowering urine calcium and leading to a consequent fall in calcium oxalate and phosphate supersaturation. Two double-blind, randomized, prospective, and placebo-controlled trials using either hydrochlorothiazide or chlorthalidone, in doses of 25 to 50 mg/day, showed a significantly lower rate of recurrent stone formation only after 3 years (up to 25%) compared with placebo (up to 55%). Surprisingly, these studies were performed in patients not categorized according to urinary derangement, and response to therapy was independent of baseline level of urinary calcium. As for the effects on bone, small studies conducted by our group and others suggested additional benefits of thiazides on BMD, as one would expect. On the other hand, thiazides' adverse effects, which are often dose related (e.g., sexual impotence, potassium wasting, raised serum cholesterol, glucose tolerance) are common; thus, intolerance should be kept in mind.

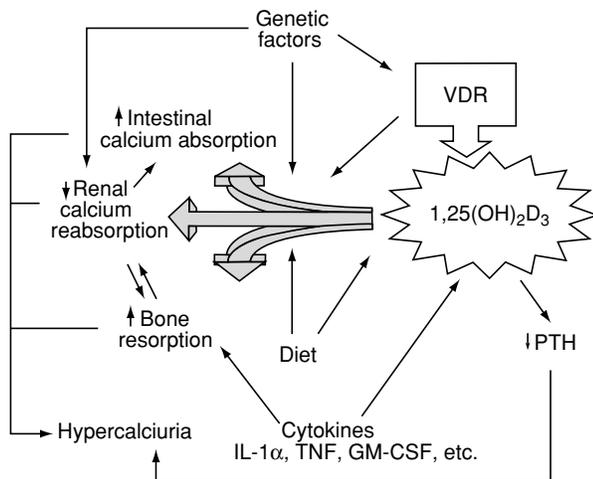
## GENETICS OF HYPERCALCIURIA

In the majority of the family-based studies, increased calcium excretion is the most common phenotype associated with kidney stone formation. According to a study by Scheinman in 1999, approximately 40% of patients with IH have a family history of nephrolithiasis. The X-linked recessive nephrolithiasis (Dent's disease) is a rare monogenic kidney stone disease affecting males for which a responsible gene has already been identified, a mutated *CLCN-5* chloride channel that determines hypercalciuria, hematuria, low-molecular-weight proteinuria, nephrocalcinosis, and nephrolithiasis with eventual renal failure. *CLCN-5* is a member of the family of voltage-gated chloride channels that also includes *CLC-Kb*, which is disrupted in the largest group of patients with Bartter syndrome, another hypercalciuric condition. Nevertheless, mutation analysis of the *CLCN-5* gene was normal among patients with IH, and *CLCN-5* sequence was shown to be normal in the GHS rat.

Although several authors have reported families in which IH appears to be inherited as a single

autosomal dominant genetic disorder, there are many reasons to conclude that IH should instead be attributed to a complex polygenic trait. The distribution curve of calcium excretion in stone formers is similar in shape to that in controls, albeit shifted toward higher rates of excretion, probably resulting from the influence of several factors. Failure in attempts to categorize IH patients under distinct subtypes probably illustrates that different mechanisms participate in producing hypercalciuria in different patients. Some candidate genes for IH have been suggested by their known physiology, including those encoding the adenosine triphosphatase (ATPase), calbindin (28 kDa), osteocalcin, IL-1 $\alpha$ , IL-1 $\beta$ , renal sodium-dependent phosphate transporter, chloride channels, VDR, calcium-sensing receptor, and 1 $\alpha$ -hydroxylase of vitamin D. The multifaceted physiology of IH may reflect the combined effects of polymorphisms in several genes. Thus, in some patients presenting features of absorptive hypercalciuria, overexpression of VDR and/or overactivity of the renal 1 $\alpha$ -hydroxylase could underlie the excessive intestinal calcium absorption. In a 1998 study, Alain Bonnardeaux and co-workers found no linkage between the putative 1 $\alpha$ -hydroxylase of vitamin D locus and quantitative traits associated with IH. Shortly thereafter, the same authors suggested evidence for a susceptibility gene near the VDR locus in idiopathic calcium stone formation, and quantitative trait linkage (QTL) analysis of urinary calcium yielded linkage to some, but not all, of the markers. In a recent evaluation by our group, BsmI VDR polymorphism was not associated with bone loss in hypercalciuric CSF patients. Genetic variants of the calcium-sensing receptor gene were not linked to biochemical markers of IH. In a 1999 study, Berenice Reed and colleagues found linkage to chromosome 1q23–24 in three families with absorptive hypercalciuria. In more recent work, the same group reported the occurrence of base-pair substitutions in the soluble adenylate cyclase gene that segregate with the disease in the same three kindreds and also reported that these changes were associated with absorptive hypercalciuria and low spinal bone density in a population of patients. However, there is still no evidence that substitutions or mutations in this gene cause absorptive hypercalciuria. It is likely that different genetic mechanisms may predominate in different patient populations. Technical advances will shed light on our understanding of the role of genetics in hypercalciuria.

As is evident from the current knowledge, the pathophysiology of hypercalciuria cannot be explained by a single mechanism or theory. [Figure 1](#) presents a



**Figure 1** Hypothetical mechanisms for idiopathic hypercalciuria. Increased intestinal calcium absorption (mediated or not by  $1,25(\text{OH})_2\text{D}_3$ ), decreased renal tubular calcium reabsorption, and enhanced bone resorption are potential mechanisms representing a systemic abnormality in calcium homeostasis. Experimental and genetic data point to a possible pivotal involvement of vitamin D receptor (VDR) and/or  $1,25(\text{OH})_2\text{D}_3$  under the influence of diet as well. The role of cytokines is crucial in bone demineralization. Disordered control of  $1,25(\text{OH})_2\text{D}_3$  may further suppress parathyroid hormone (PTH). There is a complex interrelationship among all of these factors, and genetic influence seems unequivocal. IL- $1\alpha$ , interleukin-1; IL-6, interleukin-6; TNF, tumor necrosis factor; GM-CSF, granulocyte macrophage-colony-stimulating factor.

hypothetical view of the pathophysiological events leading to idiopathic hypercalciuria in CSF patients.

## See Also the Following Articles

Hypercalcemia and Hypercalcemia Treatment • Kidney Stones • Parathyroid Hormone (PTH) • Vitamin D

## Further Reading

- Broadus, A. E., Insogna, K. L., Lang, R., Ellison, A. F., and Dreyer, B. E. (1984). Evidence for disordered control of  $1,25$ -dihydroxy-vitamin D production in absorptive hypercalciuria. *N. Engl. J. Med.* **311**, 73–80.
- Bushinsky, D. A. (1996). Genetic hypercalciuric stone forming rats. *Seminars Nephrol.* **16**, 448–457.
- Bushinsky, D. A. (1999). Calcium, magnesium, and phosphorus: Renal handling and urinary excretion. In “Primer on the Metabolic Bone Diseases and Disorders of Mineral

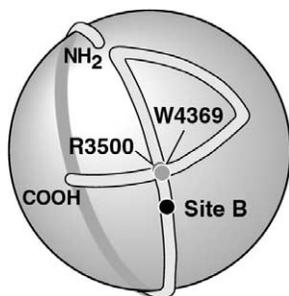
Metabolism” (M. J. Favus, ed.), 4th ed., pp. 67–74. Lippincott Williams & Wilkins, Philadelphia.

- Coe, F. L., Favus, M. J., Crockett, T., Strauss, A. L., Parks, J. H., Porat, A., Gantt, C. L., and Sherwood, L. M. (1982). Effects of low calcium diet on urine calcium excretion, parathyroid function, and serum  $1,25(\text{OH})_2\text{D}_3$  levels in patients with idiopathic hypercalciuria and in normal subjects. *Am. J. Med.* **72**, 25–32.
- Curhan, G. C., Willet, W. C., Rimm, E. B., and Stampfer, M. J. (1983). A prospective study of dietary calcium and other nutrients and the risk of symptomatic kidney stones. *N. Engl. J. Med.* **328**, 833–838.
- Hebert, S. C. (1996). Extracellular calcium-sensing receptor: Implications for calcium and magnesium handling in the kidney. *Kidney Int.* **50**, 2129–2139.
- Heilberg, I. P. (2000). Update on dietary recommendations and medical treatment of renal stone disease. *Nephrol. Dial. Transplant.* **15**, 117–123.
- Heilberg, I. P., Martini, L. A., Draibe, S. A., Ajzen, H., Ramos, O. L., and Schor, N. (1996). Sensitivity to calcium intake in calcium stone forming patients. *Nephron* **73**, 145–153.
- Heilberg, I. P., Martini, L. A., Szejnfeld, V. L., Carvalho, A. B., Draibe, S. A., Ajzen, H., Ramos, O. L., and Schor, N. (1994). Bone disease in calcium stone-forming patients. *Clin. Nephrol.* **42**, 175–182.
- Heilberg, I. P., Teixeira, S. H., Martini, L. A., and Boim, M. A. (2002). Vitamin D receptor gene polymorphism and bone mineral density in hypercalciuric calcium stone-forming patients. *Nephron* **90**, 51–57.
- Hodgkinson, A., and Pyrah, L. N. (1958). The urinary excretion of calcium and inorganic phosphate in 344 patients with calcium stone of renal origin. *Br. J. Surg.* **46**, 10–18.
- Lauderdale, D. S., Thisted, R. A., Wen, M., and Favus, M. J. (2001). Bone mineral density and fracture among prevalent kidney stone cases in the Third National Health and Nutritional Examination Survey. *J. Bone Mineral Res.* **16**, 1893–1898.
- Lemann, J., Jr., Worcester, E. M., and Gray, R. W. (1991). Hypercalciuria and stones. *Am. J. Kidney Dis.* **27**, 386–391.
- Martini, L. A., Cuppari, L., Colugnati, F. A. B., Szejnfeld, V. L., Schor, N., and Heilberg, I. P. (2000). High sodium chloride intake is associated with low bone density in calcium stone forming patients. *Clin. Nephrol.* **54**(2), 85–93.
- Pacifici, R., Rothstein, M., Rifas, L., Lau, K., Baylink, D. J., Avioli, L. V., and Hruska, K. (1991). Increased monocyte interleukin-1 activity and decreased vertebral bone density in patients with fasting idiopathic hypercalciuria. *J. Clin. Endocrinol. Metab.* **71**, 138–145.
- Pak, C. Y. C., Kaplan, R., Bone, H., Townsend, J., and Waters, O. (1975). A simple test for the diagnosis of absorptive, resorptive, and renal hypercalciurias. *N. Engl. J. Med.* **292**, 497–500.
- Scheinman, S. J. (1999). Nephrolithiasis. *Seminars Nephrol.* **19**, 381–388.
- Weisinger, J. R., Alonso, E., Bellorin-Font, E., Blasini, A. M., Rodriguez, M. A., Paz-Martinez, V., and Martinis, R. (1996). Possible role of cytokines on the bone mineral loss in idiopathic hypercalciuria. *Kidney Int.* **49**, 244–250.

every 700 people in several Caucasian populations in North America and Europe, FDB is one of the most common single-gene defects known to cause an inherited abnormality. The genetic disorder FDB is caused by a single base substitution (G to A) at nucleotide 10708 in exon 26 of the apoB100 gene (2p23.24), creating an arginine to glutamine substitution in the codon for amino acid 3500. Nearly everyone with FDB is of European descent; in most cases, the CGG-to-CAG mutation in the codon for amino acid 3500 in apoB is on a chromosome with a rare haplotype at the apoB locus, suggesting that most probands descended from a common ancestor who lived in Europe about 6750 years ago.

Except for a few cases in which tryptophan substitutes for arginine-3500, cysteine for arginine-3531, or tryptophan for arginine-3480, extensive searches have not revealed any other apoB100 mutations that cause defective receptor binding of LDL. All mutations are located within a stretch of 51 amino acids and result in the loss of an arginine. However, none of these mutated residues appears to be directly involved in binding to the LDLR given that site-directed mutagenesis has indicated that residues 3359 to 3369 in apoB100 are responsible for LDL-receptor interaction.

How do these mutations give rise to the defective LDLR binding? Immunoelectron microscopy studies have shown that the first 89% of apoB100 enwraps the LDL particle like a belt and that the carboxyl-terminal 11% constitutes a bow that crosses over the belt, bringing the carboxyl-terminal portion of apoB100 close to amino acid 3500 (Fig. 1). Recent data indicate that the bow is stabilized by an interaction between arginine-3500 and tryptophan-4369. It has also been proposed that arginine-tryptophan interactions are crucial during the conversion of VLDL to LDL for positioning apoB100's carboxyl



**Figure 1** The first 89% of apoB100 enwrapping the LDL particle like a belt, with the carboxyl-terminal 11% constituting a "bow" that crosses over the belt. The bow is stabilized by an interaction between arginine-3500 (R3500) and tryptophan-4369 (W4369). Site B (residues 3359–3369) is the LDL receptor-binding site.

tail to permit apoB100 on LDL to bind normally to the receptor. A disturbed refolding process gives rise to the two characteristics of FDB: a defective conformation of apoB100, as demonstrated by binding studies with monoclonal antibodies and C nuclear magnetic resonance analysis, and hypercholesterolemia due to ligand-binding-defective apoB100.

## CLINIC

The first disorder, familial hypercholesterolemia, is well studied. LDL cholesterol is elevated to about twice its normal levels in heterozygous FH. In the latter, familial defective apolipoprotein B, LDL is also elevated, but not to the same degree. Furthermore, the clinical consequences of heterozygous FH tend to be more severe than those of FDB. However, as they are presented in the clinic, the patients with FDB and FH are very similar. The first clinical evidence of the disease often is tendon xanthoma, that is, a tough, noncancerous, yellow-colored deposit of fat that develops beneath the skin. Xanthomas can occur anywhere on the body but are commonly found on the knees, elbows, hands, feet, buttocks, joints, and tendons. The tendency to develop xanthoma is quite variable; xanthomas may appear during adolescence or young adulthood, and some patients do not develop xanthomas even during their later years.

The main clinical complication is early manifestations of coronary heart disease (CHD). In patients with heterozygous disease, the risk of CHD at 30 years of age has been estimated at 5%; this risk rises to an estimated 50% by 50 years of age. Women develop CHD about 10 years later than do men, and as in the general population, other risk factors, such as tobacco smoking and hypertension, further increase the risk of CHD. Without proper treatment, only about 15% of males with heterozygous FH reach 65 years of age without an ischemic coronary event. With modern treatment, these risks seem to be substantially reduced. The severity of the CHD is determined by the presence of other risk factors, but there also seems to be a variation between families and between patients with different mutations of the receptor.

In patients with homozygous disease, the clinical course is more malignant. Xanthomas often develop during early childhood, and at least half of these patients have CHD at 20 years of age. The diagnosis of heterozygous FH can be detected already in cord blood or during childhood, when patients have elevated plasma cholesterol in relation to age and

gender. Beyond 20 years of age, plasma cholesterol usually is in the range of 8 to 12 mmol/L. Plasma cholesterol in homozygotes ranges between 15 and 25 mmol/L.

Heterozygous FH and FDB can, in most cases, be treated effectively with lipid-lowering drugs. The most potent cholesterol synthesis inhibitors, 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase inhibitors (statins), may reduce plasma cholesterol levels by as much as 50 to 60%. When therapeutic goals are not reached, statins may be combined with other drugs such as bile acid binding resins, nicotinic acid, and inhibitors of dietary cholesterol absorption.

### See Also the Following Articles

ABCA1 Defects • Abetalipoproteinemia • Dysbetalipoproteinemia and Type III Hyperlipidemia • Familial Low Cholesterol Syndromes, Hypobetalipoproteinemia • Hepatocyte Growth Factor

### Further Reading

- Boren, J., Lee, I., Zhu, W., Arnold, K., Taylor, S., and Innerarity, T. L. (1998). Identification of the low density lipoprotein receptor-binding site in apolipoprotein B100 and the modulation of its binding activity by the carboxyl terminus in familial defective apo-B100. *J. Clin. Invest.* **101**, 1084–1093.
- Boren, J., Ekström, U., Ågren, B., Nilsson-Ehle, P., and Innerarity, T. L. (2001). The molecular mechanism for the genetic disorder familial defective apolipoprotein B100. *J. Biol. Chem.* **276**, 9214–9218.
- Chatterton, J. E., Phillips, M. L., Curtiss, L. K., Milne, R., Fruchart, J. -C., and Schumaker, V. N. (1995). Immunoelectron microscopy of low density lipoproteins yields a ribbon and bow model for the conformation of apolipoprotein B on the lipoprotein surface. *J. Lipid Res.* **36**, 2027–2037.
- Gaffney, D., Reid, J. M., Cameron, I. M., Vass, K., Caslake, M. J., Shepherd, J., and Packard, C. J. (1995). Independent mutations at codon 3500 of the apolipoprotein B gene are associated with hyperlipidemia. *Arterioscler. Thromb. Vasc. Biol.* **15**, 1025–1029.
- Garcia, C. K., Wilund, K., Arca, M., *et al.* (2001). Autosomal recessive hypercholesterolemia caused by mutations in a putative LDL receptor adaptor protein. *Science* **292**, 1394–1398.
- Goldstein, J. L., Brown, M. S., Anderson, R. G. W., Russell, D. W., and Schneider, W. J. (1985). Receptor-mediated endocytosis: Concepts emerging from the LDL receptor system. *Annu. Rev. Cell Biol.* **1**, 1–39.
- Innerarity, T. L. (1990). Familial hypobetalipoproteinemia and familial defective apolipoprotein B<sub>100</sub>: Genetic disorders associated with apolipoprotein B. *Curr. Opin. Lipidol.* **1**, 104–109.
- Lund-Katz, S., Innerarity, T. L., Arnold, K. S., Curtiss, L. K., and Phillips, M. (1991). CC NMR evidence that substitution of glutamine for arginine 3500 in familial defective apolipoprotein B-100 disrupts the conformation of the receptor-binding domain. *J. Biol. Chem.* **266**, 2701–2704.
- Myant, N., Forbes, S., Day, I., and Gallaghers, J. (1997). Estimation of the age of the ancestral arginine3500-to-glutamine mutation in human apoB-100. *Genomics* **45**, 78–87.
- Pullinger, C. R., Hennessy, L. K., Chatterton, J. E., Liu, W., Love, J. A., Mendel, C. M., Frost, P. H., Malloy, M. J., Schumaker, V. N., and Kane, J. P. (1995). Familial ligand-defective apolipoprotein B: Identification of a new mutation that decreases LDL receptor binding affinity. *J. Clin. Invest.* **95**, 1225–1234.
- Rauh, G., Keller, C., Schuster, H., Wolfram, G., and Zöllner, N. (1992). Familial defective apolipoprotein B-100: A common cause of primary hypercholesterolemia. *Clin. Invest.* **70**, 77–84.
- Villegier, L., Abifadel, M., Allard, D., *et al.* (2002). The UMD-LDLR database: Additions to the software and 490 new entries to the database. *Hum. Mutat.* **20**, 81–87.
- Weisgraber, K. H., Innerarity, T. L., Newhouse, Y. M., Young, S. G., Arnold, K. S., Krauss, R. M., Vega, G. L., Grundy, S. M., and Mahley, R. W. (1988). Familial defective apolipoprotein B-100: Enhanced binding of monoclonal antibody MB47 to abnormal low density lipoproteins. *Proc. Natl. Acad. Sci. USA* **85**, 9758–9762.
- Yamamoto, A., Kamiya, T., Yamamura, T., *et al.* (1989). Clinical features of familial hypercholesterolemia. *Arteriosclerosis* **9**, 166–174.

weakness, and osteoporosis. Thromboembolic phenomena, including deep venous thrombosis and pulmonary embolism, may occur and can be part of the initial presentation of Cushing's syndrome.

## ETIOLOGY

Exogenous administration of glucocorticoids, mostly iatrogenic or (rarely) factitious (self-induced), accounts for the majority of cases of ACTH-independent Cushing's syndrome. Prednisone often causes Cushing's syndrome when it is used as an immunosuppressive and anti-inflammatory agent for a myriad of nonendocrine diseases. Cushing's syndrome can also be caused by other oral, injected, topical, and inhaled glucocorticoids as well as by medroxyprogesterone acetate. Surreptitious or factitious intake of a glucocorticoid can occur and should be considered during the early stages of the diagnostic process. The detection of synthetic glucocorticoids in the plasma or urine by high-pressure liquid chromatography (HPLC) is effective in excluding this disorder.

If exogenous glucocorticoid administration is excluded, the diagnosis of endogenous Cushing's syndrome is confirmed by demonstrating the presence of cortisol hypersecretion. Endogenous Cushing's syndrome is more common in women than in men, with an incidence of approximately 2 to 4 new cases per 1 million population per year in women. It can result from autonomous cortisol hypersecretion by cortisol-secreting adrenal tumors, adrenal hyperplasia (macro- and micronodular), or excess ACTH or (rarely) corticotropin-releasing hormone (CRH) secretion.

Once the diagnosis of endogenous Cushing's syndrome is made, the clinician must determine whether the patient has primary adrenal disease or an ACTH-secreting tumor (ACTH dependent or ACTH independent) (Table I). ACTH-dependent Cushing's syndrome accounts for approximately 85% of endogenous cases. In the great majority of these cases (80%), the cause is autonomous pituitary ACTH secretion and is referred to as Cushing's disease. In the remaining 20%, the source of ACTH or (rarely) CRH secretion is ectopic. The molecular pathophysiology of ACTH-secreting tumors remains elusive. Unlike the case of growth hormone-secreting adenomas, abnormalities of the G proteins are not frequent in corticotropinomas. Nonetheless, approximately 50% of these tumors overexpress the cytoplasmic form of the p53 tumor suppressor gene.

ACTH-independent Cushing's syndrome accounts for approximately 15% of endogenous cases and is

**Table I Classification of Hypercortisolism**

<b>Physiologic (adaptive) states</b>
Stress
Pregnancy
Chronic strenuous exercise
Malnutrition
<b>Pathophysiological states</b>
<b>Cushing's syndrome</b>
ACTH dependent (85%)
Pituitary adenoma (80%)
Ectopic ACTH (20%)
Ectopic CRH (rare)
ACTH independent (15%)
Adrenal adenoma
Adrenal carcinoma
Micronodular adrenal disease (rare)
Massive macronodular adrenal disease (rare)
Ectopic adrenal adenoma (rare)
<b>Psychiatric disorders</b>
Melancholic depression
Obsessive-compulsive disorder
Chronic active alcoholism
Panic disorder
Anorexia nervosa
Narcotic withdrawal
<b>Complicated diabetes mellitus</b>
<b>Glucocorticoid resistance</b>
<b>Obesity</b>

more often caused by benign cortisol-secreting adrenal adenomas, adrenocortical carcinomas, or (rarely) ectopic (extra-adrenal) adrenocortical tumors. Cortisol-secreting adrenal tumors are monoclonal in origin, but their pathogenesis remains largely unknown. Abnormalities of the p53 tumor suppressor gene or of the inhibitory subunit of G proteins and overexpression of insulin-like growth factor-2 (IGF-2) were identified in a subset of these neoplasms and might be implicated in their pathogenesis. It is noteworthy that, unlike the case of the thyroid-stimulating hormone (TSH) receptor in hyperfunctioning thyroid adenomas, the ACTH receptor, which also couples to G proteins and adenylyl cyclase, does not appear to be oncogenic.

Several genetic syndromes have been associated with adrenocortical tumors. These include the Li-Fraumeni syndrome (due to p53 mutations), the multiple endocrine neoplasia type 1 syndrome (due to mutations of the *menin* gene), the Beckwith-Wiedemann syndrome, and Carney's syndrome or complex. The latter is a rare autosomal dominant

disorder characterized by mucocutaneous pigmented lentiginos, blue nevi, and multiple endocrine tumors. Overt Cushing's syndrome caused by bilateral micronodular adrenal hyperplasia occurs in 25% of all patients with Carney's complex, but histological changes in the adrenal cortex are found at autopsy in nearly every patient. Testicular (Sertoli cell), pituitary, and thyroid neoplasms can also occur. Nonendocrine tumors such as cutaneous, mammary, and atrial myxomas, as well as psammomatous melanotic schwannomas, may also be present. Mutations of *PRKAR1A*, an apparent tumor suppressor gene that encodes the protein kinase A regulatory 1 $\alpha$ -subunit, are present in some families with Carney's complex.

### INITIAL LABORATORY SCREENING: BIOCHEMICAL DIAGNOSIS

Because signs and symptoms of Cushing's syndrome are not specific, appropriate documentation and verification of sustained endogenous hypercortisolemia is the cornerstone for the detection and diagnosis of this syndrome.

Determination of 24-h urinary-free cortisol (UFC) excretion with immunometric techniques or HPLC, gas chromatography coupled with mass spectrometry, or tandem mass spectrometry is often used as a screening test. Because one of four measurements may be normal in as many as 15% of patients with Cushing's syndrome, two or three 24-h samples should be obtained. The sensitivity of UFC in the evaluation of Cushing's syndrome is 95 to 100%, and its specificity is 94 to 98%. UFC corrected for body surface area is not altered by age, sex, body mass index (not exceeding 35–40 kg/m<sup>2</sup>), moderate alcohol intake, or smoking. Values consistently in excess of 300  $\mu$ g/day are highly suggestive of Cushing's syndrome. When assays for UFC excretion are not available, measurement of urinary 24-h 17-hydroxysteroids can be of help. These compounds include all cortisol metabolites with a 17-dihydroxyacetone side chain; thus, they give an indirect measure of the rate of cortisol secretion. Appropriate urine collection and determination of the creatinine excretion rate are essential for the interpretation of this test. In addition, UFC may be higher than the normal limits in up to 40% of patients with depression, anxiety disorder, obsessive-compulsive disorder, sleep apnea, morbid obesity, polycystic ovary syndrome, poorly controlled diabetes mellitus, familial resistance to glucocorticoids, hyperthyroidism, and alcoholism. Urinary cortisol excretion may be falsely elevated in those patients treated with

cortisone or hydrocortisone (topical or oral) and in women treated with vulvovaginal applications of hydrocortisone.

The 1-mg overnight dexamethasone suppression test (DST) is a simple screening tool for Cushing's syndrome, with reported sensitivity of 95 to 98% and specificity of 87%. During this procedure, a plasma cortisol measurement at 0900 h, following a single dose of 1 mg dexamethasone taken at midnight, is obtained. In children, the dose of dexamethasone that should be employed is 15 mg/kg body weight. An alternative screening method involves the determination of a single plasma or serum cortisol level following 48 h of dexamethasone (0.5 mg) every 6 h. Overall, the 2-day test and the overnight 1-mg DST can be done in outpatients and appear to have comparable sensitivities. Patients with obsessive-compulsive disorder, those with Alzheimer's dementia, and alcoholics experiencing acute alcohol withdrawal may show nonsuppressed (false-positive) overnight DST. Moreover, false-positive DSTs have been reported following weight loss and after sleep deprivation. Patients who are treated with medications that enhance the hepatic metabolism of dexamethasone, such as phenytoin, barbiturates, rifampicin, and carbamazepine, may also have false-positive results.

Patients with Cushing's syndrome have elevated nocturnal cortisol levels as compared with normal controls. The disruption of normal diurnal variation in serum cortisol secretion makes its measurement at midnight a simple screening tool for Cushing's syndrome. A measurement of plasma cortisol between 11 PM and midnight (late evening serum cortisol) reliably documents the loss of the normal circadian rhythm of cortisol secretion in patients with Cushing's syndrome. The sensitivity and specificity of this test in distinguishing patients with pseudo-Cushing's states from patients with Cushing's syndrome vary from 89 to 95% and from 97 to 100%, respectively. Normal individuals who have irregular sleeping patterns or who have recently crossed many time zones may have abnormal results. An alternative to the serum midnight cortisol test is the bedtime salivary cortisol test (which measures the free hormone fraction), which has a sensitivity of approximately 93% and a specificity of approximately 100% as well as the advantage of being an outpatient procedure.

### Distinguishing Mild Cushing's Syndrome from Pseudo-Cushing's States

The diagnosis of Cushing's syndrome may be difficult given that hypercortisolism can occur in several

disorders other than Cushing's syndrome (Table I). Patients with severe isolated obesity or obesity associated with the polycystic ovary syndrome can have increased cortisol secretion. As many as 80% of patients with major depressive disorders have increased cortisol secretion. The abnormal cortisol secretion possibly results from HPA axis hyperactivity and disappears after remission of depression. Chronic alcoholism is an uncommon cause of pseudo-Cushing's syndrome. Alcohol abstinence results in normalization of the hormonal abnormalities. Critically ill patients who are stressed may have increased cortisol secretion.

The dexamethasone-CRH stimulation test distinguishes patients with pseudo-Cushing's syndrome from those with Cushing's syndrome. Thus, most patients with Cushing's syndrome (80-90%) show inadequate suppression to low-dose dexamethasone (0.5 mg every 6 h for 2 days), in contrast to the normal responses of pseudo-Cushing's patients. In addition, patients with Cushing's disease (85%) have a "normal" or exaggerated ACTH response to CRH. When these tests are considered individually, their diagnostic accuracy in the differential diagnosis of mild hypercortisolism does not exceed 80%. However, when they are used sequentially, dexamethasone suppression (0.5 mg every 6 h for 2 days) and the ovine CRH (Dex-CRH) stimulation test can distinguish Cushing's disease from pseudo-Cushing's state. In the former, the pituitary corticotrope is appropriately restrained by glucocorticoid feedback and does not respond to CRH; in the latter, the corticotrophic tumor is generally resistant to this dose of dexamethasone and responds to CRH. The criterion used for the diagnosis of Cushing's disease is a 15-min cortisol level greater than 38 nmol per liter after the CRH injection. In a limited number of patients with Cushing's syndrome, sensitivity and specificity for the dexamethasone-CRH test were found to be 100%. However, the sensitivity was lower (90%) when comparing normal individuals to patients with Cushing's syndrome. Moreover, the accuracy of the dexamethasone-CRH test in patients with episodic hormonogenesis has not been tested. Therefore, this test should be reserved for those patients with

mild hypercortisolemia who fail to respond to 1 mg of overnight dexamethasone.

### Interpretation of Screening Tests

Cushing's syndrome is generally excluded if the response to the single-dose dexamethasone suppression test and the appropriately collected 24-h UFC measurements are normal. However, one should bear in mind that cortisol hypersecretion may be intermittent and periodic in 5 to 10% of patients with Cushing's syndrome of any etiology. Documenting loss of diurnal variation of plasma cortisol would support the diagnosis of Cushing's syndrome and vice versa. Obtaining more than single morning and evening blood draws increases the diagnostic value of the test because a significant variability of cortisol levels may be present. Another strategy involves close monitoring of the patient over the course of a few months. Although true hypercortisolism will persist and possibly produce further symptomatology, the hypercortisolism of pseudo-Cushing's states will be corrected spontaneously (Table II).

### DIFFERENTIAL DIAGNOSIS OF CUSHING'S SYNDROME: ACTH DEPENDENT VERSUS ACTH INDEPENDENT

Once the diagnosis of endogenous Cushing's syndrome has been made, the source of excess cortisol should be found. A stepwise sequential diagnostic approach is essential to successfully differentiate ACTH-dependent from ACTH-independent Cushing's syndrome and its subtypes. Such an approach involves the determination of serum ACTH levels, the response of the HPA axis to dynamic endocrine testing, and the performance of adequate imaging studies. The serum ACTH level distinguishes ACTH-dependent from ACTH-independent Cushing's syndrome, whereas the high-dose dexamethasone suppression test (8 mg overnight), the CRH stimulation test, and bilateral inferior petrosal sinus sampling (BIPSS) are used to

Table II Differential Diagnosis of Hypercortisolism

	UFC	1-mg DST test	Dex/CRH test	Clinical features (Cushingoid appearance)	Natural history
<b>Pseudo-Cushing's</b>	Mild elevation (< 3 × normal)	Variable	Cortisol suppressed	None or mild	Remission
<b>Cushing's</b>	Mild to severe (>3 × normal)	Variable	Cortisol not suppressed	Mild to severe	Progression

distinguish Cushing's syndrome due to ectopic ACTH secretion from Cushing's disease. Because they may compromise the diagnostic accuracy of the tests, all adrenal-blocking agents should be discontinued for at least 6 weeks prior to testing.

Early-morning measurement of blood ACTH and cortisol concentration distinguishes ACTH-dependent from ACTH-independent Cushing's syndrome. Circulating ACTH is typically suppressed in adrenal cortisol-secreting tumors, micronodular adrenal disease, and autonomously functioning macronodular adrenals. In contrast, plasma ACTH concentrations are normal or elevated in Cushing's disease and in those patients with ectopic ACTH and CRH secretion. Serum CRH levels should be measured if ectopic CRH secretion is suspected. Although the determination of plasma ACTH levels is often helpful, normal ACTH levels can be found in a subset of patients with Cushing's disease, those with ectopic ACTH secretion, and those with adrenal tumors. Similar overlap of baseline ACTH levels may also be seen in children and adolescents with Cushing's syndrome. The discrepancies in the diagnostic value of serum ACTH in the differential diagnosis of Cushing's syndrome are probably a reflection of different assay techniques.

### DIFFERENTIAL DIAGNOSIS OF ACTH-DEPENDENT CUSHING'S SYNDROME: PITUITARY VERSUS ECTOPIC

Approximately 80% of Cushing's syndrome cases are caused by an ACTH-secreting pituitary neoplasm and less often from an ectopic site, usually a thoracic or abdominal tumor. Ectopic ACTH secretion is a rare cause of Cushing's syndrome. Because of its low prevalence, the pretest probability of ectopic ACTH secretion is low. Therefore, one must consider confirming the diagnosis with at least two biochemical tests prior to proceeding with imaging studies intended to localize the tumor. The strategies used to differentiate the subtypes of ACTH-dependent Cushing's syndrome include biochemical tests (CRH stimulation test and high-dose dexamethasone suppression test), imaging tests (magnetic resonance imaging [MRI], computed tomography [CT], radio-nuclide scans), and venous angiography with sampling of the inferior petrosal sinuses for ACTH before and after CRH administration.

#### CRH Stimulation Test

Most patients with Cushing's disease respond to CRH with increases in plasma ACTH and cortisol, whereas

patients with ectopic ACTH production do not. A mean cortisol increase at 30 and 45 min of greater than 20% following a bolus infusion of ovine CRH (1  $\mu$ g/kg intravenously) has a sensitivity of 91% and a specificity of 88%. Similarly, an increase of mean ACTH concentrations at 15 and 30 min after CRH by at least 35% above the mean basal values achieves a sensitivity of 91% and a specificity of nearly 100%. The CRH test is rapidly substituting the classic tests of dexamethasone suppression and metyrapone stimulation because it is simple, brief, reliable, and more economical.

#### High-Dose Dexamethasone Suppression Test (Liddle's Test)

This test has been used extensively in the differential diagnosis of ACTH-dependent Cushing's syndrome. Patients with Cushing's disease are sensitive to glucocorticoid inhibition only at high doses of dexamethasone (2.0 mg every 6 h for 2 days). In contrast, patients with ectopic ACTH syndrome or cortisol-secreting adrenal tumors fail to respond to the same dose. The classic Liddle's criterion for a positive response consistent with Cushing's disease is a greater than 50% drop in 17-hydroxysteroid excretion on day 2 of high-dose dexamethasone (80% diagnostic accuracy). However, the diagnostic accuracy of the test increases to 86% by measuring both UFC and 17-hydroxysteroid excretion and by requiring greater suppression of both steroids (> 64 and 90%, respectively, for 100% specificity).

#### Overnight 8-mg Dexamethasone Suppression Test

A simple, reliable, and inexpensive alternative to Liddle's dexamethasone suppression test is the overnight 8-mg dexamethasone suppression test. The advantages are its outpatient administration and the avoidance of errors due to incomplete urine collection. The diagnostic accuracy of this overnight test may be similar to that of the standard Liddle's dexamethasone suppression test.

#### Imaging Evaluation

Imaging techniques can help to clarify the etiology of hypercortisolism. These usually include MRI of the pituitary and CT scanning of the adrenal glands. CT and MRI scans of the chest and abdomen and isotope scans are also employed when tumors secreting ectopic ACTH are sought.

CT is the preferred diagnostic method in the detection of adrenal tumors. Most adrenal adenomas causing hypercortisolism are larger than 2 cm in diameter. Tumors with diameters greater than 6 cm, those containing areas of necrosis, and those with local spread should raise the suspicion of malignancy.

MRI is as sensitive as CT for the visualization of both normal and enlarged adrenal glands. T2-weighted MRIs may prove to be helpful in differentiating between malignant and benign adrenocortical neoplasms. Bilateral enlargement of the adrenal gland with preservation of a relatively normal overall glandular configuration is observed in both Cushing's disease and ectopic ACTH production. Approximately 10 to 15% of patients with ACTH-dependent Cushing's syndrome demonstrate bilateral nodules (macronodular hyperplasia).

CT and MRI scans have largely superseded the need for iodocholesterol scans in the evaluation of patients with Cushing's syndrome. However, the iodocholesterol scan can occasionally be useful in distinguishing between ACTH-dependent (bilateral uptake) and ACTH-independent (unilateral uptake) macronodular adrenals or in localizing ectopic adrenal tissue (adrenal rest) or an adrenal remnant causing recurrent hypercortisolism after bilateral adrenalectomy.

MRI scanning is the imaging procedure of choice to visualize pituitary adenomas. The large majority of pituitary ACTH-secreting tumors are microadenomas with diameters less than 10 mm. The ACTH-secreting adenomas are best demonstrated on coronal T1-weighted images as foci of reduced signal intensity within the pituitary gland. However, on unenhanced scans, ACTH-producing adenomas are detected in only 40% of patients with Cushing's disease. An additional 15 to 20% of microadenomas are visualized with injection of contrast material (gadolinium-DTPA) and repeat T1-weighted coronal scan immediately after the injection (combined MRI sensitivity of 55–60%). Although still experimental, it appears that spoiled gradient recalled acquisition MRI might be superior to conventional postcontrast spin echo MRI for detection of adrenocorticotropin-secreting pituitary tumors. CT scanning with infusion of contrast demonstrates pituitary microadenomas in less than 20% of patients with bona fide lesions on subsequent surgery. Thus, pituitary CT should be performed if necessary only to demonstrate bony anatomy prior to transsphenoidal surgery.

If suppression and/or stimulation tests are suggestive of ectopic ACTH production, radiological imaging of the chest and abdomen should be undertaken.

ACTH-producing thymic carcinoids and pheochromocytomas are generally apparent by CT at the initial presentation. Patients with a negative CT should undergo MRI of the chest and abdomen using T2-weighted images. A body scan following the injection of the radiolabeled somatostatin analogue octreotide might be helpful and disclose the tumor site, which can be reexamined by CT or MRI. However, a significant number of small ectopic tumors may remain elusive. In these cases, reassessments of the chest with MRI every 3 to 6 months are indicated.

Distinguishing Cushing's disease from the ectopic ACTH syndrome frequently presents a major diagnostic challenge. Both pituitary microadenomas and ectopic ACTH-secreting tumors may be radiologically occult and may have similar clinical and laboratory features. Bilateral inferior petrosal venous sinus and peripheral vein catheterization with simultaneous collection of samples for measurement of ACTH before and after CRH administration is one of the most specific tests available to localize the source of ACTH production.

Venous blood from the anterior pituitary drains into the cavernous sinus and subsequently into the superior and inferior petrosal sinuses. Catheters are led into each inferior petrosal sinus via the ipsilateral femoral vein. Samples for measurement of plasma ACTH are collected from each inferior petrosal sinus and a peripheral vein both before and after injection of 1  $\mu\text{g}/\text{kg}$  body weight of ovine or human CRH. Patients with ectopic ACTH syndrome have no ACTH concentration gradient between either inferior petrosal sinus and the peripheral sample. A ratio greater than or equal to 2.0 in basal ACTH samples between either or both of the inferior petrosal sinuses and a peripheral vein is highly suggestive of Cushing's disease (95% sensitivity and 100% specificity). Stimulation with CRH during the procedure, with the resulting outpouring of ACTH, increases the sensitivity of BIPSS for detecting corticotropin adenomas to 100% when the peak ACTH central-to-peripheral ratio is greater than or equal to 3.0. Petrosal sinus sampling must be performed bilaterally and simultaneously because the sensitivity of the test falls to less than 70% with unilateral catheterization.

BIPSS is technically difficult and, like all invasive procedures, can never be risk free even in the most experienced hands. It should be reserved only for patients with a clear diagnosis of hypercortisolism and a negative or equivocal MRI of the pituitary or for those patients with a positive pituitary MRI but equivocal dynamic endocrine testing.

## UNUSUAL FORMS OF HYPERCORTISOLISM

### Periodic Cushing's Syndrome

Occasionally, cortisol production in Cushing's syndrome might not be constantly increased but instead may fluctuate in a "periodic" infradian pattern, ranging in length from days to months. This relatively rare phenomenon of periodic, cyclic, or episodic hormonogenesis has been described in patients with Cushing's disease, ectopic ACTH-secreting tumors, cortisol-secreting adrenal tumors, and micronodular adrenal disease.

Patients with periodic hormonogenesis may have consistently normal 24-h UFC and paradoxically "normal" responses to dexamethasone in the presence of clinical stigmata of Cushing's syndrome. To establish the diagnosis in such patients, several weekly 24-h UFC determinations for a period of 3 to 6 months may be necessary.

### Occult Ectopic ACTH Syndrome

Occult ectopic ACTH syndrome can mimic the clinical and biochemical behavior of Cushing's disease. In some circumstances, despite extensive localization studies, the source of ACTH secretion remains elusive. In such cases, the absence of a central-to-peripheral ACTH gradient before and after administration of ovine CRH in BIPSS rules out Cushing's disease. After excluding an ACTH-secreting pituitary adenoma, imaging studies with special emphasis on the lungs, thymus, pancreas, adrenal medulla, and thyroid should be obtained given that most described ectopic ACTH-secreting tumors have been found in these organs. Thymic vein sampling for measurement of ACTH concentrations can be of help in localizing the tumor to the thorax but not necessarily to the thymus.

### Cushing's Syndrome during Pregnancy

The physiological changes that occur during pregnancy may make the diagnosis of Cushing's syndrome more complicated. During normal pregnancy, human placental CRH mRNA transcription and CRH plasma levels increase significantly during the third trimester of pregnancy. A small progressive rise in plasma ACTH and a two- to threefold increase in plasma total and free cortisol also occur. UFC is also elevated during pregnancy, especially between the 34th and 40th weeks of gestation, and the suppression of cortisol by dexamethasone may be blunted.

Nonetheless, the diurnal rhythm in serum cortisol is maintained.

### Nonhypercortisolemic Cushing's Syndrome (Glucocorticoid Hypersensitivity Syndrome)

Increased tissue sensitivity to cortisol has been reported in two patients with signs of Cushing's syndrome, with normal or low cortisol secretion, in whom iatrogenic corticosteroid administration was excluded. However, the molecular mechanisms underlying this entity remain elusive because mutational analysis of the glucocorticoid receptor did not reveal any abnormalities.

### Primary or Sporadic Cortisol Resistance Syndrome

Glucocorticoid resistance is a rare familial or sporadic condition caused by mutations of the glucocorticoid receptor that results in generalized or partial end organ insensitivity to physiological glucocorticoid concentrations. The diagnosis should be considered in patients with elevated urine and plasma cortisol and ACTH levels that are not suppressed by dexamethasone. Plasma cortisol has a circadian rhythm similar to that of normal individuals, albeit at elevated concentrations, and responds normally to stress tests such as insulin-induced hypoglycemia. In addition, these patients frequently have elevated plasma and urinary adrenal androgens and mineralocorticoids, resulting in hyperandrogenism and hypertension.

## TREATMENT

### Cushing's Disease

The treatment of choice for Cushing's disease is selective transsphenoidal microadenectomy, a procedure with a cure rate approaching 95% on the first exploration by expert pituitary neurosurgeons. Failure of surgery at the first exploration may be followed by a repeat procedure with a 50% chance of cure. Success is defined as a drop of serum cortisol and UFC to an undetectable level during the immediate postoperative period. A successful outcome can also be predicted by lack of cortisol response to ovine CRH when the test is performed 7 to 10 days after surgery.

For patients who are not cured by transsphenoidal resection or in those where a tumor is not found, pituitary irradiation is the next treatment option. Although data on long-term follow-up are still scant,

stereotactic radiotherapy provides less irradiation to neuronal tissues and may offer similar or superior results when compared with conventional radiotherapy. Conventional linear accelerator with 4500 to 5000 rad delivered over a period of 6 weeks will correct the hypercortisolism in approximately 45% of adults and 85% of children, usually within 3 to 12 months of administration. Addition of the adrenolytic agent mitotane improves the correction of hypercortisolism produced by pituitary radiotherapy. Pituitary irradiation also decreases the occurrence of Nelson's syndrome in patients not cured by irradiation who require bilateral adrenalectomy. If irradiation fails to normalize cortisol secretion, adrenal enzyme inhibitors (medical adrenalectomy) can be used to ameliorate the hypercortisolism. Surgical bilateral total adrenalectomy with lifelong daily glucocorticoid and mineralocorticoid replacement therapy is the final definitive cure.

### Ectopic ACTH

Once the source of ectopic ACTH is identified, surgery is warranted. If surgery is contraindicated, medical therapy with adrenal enzyme inhibitor is the treatment of choice. Ketoconazole blocks adrenal steroidogenesis at several levels by inhibiting the C17–20 lyase, 11 $\beta$ -hydroxylase, 17-hydroxylase, and 18-hydroxylase. Reversible side effects, including elevations of hepatic transaminases and gastrointestinal irritation, may occur and may be dose limiting. In this case, metyrapone can be added to achieve normocortisolemia. Hypercortisolism is usually easily controlled within a few days with 200 to 400 mg of ketoconazole per day and/or 250 to 750 mg of metyrapone three times per day. Other blocking agents that may be used alone or in combination with ketoconazole and/or metyrapone include aminoglutethamide and mitotane. Etomidate, a hypnotic imidazole derivative, may also be used safely as a parenteral agent.

If the source of ACTH secretion is not localized, repeat searches for the tumor should be undertaken every 6 to 12 months. If by 2 years the tumor has escaped detection, bilateral adrenalectomy could be considered. This may have to be done earlier in developing children in whom ketoconazole and the other medications may interfere with growth and pubertal progression.

### Adrenal Tumors

Unilateral or bilateral benign tumors of the adrenals should be surgically resected. Laparoscopic adrena-

lectomy is the procedure of choice nowadays. Aggressive and repetitive open surgery provides the only chance for cure or prolonged survival in adrenal carcinomas. An anterior transabdominal approach with careful examination of the liver and pararenal structures should be performed. Mitotane may be added to maximally tolerated levels or toxicity when complete resection of the tumor is unsuccessful.

### Corticosteroid Replacement

Glucocorticoid replacement should be started after a successful pituitary adenomectomy or a complete resection of an ACTH-secreting ectopic or unilateral cortisol-producing adrenal tumor. Hydrocortisone should be replaced at a rate of 10 to 12 mg/m<sup>2</sup>/day by mouth, with appropriate increases in minor stress (2-fold) and major stress (10-fold) for appropriate lengths of time. The recovery of the suppressed HPA axis can be monitored with a short ACTH test every 3 months. When the 30-min plasma cortisol exceeds 18  $\mu$ g/dl, hydrocortisone can be discontinued. After a bilateral adrenalectomy, corticosteroid replacement will be necessary for life and includes both glucocorticoids and mineralocorticoids.

### CONCLUSIONS

The diagnosis of Cushing's syndrome is suspected on a clinical basis and requires the demonstration of pathological hypercortisolism. Screening tests include the determination of 24-h urine cortisol excretion, the 1-mg dexamethasone suppression test, and the measurement of late-night serum cortisol levels. In mild cases, distinction from the hypercortisolism of pseudo-Cushing's states may prove to be difficult. A dexamethasone-CRH test or close monitoring of the patient may be helpful. Excess endogenous glucocorticoid production can be ACTH dependent or ACTH independent. Distinction between ACTH-dependent and ACTH-independent hypercortisolism is made on the basis of basal and CRH-stimulated plasma ACTH determinations and adrenal CT. Most cases of primary adrenal Cushing's syndrome can be ruled out on the basis of undetectable basal and/or CRH-stimulated plasma ACTH and the absence of an identifiable lesion on imaging studies. ACTH-dependent Cushing's syndrome can then be differentiated on the basis of CRH testing and imaging studies. A discrete pituitary lesion on imaging and a standard CRH test with results consistent with such lesions are sufficient to proceed with transsphenoidal surgery. If

no visible pituitary lesion is present or if the CRH test is equivocal, simultaneous BIPSS with CRH administration is necessary to distinguish between a pituitary source and an ectopic source. Surgery is the treatment of choice for all types of Cushing's syndrome. Radiation therapy, radiosurgery, and medical adrenalectomy with adrenolytics and adrenal enzyme inhibitors are effective adjuvant treatments. Bilateral adrenalectomy is reserved for those patients who have failed all other forms of treatments.

### See Also the Following Articles

ACTH (Adrenocorticotrophic Hormone) • Collagen Metabolism Disorders • Glucocorticoid Resistance Syndromes and States • Hyperandrogenism, Functional • Mineralocorticoids and Mineralocorticoid Excess Syndromes • Pituitary Tumors, ACTH-Secreting • Stroke

### Further Reading

- Cushing, H. (1932). The basophil adenomas of the pituitary body and their clinical manifestations (pituitary basophilism). *Bull. Johns Hopkins Hospital* **50**, 137–195.
- Doppman, J. L., Nieman, L., Miller, D. L., Pass, H. I., Chung, R., Cutler, G. B., Jr., Schauf, M., Chrousos, G. P., Norton, J. A., Zirrsman, H. A., Oldfield, E. H., and Loriaux, D. L. (1989). Ectopic adrenocorticotrophic hormone syndrome: Localization studies in 28 patients. *Radiology* **172**, 115–124.
- Gold, P. W., Loriaux, D. L., Roy, A., Kling, M. A., Calabrese, J. R., Kellner, C. H., Nieman, L. K., Post, R. M., Pickar, D., Gallucci, W., Avgerinos, P., Paul, S., Oldfield, E. H., Cutler, G. B., Jr., and Chrousos, G. P. (1986). Responses to corticotropin-releasing hormone in the hypercortisolism of depression and Cushing's disease: Pathophysiologic and diagnostic implications. *N. Engl. J. Med.* **314**, 1329–1335.
- Koch, C. A., Pacak, K., and Chrousos, G. P. (2002). The molecular pathogenesis of hereditary and sporadic adrenocortical and adrenomedullary tumors. *J. Clin. Endocrinol. Metab.* **87**, 5367–5384.
- Nieman, L. K. (2001). Cushing's syndrome. In "Textbook of Endocrinology" (L. J. Degroot and L. J. Jameson, eds.), Vol. 2, pp. 1691–1715, W. B. Saunders, Philadelphia.
- Nieman, L. K., Oldfield, E. H., Wesley, R., Chrousos, G. P., Loriaux, D. X., and Cutler, G. B., Jr. (1993). A simplified morning ovine corticotropin-releasing hormone stimulation test for the differential diagnosis of ACTH-dependent Cushing's syndrome. *J. Clin. Endocrinol. Metab.* **77**, 1308–1312.
- Oldfield, E. H., Chrousos, G. P., Schulte, H. M., Loriaux, D. L., Schaaf, M., and Doppman, J. (1985). Preoperative lateralization of ACTH-secreting pituitary microadenomas by bilateral and simultaneous inferior petrosal venous sinus sampling. *N. Engl. J. Med.* **312**, 100–103.
- Papanicolaou, D. A., Yanovski, J. A., Cutler, G. B., Jr., Chrousos, G. P., and Nieman, L. K. (1998). A single midnight serum cortisol measurement distinguishes Cushing's syndrome from pseudo-Cushing states. *J. Clin. Endocrinol. Metab.* **83**, 1163–1167.
- Torpy, D. J., Chen, C. C., Mullen, N., et al. (1999). Lack of utility of (111)In-pentetreotide scintigraphy in localizing ectopic ACTH producing tumors: Follow-up of 18 patients. *J. Clin. Endocrinol. Metab.* **84**, 1186–1192.
- Tsigos, C., and Chrousos, G. P. (1996). Differential diagnosis and management of Cushing's syndrome. *Annu. Rev. Med.* **47**, 443–461.
- Yanovski, J. A., Cutler, G. B., Jr., Chrousos, G. P., and Nieman, L. K. (1993). Corticotropin-releasing hormone stimulation following low-dose dexamethasone administration: A new test to distinguish Cushing's syndrome from pseudo-Cushing's states. *J. Am. Med. Assoc.* **269**, 2232–2238.

tertiary centers is estimated to be 1.5%. The majority of men with Klinefelter's syndrome are undiagnosed. When suspected, the diagnosis may be confirmed by the presence of the chromosomal aberration in a karyotype analysis of peripheral blood lymphocytes. Buccal smears, which are examined for the presence of Barr bodies, can be used as a screening test.

Klinefelter's syndrome is characterized by small firm testes with absent spermatogenesis. Additional clinical manifestations, which are more variably present, are gynecomastia, eunuchoid body habitus, micropenis, and sparse body hair. XXY males have above average height from 5 years of age onward, with grown males tending to be tall. In addition, they have eunuchoidal skeletons, characterized by an arm span that exceeds height, narrow shoulders, and an elevated ratio of lower to upper body length. The increased length of the lower body is present before puberty; therefore, it is secondary to the karyotype itself rather than to hormonal levels. Gynecomastia results from a decreased ratio of testosterone to estrogen. The hips of these males tend to be wide from a more feminine fat distribution.

Boys with Klinefelter's syndrome enter normal puberty and tend to have low testosterone levels during late adolescence and adulthood. Classically, gonadotropins become elevated by 12 to 14 years of age, with a plateau of testosterone concentrations in the subnormal range by 16 years of age. Testicular biopsy reveals hyalinization and fibrosis of the seminiferous tubules with a reduced number of germ cells. Loss of seminiferous tubular function and Sertoli cells prevents feedback inhibition and leads to an increase in FSH secretion. Klinefelter's patients have impairment in Leydig cell function, leading to elevations in LH. Virtually all men with Klinefelter's syndrome are infertile; however, there are isolated reports of confirmed paternity. In 1998, assisted reproductive technology, involving testicular extraction and intracytoplasmic sperm injection from two 47,XXY Klinefelter's males, resulted in the delivery of healthy infants.

The mean intelligence level of 46,XXY males is slightly below that of controls, although there is a wide range of intelligence quotients (IQs). Verbal IQ is usually lower than performance IQ. For every additional X chromosome, there is an approximate 10- to 15-point drop in the IQ. Language deficits may be the cause of the academic difficulties and social withdrawal that occurs during the school years in some affected males. None of the chromosomal variants manifests any shortening of life expectancy.

Klinefelter's syndrome is associated with other medical conditions. XXY males have a rate of breast cancer that is 20 times higher than that of the general male population. They have higher rates of germ cell tumors and of autoimmune diseases such as systemic lupus erythematosus, rheumatoid arthritis, and Sjogren's syndrome. They also have higher rates of diabetes mellitus and hypothyroidism. They are prone to venous stasis, varicose veins, deep venous thromboses, and pulmonary emboli. Hypogonadism makes them susceptible to osteoporosis. There is a high incidence of taurodontism, characterized by an enlarged pulp chamber and thinning of the tooth surface, thereby making teeth prone to early decay.

Testosterone replacement enhances the development of secondary sex characteristics, causes a more masculine fat distribution, has beneficial effects on the libido, and promotes behavioral and work performance. Testosterone therapy does not affect testicular size, spermatogenesis, and fertility. It usually does not reverse gynecomastia, which can be corrected with surgical reduction.

## XX Males

The XX male syndrome occurs in 1 in 20,000 to 24,000 births. XX males are phenotypically similar to Klinefelter's males. They have small firm testes, azoospermia, infertility, and seminiferous tubule hyalinization. Gynecomastia is present in about one-third of the cases. These patients have elevated LH and FSH levels, whereas testosterone levels are low. However, men with this disorder differ phenotypically from Klinefelter's males in that they have below average height, they do not have intellectual impairment, and they have hypospadias more commonly. XX males occasionally have severe genital ambiguity.

Frequently, XX males have a genetic translocation involving the testis-determining gene, SRY. SRY, which encodes a DNA-binding protein, is normally located near the pseudo-autosomal region of the short arm of the Y chromosome. In most XX males, SRY has been incorporated into the X chromosome or another autosome. This is thought to result from aberrant recombination between the X and Y chromosomes. Two-thirds of XX males have detectable Y chromosomal sequences in the distal region of the X chromosome. Some XX males have a mosaic pattern of expression, with a few cell lines expressing 46,XY and the rest expressing 46,XX. Interestingly, all known XX males with hypospadias and genital ambiguity have been found to lack the SRY gene. Hence, the disorder is heterogeneous, and some cases remain

unexplained. Although most cases are sporadic, there are familial cases of 46,XX males, with all familial cases having genital ambiguity.

### Myotonic Dystrophy

Testicular atrophy occurs in 75% of men with myotonic dystrophy, an autosomal dominant disorder. Pathological examination of the testes reveals seminiferous tubule degeneration with Leydig cell preservation, although Leydig cell function is obviously impaired. As a result, patients usually have high LH and FSH levels along with low levels of testosterone.

### Inherited Disorders of LH and FSH

Mutations in the  $\alpha$ -chains of the gonadotropins have not been described in humans. Mutations in the  $\beta$ -chains of LH and FSH have been described. The hormonal profile in disorders caused by these mutations can often be suggestive of hypergonadotropic hypogonadism, with low levels of testosterone accompanied by high levels of one or both gonadotropins.

One LH  $\beta$ -mutation has been reported, with the affected male having delayed puberty, low testosterone, and arrested spermatogenesis. This patient harbored a missense mutation in the LH beta gene. The mutation permitted hormone synthesis and immunoreactivity; however, it prevented binding of the hormone to its receptor. Consequently, the patient's LH level was elevated by radioimmunoassay and his FSH level was normal. Treatment with human chorionic gonadotropin (hCG) caused increased testosterone levels, testicular enlargement, virilization, and an improved sperm count; however, the patient remained infertile. At 44 years of age, his gonadotropin levels were uniformly elevated.

In 1998, the first two men harboring homozygous mutations in the FSH beta gene were described. One of the men, who was found to have a two-base pair deletion in FSH beta, presented with delayed puberty, small testes, and azoospermia. He had undetectable serum FSH levels by radioimmunoassay, high serum LH levels, and low serum testosterone. Although the other man had undergone normal puberty, he had azoospermia. He was found to have a missense mutation in the FSH beta gene. This individual had absent FSH along with normal LH and testosterone.

### LH and FSH Resistance

The LH and FSH receptors belong to the G protein-coupled class of receptors that have

seven-transmembrane domains. Mutations of the LH and FSH receptors may result in a failure of the receptor to bind ligand or in a failure of the receptor activation cascade, which is mediated by cyclic AMP (cAMP).

LH resistance, caused by inactivating mutations of the LH receptor, is a rare cause of hypergonadotropic hypogonadism. Missense mutations, nonsense mutations, insertions, and deletions have been reported. These loss-of-function mutations lead to an entity known as Leydig cell hypoplasia, characterized by the complete or relative absence of Leydig cells. Patients who are homozygous for LH receptor mutations have an impaired or absent response to hCG or LH, the natural ligands for the LH receptor.

The clinical manifestations of LH receptor mutations are variable. In milder forms, patients may have varied phenotypes such as micropenis, severe hypospadias, hypogonadism without sexual ambiguity, and isolated infertility. In its most severe form, the syndrome is characterized by male pseudohermaphroditism, with female or ambiguous genitalia, low testosterone levels, high LH levels, absence of male secondary sexual characteristics, and a lack of response to hCG or LH challenge.

Only one complete loss-of-function mutation in the FSH receptor has been described; this mutation impairs receptor function by causing a marked reduction in ligand binding and a complete inhibition of signal transduction. Five men homozygous for the FSH mutation were identified in the Finnish population. All had normal testosterone levels, normal or slightly elevated LH levels, moderately elevated FSH levels, and slightly to markedly reduced testicular volume. Two of the men had successfully fathered two children each. However, all five men had abnormal semen parameters, including oligospermia, low sperm volume, and teratozoospermia. Partial functional impairment of the FSH receptor has been described in women but not in men.

## DEVELOPMENTAL CAUSES

### Cryptorchidism or Anorchia

Males with cryptorchidism or anorchia often have hypergonadotropic hypogonadism. Cryptorchidism is caused by a disruption of normal testicular descent. The most common sites for cryptorchid testes are the neck of the scrotum or just outside the external inguinal ring. Although the etiology of cryptorchidism is multifactorial, it is believed to be a defect in prenatal androgen secretion from either decreased

pituitary gonadotropin stimulation or impaired placental production of gonadotropins. Hormonal levels are usually normal at birth, and only later during infancy is the production of androgens impaired. Deficient androgen production with elevated gonadotropins may be caused by elevated intra-abdominal temperature or by an intrinsic testicular defect.

Cryptorchidism occurs in 4 to 5% of males at birth, although more than half of the cases resolve by 6 to 12 months of age. Men with a past history of cryptorchidism have a higher risk of infertility. In addition, these men have an approximately 5- to 10-fold risk of malignant testicular tumors compared with the general male population. Of note, there is an increased risk of malignancy in the contralateral testis in unilaterally cryptorchid men. Treatment of cryptorchidism with hCG or luteinizing hormone-releasing hormone (LHRH) has a success rate of 10 to 20%. However, surgical treatment is more effective in treating cryptorchidism and prevents testicular germ cell degeneration. Testosterone synthesis becomes impaired within a few months of birth; therefore, early correction is imperative. In infants, surgery is usually recommended between 1 and 2 years of age. Electron microscopic studies showing evidence of germ cell degeneration in humans at 6 to 12 months of age suggest that orchiopexy should be performed even earlier.

There is an association between low birthweight and cryptorchidism as well as between complex genital malformations and cryptorchidism. Cryptorchidism is the most common cause of azoospermia. Fully 89% of untreated cryptorchid patients develop azoospermia. The incidence of azoospermia in unilaterally cryptorchid patients is 13% regardless of whether they are treated. There is a correlation between maldescent of one testis and poor spermatogenesis in the contralateral testis.

Patients without palpable or visible testes may have either cryptorchidism or complete anorchia. To distinguish between the two, hCG may be administered intramuscularly. In cryptorchid males, a rise in plasma testosterone should ensue. In anorchid males, there will be no rise in testosterone levels. In addition, anorchid males have undetectable levels of anti-Müllerian factor during childhood. Therefore, in prepubertal males, measurable levels of anti-Müllerian factor indicate the presence of testicular tissue. However, during puberty, anti-Müllerian factor becomes undetectable in all males; therefore, this assay is not useful in peri- or postpubertal males. Whereas some cryptorchid males might not have elevated basal levels of gonadotropins, in anorchid males

basal gonadotropin levels are uniformly high. Anorchid males have otherwise normal male genital development. They are thought to have had testes during early fetal life that underwent degeneration during the 13th week of gestation, allowing normal genital development.

## ACQUIRED CAUSES

### Chemotherapy

Testicular damage in response to cytotoxic drugs was first described in 1948. Of 30 men studied at autopsy after being treated with nitrogen mustard, 27 were found to have azoospermia. The testes are more sensitive to damage from chemotherapy or radiation than are the ovaries. The germinal epithelium is more sensitive to the damaging effects of chemotherapy than are Leydig cells. Gonadal dysfunction from chemotherapeutic agents is variable. Patients may have any or all of the following: elevated LH and FSH levels, azoospermia, and evidence of testicular damage on pathology specimens. Testosterone is usually in the normal range. Testicular damage as a result of chemotherapeutic agents is usually dose dependent. The drugs known to be gonadotoxic are listed in [Table I](#).

### Radiotherapy

The testis is one of the most radiosensitive organs. The more immature a testicular cell, the more sensitive it is to radiation damage. Thresholds have been established for the amount of radiation that will cause specific testicular cell damage. Recovery of spermatogenesis depends on the presence of intact type A spermatogonia and, therefore, on the dose of radiation received. The Leydig cells are more resistant to

**Table I** Gonadotoxic Drugs

Alkylating agents	Busulfan
	Carmustine
	Chlorambucil
	Cyclophosphamide
	Lomustine
	Melphalan
	Mustine
Antimetabolites	Cytarabine
Vinca alkaloids	Vinblastine
Others	Procarbazine
	Cisplatin

Adapted from Howell and Shalet (1998).

radiation damage than is the germinal epithelium, although Leydig cell damage can occur with high radiation doses that are sometimes used in the treatment of testicular cancer.

### Mumps and Viral Orchitis

Mumps is the most common cause of viral orchitis, followed by echovirus, lymphocytic choriomeningitis virus, and group B arbovirus. The incidence of mumps orchitis has declined dramatically since the introduction of the mumps vaccination. It is rare in prepubertal boys; however, it develops in 15 to 35% of adolescent boys and men infected with the mumps virus. It usually presents with marked scrotal swelling and fever. Most patients have typical mumps parotitis preceding the scrotal manifestations. Orchitis results from direct invasion of the testes by the virus. Acutely, patients have elevated plasma FSH and LH along with decreased testosterone levels. After testicular swelling has resolved, gonadal function and the hormonal profile may return to normal. However, in one-third of men with orchitis, the affected testis may atrophy, causing persistently increased gonadotropins and low testosterone as well as azoospermia. In one-tenth of patients, both testes are affected. In some cases of unilateral involvement, atrophic changes may be observed in the contralateral and seemingly unaffected testis.

Atrophy is a result of destruction by the virus and ischemia from edema. On pathological examination, atrophic testes show progressive tubular sclerosis and hyalinization. Atrophy is usually present within 1 to 6 months after inflammation subsides. Approximately 50% of patients with unilateral orchitis have azoospermia. However, within 1 to 2 years, the sperm count returns to normal in three-quarters of those individuals. On the other hand, sperm counts in patients with bilateral orchitis return to normal in only one-third of patients.

### Aging

The effect of aging on the hypothalamic–pituitary–gonadal axis appears to be primarily at the level of the testicle. Necropsy studies have revealed an age-related decrease in the number of sperm as well as increased morphological sperm abnormalities. Normal aging in men is accompanied by a slow but statistically significant decline in androgens, particularly in free testosterone levels. There is estimated to be a 1 to 2% annual decline in levels of testosterone beginning during the fourth decade of life. This decline is not

present in every individual, and it varies considerably among individuals. Of note, the drop in testosterone levels in aging males is so gradual that the majority of elderly men still have testosterone levels within the normal range. The hormonal profile found in hypogonadal aging men is typical of hypergonadotropic hypogonadism. FSH tends to increase at a more rapid rate than does LH. Aging is also associated with increased testosterone binding to sex hormone-binding globulin (SHBG), which lowers the levels of free, or biologically active, testosterone.

Older men commonly report symptoms of depression, impaired memory, irritability, poor concentration, and decline in sexual interest and function. As men age, they experience decreasing muscle mass, increasing central adiposity, and increasing incidence of osteoporosis and bone fractures. Although these clinical signs and symptoms are typically present in young hypogonadal men, they are not clearly attributable to hypogonadism in older men. Hypogonadal young men respond well to testosterone replacement, with resolution or amelioration of their symptoms. In aging men with normal or low-normal levels of testosterone, the benefit of replacement therapy is much less obvious.

Fewer than 5% of men over 60 years of age are strictly hypogonadal. However, if one defines hypogonadism as a subnormal level of non-SHBG-bound testosterone, the percentage of men over 60 years of age who are hypogonadal may be as high as 50%. The goals of testosterone replacement in aging men include improved libido and erectile function, improved psychological well-being, increased lean body mass and muscle strength, preserved bone mass, and increased stamina. Although it is effective in restoring libido, testosterone supplementation in older men rarely reverses erectile dysfunction. In some studies, testosterone replacement has been shown to improve parameters of bone formation and resorption in aging men. However, other studies have shown no effect. Although testosterone therapy dramatically improves lean body mass in young hypogonadal men by 9 to 19%, it has a more modest effect in healthy older men with low to low-normal testosterone levels.

There is general concern regarding the effects of supplemental testosterone on disorders of the cardiovascular system and the prostate. Testosterone can adversely affect the cholesterol profile as well as other cardiovascular risk factors. Nonaromatizable testosterone preparations, including the oral agents oxandrolone and methyltestosterone, decrease high-density lipoprotein (HDL) cholesterol levels, increase low-density lipoprotein (LDL) cholesterol levels, and

have even been reported to lead to myocardial infarction and stroke in young men. Testosterone may increase cardiovascular risk by causing a propensity for vascular constriction via endothelin. It also promotes platelet aggregation and thrombus formation. On the other hand, most epidemiological studies indicate that higher testosterone levels in men correlate with a lower cardiovascular risk. This may be due to a secondary effect from comorbidities in elderly men with cardiovascular disease such as diabetes mellitus, cirrhosis, and renal failure, all of which are associated with hypogonadism. Studies of men treated with aromatizable testosterone agents have shown lowering of LDL cholesterol and no effect on HDL cholesterol. Furthermore, by decreasing visceral fat accumulation, testosterone may be a cardioprotective agent. The resultant cardiovascular effect in a given individual treated with testosterone replacement is unknown.

Testosterone, via conversion to dihydrotestosterone by the enzyme 5- $\alpha$ -reductase, has also been implicated as a growth stimulant for both normal and neoplastic prostatic tissue. There is widespread concern that testosterone replacement may exacerbate benign prostatic hypertrophy (BPH) and accelerate the growth of prostate cancer. However, although no causal link between testosterone administration and either prostate cancer or BPH has been demonstrated, caution is advised. Testosterone has been implicated in exacerbating cases of sleep apnea. It also stimulates erythropoietin production, thereby elevating hematocrit, hemoglobin, and red cell mass.

### Chronic Illness

Chronic illnesses are frequently associated with testicular dysfunction that mimics primary testicular failure. There have been reported cases of primary gonadal failure with hemochromatosis. Primary hypogonadism is seen in sickle cell disease and is thought

to result from direct testicular damage related to vascular trauma from sickling with its consequent impairment of blood flow. Chronic renal failure can cause primary testicular failure. Finally, AIDS may cause features of both primary and secondary hypogonadism. The mechanism for the association of poor testicular function with chronic disease has not been elucidated, but it may represent a nonspecific response of the testes.

### See Also the Following Articles

Agonadism, Male and Female • Aging and the Male Reproductive System • Delayed Puberty and Hypogonadism, Female • Delayed Puberty and Hypogonadism, Male • Hormone Replacement Therapy, Male • Kallmann's Syndrome and Idiopathic Hypogonadotropic Hypogonadism • Klinefelter's Syndrome • Müllerian Inhibiting Substance: New Insights • Testes, Embryology of • Undescended Testes

### Further Reading

- Ascoli, M., Fanelli, F., and Segaloff, D. L. (2002). The lutropin/choriogonadotropin receptor: A 2002 perspective. *Endocrine Rev.* **23**, 141–174.
- Baker, H. W. (1998). Reproductive effects of nontesticular illness. *Endocrinol. Metab. Clinics North America* **27**, 831–850.
- Gooren, L. J. (1998). Endocrine aspects of ageing in the male. *Mol. Cell. Endocrinol.* **145**, 153–159.
- Howell, S., and Shalet, S. (1998). Gonadal damage from chemotherapy and radiotherapy. *Endocrinol. Metab. Clinics North America* **27**, 927–943.
- Hutson, J. M., Hasthorpe, S., and Heyns, C. F. (1997). Anatomical and functional aspects of testicular descent and cryptorchidism. *Endocrine Rev.* **18**, 259–279.
- Smyth, C. M., and Bremner, W. J. (1998). Klinefelter syndrome. *Arch. Int. Med.* **158**, 1309–1314.
- Tenover, J. L. (1998). Male hormone replacement therapy including andropause. *Endocrinol. Metab. Clinics North America* **27**, 969–987.
- Visootsak, J., Aylstock, M., and Graham, J. M. (2001). Klinefelter syndrome and its variants: An update and review for the primary pediatrician. *Clin. Pediatr.* **40**, 639–651.

## PRECIPITATING FACTORS

Respiratory and urinary tract infections are common precipitants; others include myocardial infarction, pulmonary embolism, and acute pancreatitis. Certain antihypertensive drugs, including high-dose diuretics and beta-blockers, have been implicated; diuretics may exacerbate dehydration and potassium depletion. Other drugs include phenytoin, cimetidine, and chlorpromazine. High-dose corticosteroids have potent effects on carbohydrate and lipid metabolism, readily precipitating metabolic decompensation in predisposed individuals.

## DIAGNOSIS

Urinalysis reveals marked glycosuria with a “negative” or “trace” ketone reaction using semiquantitative nitroprusside-based test strips. Blood glucose concentration is markedly elevated, i.e., >500 mg/dl. Blood urea nitrogen is elevated and hematocrit is raised. Serum osmolality can be measured directly in the laboratory (by freezing-point depression), or the effective osmolality can be estimated from the formula

$$\text{Sodium(mEq/L)} + \frac{\text{Glucose(mg/dl)}}{18} = \text{mOsm/kg H}_2\text{O}$$

**Table I** Guide to Initial Treatment of Hyperosmolar Nonketotic Hyperglycemia in Adults

### Fluids and electrolytes

#### Volumes

- 1 liter per hour  $\times$  2–3, thereafter adjusted according to the degree of hydration and taking continuing polyuria into account. N.B. Use caution in patients with known renal disease or cardiovascular insufficiency; consider central venous pressure monitoring.

#### Fluids

- Isotonic saline (0.9%, 150 mEq/liter sodium chloride) initially.
- Hypotonic saline (0.45%, 75 mEq/liter sodium chloride) if serum sodium exceeds 150 mEq/liter (restrict to 1–2 liters in total and consider use of 5% dextrose with increased insulin if marked or worsening hyponatremia).
- 5% Dextrose at 1 liter every 4–6 h replaces saline when blood glucose has fallen to 250 mg/dl until patient is eating and drinking again.

### Potassium replacement

- No potassium chloride added to the first liter of fluid, unless initial plasma potassium <3.5 mEq/liter.
- Measure serum potassium every 2 h initially (see below: Other Measures)
- Thereafter, add dosages below to each liter of fluid:
  - If plasma  $\text{K}^+$ :
    - <3.5 mEq/liter, add 40 mEq potassium chloride (severe hypokalemia may require more aggressive replacement with careful monitoring of serum potassium concentration).
    - 3.5–5.5 mEq/liter, add 20 mEq potassium chloride.
    - >5.5 mEq/liter, add no potassium chloride to the infusion.

### Insulin

- 5–10 U/h (average 6 U/h in adults) by continuous intravenous infusion initially until blood glucose has fallen to 250 mg/dl. Thereafter, adjust rate (usually 1–4 U/h required in the absence of severe infection) together with dextrose infusion to maintain serum glucose between 100 and 200 mg/dl until patient is eating and drinking again.
- Thereafter, change to an appropriate subcutaneous insulin regimen.
- Review continuing need for insulin therapy 2–3 months after full recovery.

### Other measures

- Search for and treat precipitating cause (e.g., infection, myocardial infarction).
- Systemic hypotension usually responds to adequate fluid replacement with crystalloids.
- Pass bladder catheter if level of consciousness is impaired or no urine is passed within 2 h of the start of therapy.
- Continuous electrocardiographic monitoring may warn of hyperkalemia or hypokalemia (serum potassium should be measured hourly if <3.5 or >5.5 mEq/liter).
- Consider cranial computed tomography imaging to exclude other pathology (e.g., cerebral hemorrhage, venous sinus thrombosis) if the level of consciousness remains impaired following correction of hyperosmolar state.
- Treat acute thrombo-embolic complications with anticoagulant doses of heparin.
- Update clinical and biochemical progress using a purpose-designed flowchart.

Total body sodium is reduced, yet blood sodium concentration at presentation can be low, normal, or high, depending on the concomitant water deficit. Hyperglycemia may depress sodium concentration (correction factor: add 1.6 mEq sodium per 100 mg plasma glucose >100 mg/dl to the measured serum sodium concentration). Hypertriglyceridemia may result in pseudo-hyponatremia; plasma should be examined for turbidity. Dehydration is usually severe, with an average deficit in adults of approximately 8–10 liters. Arterial plasma pH is >7.30 and bicarbonate is >15 mEq/liter. However, a degree of acidosis may result from impaired renal excretion of H<sup>+</sup> ions allied to tissue hypoperfusion. Urinary losses of potassium are approximately 400–700 mEq in adults.

## TREATMENT AND COMPLICATIONS

### Initial Emergency Treatment

Successful management depends on early recognition of the diagnosis, general intensive care of the unconscious patient, and prompt treatment of underlying causes. The principal elements of initial treatment include the following: (1) intravenous rehydration, with saline; (2) insulin therapy, by continuous intravenous infusion; and (3) replacement of electrolytes, especially potassium (Table I).

### Longer-Term Management

Following recovery, in the medium to long term, oral anti-diabetic agents may control type 2 diabetes, thereby permitting safe withdrawal of insulin treatment. The need for insulin therapy should be reviewed 2–3 months after recovery. Steps should be taken to avoid recurrences wherever possible.

## Complications and Outcomes

Nontraumatic rhabdomyolysis in patients with greater degrees of hyperosmolarity may precipitate acute renal failure. There is a high frequency of thrombo-embolic complications. The average mortality rate is approximately 15%, higher than for patients with diabetic ketoacidosis. The acute outcome tends to be worse for individuals of advanced age or with serious comorbidity.

## See Also the Following Articles

Diabetes, Type 2 • Hypoglycemia • Insulin Secretion: Functional and Biochemical Aspects

## Further Reading

- American Diabetes Association (2001). Hyperglycemic crises in patients with diabetes mellitus. *Diabetes Care* **24**, 154–161.
- Cruz-Caudillo, J. C., and Sabatini, S. (1995). Diabetic hyperosmolar syndrome. *Nephron* **69**, 201–210.
- Delaney, M. F., Zisman, A., and Kettyle, W. M. (2000). Diabetic ketoacidosis and hyperglycemic hyperosmolar nonketotic syndrome. *Endocrinol. Metab. Clin. N. Am.* **29**, 683–705.
- Ennis, E. D., Stahl, E. J. B. V., and Kreisberg, R. A. (1994). The hyperosmolar hyperglycemic syndrome. *Diabetes Rev.* **2**, 115–126.
- Kitabchi, A. E., Umpierrez, G. E., Murphy, M. B., Barrett, E. J., Kreisberg, R. A., Malone, J. I., et al. (2001). Management of hyperglycemic crises in patients with diabetes. *Diabetes Care* **24**, 131–153.
- Krentz, A. J., and Nattrass, M. (2003). Acute metabolic complications of diabetes mellitus: Hyperosmolar non-ketotic syndrome. In "Textbook of Diabetes," (J. C. Pickup and G. Williams, eds.), pp. 32.15–32.19. Blackwell Science, Oxford, UK.
- Magee, M. F., and Bhatt, B. A. (2001). Management of decompensated diabetes: Diabetic ketoacidosis and hyperglycemic hyperosmolar syndrome. *Crit. Care Clin.* **17**, 75–106.
- Matz, R. (1997). Hyperosmolar nonacidotic diabetes (HNAD). In "Diabetes Mellitus: Theory and Practice" (D. Porte, Jr., and R. S. Sherwin, eds.), 5th ed., pp. 845–860. Elsevier, Amsterdam, The Netherlands.

**Table 1** Causes of Primary Hyperparathyroidism

Benign tumors
Single adenoma
Double adenomas
Multigland disease
Multiple endocrine neoplasia type 1 syndrome (MEN1)
Multiple endocrine neoplasia type 2A syndrome (MEN2A)
Familial hypocalciuric hypercalcemia (FHH)
Hyperparathyroidism–jaw tumor syndrome (HPT–JT)
Familial isolated hyperparathyroidism (FIHP)
Malignant tumors
Parathyroid carcinoma
Chemical induced
Lithium

which (>99%) are benign. The only environmental factor associated with PHPT (benign adenomas) is childhood exposure of the neck or upper chest to ionizing radiation (e.g., a common treatment for facial acne prior to 1950). The only medication known to cause PHPT is lithium. Chronic administration of lithium is associated with PHPT in at least 6% of patients who are treated with lithium for 15 years or more. Prospective studies have shown that during the first few years of lithium therapy, parathyroid hormone (PTH) levels rise, urinary calcium excretion falls, and serum calcium levels remain normal. Over time, true PHPT can develop, and the subsequent discontinuation of lithium does not reverse the hyperparathyroidism. In small surgical series, half of patients with lithium-associated PHPT have had single parathyroid adenomas and half of patients have had multigland disease.

### Single Adenomas

In most patients (80–90%), PHPT is the result of a solitary parathyroid adenoma. Parathyroid adenomas are benign monoclonal tumors. Most parathyroid adenomas are in the neck, but in some cases an abnormal gland may descend to an ectopic location (i.e., pseudoectopia), particularly when hyperparathyroidism is due to superior gland disease. Approximately 15% of all inferior parathyroid adenomas will be found within thymic tissue, and in rare cases (1–5%) an adenoma will be located ectopically in the thyroid, the mediastinum, or the angle of the jaw (undescended gland).

### Double Adenomas

Occasional patients have double adenomas, which can be asynchronous or parathyroid cysts.

### Multigland Disease

Approximately 5 to 15% of patients with PHPT have multigland disease, previously termed parathyroid hyperplasia, in which all four parathyroid glands (or more in cases where an extra gland is present) are abnormal. Parathyroid enlargement in multigland disease can be asymmetrical, and some abnormal glands may appear to be normal in size (up to 10:1 size ratios can be seen). Multigland disease occurs most commonly in patients with familial forms of PHPT that are inherited as distinct autosomal dominant traits. The age at diagnosis in hereditary PHPT is typically 20 to 35 years, three decades younger than seen in sporadic single-parathyroid adenoma patients. Five familial syndromes that are associated with PHPT have been described.

First, multiple endocrine neoplasia, type I syndrome (MEN1, MIM131100), is associated with tumors of at least 20 different tissues, most frequently of the parathyroids, pituitary, and pancreas. The parathyroid tumors are almost always benign, but tumors of the pancreas and gut (foregut carcinoid) are often malignant. The MEN1 gene located on chromosome 11 encodes *menin*, a nuclear factor that appears to function as a tumor suppressor. Nearly all patients (95%) with MEN1 will develop PHPT by 40 years of age. The average age at onset is 25 to 30 years, although patients as young as 8 years old have been reported. Genetic mutation analysis for the MEN1 gene located at 11q13 is available, but measuring the serum calcium level remains an effective, albeit late, approach to screening at-risk members of known kindreds.

Second, multiple endocrine neoplasia, type II (also termed MEN2A) syndrome (MIM 171400), is a tumor syndrome that includes medullary thyroid cancer, pheochromocytoma, and PHPT. The parathyroid tumors are almost always benign, with a mean age of diagnosis of 34 years. PHPT occurs in only 30 to 50% of patients with MEN2A but may be recognized prior to discovery of medullary thyroid cancer. MEN2A is caused by heterozygous-activating mutations in the RET protooncogene (MIM 164761) located at 10q11.2.

Third, familial hypocalciuric hypercalcemia (MIM 145980; also termed familial benign hypocalciuric hypercalcemia [FHH]) is an atypical form of PHPT in which adverse effects of hypercalcemia are rare and no specific treatment is required. Patients with FHH have hypercalcemia from birth, and in some cases affected newborns manifest life-threatening hypercalcemia, a variant previously termed “neonatal severe primary hyperparathyroidism.” Unlike other forms of

PHPT, urinary calcium excretion is low and so the risk of calcium kidney stones is not increased. In most cases (> 90%), FHH is caused by a heterozygous loss-of-function mutation in the gene encoding the calcium-sensing receptor located at 3q13.3–q21. This form of FHH has been designated as FHH1 (or FBH1) to distinguish it from much less common variants that have been mapped to 19p13.3 (termed FBH2) or linked to loci on 19q (FBH3). The 50% reduction in calcium-sensing receptors on the surface of parathyroid cells leads to secretion of PTH at higher ambient serum levels of calcium than normal, and serum levels of PTH are generally normal or only slightly elevated. A similar reduction in calcium-sensing receptors in the kidney results in increased renal reabsorption of calcium and accounts for the very low urinary calcium clearance (i.e., less than 0.01) that can be used to distinguish FHH from other forms of PHPT. In nearly all cases, patients with FHH do not develop adverse effects from either elevated levels of serum calcium or PTH. It is important to recognize when PHPT is due to FHH so that unnecessary and inappropriate surgery can be avoided.

Fourth, hyperparathyroidism–jaw tumor syndrome (HPT-JT, MIM 607393) is a rare syndrome of PHPT and ossifying jaw tumors (seen in dental radiographs). Unlike other hereditary forms of PHPT, as many as 15% of parathyroid tumors in HPT–JT are malignant. HPT–JT syndrome is caused by inactivating germline mutations in the *HPRT2* gene. Similar somatic mutations have been identified in approximately 5% of sporadic parathyroid adenomas with cystic features. This gene, which is located at 1q25–31, encodes a putative tumor suppressor protein that has been termed parafibromin.

Fifth, familial isolated hyperparathyroidism (FIHP) is another rare syndrome. Affected individuals do not have additional endocrine tumors or other syndromic features. Germline mutation of *HPRT2*, the gene responsible for the HPT–JT syndrome, has been identified in some cases of FIHP.

### Parathyroid Cancer

Malignant parathyroid tumors comprise less than 1% of cases of PHPT. These patients tend to be younger than patients with benign tumors and present with severe symptomatic hypercalcemia. Patients with parathyroid carcinoma often will have both skeletal and renal complications of PHPT, and many (50%) will have a palpable neck mass.

### CLINICAL PRESENTATION OF PHPT

In the United States and other developed nations, PHPT is most commonly diagnosed after the incidental discovery of symptomatic hypercalcemia on a multichannel chemistry panel. Therefore, at the time of diagnosis, few patients will manifest the specific clinical signs and symptoms that have been classically ascribed to PHPT and that are still common in patients who live in regions of the world where periodic measurement of the serum calcium level is not routine. These clinical features may be recalled through the following mantra: stones (nephrolithiasis), bones, (osteoporosis), abdominal groans (pancreatitis, peptic ulcer disease, constipation), and psychic overtones (depression, lethargy, poor concentration). Muscle weakness resulting from a neuropathic myopathy or arthritis resulting from pseudogout can also occur.

### Nephrolithiasis and Nephrocalcinosis

The incidence of kidney stones (nephrolithiasis) in PHPT has declined over the past 25 years and is now considered to be 15 to 20%. PHPT is an uncommon cause of kidney stones in general, and only about 5% of unselected patients with a renal stone will be found to have PHPT. Kidney stones in patients with PHPT are similar in composition to calculi from nonhyperparathyroid patients and may contain pure calcium oxalate, pure calcium phosphate, or a mixture of calcium phosphate and calcium oxalate. Most studies show, however, that calcium phosphate stones are more common in patients with PHPT than in the general stone-forming population. The factors contributing to stone formation include increased urinary excretion of calcium and phosphate, reduced urinary excretion of crystallization inhibitors (e.g., pyrophosphate, citrate, magnesium), hyperuricosuria, and alkaline urine due to mild renal tubular acidosis. Successful parathyroid surgery reduces the risk of recurrent kidney stones in approximately 90% of patients; persistent postoperative hypercalciuria is typically found in those patients who continue to form stones after surgery. Calcium phosphate crystals may form throughout the renal parenchyma, a process termed nephrocalcinosis. Nephrocalcinosis is less common than nephrolithiasis (e.g., 5%) and might not be associated with formation of stones or development of renal insufficiency. Nephrocalcinosis does not resolve after correction of PHPT.

## Bones

Excessive PTH action activates osteoclastic bone resorption, which when severe results in radiological manifestations that collectively are termed osteitis fibrosa cystica: subperiosteal resorption of the distal digits (particularly along the radial side of the middle phalanges), erosion of the distal one-third of the clavicles, salt-and-pepper skull, and bone cysts or brown tumors. Osteitis fibrosa cystica can cause pathological fractures and pain but is distinctly uncommon today. A characteristic pattern of skeletal demineralization, in which significant reductions in cortical bone density are seen regularly with relative preservation of cancellous bone density, is more common in patients with PHPT today. This results in a pattern of bone loss that contrasts sharply with that of postmenopausal osteoporosis, specifically a more significant loss of bone density at cortical sites (e.g., distal one-third site of the forearm) than at cancellous sites (e.g., spine). The hip (composed of cancellous and cortical bone) shows less consistent reductions in bone mass. The risk of fracture is increased in PHPT and overall is about 1.8-fold. Studies of PHPT have consistently found an increased risk of forearm fractures, but the risk of hip and spine fractures appears to be more variable. After surgical correction of PHPT, bone mass increases dramatically at all sites.

## LABORATORY EVALUATION

### Serum Calcium

Hypercalcemia is the primary biochemical hallmark of PHPT and is an essential diagnostic criterion. Several determinations of serum calcium may be required to confirm hypercalcemia because it may occur only intermittently. Routine determination of the serum calcium value measures the concentration of total serum calcium, which can sometimes provide misleading information. For example, hypercalcemia may result from elevated or abnormal circulating proteins or from hemoconcentration induced by prolonged application of the tourniquet. By contrast, total serum calcium can be normal in patients with PHPT who have abnormally low concentrations of serum albumin. Several algorithms have been proposed to correct the measured total serum calcium for a low level of serum albumin (e.g.,  $[(4.0 - \text{measured albumin in mg/dl}) \times 0.8]$ ), but none is entirely satisfactory. Under conditions where abnormal circulating proteins may reduce the reliability of determining the total serum calcium concentration, it is advisable to measure

the ionized serum calcium concentration directly. If both the total and ionized serum calcium concentrations are normal, it is difficult to make a definitive diagnosis of PHPT. Rare patients with PHPT and coexisting vitamin D deficiency (typically with 25-hydroxy-vitamin D levels  $< 20$  ng/ml) may have serum calcium levels that are consistently normal. In these patients, a diagnosis of PHPT might not be obvious until correction of vitamin D deficiency permits development of hypercalcemia.

### Parathyroid Hormone

The immunometric and chemiluminescent assays for intact PTH are currently the preferred formats for measuring circulating concentrations of biologically active PTH. These assays are highly sensitive and have replaced older immunoassays that were directed against midmolecule or carboxy-terminal epitopes of PTH. Assays for intact PTH provide outstanding diagnostic accuracy for PHPT, even in patients with renal insufficiency. Serum levels of intact PTH are elevated or high normal in nearly 90% of patients with PHPT. Although the upper limit of normal for intact PTH is considered to be 65 pg/ml, in individuals under 45 years of age the upper limit of normal has been determined to be closer to 45 pg/ml given that levels of PTH normally rise with age. Although assays for intact PTH had been considered specific for the full-length PTH 1–84 molecule, recent studies have determined that intact PTH assays also detect circulating PTH fragments that consist of the 7–84 sequence. These truncated fragments possess inhibitory activity and can comprise as much as 35% of circulating “intact” PTH immunoreactivity. Newer assays that appear to be specific for only PTH 1–84 (“whole” PTH assay) have been developed and may provide improved diagnostic discrimination. Intact (or whole) PTH is elevated or normal in PHPT and is suppressed or undetectable in other causes of hypercalcemia (e.g., malignancy, granulomatous disease, vitamin D intoxication). Importantly, currently available assays for PTH do not cross-react with parathyroid hormone-related protein (PTHrP), the cause of hypercalcemia in most cases of malignancy-associated hypercalcemia in solid tumors.

### Serum Phosphate, Chloride, Vitamin D, and Bone Turnover Markers

The serum phosphate level is usually in the low or low-normal range because of reduced renal tubular

reabsorption of phosphorus. This change in renal handling of phosphorus can be calculated as the tubular reabsorption of phosphorus (TRP) or as the renal tubular maximum reabsorption of phosphate per liter of glomerular filtration rate (TmP/GFR) from plasma and urine creatinine and phosphate measurements. A mild hyperchloremic metabolic acidosis (Cl  $>103$  mEq/L) is common. Serum levels of 25-hydroxy-vitamin D are usually normal or low normal, whereas serum levels of calcitriol (1,25-dihydroxy-vitamin D) are at the upper limit of normal in most patients with PHPT, with levels that are frankly elevated in one-third of patients. Most patients with PHPT have an elevated rate of bone turnover. This can be documented by measuring markers of bone resorption, such as serum C-telopeptide or N-telopeptide (best measured on fasting blood specimens) or urinary N-telopeptide (best measured using a first- or second-morning urine sample), which reflect the breakdown products of type I collagen. Markers of bone formation, such as the bone-specific alkaline phosphatase isoenzyme, type I procollagen peptide, and serum osteocalcin, are also elevated in many patients with PHPT. These markers can be elevated in patients with even mild PHPT who have normal serum levels of total alkaline phosphatase.

### Urinary Excretion of Calcium and Magnesium

A 24-h urine collection should be obtained for determination of calcium, creatinine, and sodium (Table II). Hypercalciuria is present in 40% of patients with PHPT and likely reflects increased gastrointestinal absorption of calcium as well as increased bone resorption. The renal tubular handling of calcium and sodium is related such that factors that promote natriuresis also

increase urinary calcium excretion. Therefore, patients with an elevated dietary intake of salt are at increased risk for hypercalciuria. If the urine calcium and sodium both are high, repeating the collection on a reduced sodium intake (i.e., less than 100 mmol sodium) may be useful. Patients with PHPT also have an increased renal clearance of magnesium. In contrast, patients with FHH have unexpectedly normal or low urinary calcium excretion (generally less than 100 mg/24 h) with a fractional excretion of calcium (i.e., renal calcium-to-creatinine clearance ratio) that is typically less than 0.01 and relative hypomagnesuria. Vitamin D deficiency is also associated with low urine calcium levels (e.g.,  $<100$  mg/24 h), and coexistence of vitamin D deficiency with PHPT can cause a low enough urinary calcium level to suggest FHH. In this case, repletion with vitamin D and repeating serum and 24-h urine calcium measurement can clarify the correct diagnosis.

## IMAGING STUDIES

### Skeletal Evaluation

Skeletal radiographs to search for signs of osteitis fibrosa cystica are not recommended except in patients with symptomatic bone pain or pathological fractures. Subperiosteal bone resorption is most readily detected in the distal phalanges using high-sensitivity X-ray film; in advanced cases, evidence of excessive bone resorption may also be present in the skull or distal clavicles. In most patients with PHPT, bone resorption is more subtle and evaluation of bone density in the hip, spine, and radius by dual-energy X-ray absorptiometry (DEXA or DXA) will be necessary to appreciate skeletal demineralization. PHPT causes a preferential loss of bone density at sites that are enriched in cortical bone (distal radius). DXA should be repeated annually in patients who are not referred for surgery.

### Urological System

In selected cases, a single abdominal radiograph or an ultrasound study can be useful in assessing the presence of nephrolithiasis or nephrocalcinosis. Although computed tomography (CT) is more sensitive for detecting small stones and nephrocalcinosis, the higher levels of radiation associated with CT compared with other techniques make its routine use imprudent.

### Parathyroid Imaging

Technetium-99m (Tc-99m) sestamibi with or without single photon emission computed tomography

**Table II** Differential Diagnosis of Elevated Parathyroid Hormone

Diagnosis	Serum calcium	Intact parathyroid hormone	Urine calcium
Primary hyperparathyroidism	H	H or N	H or N
Familial hypocalciuric hypercalcemia	H	H or N	L
Secondary hyperparathyroidism	N or L	H	H, N, or L

Note. H, high; N, normal; L, low.

(SPECT), ultrasound (US), CT, and magnetic resonance imaging (MRI) all are noninvasive techniques that can detect enlarged or abnormally functioning parathyroid tissue. Tc-99m sestamibi appears to be the most sensitive and specific of these techniques and is commonly used to identify a parathyroid lesion in patients who are candidates for minimally invasive parathyroid surgery. In contrast, parathyroid imaging does not appear to improve the outcome of a standard bilateral neck exploration when it is performed by an experienced operator on a patient who has not had previous parathyroid surgery. Tc-99m sestamibi imaging is typically performed using early and delayed images to take advantage of the fact that the thyroid gland discharges the radionuclide more rapidly than does the abnormal parathyroid gland. Alternatively, Tc-99m sestamibi can be combined with a second imaging agent, such as  $^{123}\text{I}$ , so that the thyroid gland image can be “subtracted” from the image obtained with Tc-99m sestamibi. Imaging with Tc-99m sestamibi can identify a pathological parathyroid gland in 70 to 80% of patients with PHPT.

## DIAGNOSIS

Elevated serum calcium and elevated or normal PTH are required for the diagnosis of PHPT (Table II). The intact PTH (iPTH), the most sensitive assay, but can be normal in up to 10 to 15% of patients with surgically proven PHPT. In these cases, a normal serum iPTH value in the presence of hypercalcemia indicates nonsuppressible or autonomous parathyroid tissue (e.g., adenoma). However, a serum iPTH level that is below 35 pg/ml is strong evidence that hypercalcemia is due to a nonparathyroid etiology (e.g., malignancy, granulomatous disease, vitamin D intoxication). The 24-h urine excretion of calcium may be elevated or normal in PHPT, but a calcium/creatinine clearance ratio that is less than 0.01 should raise suspicion of possible FHH.

## TREATMENT: WHEN TO REFER TO SURGERY

Surgical removal of hyperfunctioning parathyroid tissue remains the only definitive treatment of PHPT. However, patients with neither signs nor symptoms of PHPT, defined as asymptomatic PHPT, might not require surgery. The management of asymptomatic PHPT was addressed by participants

of a National Institutes of Health (NIH) Consensus Conference in 1990 and was subsequently updated in 2002. The NIH Consensus Conference produced a set of guidelines to help clinicians decide which patients should be referred for surgery and which patients could be placed under medical surveillance. The 2002 workshop led to a new set of guidelines for managing patients with PHPT:

- Serum calcium  $\geq 1$  mg/dl above normal
- Kidney stone or extreme hypercalciuria (24-h urine calcium  $> 400$  mg/day)
- Creatinine clearance reduced by 30% compared with age-matched normal individuals
- Reduced bone density compared with young normals at peak bone mass (T score  $< -2.5$ ) at hip, lumbar spine, or distal radius
- Patient age less than 50 years
- Patients for whom medical surveillance is neither desirable nor possible (e.g., patients for whom serial monitoring is impractical or unacceptable)

Other factors to consider when deciding whether to recommend surgical intervention include neuropsychological abnormalities (e.g., weakness, depression) and onset of menopause (when accelerated bone loss is anticipated). Using these guidelines, approximately half of patients with PHPT have one or more indications for surgery. In addition, surgery is always an acceptable choice for patients who prefer surgery to conservative follow-up after a definitive diagnosis of PHPT is made.

Minimally invasive parathyroid surgery under local or regional anesthesia has become a popular alternative to conventional surgery for many patients with PHPT, particularly those who are poor candidates for general anesthesia. The success of a minimally invasive surgical procedure is dependent on positive preoperative localization of a parathyroid lesion using a noninvasive imaging technique (e.g., Tc-99m sestamibi imaging). Intraoperative monitoring of serum PTH levels with a modified rapid iPTH assay is now performed at many centers and is considered by many as a useful adjunct to both minimally invasive and classical parathyroid surgery. A fall of at least 50% from baseline within 15 min of removal of hyperfunctioning parathyroid tissue is indicative of resolution of clinical signs in 95% of cases. In the case of solitary adenoma, iPTH is not only a biochemical substitute for histological examination but also a predictive test of normocalcemia.

## NONSURGICAL MANAGEMENT

### Chronic Medical Management

#### **General Measures**

Conservative management consisting of medical surveillance is appropriate for patients who are asymptomatic and have no surgical indications. Patients should be advised that dehydration could worsen hypercalcemia and so they should drink at least 2 L of water per day. Diuretics should be used with caution. Thiazide diuretics are particularly troublesome because they reduce renal calcium excretion through a direct effect on the distal renal tubule and can also cause dehydration. Although loop diuretics (e.g., furosemide) can induce calciuria at high doses, at typical oral doses their primary effect is to reduce extracellular volume. A reduction in extracellular volume can decrease the glomerular filtration rate and thereby reduce renal calcium excretion, further increasing the serum calcium level. The dietary calcium intake should not be restricted but should be limited to a moderate intake of 1000 mg/day. Patients should avoid very low-calcium diets that can worsen a negative calcium balance and that may further increase PTH secretion. Vitamin D supplementation of 400 IU/day can help to prevent bone loss and does not exacerbate hypercalcemia unless there is an underlying vitamin D deficiency; higher doses and other forms of vitamin D (e.g., calcitriol) should be avoided. Because immobilization can increase bone resorption and result in hypercalciuria or hypercalcemia, patients should be instructed to avoid protracted bed rest and to maintain (or increase if possible) their general level of activity.

#### **Bisphosphonates**

Bisphosphonates are analogues of inorganic pyrophosphate that adsorb to the hydroxyapatite matrix of bone and inhibit osteoclast-mediated bone resorption. Bisphosphonate compounds can be divided into two distinct pharmacological classes with different mechanisms of action depending on whether they contain a nitrogen atom(s) in their side chains. Non-nitrogen-containing bisphosphonates, including etidronate, clodronate, and tiludronate, are metabolized intracellularly to cytotoxic, nonhydrolyzable analogues of adenosine triphosphate (ATP). Nitrogen-containing bisphosphonates, including alendronate, ibandronate, pamidronate disodium, risedronate, and zoledronic acid, inhibit protein prenylation and are more potent inhibitors of osteoclast-mediated bone resorption. Zoledronic acid is a

new-generation, heterocyclic, nitrogen-containing bisphosphonate and is the most potent inhibitor of bone resorption identified to date. The more powerful oral and parenteral bisphosphonates can acutely lower serum and urinary calcium values and can increase PTH in patients with PHPT. Short-term use of oral bisphosphonates lowers serum calcium levels, but during chronic use the serum and urinary calcium levels tend to increase over time and generally return to their pretreatment elevations; PTH levels may remain higher than at pretreatment. Despite a lack of persistent beneficial effect on serum levels of calcium and PTH in PHPT, the newer, more potent oral bisphosphonates (e.g., alendronate, risedronate) demonstrate sustained efficacy in reducing bone resorption that can result in moderate increases in bone mass (3–8% over 2 years). The ability to prevent (or reverse) bone loss makes bisphosphonates useful adjuncts in the medical management of patients with PHPT who are not candidates for surgery.

#### **Estrogens, SERMs, and Progesterone**

Estrogen reduces bone turnover in both normal postmenopausal women and those with PHPT. Estrogen therapy can reduce serum calcium levels in patients with PHPT without increasing PTH secretion; this is in contrast to bisphosphonates, which increase PTH levels. In addition, therapy with estrogen can increase bone mineral density, with significant increases after 2 years of treatment in both the lumbar spine (4–5%) and femoral neck (3–4%). Estrogen therapy appears to be most useful in asymptomatic postmenopausal women with PHPT and mild osteopenia but with no other indications for surgery. However, the increased risks of stroke, myocardial infarction, thromboembolic disease, and breast cancer must be carefully considered before recommending long-term use of estrogen therapy. These concerns have raised interest in the potential use of selective estrogen receptor modulators (SERMs) as an alternative to estrogen in women with PHPT. Preliminary studies performed in small numbers of women with mild PHPT who had refused either surgery or estrogen have produced encouraging results. Treatment with raloxifene (60 or 120 mg/day for 12 months) increased bone mineral density by 3.4% at lumbar spine and by 2.5% at femur neck. Although total calcium levels decreased modestly in all patients after 12 months of treatment, serum levels of ionized calcium and iPTH returned close to baseline values after 12 months. Markers of bone resorption showed sustained reductions over the 12 months of treatment.

Progestins vary in their effect on bone metabolism in PHPT. Norethindrone, a progestin with relatively high androgenic activity, has effects that are similar to but less potent than those of estrogen, but less androgenic progestins (e.g., medroxyprogesterone acetate) have no effect on bone metabolism in PHPT.

### **Calcimimetic Agents**

Calcimimetic drugs are in development as a therapy for primary and secondary hyperparathyroidism. These agents directly reduce PTH secretion through their effect on the calcium-sensing receptor on the parathyroid cell. The calcimimetics increase the sensitivity of the calcium-sensing receptor to ambient concentrations of extracellular calcium, thereby increasing the suppressive effect of hypercalcemia on PTH secretion. Clinical trials of NPS568, a first-generation calcimimetic, and AMG073, a longer acting and more potent second-generation agent, have yielded very promising results as potential medical treatments for PHPT. These agents can decrease both serum calcium and PTH levels in a dose-dependent manner and, when used over long periods of time, can also reduce parathyroid size. Although calcimimetics are not yet generally available, they can be obtained on a compassionate need basis for patients with severe PHPT due to parathyroid cancer.

### **Phosphate**

In the past, oral phosphate was a common treatment for patients with PHPT who required medical therapy, particularly when there was a risk of recurrent nephrolithiasis. Phosphate reduces serum and urine calcium levels via suppression of 1,25-dihydroxyvitamin D, which in turn reduces calcium absorption from the gut. However, phosphate also increases PTH secretion, which is undesirable in patients with PHPT. Phosphate is contraindicated in patients with renal insufficiency and its dose is limited by the side effect of diarrhea. The frequent occurrence of gastrointestinal intolerance and the concern that patients who develop an elevated calcium-phosphate product will develop soft tissue (metastatic) calcification have reduced enthusiasm for long-term phosphate therapy, and the introduction of powerful oral and intravenous bisphosphonates has largely replaced phosphate as a therapy for PHPT. Phosphate is no longer recommended for long-term use, but in selected cases of PHPT, short-term phosphate

treatment (1–2 g/day in three or four divided doses) can be beneficial.

### **Urgent Management for Severe PHPT or Hypercalcemic Crisis**

Patients who have marked hypercalcemia (i.e., serum calcium level > 13 mg/dl) or milder hypercalcemia (i.e., serum calcium level > 12 mg/dl) with symptoms (e.g., altered mental status) or a shortened QT interval on EKG should be treated promptly. Intravenous infusion of 0.9% saline should be administered immediately at 200 to 800 ml/h, depending on the severity of symptoms and cardio-renal status, to maintain a normal to expanded state of hydration and to induce calciuresis. Hydration in conjunction with a loop diuretic requires careful monitoring to avoid dehydration. This treatment will be sufficient in many cases of mild to moderate hypercalcemia, but patients with severe hypercalcemia will require an additional intervention to inhibit bone resorption so as to eliminate the primary source of excess serum calcium. A wide variety of agents have been used to reduce bone resorption, including phosphate, mithramycin, calcitonin, gallium nitrate, and bisphosphonates. Administration of an intravenous bisphosphonate (e.g., pamidronate, zoledronic acid), coupled with adequate hydration, effectively normalizes serum calcium in the majority of patients with severe hypercalcemia. Pamidronate is given as a single infusion of 60 to 90 mg over 12 to 24 h in 500 ml of 0.9% saline. Zoledronic acid is more potent than pamidronate, and comparison studies between these two agents in hypercalcemia of malignancy show that patients treated with zoledronic acid had a more rapid normalization of serum calcium levels than did those in the pamidronate disodium group. Mean serum calcium levels at days 4, 7, and 10 were significantly lower in patients treated with 4 or 8 mg of zoledronic acid than in patients treated with pamidronate disodium, and the median time to relapse, median duration of response, and time to relapse all were longer in the zoledronic acid group than in the pamidronate disodium group. Zoledronic acid (4–8 mg) is infused over 5 to 15 min. Restoration of eucalcemia occurs within days and may last from 1 week to several weeks. Treatment may be repeated as needed. The most common adverse events are transient fever and hypophosphatemia, with these events typically being associated with bisphosphonate therapy. Neck surgery should be undertaken as soon as an elevated or

inappropriately normal iPTH is confirmed and the patient is stable.

### See Also the Following Articles

Bisphosphonates • Hypercalcemia and Hypercalcemia Treatment • Hypoparathyroidism • Hyperparathyroidism, Primary • Multiple Endocrine Neoplasia (MEN) Type 2 • Osteoporosis in Older Women • Parathyroid Cancer • Parathyroid Hormone (PTH) • SERMS (Selective Estrogen Receptor Modulators) • Vitamin D

### Further Reading

- Antonucci, D. M., and Shoback, D. (2002). Calcimimetics in the treatment of primary hyperparathyroidism. *J. Bone Min. Res.* **17**, N141–N145.
- Bilezikian, J. P., Potts, J. T., Fuleihan, G. E., Kleerekoper, M., Neer, R., Peacock, M., Rastad, J., Silverberg, S. J., Udelsman, R., and Wells, S. A. (2002). Summary statement from a workshop on asymptomatic primary hyperparathyroidism: A perspective for the 21st century. *J. Clin. Endocrinol. Metab.* **87**, 5353–5361.
- Hsu, S. C., and Levine, M. A. (2002). Primary hyperparathyroidism in children and adolescents: The Johns Hopkins Children's Center experience 1984–2001. *J. Bone Min. Res.* **17**, N44–N50.
- Marx, S. J., Simonds, W. F., Agarwal, S. K., Burns, A. L., Weinstein, L. S., Cochran, C., Skarulis, M. C., Spiegel, A. M., Libutti, S. K., Alexander, H. R., Chen, C. C., Chang, R., Chandrasekhara, S. C., and Collins, F. S. (2002). Hyperparathyroidism in hereditary syndromes: Special expressions and special managements. *J. Bone Min. Res.* **17**, N37–N43.
- Melton, L. J. (2002). The epidemiology of primary hyperparathyroidism in North America. *J. Bone Min. Res.* **17**, N12–N17.
- Silverberg, S. J., and Bilezikian, J. P. (2001). Primary hyperparathyroidism. In "Endocrinology and Metabolism" (K. L. Becker, ed.), 3rd ed., pp. 564–573. Lippincott Williams & Wilkins, Philadelphia.
- Strewler, G. J. (2000). Medical approaches to primary hyperparathyroidism. *Endocrinol. Metab. Clin. North America* **29**, 523–539.

Because the dietary intake of phosphate is variable, the kidney plays an important role in maintaining phosphate balance by adjusting excretion to match dietary intake. Approximately 80 to 90% of plasma phosphate is unbound and freely filtered by the glomerulus. Under conditions of a normal phosphate diet, approximately 80% of the filtered phosphate is reabsorbed and 20% is excreted in the urine. Virtually all of the reabsorption occurs in the proximal tubule, consistent with the exclusive localization of the phosphate transporter to this segment of the nephron.

A number of hormones and other factors have been shown to regulate the renal excretion of phosphate. However, the major factors appear to be the level of intake of phosphate and PTH. Dietary phosphate has a major influence on the urinary excretion of phosphate. Thus, an increase in phosphate intake decreases reabsorption, resulting in increased excretion of the anion. Conversely, low-phosphate diets increase renal reabsorption of phosphate such that urinary excretion may fall to zero. These adaptations in the reabsorption of phosphate can occur in the absence of measurable changes in plasma phosphate and in the absence of PTH. The factor(s) responsible for this remarkable adaptive change in the kidney to dietary phosphate intake is unknown.

In addition to diet, PTH plays a major role in the regulation of phosphate excretion. PTH has a direct effect on the proximal tubule to decrease phosphate reabsorption. Thus, parathyroidectomy decreases phosphate excretion, whereas administration of PTH increases phosphate excretion. Plasma phosphate levels, in turn, can regulate PTH secretion. Elevation of plasma phosphate increases PTH synthesis/secretion by two mechanisms. Phosphate can complex with calcium, resulting in a reduction of plasma calcium. The resulting fall in plasma calcium is a major stimulus for PTH secretion from the parathyroid glands. In addition, recent studies have shown that high plasma phosphate per se can increase PTH synthesis by the parathyroid gland. This feedback mechanism, which under normal physiological conditions functions to maintain plasma phosphate levels, plays a major role in the hyperphosphatemia of chronic renal failure.

## CAUSES OF HYPERPHOSPHATEMIA

There are a number of causes of hyperphosphatemia. For example, hereditary or acquired hypoparathyroidism, in which circulating levels of PTH are low or absent, results in the development of hyperphosphatemia, most likely due to increased renal reabsorption

of phosphate. Elevated levels of growth hormone, as seen in acromegaly, are also associated with elevated plasma phosphate levels due to increased renal absorption. Other causes of hyperphosphatemia include the release of phosphate from the large intracellular pool as a result of cell injury or cell death as occurs in tumor lysis syndrome or rhabdomyolysis.

However, perhaps the most common cause of chronic hyperphosphatemia, and the one with the most dire consequences for the patient, is that associated with chronic renal disease. When renal function is compromised either experimentally or by disease, there is a compensatory enlargement of the remaining nephrons and an increased rate of filtration per nephron. To remain in balance, the phosphate excretion per nephron must also increase. Increased levels of PTH appear to mediate the increased excretion of phosphate per nephron in early renal disease. Thus, normal plasma phosphate levels are maintained, but at the expense of elevated PTH levels. As renal function progressively declines, increasingly higher levels of PTH are needed to maintain phosphate homeostasis. In advanced stages of renal disease in which the kidney's excretory function is markedly reduced, the elevated levels of PTH are unable to maintain normal phosphate levels and hyperphosphatemia becomes evident. The consequences of hyperphosphatemia are numerous, with the most important being the development of secondary hyperparathyroidism, uremic bone disease, and the promotion of vascular and visceral calcifications (Fig. 1).

Secondary hyperparathyroidism is a common complication in renal failure patients. Hyperphosphatemia contributes to elevated levels of PTH by at least three mechanisms. First, phosphate by itself appears to increase PTH synthesis by the parathyroid gland by

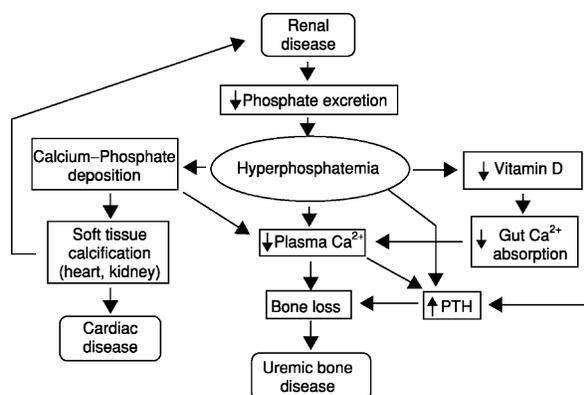


Figure 1 Development and consequences of hyperphosphatemia in chronic renal disease.

posttranslational mechanisms. Second, high levels of plasma phosphate can lead to the precipitation of calcium phosphate in soft tissues, resulting in a decrease in plasma calcium, which is a major signal for PTH release. Finally, vitamin D<sub>3</sub> is a major inhibitor of PTH gene transcription and also promotes intestinal calcium absorption. The kidney is the major source of the enzyme 1 $\alpha$ -hydroxylase, which is responsible for converting 25(OH)-vitamin D to the active form, 1,25(OH)<sub>2</sub>-vitamin D<sub>3</sub>. Not only is 1 $\alpha$ -hydroxylase activity decreased in renal disease because of the reduction in renal mass, but high levels of phosphate can also inhibit the enzyme activity. Therefore, hyperphosphatemia, either directly or indirectly, can attenuate major negative feedback mechanisms aimed at reducing circulating PTH. Furthermore, with low levels of vitamin D<sub>3</sub>, intestinal calcium absorption is impaired, and this can also contribute to the hypocalcemia. All of these abnormalities related to phosphate, calcium, PTH, and vitamin D<sub>3</sub> metabolism in chronic renal disease patients result in nearly all such patients having some degree of abnormal bone metabolism, generally known as uremic bone disease or renal osteodystrophy.

As mentioned previously, high levels of plasma phosphate can complex with calcium, resulting in the deposition of calcium-phosphate crystals in soft tissues. Recent attention has been directed toward the consequences of soft tissue calcification. It is now clear that hyperphosphatemia and an elevated calcium-phosphorus product (Ca  $\times$  P) can promote visceral and vascular calcification and are linked to increased cardiovascular mortality. Cardiovascular disease accounts for nearly 50% of all deaths in dialysis patients, a percentage that is markedly higher than that in the general population. A recent analysis of dialysis patients revealed that the relative risk of death increased in proportion to elevations in the Ca  $\times$  P product. In another study of dialysis patients, the prevalences of mitral and aortic valve calcification were markedly higher (44.5 and 54.0%, respectively) than those in the control populations (10.0 and 4.3%, respectively). There is also evidence that elevated PTH levels may contribute to cardiovascular morbidity and mortality through their effects on arteriolar wall thickening and myocardial interstitial fibrosis. All of these recent findings have led to recommendations for the tighter control of plasma phosphate, calcium, and PTH levels in patients with chronic renal disease, especially in the dialysis population.

## TREATMENT OF HYPERPHOSPHATEMIA

The preceding studies point to the importance of dietary phosphate intake and plasma phosphate in renal disease progression as well as the complex interaction among phosphate, PTH, calcium, and vitamin D in the development of secondary hyperparathyroidism and uremic bone diseases. Nonpharmacological approaches to the treatment of secondary hyperparathyroidism include surgical parathyroidectomy. In end stage renal failure patients, phosphate removal by dialysis is limited due to complex elimination kinetics; therefore, additional pharmacological means are needed to control plasma phosphate. Current and potential pharmacological approaches to the management of hyperphosphatemia and its consequences can be grouped into three categories: (1) modification or inhibition of dietary phosphate intake, (2) suppression of elevated PTH levels, and (3) vitamin D<sub>3</sub> analogues.

### Modification of Dietary Phosphate Intake

Long-term restriction of phosphate intake by dietary means is limited by the difficulty of maintaining adequate nutrition on such a diet in patients with chronic renal disease and by poor patient compliance. Accordingly, phosphate binders, which are compounds that bind ingested phosphate and so prevent its intestinal absorption, are required in the majority of patients for adequate control of plasma phosphate. Aluminum-containing phosphate binders (e.g., aluminum hydroxide) are effective at inhibiting intestinal phosphate absorption but are little used due to the accumulation of aluminum in the brain and bone as well as the resulting side effects of osteomalacia, dementia, and myopathy. Calcium-based binders (e.g., calcium carbonate, calcium acetate) have largely replaced the aluminum-based binders; they effectively bind intestinal phosphate and help to control plasma phosphate. However, the large amounts of calcium ingested with calcium-based binders can lead to hypercalcemia, especially in patients on concurrent vitamin D therapy, and may worsen soft tissue calcification.

Newer calcium- and aluminum-free phosphate binders that effectively control plasma phosphate without the side effects associated with calcium- and aluminum-containing binders have been developed. Sevelamer hydrochloride is a cross-linked polyallylamine approved by the Food and Drug Administration as a phosphate binder. In clinical

studies on hemodialysis patients, sevelamer hydrochloride reduced plasma phosphate and the  $\text{Ca} \times \text{P}$  product with a potency similar to that seen with calcium-containing binders. Sevelamer also significantly reduced plasma PTH and cholesterol levels. Additional phosphate binders based on ferric-containing compounds are under investigation.

### Inhibition of Phosphate Transport

A little-discussed but potential therapeutic target to control hyperphosphatemia is the development of compounds that inhibit the renal and/or intestinal phosphate transporters in a way similar to the inhibition of the  $\text{Na-K-Cl}$  transporter by furosemide. Theoretically, such compounds should both decrease intestinal phosphate absorption and increase renal phosphate excretion during the early phases of renal failure. When little functional renal mass is present in severe renal failure or end stage renal disease, compounds selective for the intestinal phosphate transporter would function in a way similar to that of phosphate binders. Little has been published in this area except for studies on nickeritrol and various monophosphonates. Nickeritrol is a pro-drug of nicotinic acid, which is used to improve peripheral circulation and plasma lipid profiles. Experiments in animals and hemodialysis patients have shown that nickeritrol reduced plasma phosphate levels by inhibiting intestinal phosphate absorption. The monophosphonate phosphonoformic acid (foscarnet), which was originally developed as an antiviral agent, has been shown to be a relatively weak but selective inhibitor of  $\text{Na}^+$ -dependent phosphate transport across renal and intestinal brush border membranes. Furthermore, fosfarnet is phosphaturic in rats when administered parenterally or orally. In a rat model of chronic renal failure, oral administration of fosfarnet for 48 h resulted in an increase in phosphate excretion and in a small but significant decrease in plasma phosphate. Longer term treatment with fosfarnet (8 weeks) in this animal model also increased phosphate excretion, but no consistent effects on plasma phosphate or on the rate of decline in renal function were observed. Although fosfarnet did not alter the course of renal failure in these studies and is nephrotoxic at high doses, the results are encouraging and suggest that inhibiting renal and especially intestinal phosphate transporters may be a viable therapeutic target to control hyperphosphatemia. However, longer term

studies with these compounds are needed to address issues of side effects and tolerability.

### PTH Suppression

The discovery and cloning of the extracellular calcium-sensing receptor (CaR) and elucidation of its role in divalent cation metabolism has opened up new therapeutic areas for the treatment of hyperparathyroidism and other mineral metabolism disorders. The CaR is a low-affinity G protein-coupled receptor present on parathyroid cells, renal cells, and elsewhere in the body. In the parathyroid gland, activation of the CaR by small increases in extracellular calcium results in a decrease in PTH secretion. Calcimimetics, compounds that enhance the affinity of the CaR for calcium, have been synthesized and are undergoing evaluation for the treatment of primary and secondary hyperparathyroidism. One of these compounds, NPS 568, has been shown to halt the progression of secondary hyperparathyroidism in uremic rats and to lower PTH levels in hemodialysis patients with mild or severe secondary hyperparathyroidism. Another compound of this class, AMG 073, has been shown to reduce plasma PTH, calcium, phosphate, and the  $\text{Ca} \times \text{P}$  product in hemodialysis patients with secondary hyperparathyroidism. Although further clinical assessment is necessary, these compounds may provide an additional option for the treatment of secondary hyperparathyroidism in chronic renal disease.

### Vitamin D Analogues

Treatment with vitamin  $\text{D}_3$  and vitamin D analogues has long been used for controlling various aspects of uremic bone disease. The most active metabolite of vitamin D, calcitriol ( $1,25(\text{OH})_2\text{D}_3$ ), has been used in the treatment of secondary hyperparathyroidism. Intravenous calcitriol effectively suppresses PTH levels in patients with secondary hyperparathyroidism. However, calcitriol also has potent effects to increase intestinal calcium and phosphate absorption and calcium mobilization from bone. Therefore, therapeutic doses of calcitriol often result in hypercalcemia and a worsening of hyperphosphatemia, especially in patients taking calcium-containing phosphate binders. Accordingly, newer analogues of vitamin D that retain the therapeutic effects on PTH levels with less calcemic and phosphatemic activity are

under investigation. Paricalcitol (19-nor-1,25-dihydroxy-vitamin D<sub>2</sub>) is a recently approved injectable calcitriol analogue that decreased PTH levels by approximately 60% in hemodialysis patients, with only a few incidences of hypercalcemia that occurred in patients whose PTH levels decreased to very low levels. Oral 1 $\alpha$ -hydroxy-vitamin D<sub>2</sub>, which has also been recently approved, decreased PTH levels with a low incidence of hypercalcemia or hyperphosphatemia. Finally, 22-oxacalcitriol, which is still under investigation, was as effective as calcitriol at reducing PTH levels in mice and rats but was less calcemic.

## SUMMARY

A major cause of chronic hyperphosphatemia is progressive renal disease. In end stage renal failure patients, appropriate levels of plasma phosphate cannot be achieved by dialysis without the aggressive use of phosphate binders and vitamin D analogues. With the growing awareness that abnormal phosphate metabolism and secondary hyperparathyroidism play a key role in the morbidity and mortality of these patients, new therapeutic approaches are needed. The recent introduction of noncalcemic phosphate binders and vitamin D analogues partially addresses this issue. Initial results with novel agents, such as calcimimetics, to reduce PTH levels are promising and await full

clinical testing. Other approaches that could prove to be fruitful include inhibition of intestinal and renal inorganic phosphate transporters and the development of PTH receptor antagonists.

## See Also the Following Articles

Kidney Disease in Diabetes • Parathyroid Hormone (PTH) • Vitamin D

## Further Reading

- Block, G. A., Hulbert-Shearon, T. E., Levin, N. W., and Port, F. K. (1998). Association of serum phosphorus and calcium  $\times$  phosphate product with mortality risk in chronic hemodialysis patients: A national study. *Am. J. Kidney Dis.* **31**, 607–617.
- Block, G. A., and Port, F. K. (2000). Re-evaluation of risks associated with hyperphosphatemia and hyperparathyroidism in dialysis patients: Recommendations for a change in management. *Am. J. Kidney Dis.* **35**, 1226–1237.
- Edwards, R. M. (2002). Disorders of phosphate metabolism in chronic renal disease. *Curr. Opin. Pharmacol.* **2**, 171–176.
- Llach, F. (1999). Hyperphosphatemia in end-stage renal disease patients: Pathophysiological consequences. *Kidney Int.* **56**, S31–S37.
- Malluche, H. H., and Monier-Faugere, M-C. (2000). Hyperphosphatemia: Pharmacologic intervention yesterday, today, and tomorrow. *Clin. Nephrol.* **54**, 309–317.
- Ritz, E., and Hergesel, O. (2001). Compounds in development to combat hyperphosphatemia. *Expert Opin. Invest. Drugs* **10**, 2185–2190.

diuretics, ACE inhibitors, angiotensin receptor antagonists, and vasodilators and is decreased by beta-blockers and centrally acting agents. Treatment with these drugs should be interrupted before renin levels are assessed, although this may be difficult in patients with severe hypertension. In such cases, alpha-blockers and calcium antagonists may be continued because they have little effect on renin release.

### Acute Renin-Suppression Tests

Acute renin-suppression tests interrupt the renin-angiotensin system, showing that this system is involved in the control of hypertension. The ACE inhibitor captopril is generally administered orally due to its rapid onset of action (20 min) and maximum length of action (1–3 h). BP should consistently decrease in patients with renin-dependent hypertension.

### When Is It Useful to Measure Renin?

In patients with essential hypertension, renin determination may help doctors to select the most effective first-line treatment. Patients in the highest renin index quartile show the greatest response to beta-blockers and those in the lowest quartile respond better to diuretics. However, considering the fluctuations in renin levels in individual patients and the relatively high cost, the therapeutic value of measuring plasma renin during routine investigation of mild to moderate essential hypertension is limited (see [Table I](#)). Renin should be measured in patients prone to secondary hypertension. Such patients can be identified by their medical history, signs, and symptoms, or the presence of hypokalemia, hyperazotemia, or asymmetrical kidneys (see [Table III](#)). Renin should also be measured in the small proportion of patients with severe untreated hypertension or those in whom hypertension is resistant to conventional treatment. In such cases, an unsuspected form of secondary hypertension or a pronounced drug-induced stimulation of the renin-angiotensin system may be found. Such findings may indicate that additional etiologic evaluation is required or that an ACE inhibitor, an angiotensin receptor antagonist, or a beta-blocker should be added to the treatment regimen.

## RENIN-DEPENDENT FORMS OF HYPERTENSION

### Primary Hypertension

High renin levels are found in approximately one in six patients with benign essential hypertension

(see [Table II](#)). The purpose and efficiency of screening for high renin hypertension in this setting is discussed above. Most patients with malignant hypertension have high renin levels. The renin concentration is very high in patients with hyponatremic hypertensive syndrome, leading to hyperangiotensinemia with thirst, polyuria, and hyponatremia and secondary aldosteronism with potassium wasting and hypokalemia. Malignant hypertension may be primary or, as in 45% of cases, the consequence of an underlying renal or adrenal disease.

### Renal Artery Stenosis

Renal ischemia associated with renal artery stenosis (RAS) is the most frequent condition occurring with renin-dependent hypertension. Uni- or bilateral RAS can cause renovascular hypertension, a form of hypertension that can be reversed by nephrectomy or renal revascularization. Approximately two-thirds of stenoses are due to atherosclerosis, usually in patients over the age of 50 years. One-third are due to fibromuscular dysplasia, mostly in young female patients. The standard procedure for diagnosing RAS is catheter angiography. For screening purposes, this invasive procedure can be replaced by computed tomography (CT) angiography, magnetic resonance angiography, or duplex Doppler sonography. These techniques estimate the frequency of RAS at 1% in unselected patients with hypertension and at 10 to 30% in patients with drug-resistant hypertension or with atherosclerosis elsewhere. The standard for defining renovascular hypertension is a favorable BP outcome following revascularization. This definition is retrospective and the actual frequency of the condition is unknown because hypertension reversal is dependent on several parameters, such as the patient's age, hypertension duration, stenosis etiology and grade, parenchymal consequences of hypertension, the feasibility of revascularization, and the risk of restenosis.

Considering the risks associated with renal artery surgery or percutaneous angioplasty, attempts have been made to design tests to select RAS patients who have truly renin-dependent hypertension. In such patients, the captopril test is expected to induce a sharp decrease in BP. The captopril test, with concurrent determination of plasma renin levels, is predicted to induce a homeostatic increase in renin secretion from the stenotic kidney. Although the captopril test has been analyzed in numerous patients at risk of having renovascular hypertension, a complete analysis is not possible because of multiple inconsistencies in patient selection, standards, test procedures, and cutoff points.

Table I When to Measure Renin?

Setting	Aim	Comments
Mild or moderate untreated hypertension	Determination of renin subgroup for prognostic or therapeutic purposes	Classification of individual patients into low- or high-renin subgroups probably not cost-effective in this setting
Severe untreated hypertension (diastolic BP consistently >110 mm Hg) and/or hypokalemic hypertension	Guidance for subsequent etiologic investigations	High renin levels suggest that CT angiography, magnetic resonance angiography, or duplex Doppler sonography should be performed, whereas low renin levels suggest adrenal investigations
Drug-resistant hypertension	Guidance for antihypertensive treatment adaptation	Discontinue current treatment (calcium channel antagonists and alpha-blockers may be continued) and determine plasma renin levels; high renin levels suggest that a beta-blocker, an ACE inhibitor, or an angiotensin II receptor antagonist should be added to the previous treatment regimen
Hypertension with renal failure	Guidance for antihypertensive treatment adaptation	Low or normal renin levels suggest that the daily dose of loop diuretic should be increased; high renin levels suggest that an ACE inhibitor should be added
Hypertension with renal artery stenosis or unilateral small kidney or kidney tumor	Evaluation of asymmetrical renin secretion before nephrectomy or renal artery angioplasty	A RVR ratio $\geq 1.5$ is generally associated with a favorable BP outcome following nephrectomy or renal artery angioplasty

Simultaneous determination of renin in both renal veins has been used to predict BP outcome following revascularization in patients with unilateral RAS. If the renal vein renin (RVR) ratio (i.e., the ratio between the RVR levels on the stenotic and nonstenotic sides) exceeds 1.5, the stenosis is assumed to cause renin-dependent hypertension amenable to surgical or angioplastic cure. The test may be performed after acute oral captopril administration. Postcaptopril renal scintigraphy is a less invasive alternative to RVR determination and is more accurate than the captopril–renin test. Postcaptopril scintigraphy is the most appropriate test in unilateral RAS. RVR determinations should be performed only in complex cases such as branch stenoses or renovascular disease associated with focal renal parenchymal infarction.

### Other Conditions with Renal Ischemia

A few conditions characterized by chronic or acute intrarenal ischemia may also induce renin-dependent hypertension. These include systemic diseases such as polyarteritis nodosa and scleroderma, in which BP and renal outcomes are greatly improved by ACE inhibitors, and segmental renal infarction. Segmental renal infarction may result from renal embolism, renal artery thrombosis or dissection, or *in situ* thrombosis in cases with coagulation disorders. It usually presents as an abrupt-onset malignant hypertension with hyponatremic hypertensive syndrome. The condition

may revert spontaneously to normotension because the ischemic renal tissue, which releases large amounts of renin, may subsequently turn into a silent renal scar.

### Renal Tumors and Cysts

Primary reninism is a rare, purely renin-dependent form of hypertension. It may be caused by juxtaglomerular cell tumors, malignant kidney tumors (nephroblastoma, adenocarcinoma), or epithelial or soft tissue tumors. Primary reninism is associated with very high plasma renin concentrations and by the absence of RAS and intrarenal ischemia (as determined by renal angiography). A small, hypodense renal cortical tumor (juxtaglomerular cell tumor) or a larger renal or extrarenal tumor is then detected by CT scan. BP usually drops during ACE inhibition and hypertension can be cured by tumor resection, although it may recur if metastases develop in malignant cases. In the 43 reported cases of juxtaglomerular cell tumors, hypertension was severe [median BP (range) 210/114 mm Hg (160–260/115–170 mm Hg)] and associated with hypokalemia [2.9 mmol/liter (2–3.9 mmol/liter)]. PRA or renin concentration was markedly elevated in all cases (medians cannot be provided because renin was reported as PRA or active renin concentration, before or after the conversion of prorenin into renin). CT scanning revealed small renal tumors [diameter 25 mm (8–50 mm)].

**Table II Etiology, Presentation, and Key Findings in Renin-Dependent Hypertension**

<b>Underlying disease</b>	<b>Usual presentation</b>	<b>Key tests and findings</b>
<b>Primary hypertension</b>		
Benign	Nonspecific; 10–15% of patients with primary hypertension belong to the high-renin subgroup	Larger than average response to beta-blockers or ACE inhibitors
Accelerated or malignant	Headache, thirst, weight loss, diastolic BP usually >140 mm Hg with hypokalemia, azotemia, proteinuria	Funduscopy: hemorrhages, exudates, papilledema; an underlying renal disease is present in ≈50% of cases
<b>Renal ischemia</b>		
Renal artery stenosis	Recent, progressive, and/or severe hypertension; hypertension in young females (fibromuscular dysplasia) or in patients with angina or arteritis (atherosclerosis); hypertension with ACE inhibitor-induced rise in plasma creatinine	Renal angiogram: renovascular (curable) hypertension is probably present if reduction in artery diameter (1) exceeds 75% or is (2) between 50 and 75% with a positive captopril test, a lateralized captopril renography, or a RVR ratio $\geq 1.5$
Renal infarction	Hyponatremic hypertensive syndrome: lumbar pain and hematuria followed by acute hypertension, polyuria, hypokalemia, hyponatremia, and very high renin concentrations	Segmental renal infarction is detected by CT scan and its cause is determined by renal angiogram (embolism, thrombosis, renal artery dissection, or occlusion)
Polyarteritis nodosa	Systemic necrotizing vasculitis	Widespread microaneurysms on renal angiogram
Systemic scleroderma	Skin thickening, Raynaud's phenomenon, progressive hypertension, and azotemia	Clinical presentation; ACE inhibitors greatly improve scleroderma crisis
<b>Renal tumors and cysts</b>		
Juxtaglomerular cell tumor	Severe hypokalemic hypertension with very high prorenin and active renin levels	Normal renal angiogram with a small cortical tumor on CT scan
Other renal renin-producing tumors	Hypertension with renal cell carcinoma	Primary reninism is present and BP returns to normal following nephrectomy in ≈5% of cases
Compressive cysts and tumors	Hypertension with ultrasound scan evidence of a large cyst or tumor	Improvement in BP following cyst drainage or tumor resection may be expected if RVR ratio >1.5
Polycystic kidney	Family history of renal cysts, hypertension, and azotemia	Ultrasound scan evidence of >3 bilateral cysts in a person with a positive family history
Unilateral nonvascular small kidneys	Hypertension with asymmetrical kidneys and frequently a history of recurrent urinary tract infection	Pyelographic or ultrasound scan evidence of pyelonephritic scarring, reflux nephropathy, and/or segmental hypoplasia
<b>Extrarenal tumors</b>		
Pheochromocytoma	Paroxysmal hypertension with vasomotor symptoms	High urinary catecholamine metabolites, adrenal tumor
Ectopic primary reninism	Malignant tumor with hypertension and/or hypokalemia	Rare cases of lung, liver, pancreas, or ovary cancers

Some large non-renin-secreting cysts and tumors may be found in cases of renin-dependent hypertension as the mechanism of renin stimulation is renal artery compression. In selected cases, with a lateralized RVR ratio, cyst drainage or tumor resection may improve or cure hypertension.

### Unilateral Nonvascular Small Kidneys

Pyelonephritic scarring associated with urinary tract infection and vesicoureteric reflux can cause childhood hypertension and progressive degradation of renal function. PRA is often high and has been shown to be the cause of hypertension. High renin concentrations may also occur in cases of renal hypoplasia. In cases with a very small unilateral kidney and a lateralized RVR ratio, unilateral nephrectomy may improve hypertension.

### Extrarenal Tumors

The high BP levels associated with pheochromocytoma are caused by high plasma catecholamine concentrations both directly, through stimulation of vascular  $\alpha$ -adrenergic receptors, and indirectly, through renin activation mediated by the adrenergic stimulation of juxtaglomerular cells and renal vasoconstriction. Consequently, ACE inhibition may be used to control BP before surgery.

As mentioned above, rare cases of primary reninism may be due to extrarenal tumors.

## HYPERRENINEMIA WITH NORMAL BLOOD PRESSURE LEVELS

In most conditions with normal BP levels, hyperreninemia is a homeostatic response to reduced renal perfusion pressure or plasma flow. These conditions include dehydration, hemorrhage, diuretic use or pharmacological vasodilation, cardiac failure, and reduced plasma volume due to hypoproteinemia in patients with nephrotic syndrome or kidney failure. These conditions have numerous signs and symptoms that cannot be listed here and the assessment of the renin-angiotensin system has no added diagnostic

value. Renin levels are related to survival in patients with cardiac failure; high levels are associated with a poor prognosis. The prognostic value of atrial and brain natriuretic peptides, however, is higher than that of renin; thus, renin is rarely used to assess prognosis.

Bartter's syndrome, or inherited hypokalemic metabolic alkalosis, is an autosomal-recessive disorder characterized by salt wasting, insensitivity to the vasoconstrictive effects of angiotensin II, and consequently high renin levels with normal BP, hyperaldosteronism, and hypokalemia. Gitelman's syndrome is a type of Bartter's syndrome in which patients have hypomagnesemia and hypocalciuria. Advances in the field of molecular genetics have demonstrated that this disease is caused by four genetically distinct abnormalities that result from mutations in renal electrolyte transporters and channels.

### See Also the Following Articles

Bartter's Syndrome • Captopril • Hypertension, Renin and • Hyporeninemic Hypoaldosteronism • Renal Vein Renin • Renin • Tissue Renin-Angiotensin-Aldosterone System

### Further Reading

- Blaufox, M. D., Lee, H. B., Davis, B., Oberman, A., Wassertheil-Smoller, S. D., and Langford, H. (1992). Renin predicts diastolic blood pressure response to nonpharmacologic and pharmacologic therapy. *J. Am. Med. Assoc.* **267**, 1221–1225.
- Corvol, P., Pinet, F., Plouin, P. F., Bruneval, P., and Ménard, J. (1994). Renin-secreting tumors. *Endocrinol. Metab. Clin. N. Am.* **23**, 255–270.
- Derckx, F. H. M., and Schalekamp, M. A. D. H. (1994). Renal artery stenosis and hypertension. *Lancet* **344**, 237–239.
- Gaul, M. K., Linn, W. D., and Mulrow, C. D. (1989). Captopril-stimulated renin secretion in the diagnosis of renovascular hypertension. *Am. J. Hypertens.* **2**, 335–340.
- Laragh, J. (2001). Laragh's lessons in pathophysiology and clinical pearls for treating hypertension. *Am. J. Hypertens.* **14**, 186–194.
- Plouin, P. F., Corvol, P., Guyene, T. T., and Ménard, J. (1997). Clinical investigation of the renin-angiotensin-aldosterone system. In "Oxford Textbook of Clinical Nephrology" (A. M. Davison, S. Cameron, J. P. Grunfeld, D. N. S. Kerr, E. Ritz, and C. G. Winearls, eds.), pp. 1425–1432. Oxford University Press, Oxford, UK.
- Shaer, A. J. (2001). Inherited primary renal tubular hypokalemic alkalosis: A review of Gitelman and Bartter syndromes. *Am. J. Med. Sci.* **322**, 316–332.

Table I Trials in Diabetic Hypertensives to Establish the Goal of Therapy

Study	Patients	Drugs	Reduction			
			Blood pressure (mm Hg)	CHD/CHF (percentage)	Stroke (percentage)	Mortality (percentage)
UKPDS <sup>a</sup>	1148	ACEI or beta blocker + CCB, diuretic	10/5 down to 144/82	21/56	44	18
HOT <sup>b</sup>	1501	(1) CCB (2) ACEI (40%) (3) beta blocker (28%) (4) Diuretic (22%)	4/4 down to 140/81	38	30	43
ABCD <sup>c</sup>	470	Enalapril or nisoldipine	6/8 down to 132/78			51
UKPDS <sup>d</sup>	3642	Various drugs	Systolic blood pressure from >160 (168) to <120 (114)	48	77	60

Note. CHD, coronary heart disease; CHF, congestive heart failure.

<sup>a</sup>UKPDS, United Kingdom Prospective Diabetes Study. *Br. Med. J.* 17, 703 (1998).

<sup>b</sup>HOT, Hypertension Optimal Treatment. *Lancet* 351, 1755 (1998).

<sup>c</sup>ABCD, Appropriate Blood Pressure Control in Diabetes. *N. Engl. J. Med.* 338, 645 (1998).

<sup>d</sup>Adler *et al.* *Br. Med. J.* 321, 412 (2000).

blockers (ARBs). All of these trials showed impressive reductions of cardiovascular morbidity and mortality or of progression of renal damage by the active drug regimen compared with placebo.

In the Heart Outcomes Prevention Evaluation (HOPE) trial, the 3600 high-risk diabetics achieved as much protection as did the nondiabetics by the addition of the ACEI ramipril. As summarized by Hostetter, three trials have shown renoprotection by ARBs, so these drugs can be added to the list of agents appropriate for those with diabetic nephropathy.

## EVIDENCE THAT DRUGS DO NOT DIFFER IN THEIR PROTECTIVE EFFECTS

Nearly 5000 diabetic hypertensives have been included in RCTs that compare one agent against another (Table II). The overall data from these comparative trials of diabetic hypertensives provide no clear proof that one class of drug is better than another, although

CCB-based therapy was associated with fewer events (Table II). However, there is considerable evidence that ACEIs are better than any other class of antihypertensives at protecting against one of the most common and serious complications seen in diabetic hypertensives: progressive renal damage leading to end stage renal disease. The September 2000 consensus report of the National Kidney Foundation authored by Bakris and colleagues states, "Antihypertensive regimens should include an ACE inhibitor in order to provide maximum cardiovascular and renal benefits in this population." This preference for an ACEI, or for an ARB if the patient cannot tolerate the ACEI because of a cough, seems to be unequivocal for renal protection. Whether it extends to all other cardiovascular events remains to be seen.

## THE NEED FOR MULTIPLE DRUGS

To achieve the desired reduction in blood pressure to a level of 130/80 mm Hg or lower, most diabetic hypertensives will require two, three, or four

Table II Comparative Trials in Diabetic Hypertensives

Primary drug	Patients	Total events per 1000 patients			
		Coronary heart disease	Congestive heart failure	Stroke	Mortality
Diuretic and/or beta blocker	1903	75	27	62	125
ACEI	1368	84	25	66	120
CCB	1657	69	28	53	82

antihypertensive drugs. As noted in the consensus report by Bakris and colleagues,

*Data from several major recent trials, including UKPDS and HOT, demonstrate ... that attainment of these lower goal blood pressures is virtually impossible to achieve with monotherapy. The majority of the time addition of multiple antihypertensive medications including a diuretic, CCB, or any similar combination is required to achieve these lower blood pressure goals.*

The burden of such multiple drugs, both financial and beyond, can be lessened by intensive attention to lifestyle modifications, including weight reduction, physical activity, and moderation of sodium, protein, and alcohol intake. These maneuvers may also improve control of hyperglycemia and dyslipidemia, which also demand attention. Despite the need for such intensive therapy, the benefits may well exceed the miseries and costs of uncontrolled diabetes and hypertension.

## CONCLUSION

The following guidelines seem to be appropriate for the management of hypertension in diabetic patients:

- The target of blood pressure reduction should be lower than 130/80 mm Hg.
- All antihypertensive drugs are beneficial compared with placebo.
- Multiple drugs will usually be needed to achieve the target.
- The choice of drugs should always include an ACEI (or an ARB if the ACEI is not tolerated) and a diuretic. If additional therapy is needed, a CCB, a beta blocker, or an alpha blocker may be used.
- Attention should be given to lifestyle changes (e.g., weight reduction; regular exercise; moderation of sodium, protein, and alcohol) as well as control of hyperglycemia, dyslipidemia, and proteinuria.

As the population grows older and fatter, diabetes and hypertension will become increasingly common.

It is hoped that their serious consequences can be reduced by such an approach.

## See Also the Following Articles

Hypertension, Endocrine • Hypertension, Neurogenic • Hypertension, Overview

## Further Reading

- Adler, A. I., Stratton, I. M., Neil, H. A. W., Yudkin, J. S., Matthews, D. R., Cull, C. A., Wright, A. D., Turner, R. C., and Holman, R.R., for the U.K. Prospective Diabetes Study Group. (2000). Association of systolic blood pressure with macrovascular and microvascular complications of type 2 diabetes (UKPDS 36): Prospective observational study. *Br. Med. J.* **321**, 412–419.
- Bakris, G. L., Williams, M., Dworkin, L., Elliott, W. J., Epstein, M., Toto, R., Tuttle, K., Douglas, J., Hsueh, W., and Sowers, J., for the National Kidney Foundation Hypertension and Diabetes Executive Committees Working Group. (2000). Preserving renal function in adults with hypertension and diabetes: A consensus approach. *Am. J. Kidney Dis.* **36**, 646–661.
- Boulé, N. G., Haddad, E., Kenny, G. P., Wells, G. A., and Sigal, R. J. (2001). Effects of exercise on glycemic control and body mass in type 2 diabetes mellitus: A meta-analysis of controlled clinical trials. *J. Am. Med. Assoc.* **286**, 1218–1227.
- Facchini, F. S., Hua, N., Abbasi, F., and Reaven, G. M. (2001). Insulin resistance as a predictor of age-related diseases. *J. Clin. Endocrinol. Metab.* **86**, 3574–3578.
- Grossman, E., Messerli, G. E., and Goldbourt, U. (2000). High blood pressure and diabetes mellitus: Are all antihypertensive drugs equal? *Arch. Int. Med.* **160**, 2447–2452.
- Heart Outcomes Prevention Evaluation (HOPE) Study Investigators. (2000). Effects of ramipril on cardiovascular and microvascular outcomes in people with diabetes mellitus: Results of the HOPE study and MICRO-HOPE sub-study. *Lancet* **355**, 253–259.
- Hostetter, T. H. (2000). Prevention of end-stage renal disease due to type 2 diabetes. *N. Engl. J. Med.* **345**, 910–912.
- Hu, F. B., Stampfer, M. J., Solomon, C. G., Liu, S., Willett, W. C., Speizer, F. E., Nathan, D. M., and Manson, J. E. (2001). The impact of diabetes mellitus on mortality from all causes and coronary heart disease in women. *Arch. Int. Med.* **161**, 1717–1723.
- Parving, H-H., Hommel, E., Jensen, B. R., and Hansen, H. P. (2001). Long-term beneficial effect of ACE inhibition on diabetic nephropathy in normotensive type 1 diabetic patients. *Kidney Intl.* **60**, 228–234.

**Table 1** Metabolic and Cardiovascular Risk Factors with Insulin Resistance and/or Visceral Obesity

---

- Increased systolic and diastolic blood pressure
- Hyperinsulinemia/Insulin resistance
- Endothelial dysfunction
- Low high-density lipoprotein cholesterol levels
- High triglyceride levels
- Increased Apo lipoprotein B levels
- Small, dense low-density lipoprotein levels
- Increased fibrinogen levels
- Increased plasminogen activator inhibitor-1 levels
- Hypercoagulability (increased PAI-1/plasminogen activator ratio)
- Increased blood viscosity
- Increased C-reactive protein and other inflammatory markers
- Microalbuminuria
- Absent nocturnal dipping of blood pressure and heart rate
- Increased uric acid levels
- Increased left ventricular hypertrophy
- Premature atherosclerosis (coronary artery disease and stroke)

---

The mechanisms for HTN in the metabolic syndrome are multiple and include hyperinsulinemic activation of the sympathetic nervous system (SNS), increased renal sodium retention, overexpression of the rennin-angiotensin-aldosterone system (RAAS), and a decrease in the vasodilatory properties of INS and INS-like growth factor-1 (IGF-1). These peptides have functional and receptor homology in cardiovascular tissue as well as in adipose and skeletal muscle tissue. INS resistance appears to correlate with tissue resistance to IGF-1. Resistance to the actions of these peptides is associated with impairment of vascular nitric oxide (NO) production, which results in decreased vasodilatory response in INS-resistant, obese, and hypertensive persons. Also, a strong relationship exists between the pressor effects of angiotensin II and HTN. Angiotensin II interferes with signaling through the metabolic (PI-3 kinase) pathway (in part responsible for INS/IGF-1 activity) in vascular smooth muscle cells. Overexpression of the RAAS may interfere with INS/IGF-1-induced NO production. Furthermore, angiotensin-converting-enzyme inhibitors (ACEIs) prevent INS-induced HTN and may improve INS resistance, as was suggested by two large trials. Improvement in INS resistance may have played a role in the cardiovascular benefits demonstrated in the HOPE and CAPP trials. ACEIs also increase NO release by increasing bradykinin via reduction of its degradation, thereby adding to their vasodilatory effects.

The increased free-fatty acids (FFAs) in INS-resistant/obese persons may contribute to vascular

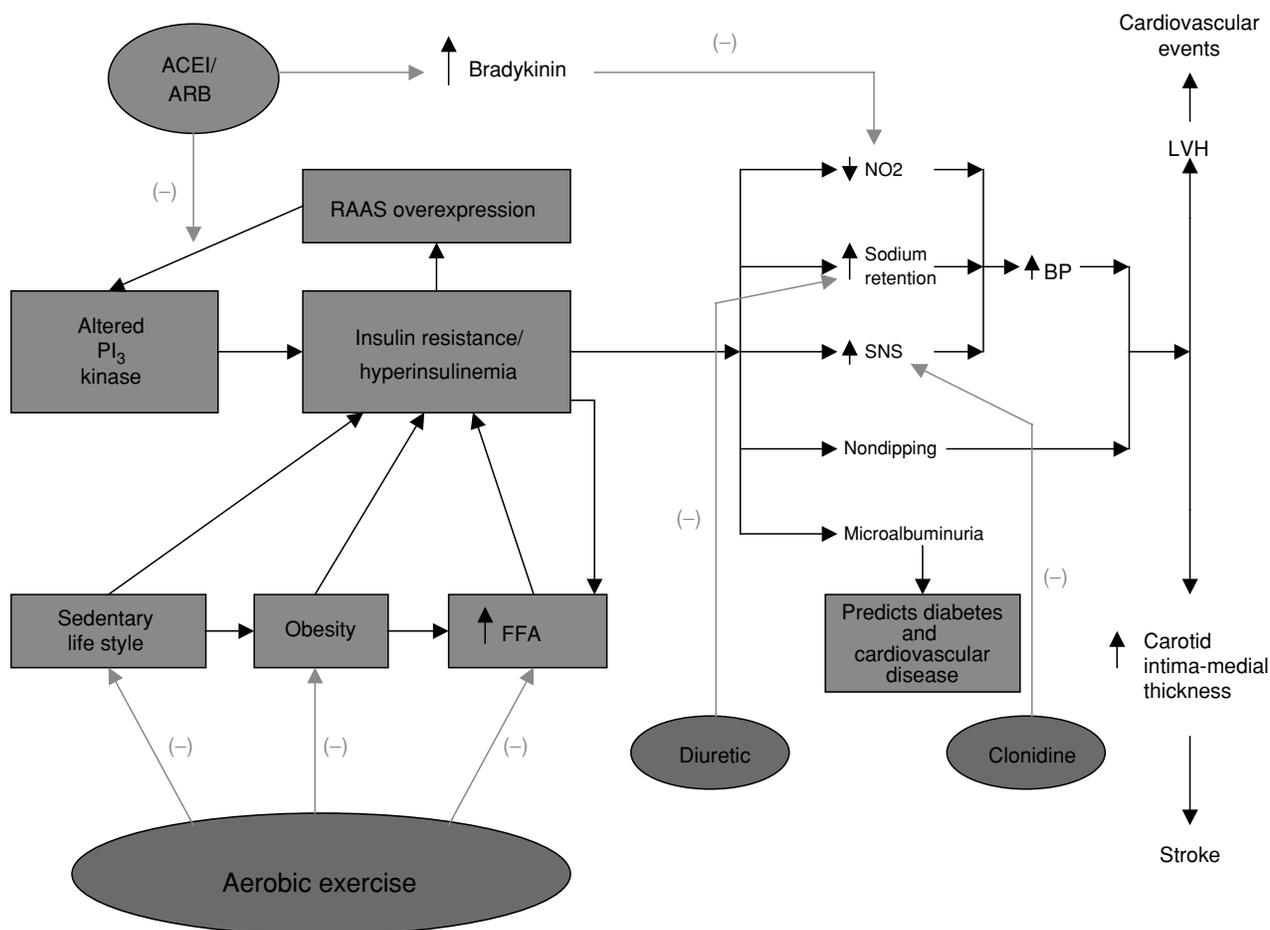
resistance to INS and, therefore, to its vasodilatory properties. Aerobic exercise reduces FFAs by increasing type 1 (INS-sensitive) muscle that uses primarily FFAs, leading to decreased truncal obesity, correction of INS resistance, and lowered BP.

## DIABETES

The cardiometabolic syndrome, a precursor of diabetes, shares similar pathology and traits. It goes to reason that HTN is more prevalent in diabetics. HTN is twice as frequent in diabetics than in nondiabetics and accounts for 75 % of CVD risk. Plasma INS and BP correlated directly in patients with diabetes. A study of 24 adult hypertensive patients showed that INS resistance, hyperglycemia, and hyperinsulinemia cosegregate. Characteristics of HTN in diabetics include increased sensitivity to salt and volume expansion; accelerated atherosclerosis, causing premature stiffness of larger arteries and resulting in isolated systolic HTN at a relatively younger age; and loss of drop of nocturnal BP and heart rate, associated with excessive CVD risk. All of these characteristics are associated with microalbuminuria in type 2 diabetics (Fig. 1).

The characteristics of diabetes and its sequelae require special consideration in treating associated HTN. The Systolic Hypertension in the Elderly Program (SHEP) demonstrates that HTN in diabetics is less likely to respond to monotherapy. The Joint National Committee (JNC) VI guidelines classify HTN in diabetes as high risk or complicated, and drug therapy and lifestyle changes are recommended at the onset. The American Diabetic Association and the Canadian Hypertension Society have adopted a BP goal of 130/80 or lower and may require two or more drugs. The JNC VII guidelines adopted 130/80 or lower as the goal for diabetics and also reported that two or more drugs may be necessary. For diabetics who excrete at least 1 g of protein in urine, the target BP is 125/75 or lower.

Various classes of antihypertensives compile the armamentarium for treating hypertensive diabetics. ACEIs attenuate albuminuria and progression to renal disease, and they may also improve INS sensitivity. They are recommended as firstline therapy in diabetics with HTN, proteinuria, and/or heart failure. Angiotensin receptor blockers (ARBs) have a lower incidence of cough and angioedema but may be equal to ACEIs for diabetic renal protection. Beta blockers are appropriate therapy in diabetics, particularly in those with ischemic heart disease. In the U.K.



**Figure 1** The cardiometabolic syndrome. RAS, renin-angiotensin-aldosterone system; FFA, free fatty acids; BP, blood pressure; SNS, sympathetic nervous system; LVH, left ventricular hypertrophy; ACEI, angiotensin-converting enzyme inhibitor; ARB, angiotensin receptor blocker; NO<sub>2</sub>, nitric oxide. Minus sign (-) indicates inhibition of pathological process.

Prospective Diabetes Study, the beta blocker atenolol reduced microvascular complications, strokes, and diabetes-related deaths. It had significant long-term effects on CVD in hypertensive patients with diabetes and is equal in efficacy to captopril in reducing micro- and macrovascular complications of diabetes. In nondiabetic African Americans with nephropathy, dihydropyridine calcium channel blockers (CCBs) adversely affected renal function compared with ACEIs and beta blockers. The nondihydropyridine CCBs may be more beneficial than dihydropyridine channel blockers in reducing proteinuria. However, the addition of an ACEI with a dihydropyridine CCB has been shown to reduce both proteinuria and CVD event rate. Low-dose diuretics are effective in diabetics because of salt sensitivity and its associated volume expansion, especially in African Americans, without adversely affecting lipid or glucose metabolism. For isolated systolic HTN, the JNC VII

guidelines recommend diuretics for their reduced rates of CVD, strokes, and all-cause mortality. Although alpha<sub>1</sub> blockers may improve lipid profile and enhance INS sensitivity, the increased propensity of some diabetics to orthostatic hypotension leads to avoidance of alpha<sub>1</sub> blockers as firstline agents for treatment of HTN associated with diabetes. Furthermore, in the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT), doxazosin seemed to be more associated with an increased incidence of congestive heart failure (CHF) than did chlorthalidone. Those diabetics with orthostasis and hyperadrenergic symptoms may benefit from low-dose clonidine.

### PRIMARY HYPERALDOSTERONISM

The incidence of primary hyperaldosteronism in essential HTN is 10 to 20%. Its diagnostic importance

is derived from the benefits of successful treatment given that reduced morbidity and mortality of heart failure are noted with low-dose spiranolactone. Screening for primary hyperaldosteronism (PA) should be considered in all patients with refractory or severe HTN, with spontaneous or provoked (with diuretics) hypokalemia, or requiring excessive potassium supplements. This condition is not necessarily associated with severe HTN, and hypokalemia is a poor screen. Approximately 61 to 70% of all cases have normal potassium. The ratio of aldosterone to plasma-rennin activity (A/R) is a good screen but may be of limited use with low-renin HTN. False-positive results with A/R ratio may occur with renal impairment due to suppression by sodium retention, and aldosterone may be increased by the associated hyperkalemia. Beta blockers, methyldopa, and clonidine may suppress renin and give false-positive results. Drugs that stimulate renin, such as diuretics, ACEIs, and dihydropyridine CCBs, falsely lower the A/R and may be useful given that a failure of renin to increase may be evidence of primary aldosteronism. False-negative results may occur with severe sodium restriction, pregnancy, and malignant HTN due to excessive renin and may also occur in patients with hypokalemia. In patients with PA, ACEIs do not appear to stimulate renin and might not cause false-negative results as was previously thought. This may depend on whether PA is dependent on angiotensin II. Verapamil and hydralazine have little effect, whereas prazosin has no effect on A/R. The sensitivity and specificity of an A/R ratio of at least 35 are increased with postcaptopril testing. The diagnosis may be confirmed by a high 24-h urinary aldosterone on a high-sodium diet, acute saline loading, or fludrocortisone suppression.

Once the diagnosis is confirmed, it is necessary to distinguish between a unilateral adenoma and a bilateral hyperplasia. Computed topography (CT) scanning of the adrenals can help. A solitary tumor greater than 1 cm in a young patient is an indication for surgery. However, in many cases, the CT scans appear to be normal. A nondiagnostic scan or an adrenal mass less than 1 cm requires bilateral adrenal vein sampling to demonstrate and lateralize an adenoma, especially because CT scans miss 50% of adenomas and can incorrectly lateralize the mass. Lateralization based on an elevated aldosterone/cortisol ratio associated with contralateral suppression is more sensitive and specific than imaging and is an indication for unilateral adrenalectomy. For nonlateralization, or if sampling is not available, low-dose spiranolactone is the treatment of choice.

Mineralocorticoid excess induces an increase in the number of epithelial sodium channels (ENaCs) that reabsorb sodium in the collecting duct of the kidney. HTN is caused by sodium and water retention by the kidney with expansion of the extracellular fluid compartment. Direct effects of aldosterone on the central nervous system may contribute to HTN. Aldosterone causes cardiac and vascular fibrosis in animals, an action prevented by aldosterone receptor blockade. In the clinical setting, aldosterone antagonists improve endothelial dysfunction and decrease hospitalization, symptoms, and mortality from progressive CHF and other heart-related deaths. Eplerenone, a selective aldosterone blocker, displays similar results.

## CONGENITAL ADRENAL HYPERPLASIA

Of the enzyme abnormalities that cause congenital adrenal hyperplasia (CAH), the two that cause HTN are 11 $\beta$ -hydroxylase (CYP11B1) and 17 $\beta$ -hydroxylase (CYP17) deficiencies.

### 11 $\beta$ -Hydroxylase Deficiency

Approximately 5 to 8% of CAH cases are due to CYP11B1 deficiency. With CYP11B1 deficiency, production of cortisol and corticosterone is inhibited, and enhanced adrenocorticotropin (ACTH) secretion further increases the precursors, deoxycorticosterone (DOC) and 11-deoxycortisol (11DOC), as well as androgens. Although DOC is a weak mineralocorticoid, its normetabolites may be more potent in causing salt and fluid retention with subsequent HTN. Plasma renin activity (PRA) and aldosterone are suppressed. Hypokalemia with muscle weakness and cramping is seen in a minority of patients. The severe homozygous classic form is common in Jews from Morocco, and at least 20 mutations have been identified with varying degrees of severity. Signs of androgen excess are demonstrated in both sexes, and female infants are masculinized. Accelerated skeletal growth and short stature with acne and premature adrenarche are manifested clinically. Those with the severe classic form have mild to moderate HTN appearing during the first few years of life and have a high incidence of left ventricular hypertrophy (LVH) and/or retinopathy. Diagnosis is made by demonstrating an increase in basal and/or ACTH-stimulated serum DOC, 11DOC, and adrenal androgens or elevated levels of urinary metabolites, tetrahydro-DOC and tetrahydro-11DOC, and 17-ketosteroids in a 24-h urine

collection. Steroid replacement suppresses ACTH levels, subsequently lowering both mineralocorticoid and adrenal androgen production and so controlling BP and preventing further virilization. Potassium-sparing diuretics, such as spironolactone and amiloride, may be needed to control HTN of long standing while preventing hypokalemia. Alternatively, a calcium channel blocker may be used.

### 17 $\beta$ -Hydroxylase Deficiency

Absence of secondary sexual development differentiates CYP17 deficiency from CYP11B1 deficiency, which might not be apparent until the expected onset of puberty. Prepubertal females appear to be normal but may later fail to develop normal sexual characteristics. Genetic males are phenotypically female, with a blind-ending vagina, inguinal testes, or undescended testes, or they may present with genital ambiguity. Although cortisol is not produced, adrenal insufficiency is not seen because corticosterone, which has glucocorticoid activity, is still produced. DOC is produced in excess, resulting in low-renin HTN and hypokalemia. Overproduction of DOC in the zona fascicula, which results in sodium and fluid retention, suppresses the production of aldosterone in the zona glomerulosa. Treatment with a glucocorticoid will increase aldosterone and PRA and will normalize BP. The increase in aldosterone does not result in HTN, presumably because it is regulated by volume status.

## GLUCOCORTICOID-REMEDIAL ALDOSTERONISM

Glucocorticoid-remedial aldosteronism (GRA), also called dexamethasone-suppressible hyperaldosteronism or familial hyperaldosteronism type 1, was first described by Sutherland and colleagues in 1966. Inherited in a Mendelian autosomal dominant fashion, GRA results from a chimeric gene that fuses the 11 $\beta$ -hydroxylase regulatory gene to the coding gene of aldosterone synthase. It is a mineralocorticoid excess state characterized by low PRA with an A/R ratio greater than 30, HTN, and spontaneous hypokalemia. The BP elevation is typically severe but can masquerade as essential HTN with an increased prevalence of early, and often fatal, cerebrovascular hemorrhage. The full phenotype is seldom seen. One prospective study showed that most patients are normokalemic unless they are given potassium-wasting diuretics.

The diagnosis is supported by improved BP following dexamethasone suppression testing (DST) with a decrease in aldosterone to nearly undetectable levels, demonstrating the ACTH dependency of this condition. A plasma aldosterone of less than 4 ng/dl post-DST is considered sensitive and specific. Suppression can also occur with aldosterone-producing adenomas (APAs) and bilateral adrenal hyperplasia (BAH) but fails to fall to very low levels. Furthermore, elevated 18-hydroxy and 18-oxy cortisol in the urine of patients with GRA will help to distinguish it from APA and BAH. However, sometimes these urinary metabolites are also elevated with APA. Sustained ACTH stimulation may cause an initial rise in aldosterone followed by a decline in normal persons and patients with APA and BAH, whereas in GRA the rise in aldosterone is sustained, similar to that seen with CAH. However, in contrast to CAH, patients with GRA show a normal cortisol and 17-hydroxy urinary corticosteroid response to ACTH. Polymerase chain reaction (PCR) is simple and reliable without the pitfalls of other tests, making genetic testing essential for confirmation.

Although glucocorticoid suppression is the treatment of choice, normalization of BP might not always occur. Linear growth should be monitored in children, and manifestations of Cushing's syndrome should be avoided. Furthermore, hypoaldosteronism may occur with hypotension. Spironolactone is also effective as monotherapy, and amiloride and triamterene may be used as alternatives.

## LIDDLE'S SYNDROME

In 1963, G. W. Liddle described a syndrome of HTN, increased potassium excretion, metabolic alkalosis, decreased sodium excretion, and suppressed aldosterone. Hypokalemia is not a universal finding. Spironolactone was ineffective, but triamterene, amiloride, and salt restriction corrected hypokalemia and HTN. Renin is also suppressed by a volume-expanded state with atrophy of the juxtaglomerular apparatus. Later, it was observed that the HTN, hypokalemia, and metabolic alkalosis were ameliorated by renal transplantation with establishment of a renin response to a low-sodium diet, pointing to a disorder of the renal tubule. The absence of edema may be attributed to aldosterone escape due to suppressed sodium reabsorption in other parts of the nephron and to enhanced release of atrial natriuretic peptide in response to volume expansion. Liddle's

syndrome is an autosomal dominant-activating mutation of the renal ENaC located in the collecting ducts. The ENaC is made of three subunits:  $\alpha$ ,  $\beta$ , and  $\chi$ . Mutations resulting in truncation of the tail of either the  $\alpha$ - or  $\beta$ -subunit is responsible for functional gain of the ENaC in Liddle's syndrome. Thus, this is a salt-retaining, volume-expanded, genetic form of HTN.

### GORDON'S SYNDROME

In 1970, R. D. Gordon described a congenital renal tubule defect that includes HTN and severe hyperkalemia, in association with suppressed renin and aldosterone, that was completely reversed with dietary sodium restriction. This syndrome, which includes a normal glomerular filtration rate (GFR), can be rapidly reversed with thiazide diuretics. Also, a hyperchloremic acidosis in which bicarbonate wasting exists is the result of hyperkalemia and volume expansion. This condition is also called type 2 pseudohypoaldosteronism (type 1 is associated with sodium wasting and hypotension during infancy) because of the proposed inability to increase urinary potassium excretion with hyperkalemia despite the addition of mineralocorticoids. The pathophysiology of Gordon's syndrome remains undefined.

Gordon proposed that an inability to respond to atrial natriuretic peptide (ANP) or renal natriuretic prostaglandins may be responsible. Lack of response to ANP, which was noted by Klemm and colleagues, also explains the hyperchloremia and decreased levels of renal vasodilatory prostaglandin that is demonstrated in the urine of Gordon's syndrome patients. These result in volume expansion and suppression of renin and aldosterone to dietary salt despite high serum potassium. The low aldosterone level could be responsible for maintaining a high potassium level. Salt restriction or diuresis stimulates renin and aldosterone, thereby lowering potassium to normal.

Renal potassium excretion is not increased by mineralocorticoids or sodium loading but did increase when nonchloride forms of sodium were administered, suggesting another mechanism. Shambelan and colleagues proposed that enhanced chloride reabsorption leads to enhanced sodium, hydrogen ion, and potassium reabsorption. The inability to excrete potassium despite mineralocorticoid administration may also imply an altered clearance of potassium in the distal tubule. Decreasing potassium with sodium polystyrene sulfate corrected the hyperchloremic acidosis, probably by increasing ammoniogenesis by the nephron.

### SYNDROME OF APPARENT MINERALOCORTICOID EXCESS

The syndrome of apparent mineralocorticoid excess (AME) is a relatively rare autosomal recessive mutation occurring in consanguineous populations and presenting during childhood with HTN and hypokalemia. The hypokalemia results in hypercalciuria with nephrocalcinosis and nephrogenic diabetes insipidus, and it may also result in rhabdomyolysis. Cardiovascular accidents and death can occur during infancy and adolescence. The enzyme that confers specificity to aldosterone, by degrading other mineralocorticoids at the distal renal tubule, is deficient or absent. Cortisol is converted to its metabolite (cortisone), which has no mineralocorticoid activity, by a reversible kidney enzyme, type 2  $11\beta$ -hydroxycorticosteroid dehydrogenase ( $11\beta$ -HSD). When  $11\beta$ -HSD is deficient, cortisol is allowed unrestricted access to the mineralocorticoid receptor. Glycyrrhetic acid from licorice and its semisynthetic analogue carbenoxylone can induce AME by inhibition of the same enzyme. All mineralocorticoids are low, and cortisol acts as a stronger agonist at the aldosterone site than is usually the case. The diagnosis can be made by demonstrating an increased ratio of urinary metabolites of cortisol (tetrahydrocortisol and allo-tetrahydrocortisol) to a low urinary metabolite of cortisone (tetrahydrocortisone).

Treatment with spironolactone corrects the hypokalemia and HTN. Thiazide diuretics have been used successfully to treat nephrocalcinosis and may allow a decreased dose of spironolactone, reducing the likelihood of gynecomastia. Amiloride, beta blockers, and ACEIs can be used alternatively. In one reported case, a patient was successfully treated with an ACEI, a beta blocker, and dexamethasone. Years later; kidney transplant in the same patient normalized potassium, PRA, and BP as a result of the new kidney synthesizing the deficient enzyme. Loop diuretics should be avoided because they can aggravate hypokalemia and alkalosis.

### CUSHING'S HYPERTENSION

HTN is seen in 80% of patients with Cushing's syndrome, and cortisol is primarily responsible. Local cortisol excess is responsible for the HTN associated with AME and licorice intoxication. Although cortisol and aldosterone have equal binding affinity,  $11\beta$ -HSD, by locally changing cortisol, prevents its binding to the mineralocorticoid receptor in the

nephron-collecting tubule. Excess cortisol may overwhelm this enzyme and bind to the receptor. Indeed aldosterone and renin levels are low in Cushing's HTN, suggesting an apparent mineralocorticoid overactivity that causes sodium and volume retention. Despite this, hypokalemia, metabolic alkalosis, and edema are more often associated with ectopic secretion of ACTH, reflecting its very high levels. Elevated levels of corticosterone and deoxycorticosterone associated with ectopic ACTH production may also contribute. Because ANP and decreased sodium reabsorption occur with volume expansion, thereby causing aldosterone (in this case mineralocorticoid) escape, the amount of sodium being exchanged is increased. This may explain the presence of normokalemia and the lack of edema seen with Cushing's syndrome. However, plasma volume is expanded with salt intake, but not to the point of edema, which may occur prior to the escape mechanism.

HTN in Cushing's syndrome may be due to mechanisms other than sodium and fluid retention; with dexamethasone, virtually devoid of mineralocorticoid activity, HTN is not associated with volume expansion. Dexamethasone increases total peripheral resistance but not cardiac output. Although cortisol is associated with an increase in cardiac output, pretreatment with beta blockers will decrease cardiac output but not the BP response to cortisol that is explained by an increase in total peripheral resistance. An increased response to pressor substances is noted in cortisol-treated persons. Furthermore, enhanced binding and response of angiotensin II occur in vascular smooth muscle cells (VSMCs), resulting in augmented renovascular and forearm arterial resistance. The "glucocorticoid" activity of dexamethasone and other steroids could result from increased calcium channel binding and uptake of calcium and inhibited inducible NO synthase in VSMCs and endothelium, respectively (with the former blocked by nifedipine). ACTH-dependent, but not primary adrenal, Cushing's is coupled with increased levels of 19-hydroxy androstenedione, which has pressor activity.

The pathophysiology of HTN in Cushing's syndrome is complex, with a number of possible mechanisms. Treatment of the HTN should be directed at the primary source of excess cortisol: ectopic/pituitary ACTH, adrenal, or exogenous. Calcium channel blockers or drugs that improve endothelial NO production, such as ACEIs and ARBs, may be used until a final diagnosis is reached.

## HIGH-RENIN HYPERTENSION

Conditions of secondary aldosteronism can occur in normotensive and hypertensive patients. Secondary hyperaldosteronism may be an appropriate physiological response but may be noncompensatory and detrimental in pathophysiological settings. A high incidence of unsuspected renovascular atherosclerosis is found in patients who undergo cardiac catheterization where age, blood urea nitrogen, and diabetes may predict renal artery stenosis. Other diagnostic modalities include color Doppler ultrasound, magnetic resonance imaging (MRI), magnetic resonance angiography (MRA), computed tomography (CT), computed tomography-angiography, scintigraphy, and captopril-enhanced renography or scintigraphy. Selective use of renal MRA can avoid invasive, potentially nephrotoxic conventional angiography in high-risk patients. Clinical features of renal artery stenosis include abrupt onset or accelerated HTN, acute or chronic azotemia, azotemia induced by ACEIs, asymmetric kidney size, CHF with normal ventricular function, and nonischemic ("flash") pulmonary edema. Elderly patients with generalized atherosclerosis typically have non-renin-dependent HTN, and revascularization rarely cures HTN in atherosclerotic renal artery disease, whereas young or middle-aged women with fibromuscular dysplasia are often renin dependent. Revascularization should be reserved for severe or resistant cases of HTN due to renovascular atherosclerosis. Gill and colleagues reported improved BP in a majority of resistant cases with recovery of renal function. Ischemic nephropathy is best treated before the onset of renal failure. Revascularization with surgery or stenting should be considered with creatinine less than 2 mg/dl, bilateral renal artery stenosis, absence of proteinuria, and/or evidence of end-organ injury.

Severe HTN with secondary hyperaldosteronism can occur in systemic sclerosis due to renal artery vasospasm and subsequent hyperreninemia. Nifedipine has led to normalization of BP, renin, and aldosterone.

Secondary hyperaldosteronism and HTN can result from renin-secreting tumors. Juxtaglomerular cell tumors can cause hyperreninemia, typically in teenagers. Renal vein renin failed to lateralize the tumors in 52% of cases in one study. In that study, 100% of renal masses were demonstrated by CT scan. Sonography can be helpful as well. Surgical treatment results in prompt normalization of BP, and a calcium channel blockade may control BP until surgery. Renal

cell carcinoma rarely causes hyperreninemia or secondary hyperaldosteronism.

## PHEOCHROMOCYTOMA

Pheochromocytoma (PCC), a chromafin tumor, can be life threatening, whereas early diagnosis can be life saving. A crisis can be associated with signs and symptoms of myocardial infarction, supraventricular tachycardia, or CHF. Sudden elevations in arterial pressure are common and do not correlate with levels of plasma catecholamines (norepinephrine and epinephrine [sometimes dopamine]). Even in patients with persistent elevations of BP, symptoms may be paroxysmal; the most frequent symptoms include headache, diaphoresis, and palpitations. Other symptoms may include pallor, nausea, tremor, anxiety, abdominal pain, chest pain, dyspnea, and (rarely) flushing. Occasionally, an orthostatic drop in BP is observed. A patient may be normotensive between episodes. Sometimes, paroxysms may be absent, mimicking essential HTN. This is important to recognize in that 76% of patients with PCC in one study were discovered postmortem, and 54% of these had HTN. Approximately 10% are bilateral, extra-adrenal (paragangliomas), familial, multiple (other than bilateral), or malignant. Diagnosing PCC requires a high index of suspicion. One should suspect this disease when any of those features listed in Table II apply (Table II).

The HTN in PCC is the result of increased peripheral resistance. BP elevations do not correlate well with plasma catecholamines. Clonidine, a centrally working agent that works as well for this condition as it does for essential HTN, is effective because of a

functional SNS. This implies an effect of catecholamines at the synaptic junction and not in the plasma. Small tumors release large amounts of catecholamines into the blood, whereas large tumors, which metabolize much of the catecholamines prior to their release, release small amounts. Large tumors produce greater amounts of the urinary metabolites (metanephrines).

A 24-h urinary vanillylmandelic acid (VMA) collection has poor sensitivity with a false-negative rate of 41%, whereas urinary catecholamines have a false-negative rate of 14%. However, the combination of plasma catecholamines of at least 2000 pg/ml and urinary metanephrines of at least 1.8 mg/24 h is 98% accurate with a false-negative rate of 7%. The specificity of any of these is high. Free plasma metanephrines alone have a specificity of 96% and a sensitivity of 97% for the familial form of PCC. Although the sensitivity remains high for the sporadic form, the specificity drops to 82%. Combining plasma metanephrines with urinary metanephrines increases the specificity to 89%. A plasma catecholamine level of at least 2000 pg/ml is diagnostic, and a level of less than 500 pg/ml rules out PCC. Intermediate values require further testing. A provocative test with glucagon (1 mg intravenously) can be used for levels of 500 to 1000 pg/ml and a BP under 170/100. A threefold increase or a level greater than 2000 pg/ml 3 min later is diagnostic. A clonidine suppression test is used for levels of 1000 to 2000 pg/ml and a BP above 160/90. A 50% drop from baseline, or to less than 500 pg/ml, is a negative test. This has a low false-negative rate.

MRI has slightly better sensitivity than CT, approaching 100% but without the need for contrast. A T2-weighted image of a PCC has a characteristically hyperintense image in comparison with that of the liver. Scintigraphy with <sup>131</sup>I-metaiodobenzylguanidine (MIBG) is used when biochemical testing is highly suspicious but MRI and/or CT are negative. It is also useful when multiple tumors are suspected but MRI and/or CT are negative such as in children or those with a family history of MEN/familial PCC or possible malignant spread. Although octreotide scintigraphy is also useful for visualizing chromafin tumors and has good sensitivity for locating head and neck paragangliomas, its sensitivity for detecting intra-abdominal PCCs is limited. However, it has been shown to find tumors not detected by MIBG and may be complementary to MIBG. One case report of a cardiac PCC, which was missed by MIBG and CT scan, was localized by octreotide scintigraphy and confirmed by MRI. Venous sampling should be used only when suspicion is very high and

**Table II Features That Are Suspicious for Pheochromocytoma**

- 
- Symptoms of pheochromocytoma
  - Paradoxical responses to antihypertensive drugs especially beta blockers
  - Signs and symptoms of accelerated or malignant hypertension
  - Hypertensive paroxysms with exercise, intubation, manipulation of the abdominal contents
  - Family members with MEN syndromes or familial pheochromocytoma
  - Coexisting hypertension with neurofibromatosis, von Hippel-Lindau disease, or Cushing's syndrome
  - Orthostasis without antihypertensives
  - Incidental adrenal tumors
-

all imaging studies are negative. Positron emission tomography, a new modality using 6-fluorodopamine, may be superior to other nuclear imaging.

Treatment involves avoidance of hypertensive episodes during surgery and postsurgery. The nonselective alpha blocker phenoxybenzamine allows for volume repletion but is associated with significant reflex tachycardia and orthostasis and may prolong postsurgical hypotension. After volume repletion, selective alpha<sub>1</sub> blockers such as prazosin have a shorter duration of action and do not cause reflex tachycardia. Alpha<sub>1</sub> blockers can still be associated with postoperative hypotension as well as mask the hypotension associated with the removal of the tumor. The latter is used by surgeons as an indication of complete tumor removal. Boutros and colleagues suggested holding alpha blockers and CCBs unless needed. CCBs seldom cause orthostasis or hypotension and are safe to use in normotensive patients. They also prevent catecholamine-induced spasm of the coronary arteries. By decreasing the norepinephrine-induced increases of calcium entry into the VSMCs, CCBs can prevent the hypertensive response in PCC. Beta blockers should be used only after alpha blockade.

A PCC crisis can be treated by sodium nitropruside. A beta<sub>1</sub>-selective blocker, to avoid unopposed alpha activity, can control supraventricular or reflex tachycardia. Esmolol is preferred for its rapid onset and short duration.

In MEN 2 and MEN 3, adrenal involvement is usually diffuse nodular hyperplasia, bilateral, and sometimes malignant. Pheochromocytoma can by itself cause hypercalcemia; therefore, it might not be associated with the hyperparathyroidism of MEN 2.

## PRIMARY HYPERPARATHYROIDISM

Although primary hyperparathyroidism (HPT) has been associated with HTN, the relationship between these entities is not fully understood. Whereas some studies show a possible causal connection between HPT and HTN, others do not. One study showed that parathyroidectomy (PTX) did not correct HTN, whereas another investigation reported significant BP reductions in 20 of 45 hypertensive patients with HPT following PTX.

Gennari and colleagues showed a greater pressor response to norepinephrine in those hypertensives with HPT, possibly due to the effect of the catecholamine to increase calcium entry into the VSMCs (of a hypercalcemic person). In the same study, PTX cured 8 of 10 persons with HPT of HTN with a

reduction in PRA and aldosterone, suggesting a direct effect of PTH on renin secretion. Furthermore, intracellular calcium is increased in essential HTN as well as in HPT, with the fall in BP being correlated with the decrease in intracellular calcium after PTX. Nifedipine reduced both BP and intercellular calcium in hypertensives with HPT. Parathyroid hypertensive factor (PHF) is found in some persons with essential HTN and is known to be secreted by the parathyroids of spontaneously hypertensive rats. Significantly higher PHF levels were reported in hypertensive persons with HPT than in their normotensive counterparts. In those patients undergoing a PTX, a postoperative fall in BP could be predicted by the level of PHF, suggesting that human parathyroids can secrete PHF. Although BP remained unchanged in some patients with HPT undergoing PTX, endothelium-dependent (flow-mediated) vasodilation improved postsurgery.

## ACROMEGALY

Acromegaly is associated with a higher incidence of HTN contributing to increased morbidity and mortality. Hypertensive acromegalics tend to be women, weigh more, and are older than their normotensive counterparts, suggesting risk factors similar to those of the general population. Although there was no correlation with growth hormone (GH), another study showed a direct association of BP with IGF-1 and serum sodium in only those acromegalics with a family history of HTN. Systolic BP was the largest independent risk factor for LVH, followed by GH levels. A subgroup of acromegalics with diabetes had the highest prevalence of LVH and impairment of diastolic and systolic dysfunction. Acromegaly is frequently associated with INS resistance and hyperinsulinemia and may induce HTN by stimulating renal reabsorption of sodium and SNS. The sodium and volume retention may be due, in part, to an ability of either GH or IGF-1 to activate the RAAS and suppression of ANP. Indeed PRA and aldosterone levels increased with administration of GH, and this was blocked by captopril and spironolactone. There is no consensus on the mechanism of HTN with acromegaly.

## HYPOTHYROIDISM

Hypothyroidism is associated with HTN that may be resistant to antihypertensive treatment. Treatment with thyroxine may result in a reduction in BP.

Hypothyroidism is also associated with an increase in cholesterol, triglycerides, homocysteine, and total cholesterol/HDL ratio as well as with accelerated atherosclerosis and coronary artery disease. Subclinical hypothyroidism is also associated with HTN (mostly diastolic) and lipid disorders. In one study, treatment of hypothyroidism improved coronary insufficiency, with relapse on therapeutic withdrawal. The increase in BP may be due, in part, to increased vascular resistance and aortic stiffness. A correlation of BP reduction with an associated decrease in aortic stiffness occurred with thyroxine treatment in hypothyroid patients. Stimulation of adrenal function by augmenting PRA, with an increase in aldosterone and cortisol, occurs in hypothyroidism. This also is reduced with thyroxine treatment. Indeed, the use of captopril and spironolactone was effective in treating the associated HTN. An alternative mechanism for high BP is animation of the SNS. Correction of the hypothyroid state can reduce the elevated levels of epinephrine and norepinephrine that exist with hypothyroidism along with a concurrent decline in BP. Furthermore, heart rate variability decreased with therapy, highlighting the importance of the SNS. The HTN of an underactive thyroid is also associated with an elevated total peripheral resistance, possibly due to an overactive SNS.

## HYPERTHYROIDISM

A decreased total peripheral resistance occurs in hyperthyroidism. The HTN associated with an overactive thyroid is mainly systolic and is due to increased cardiac output. This may be the result of an increased number of beta receptors in the heart making it more sensitive to catecholamines. Treatment with beta blockers can control many of the symptoms of hyperthyroidism, including elevated BP. Furthermore, treatment of the hyperthyroid state leads to a decrease in or complete control of BP in those who are younger and without essential HTN. An other cardiovascular effect of thyrotoxicosis, as demonstrated in rats, is the development of LVH, which was prevented by valsartan, suggesting that the RAAS plays a role.

### See Also the Following Articles

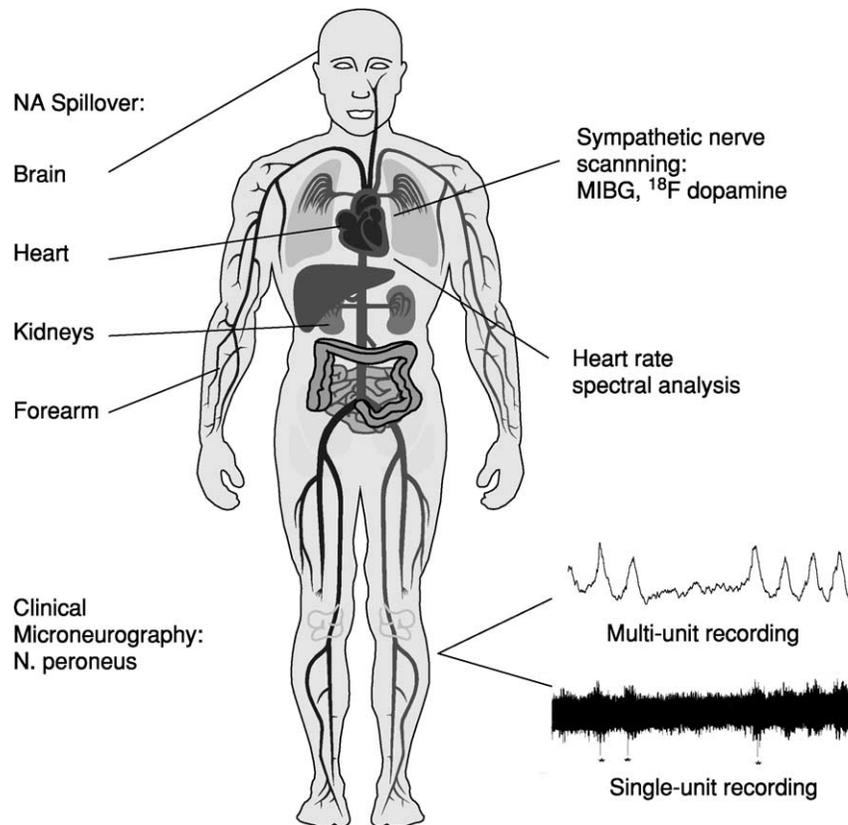
Acromegaly, Clinical Features of • Atherosclerosis • Hyperparathyroidism, Primary • Hypertension and Diabetes • Hypothyroidism, Treatment of •  $11\beta$ -Hydroxylase Deficiency • Insulin-Resistant States, Role of Free Fatty Acids

(FFA) • Mineralocorticoids and Mineralocorticoid Excess Syndromes • Neurofibromatosis • Obesity and Diabetes, Regulation of Food Intake • Pheochromocytoma • Von Hippel-Lindau Syndrome

## Further Reading

- ALLHAT Collaborative Research Group. (2000). Major cardiovascular events in hypertensive patients randomized to doxazosin vs. chlorthalidone: The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). *J. Am. Med. Assoc.* **283**, 1967–1975.
- Bondanelli, M., Abrosio, M. R., and degli Uberti, E. C. (2002). Pathogenesis and prevalence of hypertension in acromegaly. *Pituitary* **4**, 239–249.
- Botero-Velez, M., Curtis, J. J., and Warnock, D. G. (1994). Brief report: Liddle's syndrome revisited—A disorder of sodium reabsorption in the distal tubule. *N. Engl. J. Med.* **330**, 178–181.
- Cohobanian, A., Bakris, G., Black, H., et al. (2003). The seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. *J. Am. Med. Assoc.* **289**, 2560–2572.
- DeCobelli, F., Cappio, S., and Del Maschio, A. (2001). A noninvasive imaging of the abdominal vessels. *Rays* **26**, 269–276.
- Dluhy, R., and Lifton, R. (1999). Glucocorticoid-remedial aldosteronism. *J. Clin. Endocrinol. Metab.* **84**, 4341–4344.
- Gordon, R. D. (2001). Diagnostic investigations in primary aldosteronism. In "Clinical Medicine Series on Hypertension" (A. Zanchetti, ed.), pp. 101–114. McGraw-Hill International, Maidenhead, UK.
- Gordon, R. D., Geddes, R. A., Pawsey, C. G. K., and O'Halloran, M. W. (1970). Hypertension and severe hyperkalemia associated with suppression of renin and aldosterone and completely reversed by dietary sodium restriction. *Aust. Ann. Med.* **4**, 287–294.
- Lewanczuk, R. Z., and Pang, P. K. (1993). Expression of parathyroid hypertensive factor in hypertensive primary hyperparathyroid patients. *Blood Pressure* **2**(1), 4–5.
- Liddle, G. W., Bledsoe, T., and Coppage, W. S., Jr. (1963). A familial renal disorder simulating primary aldosteronism but with negligible aldosterone secretion. *Trans. Assoc. Am. Physicians* **76**, 1999–2113.
- Manger, W. M., Gifford, R. W. (1996). "Clinical and Experimental Pheochromocytoma," 2nd ed. Blackwell Science, London, UK.
- Nilson, I. L., Aberg, J., Rastad, J., and Lind, L. (1999). Endothelial vasodilatory dysfunction in primary hyperparathyroidism is reversed after parathyroidectomy. *Surgery* **126**, 1049–1055.
- Pirpiris, M., Sudhir, K., Yeung, S., et al. (1992). Pressor responsiveness in corticosteroid-induced-hypertension in humans. *Hypertension* **19**, 99–108.
- Ravichandran, R., Rengarajan, T., Reddy, K. N., and Rao, S. M. (2003). Prevalence of renovascular abnormalities in patients undergoing cardiac catheterization. *J. Assoc. Physicians India* **51**, 175–177.
- Saito, I., and Saruta, T. (1994). Hypertension in thyroid disorders. *Endocrinol. Metab. Clinics North America* **23**, 379–386.
- Savage, P. J., Pressel, S. L., Curb, J. D., et al. (1998). Influence of long term, low dose, diuretic based, antihypertensive therapy on glucose, lipid, uric acid, and potassium levels in older men and women with isolated systolic hypertension: The Systolic

- Hypertension in Elderly Program—SHEP Cooperative Research Group. *Arch. Int. Med.* **158**, 741.
- Schambelan, M., Sebastian, A., and Rector, F. C., Jr. (1981). Mineralocorticoid-resistant renal hyperkalemia without salt wasting (type II pseudohypoaldosteronism): The role of increased renal chloride reabsorption. *Kidney Int.* **19**, 716–727.
- Sowers, J. R. (1997). Insulin and insulin-like growth factor in normal and pathological cardiovascular physiology. *Hypertension* **29**, 691–699.
- Sutherland, D. J., Ruse, J. L., and Laidlaw, J. C. (1966). Hypertension, increased aldosterone secretion, and low plasma renin activity relieved by dexamethasone. *Can. Med. Assoc. J.* **95**, 1109–1119.
- Ulick, S., Wang, J. Z., Bloomenfeld, J. D., *et al.* (1992). Cortisol inactivation overload: A mechanism of mineral corticoid hypertension in ectopic adrenocorticotropin syndrome. *J. Clin. Endocrinol. Metab.* **74**, 963–967.
- White, P. C. (1996). Inherited forms of mineral corticoid hypertension. *Hypertension* **28**, 927–936.
- White, P. C. (2001). 11 beta-hydroxysteroid dehydrogenase and its role in the syndrome of apparent mineral corticoid excess. *Am. J. Med. Sci.* **322**, 308–315.
- Young, M. J., and Funder, J. W. (2002). Mineralocorticoid receptors and pathophysiological roles for aldosterone in the cardiovascular system. *J. Hypertension* **20**, 1565–1568.



**Figure 1** Clinically applicable methods for studying sympathetic activation in various regions in humans. Overflow of noradrenaline (NA) to plasma from internal organs can be measured by isotope dilution methodologies. Rates of nerve firing can be measured from postganglionic sympathetic fibers distributed to the skeletal muscle vasculature using microneurography as either multiunit or single-fiber recordings. Scanning techniques allow visualization of sympathetic nerves in the heart, enabling assessment of cardiac sympathetic nerve density. Heart rate spectral analysis has been applied as a noninvasive method for studying sympathetic function in the heart, although its validity is questioned. MIBG, meta-iodobenzylguanidine; N. peroneus, nervus peroneus.

## Plasma Noradrenaline

Although the measurement of noradrenaline (NA) plasma concentration is still widely used, it is a poor marker of overall sympathetic tone given that plasma NA concentrations depend not only on sympathetic tone and NA release but also on rates of removal of the neurotransmitter from plasma and do not provide any information on regional differentiation of sympathetic drive.

## Microneurography

Postganglionic muscle sympathetic nerve activity (MSNA) can be measured directly by placing a fine electrode into a superficial peripheral nerve, most commonly the peroneal nerve due to its easy accessibility, and recording its activity. MSNA contributes substantially to the regulation of peripheral vascular

resistance and can be identified by a characteristic spontaneous, intermittent, and pulse-synchronous burst pattern. Microneurography is unique in that it provides direct and continuous assessment of adrenergic activity to the skeletal muscle circulation.

## Noradrenaline Spillover Rate Measurement

The application of isotope dilution methodology to measure total rates of NA spillover to plasma from sympathetic nerves, adopting a measurement strategy applied in clinical endocrinology for the measurement of hormone secretion rates, eliminated the confounding effect of variation in NA plasma clearance on plasma NA concentration. The lack of clinical methods for studying human sympathetic nervous outflow in otherwise inaccessible internal organs provided an incentive for the development of

biochemical techniques for studying the rates of overflow of catecholamines from these organs to the circulation. Using radiotracer-derived methods involving the infusion of tritiated NA and regional blood sampling from the coronary sinus and jugular and renal veins, the appearance rate of NA in plasma from the heart, cortical and subcortical brain areas, and kidneys can be measured (Fig. 1).

### Heart Rate Power Spectral Analysis

Spectral analysis techniques are commonly applied as an alternative noninvasive method for studying sympathetic function in the heart. With this technique, mathematical partitioning allows identification of individual superimposed rhythms producing cyclical variation in heart rate and arterial pressure. Despite earlier enthusiasm for this approach, it is now clear that the low-frequency spectrum of heart rate variability is not a reliable measure of cardiac sympathetic nervous activity.

## FEATURES OF NEUROGENIC HYPERTENSION

There exists a typical pattern of sympathetic nervous activation, particularly in younger individuals (<40 years of age), evident in a high heart rate, elevated cardiac output, elevated renal renin release and plasma renin activity, and (in some patients) isolated systolic hypertension. An increase in muscle sympathetic nerve firing and high rates of spillover of NA from the heart and kidneys are further manifestations of the syndrome of neurogenic hypertension, whereas adrenal medullary secretion of adrenaline is normal. The increased cardiac and renal sympathetic nerve firing provides a plausible mechanism for the development of hypertension through the regulatory influence of the sympathetic nervous system on renin release, glomerular filtration rate, and renal tubular sodium reabsorption as well as on cardiac growth and pump performance.

## NEUROGENIC MECHANISMS OF BLOOD PRESSURE ELEVATION

Traditionally, a common view was that increased sympathetic nervous activity could not cause sustained elevation of blood pressure and could not be of importance in the pathogenesis of essential hypertension. The sympathetic nervous system was thought to exert minute-by-minute control over the circulation, with

long-term blood pressure control being exercised principally through control of salt and water balance by the kidneys.

### Renal Neural Mechanisms

It is now well established that neural mechanisms are involved in hypertension development through regulatory effects of the renal sympathetic nerves on renin release, glomerular filtration rate, renal tubular reabsorption of sodium, and renal pressure natriuresis. Renal renin release in particular is clearly related to activation of renal sympathetic nerves, the structural basis of which is evident in the rich postganglionic sympathetic innervation of the juxtaglomerular apparatus. Younger essential hypertensive patients with high renal sympathetic activity (manifest in elevated renal NA spillover) have "high renin essential hypertension" with markedly increased renal renin release but without reduction in renal blood flow.

### Cardiac and Vascular Neural Mechanisms

An increase in NA release from the heart in essential hypertension is likely to contribute to the hemodynamic profile of early human hypertension. The patient commonly has isolated essential hypertension, to which neural increases in stroke volume and left ventricular (LV) ejection rate, as well as a neurally mediated reduction in arterial compliance, all contribute. Experimental data suggest that NA exerts tropic effects on vascular smooth muscle cells and cardiomyocytes, thereby contributing to LV hypertrophy commonly present in hypertension. In line with these observations, clinical studies have demonstrated that increased sympathetic outflow to the skeletal muscle vasculature and NA spillover from the heart are directly related to human hypertensive LV hypertrophy, which in turn is a strong and independent risk factor for the development of arterial hypertension. Whereas neural mechanisms appear to initiate the increase in blood pressure, NA-induced LV hypertrophy may be one of the factors contributing to sustained blood pressure elevation.

## ORIGINS OF HIGH SYMPATHETIC TONE IN ESSENTIAL HYPERTENSION

### Genetics

Whether inherited higher levels of sympathetic nervous activity predispose to subsequent development of

hypertension in primary human hypertension has been little studied. In normotensive offspring of hypertensive parents, the MSNA response to mental stress was found to be more pronounced than that in offspring of normotensive parents. Normotensive young men with a family history of hypertension were reported to have higher rates of NA spillover to plasma than were young men with a negative family history of hypertension. The limited search undertaken for single-gene abnormalities involving the sympathetic nervous system in patients with high blood pressure, such as for polymorphisms of the  $\alpha_2$ -adrenergic receptor and NA transporter genes, has not disclosed abnormalities that could account for the consistent increase in spillover of NA or the impaired neuronal NA reuptake.

### Lifestyle Influences

Overweight, physical inactivity, and mental stress all have proven stimulatory effects on the sympathetic nervous system, and this might be important in mediating the blood pressure elevation. Patients with primary hypertension are commonly overweight. An excessive dietary energy load is known to stimulate the sympathetic nervous system and elevate arterial pressure. In obese individuals, a selective activation of the sympathetic nerves to the kidneys and skeletal muscle vasculature has been demonstrated, and this could be driven by high plasma levels of leptin. An additional factor contributing to sympathetic nervous overactivity in hypertensive patients, and particularly evident in obese individuals, is a sedentary lifestyle. Regularly performed physical exercise produces long-term lowering of blood pressure that is caused by inhibition of the sympathetic nerves of the kidneys. Exposure to ongoing mental stress has recently been acknowledged as a proven cause of hypertension by an Australian government body. In addition to epidemiological evidence, this body took into account the findings of sympathetic activation, adrenaline cotransmission, and activation of suprabulbar subcortical projections of brainstem noradrenergic neurons that were considered to provide biological evidence of ongoing mental stress.

### Central Nervous System Control of Sympathetic Nervous Activity in Hypertension

There is growing evidence of a disturbance in central nervous system (CNS) monoaminergic control of sympathetic outflow, and this may be the common

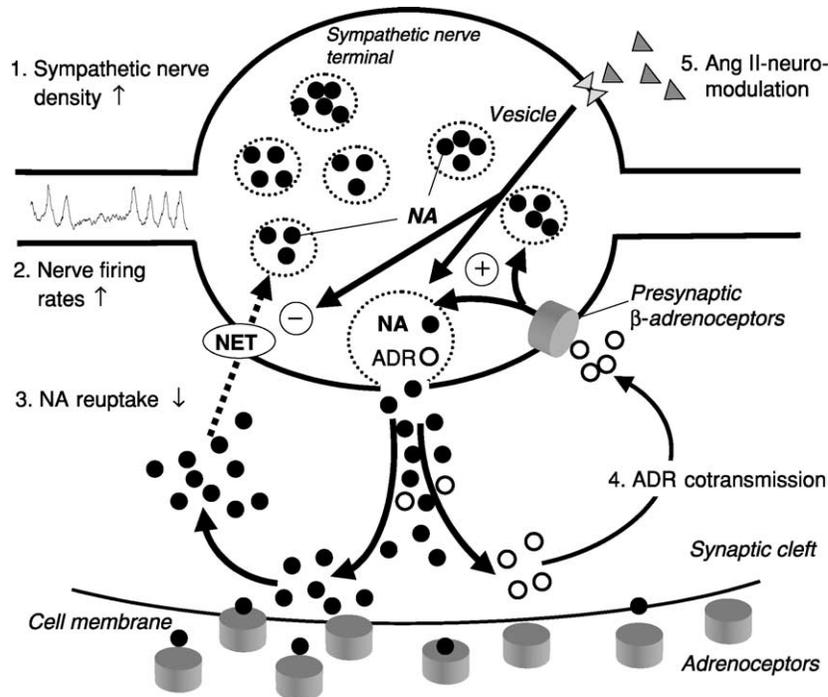
mediating mechanisms of peripheral sympathetic activation with stress, obesity, and physical inactivity. Catecholaminergic neurons releasing NA are widely distributed in the brain and are located in particular in the medulla and pons. Overflow of NA and its lipophilic metabolites, dihydroxyphenylglycol (DHPG) and 3-methoxy-4-hydroxyphenylglycol (MHPG), from the brain into the internal jugular veins is higher in patients with essential hypertension. Cerebral venous scans indicated that the increased turnover of NA in hypertensive patients is in subcortical brain regions only. In patients shown to have increased CNS NA turnover, peripheral sympathetic activity was increased. These findings suggest a primary importance in pathogenesis of increased neuronal firing in brainstem neurons projecting to forebrain noradrenergic pressor areas.

### Sympathetic Nerve Density

Spontaneously hypertensive rats exhibit an increase in the density of sympathetic innervation, probably related to increased neurotropic stimulation due to overexpression of nerve growth factor (NGF). Whether such mechanisms are operational in human hypertension has not yet been delineated. Sympathetic hyperinnervation would be expected to lead to increased interstitial concentration of NA in tissues and increased spillover of NA into the venous drainage of individual organs even at normal rates of nerve firing (Fig. 2). Intraneuronal stores of NA would also be expected to be elevated. However, preliminary data from our laboratory suggest that NA stores, estimated by measuring the specific activity of tritiated DHPG (the intraneuronal metabolite of tritiated NA) produced by the heart, are decreased in essential hypertension. Similarly, the transcatheter venoarterial gradient of NGF appears to be reduced in hypertensive individuals. Whether these findings are indicative of reduced sympathetic innervation as an adaptive response to increased sympathetic tone has been under investigation.

### Sympathetic Cotransmission

Adrenaline, the principal hormone of the adrenal medulla, is also present in low concentrations in extra-adrenal tissues largely contained within sympathetic nerves. Adrenaline within sympathetic nerves may be released with NA as a cotransmitter, facilitating the release of the major transmitter through stimulation of presynaptic  $\beta_2$ -adrenoceptors on sympathetic nerves and increasing the amount of NA released per nerve impulse (Fig. 2). It has been demonstrated that



**Figure 2** Schematic illustration of possible mechanisms underlying the increased spillover of noradrenaline (NA) to the circulation in patients with neurogenic essential hypertension. Sympathetic hyperinnervation, if present in patients with essential hypertension, would be expected to lead to increased spillover of NA (1). Increased rates of nerve firing results in enhanced release of NA from vesicles in the sympathetic nerve terminal (2). Impaired neuronal reuptake, possibly due to dysfunction of the noradrenaline transporter (NET), increases the amount of NA in the synaptic cleft and could amplify the neuronal signal (3). Corelease of adrenaline (ADR) and subsequent stimulation of presynaptic  $\beta_2$ -adrenoceptors could facilitate the release of the main neurotransmitter, NA (4). Angiotensin II (Ang II) may exert excitatory effects on the sympathetic nervous system by facilitating NA release from sympathetic nerve terminals or inhibiting NA reuptake (5).

adrenaline is released from the heart in patients with essential hypertension and that a proportionality exists between the rate of cardiac adrenaline and the rate of NA release. These findings, in conjunction with a previous report of adrenaline release from the renal sympathetic nerves in essential hypertension, provide perhaps the most direct evidence in support of the “adrenaline hypothesis” of essential hypertension, which draws on the concept that stress is a major factor in hypertension pathogenesis, with stress-induced elevations in the plasma concentration of adrenaline enlarging the pool of adrenaline present in sympathetic nerves, leading to release of adrenaline as a cotransmitter, facilitation of NA release, cardiovascular stimulation, and development of arterial hypertension.

### Impaired Neuronal Noradrenaline Reuptake

Neuronal NA reuptake may be impaired in essential hypertension, perhaps due to dysfunction of the NA

transporter, and might contribute to the development of essential hypertension (Fig. 2). We have further tested this proposition by applying specific radiotracer methods and using pharmacological blockade of NA transport. The fractional extraction of plasma-tritiated NA in passage through the heart, determined mainly by neuronal NA uptake, was found to be reduced in essential hypertensive patients, as was the cardiac release of the tritiated NA metabolite, tritiated DHGP, produced intraneuronally by monoamine oxidase after uptake of tritiated NA by the transporter. The reduction in NA transport in response to its blockade with the tricyclic antidepressant desipramine was found to be less pronounced in hypertensive patients. These findings strongly suggest that neuronal reuptake of NA is impaired in essential hypertension. By amplifying the neural signal, such a defect could constitute a neurogenic variant of essential hypertension. No mutation in the NA transporter gene that might explain this phenotype has been found.

## Angiotensin II Neuromodulation

A body of experimental evidence suggests that angiotensin II exerts excitatory effects on the sympathetic nervous system by stimulating sympathetic outflow, facilitating NA release from sympathetic nerve terminals, amplifying adrenergic receptor responsiveness to various stimuli, and inhibiting NA reuptake (Fig. 2). Despite evidence from animal and human studies to support the hypothesis of a functional cardiac renin-angiotensin system, it remains to be determined whether there is any functional effect on the cardiac sympathetic nerves in humans. The clinical course of heart failure has been associated with a progressive increase in cardiac angiotensin II formation that may contribute to the substantial increase in cardiac sympathetic drive characteristic of that condition. The limited data available in essential hypertensive patients suggest that angiotensin II, at least in the heart, does not contribute substantially to sympathetic augmentation present in essential hypertension.

## DIAGNOSTIC PITFALLS

Although a combination of elevated blood pressure and increased heart rate suggests the presence of neurogenic hypertension, sophisticated measures of regional sympathetic nervous activity are needed to confirm the diagnosis. There are several possible sources of confusion. An important example is pheochromocytoma, in which resting heart rate is often elevated. The presence of a catecholamine-producing tumor must be suspected if there is a history of episodic symptoms of palpitations, sweating, headaches, and/or chest or abdominal pain, with hypertension being paroxysmal or sustained, and then must be confirmed biochemically by testing for increased urine or plasma concentrations of catecholamines or their metabolites. These generally are substantially higher

than the concentrations typically seen in neurogenic forms of essential hypertension, in which plasma and urine NA levels are rarely elevated sufficiently to cause diagnostic uncertainty. Patients suffering from panic disorder can also cause diagnostic difficulties. During panic attacks, these patients exhibit a substantial increase in blood pressure, heart rate, sympathetic nerve activity, and epinephrine secretion but exhibit no such features between attacks. Similarly, patients with “white coat hypertension” display an alerting response to medical examination; however, in contrast to patients with neurogenic essential hypertension, these patients have normal blood pressure values on 24-h ambulatory blood pressure monitoring.

## See Also the Following Articles

Hypertension and Diabetes • Hypertension, Endocrine • Hypertension, Overview • Norepinephrine Receptors • Norepinephrine Transporter

## Further Reading

- Esler, M. (1997). Sympathetic activity in experimental and human hypertension. In “Handbook of Hypertension” (G. Mancin and A. Zanchetti eds.), “Pathophysiology of Hypertension,” vol. 17, pp. 628–673. Elsevier, Amsterdam.
- Esler, M., Jennings, G., Lambert, G., Meredith, I., Horne, M., and Eisenhofer, G. (1990). Overflow of catecholamine neurotransmitters to the circulation: Source, fate, and functions. *Physiol. Rev.* **70**, 963–985.
- Esler, M., Rumantir, M., Kaye, D., and Lambert, G. (2001). The sympathetic neurobiology of essential hypertension: Disparate influences of obesity, stress, and noradrenaline transporter dysfunction? *Am. J. Hypertens.* **14**, S139–S146.
- Floras, J. S. (1992). Epinephrine and the genesis of hypertension. *Hypertension* **19**, 1–18.
- Goldstein, D. S. (1983). Plasma catecholamines and essential hypertension: An analytical review. *Hypertension* **5**, 86–99.
- Julius, S. (1996). The evidence for a pathophysiologic significance of the sympathetic overactivity in hypertension. *Clin. Exp. Hypertens.* **18**, 305–321.

(3) total blood volume, and (4) tension generated by the walls of the largest blood vessels (termed conduit arteries) in resisting the pulse of blood ejected into the arteries by the heart.

## Vascular Compartments

All multicellular organisms, from the most primitive to humans, have one key requirement for survival: a functional pathway whereby all of the organism's cells may be reached so that nutrients and critical gases (e.g., oxygen) may be distributed and waste products (e.g., carbon dioxide) may be removed. Evolution has matched more complex organisms with comparably complexed circulatory systems. In mammals, blood, containing both cells and dissolved materials, circulates continuously through an extensive closed network of tubes (blood vessels). This vascular system actually consists of three separate but connected sections: arterial (arteries), capillaries, and venous (veins).

The arterial compartment carries blood away from the heart, beginning with the aorta and extending, in an ever-branching manner, to all of the organ systems and tissues. The arterial system is a high-pressure circuit, and the large conduit arteries receive the full pressure and flow of the blood pumped from the left ventricle of the heart. Conduit arteries serve as "feeder vessels" to an increasing number of smaller vessels ultimately connecting with the capillary compartment. Arterial pressure is highest in the aorta and other conduit vessels and declines as the vessels become smaller, becoming lowest (in the arterial circuit) at the junction with the capillary network. Conduit arteries, such as the aorta and carotid, also serve to dampen the large "pressure pulse" generated by contraction of the left ventricle and function to smooth out the pressure wave that spreads from the heart to the smallest blood vessels. One way in which they accomplish this is through the elasticity of the vessel wall, which "captures" some of the energy of the pulse of blood ejected by the heart into the arterial compartment. As the pressure wave progresses away from the heart, the conduit arterial walls relax and return the energy to the blood, supporting the arterial pressure while the heart is not contracting. When conduit arteries lose their elasticity, systolic hypertension, a disorder common in the elderly, may result.

Arterial blood vessels are complex and consist of multiple cell types and several discrete layers. In immediate contact with blood is a single cell-thick lining of endothelial cells with underlying connective tissue protein, a layer termed the intima of the vessel.

Arterial endothelial cells serve many functions, including local release of chemicals that may relax or contract the vessel wall and regulation of hemostasis and clot formation. The intima is the site of atherosclerotic plaques and has assumed increasing importance in studies to identify the causes of primary genetic hypertension. The middle layer of the arterial vessel wall is termed the media and consists primarily of multiple layers of contractile smooth muscle cells. The state of contraction (termed tone) of the smooth muscle cells imparts strength to the wall to resist the blood pressure within the lumen of the blood vessel and, at the terminal ends of the arterial compartment, comprises the "resistance elements" that determine the rate of flow of blood out of the arterial compartment into the capillary network. Systemic arterial resistance is a measure of the average tone (contraction state) of all the resistance elements of the arterial compartment. The outer layer of the artery wall is the adventitia, consisting of connective tissue cells such as fibroblasts, elastic and connective tissue proteins, nerve endings of the sympathetic branch of the autonomic nervous system (ANS), and other cell types that vary according to the size of the vessel and the organ system.

The capillary network includes the smallest blood vessels in all three compartments of the vascular system. These vessels consist of only a single-cell layer of endothelial cells with associated connective tissue proteins and no media or adventitia. This structure permits rapid diffusion of materials, in both directions, between the blood within the vessel and fluids in the adjoining tissue. This vascular compartment is primarily a passive circuit for exchange; however, it has significant importance in maintaining the amount of fluid outside of the vascular compartments and in tissues. An imbalance in fluid exchange, due to altered control of fluid dynamics between the arterial and venous compartments, contributes to tissue fluid retention in individuals with compromised heart function and in uncontrolled hypertension. Although antihypertensive drugs are not designed to work directly on the capillaries, these drugs do affect the capillaries' functioning indirectly through improved fluid dynamics and enhanced elimination of retained fluids through increased urine formation in the kidneys (diuretics).

The third compartment is the network of vessels termed the venous compartment, which collects the output of the capillary compartment. The venous compartment commences at the outflow from the capillary compartment and extends to the right atrium of the heart. There are two functions of this compartment

that directly influence arterial pressure: (1) collecting the blood from the capillary compartment and returning it to the heart (termed venous return) and (2) acting as a reservoir of blood in the body. The venous compartment is a mirror image of the arterial compartment with small veins, highly ramified, joining into successively larger vessels. Veins are complex vessels, just like arteries, and consist of an endothelial cell lining, medial smooth muscle, nerves, and multiple types of cells; however, the venous system is a low-pressure circuit, so the media and adventitia of the vessel wall are thinner than arteries. Large veins are termed capacitance vessels because they have the capacity to expand and hold large amounts of blood. When needed, contraction of the capacitance vessels (through contraction of venous smooth muscle) allows the recruitment of blood back into the arterial and capillary compartments. This is accomplished by the formation and release within the body of chemicals, which contract the smooth muscle, and through the activity of the sympathetic branch of the ANS.

Among other vascular compartments that are unique and are not addressed in this article are the pulmonary and cerebral vascular compartments. Blood vessels connecting the heart and lung constitute a special (pulmonary vascular) compartment. Whereas arteries from the left ventricle of the heart to all of the organ systems (except the lungs) carry highly oxygenated blood and veins returning to the right atrium of the heart carry low-oxygenated blood from the tissues, the converse exists in the pulmonary vascular compartment. The pulmonary artery, from the right ventricle of the heart to the lungs, carries venous blood of low oxygen content, whereas the pulmonary vein, from the lungs to the left atrium of the heart, carries highly oxygenated blood. Pulmonary hypertension is excess blood pressure in the pulmonary artery from the heart to the lung. It is not equivalent to systemic arterial hypertension, the subject of this article, and generally has other causes and treatment. For this reason, this article does not deal with this special form of hypertension. The cerebral vascular compartment is also unique in being both autoregulatory (i.e., it responds to changes of pressure within the compartment so as to maintain blood flow to the brain constant) and generally unresponsive to chemicals and neurohormones in the blood.

### Arterial Blood Pressure Gradient

Blood pressure is highest in the large conduit arteries, which leave the left ventricle of the heart, and

decreases steadily through the arterial, capillary, and venous compartments; however, the largest drop in pressure occurs on the arterial side. Maintenance of this decreasing gradient of pressure is essential to ensure adequate blood delivery to all tissues, to optimize fluid and material exchange in the capillary network, and to deliver the blood back to the right side of the heart. A disease process that interferes with any of these three functions will invariably result in a form of hypertension.

### Determinants of Levels of Arterial Pressure

Hypertension is a disorder of regulation of systemic arterial pressure, which itself is set and regulated by multiple organ systems.

Arterial pressure derives from the pumping action of the left ventricle of the heart; therefore, the level of arterial pressure at any point in the arterial vascular compartment reflects on the functioning of the left ventricle. During each contraction of the left ventricle, the highest systemic pressure generated within the arteries is termed the systolic pressure. When the heart valve controlling outflow from the left ventricle closes and the left ventricle relaxes (between beats), the arterial pressure drops as the arterial blood rapidly flows out of the arterial compartment into the capillaries. The rate of drop of pressure is controlled by the terminal arterioles and by the energy being returned to the blood with relaxation of the walls of the large conduit arteries, a process termed the windkessel effect and related directly to the elasticity (termed compliance) of the conduit arteries. The windkessel process is very much like the stretched rubber band of a slingshot rebounding and exerting force on the object being propelled. The lowest systemic arterial pressure level is reached just prior to the next contraction and is termed the diastolic pressure. Thus, systolic pressure reflects the action of the heart, resistance to outflow from the arterial compartment, and the windkessel effect, whereas diastolic pressure is set by the rate of outflow (resistance set by the arterioles) and the time between contractions (the “interbeat” interval or heart rate). At constant arteriole resistance, increasing heart rate may increase apparent diastolic pressure. Diastolic pressure also tracks systolic pressure given that an increase in systolic pressure sets a higher starting point from which the arterial pressure may descend between contractions. The pressure difference between systolic and diastolic pressure is termed the pulse pressure.

Pulse pressure is assuming greater research interest as a potential contributor to the development of systemic hypertension and damage to the arterial wall leading to atherosclerosis.

Levels of systolic and diastolic pressure are not constant over time but rather vary continuously, on a beat-by-beat basis, even during rest and sleep. Arterial pressure depends on many factors, including age, gender, body weight, level of physical conditioning, current physical activity, and behaviors of all kinds (e.g., eating, drinking). Of course, arterial pressure is also influenced by many drugs, including prescription drugs, over-the-counter drugs, and drugs of abuse. Human systemic arterial pressure is usually measured with an occlusive device (cuff) placed on one or both arms. When arterial pressure is measured in this manner, both upper and lower values are quoted (e.g., 120 over 80, systolic over diastolic). Rather than systolic and diastolic, we may also speak of mean arterial pressure (MAP), which is the average pressure between systolic and diastolic pressure. MAP, when averaged over time, is defined by the following relationship involving cardiac output (CO) and total systemic vascular resistance (TSVR):  $MAP = CO \times TSVR$ . TSVR is the sum total resistance to the flow of blood out from the arterial compartment and reflects the action of all terminal arterioles. CO is the amount of blood (in liters) pumped by the left ventricle of the heart over a full minute. This volume of blood is determined by the force of contraction of the left ventricle, the heart rate, and the amount of blood contained in the left ventricle chamber during each contraction. The latter is controlled in part by the amount of blood that returns to the heart from the venous compartment (termed venous return) and by the resistance encountered when the heart pumps the blood into the arterial circuit. Because capacitance veins influence venous return, changes in both blood volume and the degree of constriction of venous smooth muscle influence the low blood pressure in the veins and the amount of blood returned to the heart. Because CO is defined by volume of blood ejected by the left ventricle with each beat (termed stroke volume) and by heart rate, arterial pressure is determined by stroke volume, heart rate, and TSVR.

Within all organisms, arterial pressure is set and regulated by many factors, most of which are integrated through mechanisms of information exchange, both nervous system and chemical. The major system that regulates and sets arterial pressure is the ANS, which works in an integrated fashion with the central nervous system (CNS). Both branches of the ANS,

sympathetic and parasympathetic, work together in an integrated fashion to control arterial pressure. Some studies indicate that the two systems work in opposition, with one stimulating (sympathetic) and the other inhibiting (parasympathetic) to achieve regulation of arterial pressure and heart action. However, a more accurate view is that the two systems work together to achieve the ultimate goal, namely to permit the organism to survive and accomplish whatever it seeks to do. It is important to appreciate this concept to understand the importance of blood pressure dynamics. The sympathetic system is generally considered the stress-responsive branch of the ANS because it alters organ system functions to optimize an organism's response to stress, whether the stress arises externally or internally. The parasympathetic system is considered the "vegetative" branch of the ANS, regulating the most primitive and essential biological actions necessary for survival of the organism and the species. The sympathetic system (1) can increase heart rate and force of contraction; (2) can increase tension (tone) of the smooth muscle in the terminal arterioles, thereby decreasing the rate of outflow of blood from the arterial compartment and increasing systemic vascular resistance; (3) will stimulate release of chemicals from the kidney and adrenal glands that are important for control of blood volume, blood electrolytes, and contraction or relaxation of smooth muscle in the arteries and arterioles; and (4) controls a myriad of additional functions from metabolism, to functioning of eyes, to sexual functions. One of the most important functions of the sympathetic system is shifting the flow of blood between organ systems to meet the needs of the tissues. Each organ system gets a fraction of total CO; however, during some functions of the individual, one organ system might need more. This is accomplished by the CNS through a selective increase in sympathetic nerve activity to particular organ systems that do not need the flow (at that time) and a decrease in nerve activity to organ systems that need more blood. The parasympathetic system controls many organ systems so as to maintain normal homeostasis in the absence of stress. For example, the parasympathetic system slows the heart, increases gastrointestinal activity and secretion to aid digestion, facilitates elimination of waste products from the body, protects the lungs from inhaling toxic chemicals and substances, protects the retina from excessive light, and facilitates vision at short distances. Both the sympathetic and parasympathetic branches of the ANS project from the CNS to the heart; however, only the sympathetic system sends nerve projections to blood vessels.

The ANS originates in the CNS and is intimately linked, through short and long nerves, to the parts of the brain that are important in coordinating cardiovascular and respiratory functions (brainstem) as well as parts that are important for primitive and complex behaviors and even cognition. Each behavior or action of an individual requires an appropriate and selective autonomic response; otherwise, the organism could not perform the desired action. For example, “fear” generally increases sympathetic activity and diminishes parasympathetic activity. Yet, although both fear from an external threat and fear deriving from an “internal” cognitive (perceived) threat may result in activation of a sympathetic response (e.g., increased heart rate), the specific changes in autonomic functioning are not the same. Thus, one cannot generalize and say that all fear responses will have the same effect on the cardiovascular system; some may be more demanding or even more detrimental than others. The relationships between behavior and normal or abnormal cardiovascular functioning are just recently being elucidated, and such studies comprise an area of investigation termed behavioral–autonomic coupling. That such coupling is dictated by genes, and thus becomes controlled partly by inheritance, has been established recently through studies within our laboratory. Could an individual inherit genes that lead to aberrant behavioral–autonomic coupling?

Systemic arterial pressure also exhibits a diurnal rhythm that is generally higher during the awake/day period and lower during the rest/sleep period. When an individual’s arterial pressure decreases from a high during the active period to a low during the rest period, the individual may be classified as a “dipper.” Interestingly, many human hypertensives exhibit a failure to “dip” and are termed “nondippers.”

The endocrine system has direct and indirect effects in determining the levels of systemic arterial pressure. Steroids, both gonadal and adrenal cortical, exert direct influences on all of the cellular components of the arterial compartment (including smooth muscle and endothelial cells), on functioning of the kidneys that relates to retention of sodium and water, on the actions of the heart, and especially on functioning of the CNS. Endocrine systems are linked with control of diurnal (circadian) rhythm and will directly influence the CNS (including cognitive areas). Furthermore, given that every behavior must have an appropriate autonomic and cardiovascular response, it is clear that subtle endocrine-mediated changes in behavior, when exerted over an extended period, may have profound effects on the level of systemic arterial pressure.

## HYPERTENSION AS A DISORDER IN HUMANS

Because arterial pressure is determined by the functioning of the heart, the state of resistance to outflow from the arterial compartment, the state of constriction of veins, the total volume of the system, behavioral–autonomic coupling, and the endocrine state, hypertension could be the result of potentially many different factors. However, there are numerous integrated control mechanisms that function to compensate for changes in one or more of these factors so as to keep arterial pressure constant. For example, if TSVR increases and leads to increased arterial pressure, CO is reflexively decreased through homeostatic mechanisms that increase parasympathetic activity to slow heart rate and reduce contractility. This is mediated by a nervous system pathway known as the high-pressure baroreceptor reflex. Likewise, if CO drops (e.g., due to a sudden loss of blood), there is a homeostatic increase in sympathetic activity to blood vessels, resulting in strong constriction of smooth muscle in resistance arterioles and thereby a concomitant increase in TSVR, with maintenance of diastolic pressure. These homeostatic regulatory systems operate continuously and dynamically on a beat-by-beat basis.

Hypertension is considered a disorder of regulation, in part because in hypertension many homeostatic mechanisms appear to continue to operate but “reset” to a higher level of arterial pressure. Therefore, a resetting of the level of homeostasis permits sustained elevated arterial pressure. In all hypertension, the baroreceptor reflex resets to maintain arterial pressure at a higher level; thus, with initial drug therapy for hypertension, the cardiovascular system resists lowering arterial pressure. With chronic elevation of arterial pressure, there is a compensatory thickening of the smooth muscle layer in the media of arteries to resist the increased pressure within the vessel. This thickening causes the lumen of the vessel to become smaller, and this itself raises the resistance to flow (increases TSVR). If it is not reversed in time, this thickened lumen leads to a resetting of the basal state of resistance to flow. Furthermore, over time, there is a decrease in the compliance of the conduit vessels and a failure to dampen out pulse pressure. Changes also take place in the heart, termed hypertrophy, which can progress to heart failure if untreated.

Arterial pressure is termed a quantitative trait because there is not a single value for normal arterial pressure. Rather, arterial pressure is continuously changing depending on what the individual is doing; however, it tends to range about an average, with

transient excursions outside of that range. To define the state of hypertension, researchers have measured large numbers of individuals to obtain a population average as the reference value for normal pressure; however, it must be remembered that many individuals may be higher, and many may be lower, and still be categorized as normal. An individual's arterial pressure must be compared with the population average for comparable individuals. For example, gender is important with differences between age-matched males and females. Furthermore, the average arterial pressures of normally cycling females are generally lower than those of age-matched males or postmenopausal females. Blood pressure is also influenced by body mass or weight and by the extent of physical conditioning. A blood pressure that is considered normal in one individual may be considered abnormal in another individual. The well-known values of 120 mm Hg systolic and 80 mm Hg diastolic entered our vocabulary to imply normal blood pressure; however, these values are, at best, approximations for the population of adults as a whole. Biological variability results in a range of normal arterial pressures; therefore, a complete history and physical examination with appropriate and repeated testing are essential in defining the disorder of hypertension.

There is a consensus among researchers and physicians that diastolic pressures for adults should be below 90 mm Hg and that systolic pressures should not exceed 140 mm Hg. For this reason, these have become the levels for defining a state of clinical hypertension. Critical to this definition is that the pressures be measured several times, over weeks, in both the physician's office and at home (if possible) and that all potentially extenuating circumstances or influences be defined and excluded, to the extent possible. In general, for all but the elderly, diastolic pressure measurements are considered a more important guide for classifying the disorder than are systolic pressure measurements. Why? Although elevated systolic pressures (160 mm Hg or higher) alert to the possibility of hypertension and, in hypertensive patients, provide valuable information on adverse effects to the cardiovascular system, systolic pressure is very sensitive to environmental factors such as stress and anxiety. Recall that systolic pressure is determined in large part by the action of the heart, which can readily respond to states of stress and anxiety. In contrast, diastolic pressure is determined largely by systemic vascular resistance and blood vessel compliance. Therefore, a sustained elevation of diastolic pressure would more likely reflect irreversible changes in the

structure of arterial vessels, heightened vascular resistance, or a significant and chronic alteration in homeostatic regulation. An excessive diastolic pressure also will result in hypertrophy of the left ventricle, an event that could contribute to a compensatory increase in TSVR.

When does an individual become classified as having hypertension? Studies of the adult population have resulted in guidelines, and in 2003 the National Heart Lung and Blood Institute issued the Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure, the JNC7 Report ([www.nhlbi.nih.gov/guidelines/hypertension/jncintro.htm](http://www.nhlbi.nih.gov/guidelines/hypertension/jncintro.htm)). The JNC7 Report argues that healthy adults should have a target average (awake) systolic/diastolic pressure not exceeding 140/90 mm Hg, whereas that of individuals with diabetes or chronic kidney disease should not exceed 130/80 mm Hg. The JNC7 Report also addresses appropriate therapeutic interventions for individuals with measurements exceeding 140 systolic or 90 diastolic. There is little disagreement that a sustained diastolic pressure exceeding 95 mm Hg or a systolic pressure of 160 mm Hg warrants appropriate therapeutic intervention. But what about the individual whose diastolic falls within the range of 90 to 94 mm Hg, considered as a borderline hypertensive? It is increasingly accepted that such an individual should also be considered as a candidate for therapeutic intervention. However, such an individual should not necessarily be considered for antihypertensive drugs; rather, nondrug interventions may be more appropriate.

How is hypertension identified? Although many people believe that headaches and rapid heartbeat are indicators of hypertension, these symptoms are not most commonly associated with the disorder. In fact, hypertension has been termed the "silent killer" because a patient may be hypertensive for years without any evident symptoms. Within recent years, the importance of hypertension as a major risk factor for atherosclerosis, heart attack, stroke, and kidney failure has raised awareness among all health professionals to identify hypertensive individuals. Ophthalmologists and opticians often discover hypertensive patients while examining the retina (the back inner surface of the eye). The retina is highly vascularized, and chronic untreated hypertension damages the retinal vascular bed (damage that can be readily observed on examination). However, atherosclerosis and diabetes may cause similar changes in the retina, and such conditions must be excluded. Yet given the high frequency of untreated hypertensives who develop

diabetes or atherosclerosis, the possible concurrence of these disorders is more likely the norm rather than the exception.

Identification of hypertension during routine physical examination must be rigorous and thorough. A single finding of elevated diastolic or systolic pressure does not constitute sufficient proof to initiate therapy, whereas dismissal of borderline readings of heightened arterial pressure without follow-up also places the patient at risk for future disease. Accurate measurement of arterial pressure is not trivial. Nearly all methods of measurement employ a form of the sphygmomanometer technique, an indirect method using the cuff occlusion. This involves measurement of the external pressure required to occlude, totally or in part, blood flow in a readily accessible artery such as the brachial artery of the upper arm. The procedure involves wrapping a correctly sized rubber cuff around the upper arm and inflating it to stop the flow of blood through the major artery. The individual taking the reading then listens for the resumption of blood flow as the pressure is lowered; in the absence of flow, no sounds are heard. Two characteristic sounds (termed Korotkoff sounds) are monitored. The first is the start of flow in the vessel as the cuff pressure is lowered and defines systolic (maximal) pressure, whereas the second is when no flow sounds are present and indicates no restriction on flow. This defines the level of diastolic (minimal) pressure. Blood pressures are expressed in terms of millimeters of mercury because the first, and still the most accurate, sphygmomanometers use a column of liquid mercury to determine air pressure. It is essential that the cuff fit the individual; too small a cuff may give an anomalously high value, whereas too large a cuff may give a falsely low value. Arterial pressure is measured while either sitting or standing (termed upright) or when the individual is in a lying-down (termed supine) position. Measurements should be taken in a calming environment to minimize stress and anxiety effects. Although automated cuff arterial pressure measuring devices found in pharmacies and food stores may be quite reliable, they cannot be used to exclude or diagnose hypertension.

A well-described phenomenon termed white coat hypertension is an elevated arterial pressure whenever pressure is measured in a physician's office or clinical setting. This form of apparent hypertension actually reflects an overreactive cardiovascular response to the stress or anxiety associated with the environment. White coat hypertension often disappears when the pressure is taken in a less forbidding environment. The use of 24-h ambulatory monitoring of arterial

pressure is being more widely used to exclude white coat hypertension and to identify extremes of arterial pressure over a typical 24-h activity day.

The most common form of human hypertension, estimated to be responsible for more than 80% of all hypertension worldwide, is of unknown cause and is called primary or essential hypertension. Some investigators believe that essential hypertension reflects a variety of forms of hypertension, each due to a different abnormality of arterial pressure and/or volume regulation. In contrast, others believe that essential hypertension is one syndrome that can arise from a variety of different abnormalities in regulation. Essential hypertension is considered to be a genetic disorder, with a strong sensitivity to environmental factors, and includes individuals who exhibit salt-dependent or obesity-dependent forms of hypertension if no other cause can be ascertained. Genetic hypertension in humans (and rodents) is due to the inheritance of multiple genes that predispose the individual (or organism) to develop the disorder, usually after puberty. Essential hypertension is age dependent in symptom expression; however, studies indicate that elevated arterial pressures may be observed even at very early ages, leading to increased efforts to evaluate arterial pressure in children.

A much smaller percentage of individuals (estimated at 10–25%) have hypertension that is attributable to a known or identifiable cause, most often of endocrine-related origins. This form of hypertension is termed secondary. Renovascular hypertension, believed to be responsible for 5 to 8% of all human hypertension, is caused by a pathophysiological process that affects the kidney and leads to activation of the renin–angiotensin system. Often, renovascular hypertension stems from diminished or restricted arterial flow to one kidney, frequently caused by a tumor that impedes vessel flow. Removal of the tumor may reverse the hypertension provided that the elevated arterial pressure was not of sufficient duration to lead to irreversible changes in the media of arteries. Another potential cause of secondary hypertension is a rare tumor of the adrenal medulla, a pheochromocytoma, which episodically releases large quantities of the neurohormone epinephrine into the bloodstream. Such patients exhibit episodes of very high arterial pressure. Again, removal of the tumor before it metastasizes may reverse the hypertension if irreversible changes in the arteries or kidney have not occurred.

Hypertension may also result from taking drugs that exert their direct actions on the kidney, arterial or venous smooth muscle, or endocrine systems. For example, oral contraceptive-induced hypertension

became widespread shortly after the birth control pill was released for general use. This form of secondary hypertension has been attributed to sex steroid induction of the synthesis by the liver of angiotensinogen, the protein precursor of the angiotensin peptide. Changing the composition of the pill has greatly reduced the incidence of this drug-induced hypertension. The active peptide of angiotensinogen, angiotensin II, is a very potent peptide hormone whose primary functions include salt and water retention and an increase in the state of contraction of arterial smooth muscle. Excessive formation of angiotensin II may result in many actions, including increased synthesis and release of the steroid aldosterone from the adrenal cortex that directly drives salt and water retention, heightened TSVR, and blunting of the high-pressure baroreceptor reflex. Another form of secondary hypertension strikes some women during the second and third trimesters of pregnancy and poses significant risks for both the fetus and the mother. Although the causes of pregnancy-induced hypertension remain controversial, this too is a secondary form of hypertension because it generally reverses following delivery. There is evidence that these latter forms of secondary hypertension are under genetic influence; that is, inheritance of specific genes may predispose the individual to these forms of hypertension. If future genetic studies prove an inheritable tendency, the possibility for prevention should be greatly enhanced.

Malignant hypertension is a term used for patients who show extremely high blood pressure and frequently irreversible hypertension. This term does not imply a cancerous origin to the disorder; rather, patients with malignant hypertension are generally those who have had untreated high blood pressure for long periods of time or who fail to respond to drug therapy. Often, malignant hypertension is associated with severe kidney failure and/or severe changes in the arterial vascular compartment. Such hypertensive patients exhibit very high sustained arterial pressure, both systolic and diastolic, and constitute a medical emergency necessitating bold drug treatment to lower the blood pressure and prevent further organ damage and heart failure.

Today, all physicians and health care professionals are urged to identify potential hypertensive individuals through family histories. Evidence of hypertension or cardiovascular disease and death in the family history should result in annual monitoring for elevations in arterial pressure or other cardiovascular risk factors. This is especially the case for individuals whose ethnic backgrounds show increased expression

of essential or salt-sensitive hypertension or when other risk factors for hypertension (e.g., obesity) are present.

## GENETIC PREDISPOSITION TO HYPERTENSION

Essential hypertension has a major hereditary or genetic component; that is, the potential for developing hypertension may be passed from one generation to another within a family. This means that there are hypertensive-prone individuals. However, because there is also a strong effect of environment, it is not clear whether all hypertensive-prone individuals will eventually become hypertensive. In fact, there is increasing evidence that hypertension may be significantly delayed in onset or even prevented by appropriate control of the environment, diet, lifestyle, and exercise/activity. Remember that hypertension develops with age, so changes in lifestyle that delay the development of elevated arterial pressure will like directly prevent its occurrence. Is it possible that earlier generations were hypertensive prone but that their lifestyle prevented the disorder from appearing until they were quite elderly? An interesting study conducted nearly 20 years ago with a strain of rat that spontaneously develops hypertension at 6 to 12 weeks of age (due to inheritance of hypertension genes) showed that if the animals were raised from birth in a stress-free environment under controlled dietary conditions, the rate of development of hypertension was greatly delayed and nearly prevented. Environmental effects also influence the state of the endocrine system, and altered hormonal secretions may be coupled to the rate of onset and progression of arterial hypertension.

## POTENTIAL CAUSES OF ESSENTIAL HYPERTENSION

Essential hypertension is a multigene disorder with strong contributions from environmental factors on disease development. Discovering the cause is complicated by the many factors that determine the level of arterial pressure, including disturbances in ANS activity, blood volume, action of the heart as a pump, compliance and elasticity of arterial vessels, behavioral-autonomic coupling, and endocrine regulation of the CNS. A complete discussion of these factors is beyond the scope of this article; however, a brief introduction is warranted because therapy is directed at one or more of these possible causative mechanisms.

There is ample evidence from both animal and human studies indicating that an elevation of activity of the sympathetic branch of the ANS is observed in all forms of hypertension. This system is critical to arterial pressure regulation. For example, if the sympathetic branch is blocked with drugs, individuals show a loss of arterial pressure control in going from a supine position to an upright one (termed postural hypotension). In addition to effects on the heart and blood vessels, the sympathetic system is important in autonomic regulation of glucose and fat metabolism, coupling hypertension with diabetes, atherosclerosis, and obesity. Perhaps most important, the sympathetic branch is a major mechanism for stimulating secretion of the proteolytic enzyme renin into the blood by the kidney. Renin enzymatically attacks angiotensinogen circulating in the blood and releases from angiotensinogen, angiotensin I, the peptide precursor of active angiotensin II. Angiotensin I, which is inactive, is converted to active angiotensin II by the enzyme angiotensin-converting enzyme (ACE) present throughout the vascular system. Thus, the sympathetic system acts directly on the heart and arterioles and indirectly via many other mechanisms, such as angiotensin, to promote salt and water retention. All of these actions can and do result in an elevation of arterial pressure. Interestingly, angiotensin II is also a potent inhibitor of the high-pressure baroreceptor reflex, which would normally reduce sympathetic activity if arterial pressure became too high. Finally, nearly all of the effective antihypertensive drugs in use today interfere, directly or indirectly, with sympathetic activity or its physiological effects.

In addition to autonomic control, regulation of systemic arterial pressure is achieved through electrolyte (e.g., sodium, potassium, chloride) and water homeostasis. Endocrine mechanisms are central to volume control mechanisms and operate both within the CNS and in peripheral organ systems. A very common finding with established essential hypertension and many forms of secondary hypertension is abnormalities in kidney function that lead to abnormal retention of sodium and, with it, water. Here again, angiotensin is a key hormonal system; however, adrenal cortical steroid hormones, such as aldosterone and vasopressin (secreted by the pituitary), are also involved in salt and water homeostasis. Yet aberrant electrolyte regulation may be independent of the angiotensin system given that one form of hypertension is characterized by low plasma renin activity but high sodium retention and elevated (termed expanded) blood volume. This form of hypertension may reflect abnormalities in kidney function or in

endocrine hormones that regulate kidney function and excretion. Restriction of sodium (or salt) intake can lower arterial pressure in most hypertensive patients, yet severe restriction may be counterproductive because too little sodium intake actually activates the renin and angiotensin systems. There is epidemiological evidence that excessive dietary salt intake in Western societies elevates population averages of arterial pressure, and many studies have demonstrated a positive relationship between the amount of salt consumed in the diet and the incidence of hypertension. Yet we now know, from human and rodent genetic studies, that salt intake and the effect of salt on arterial pressure are also genetically controlled. Furthermore, we have evidence that genes, which control salt sensitivity of arterial pressure, are inherited independently of genes for hypertension, at least in the rodent genetic hypertension model. Therefore, salt sensitivity of arterial pressure is an inherited trait; however, coinheritance of genes for salt sensitivity with one or more genes predisposing to hypertension may exacerbate the disorder and likely leads to onset of hypertension at an earlier age.

An attractive but unproven theory is that repeated exposure to stress in everyday life activities may result in abnormal cardiovascular responses in hypertensive-prone individuals. It is possible that repeated stress leads to irreversible changes in arterial blood vessels, resulting in a sustained elevation of peripheral vascular resistance and ultimately to hypertension. Behavioral stressors have been shown to result in adjustments in CO, and repeated changes in blood pressure due to stress may cause a resetting of the baroreceptor reflex associated with hypertension. There is genetic evidence that supports the effects of stress on increases in blood pressure. Individuals with family histories of hypertension exhibit heightened cardiovascular responses to both physiological and psychological stressors compared with those with no family histories of hypertension. Although it may be intuitive to expect that repeated exposure to stressful situations would have some adverse cardiovascular effects, there is no evidence supporting a hypertensive type A personality. The potentially deleterious effects of mental stress on the cardiovascular system may reflect an impaired "coping" mechanism. A failure to habituate to, or cope with, stress appears to pose an inherent risk factor in the development of hypertension; however, it remains uncertain whether it merely exacerbates the hypertension, increases its rate of development, or increases the risk of hypertension-induced organ system damage. Finally, the majority of research suggests that stress-induced increases in

arterial pressure and changes in CO that lead to the development of hypertension, over time, are the result of increased activity of the sympathetic nervous system. One reason why it is difficult to resolve the role of stress in the development of hypertension is the complex interplay of multiple endocrine mechanisms that regulate arterial pressure and blood volume. There is also a complex genetic component in both physiological and psychological coping or adjustment to stress that is in play and has not been resolved.

Not only arterial pressure and heart rate but also metabolic factors, such as blood glucose, insulin, other neurohormones, and endocrine secretions, exhibit inherent diurnal rhythms. Humans and rodents exhibit the property of lowered arterial pressure during rest (sleep) followed by an early-morning “anticipatory” increase in heart rate and arterial pressure (dippers). Many hypertensive individuals (and genetic hypertensive rats) exhibit abnormal arterial pressure and heart rate rhythms (nondippers). This abnormal rhythm is frequently associated with left ventricular hypertrophy, coronary artery disease, and increased damage to organ systems. Nondipper essential hypertensives can exhibit increased circulating norepinephrine and renin levels, decreased sleep patterns (perhaps due to altered melatonin levels), decreased ANS function, increased albumin (protein) excretion in the urine (which may hallmark increased pressure in the kidneys and kidney damage), and increased total and low-density lipoprotein (LDL) cholesterol. However, no differences have been reported in levels of insulin, leptin, or homocysteine between dippers and nondippers. The clinical importance of the absence of dipping is still under investigation; however, our laboratory has demonstrated that, at least in rodents, dipping and nondipping are not directly associated with the inheritance of high blood pressure. Nevertheless, the inability to lower arterial pressure during rest may contribute to long-term detrimental effects of hypertension.

The central biological or circadian clock relevant to cardiovascular control is believed to be located in the suprachiasmatic nucleus (SCN) of the CNS; however, there is increasing evidence of secondary “clocks” outside of the CNS and even in blood vessels and the liver. The SCN regulates bodily functions by sending daily rhythmic messages via neurohormonal signals and by direct neuronal connectivity to all organs of the body. The peptide hormones of the SCN that are crucial to its function appear to be vasopressin, vasoactive intestinal peptide (VIP), and neurotensin. Vasopressin is believed to act on rhythm oscillations, whereas VIP and neurotensin act as “pacemakers.” All three

peptides are decreased to nearly half their levels in the SCN of hypertensive patients compared with those in normotensives. In addition, the SCN has direct influence on autonomic outflow via neuronal connections with the paraventricular nucleus (PVN) of the hypothalamus that itself connects with the nucleus tractus solitarius (NTS) in the brainstem, the primary site that receives sensory information on the level of arterial pressure in the aorta and carotid arteries. Therefore, abnormal functioning of the SCN could lead to interference in the continuous flow of information to and from the NTS and eventual alteration in arterial pressure. However, the “chicken and egg” question remains unanswered: does genetic hypertension cause a pathological functioning of the SCN and effects on circadian rhythms, or is it the reverse? Studies in which SCNs were transplanted from a hypertensive rat to a normotensive rat showed that the transplant caused hypertension in the genetically normotensive animal. This may indicate that the SCN, responding to the environment, has a significant role in genetic hypertension.

## THERAPY OF HYPERTENSION

Hypertension is classified according to the level of diastolic pressure with three categories relevant to the discussion: borderline hypertension (diastolic pressures between 90 and 94 mm Hg), mild hypertension (diastolic pressures between 95 and 104 mm Hg), and moderate to severe hypertension (diastolic pressures exceeding 104 mm Hg). These three categories are loosely defined because some patients are at greater risk than others for developing severe cardiovascular complications with any degree of elevated arterial pressure. Furthermore, essential hypertension is a chronic disorder that extends over many years, and patients may progress from mild to severe hypertension in a short time, especially with inadequate treatment. Even with treatment, most patients still progress, over years, to a state necessitating more severe therapeutic approaches. Therefore, the primary goal of all therapy is to reduce the rate of progression or to prevent it entirely.

When to initiate therapy and what types of therapy to use remain controversial. Therapy takes the form of either antihypertensive drugs (i.e., pharmacotherapy) or nonpharmacological approaches such as behavior and lifestyle modification. Drug therapy, along with surgery as needed, is frequently the choice for secondary forms of hypertension; however, with essential hypertension, whether to treat or not depends on the classification of severity and the presence or absence

of other risk factors (e.g., diabetes, atherosclerosis, obesity, heart condition). Borderline hypertension should always be addressed first by behavioral and lifestyle modification before pharmacotherapy. Mild to moderate hypertension warrants both nondrug and drug therapy. The primary goal of all therapeutic approaches must be to lower the diastolic (and systolic) pressure into the normal range or, if that cannot be reached, to as low a pressure as possible. All therapeutic approaches must be individualized to the patient.

Among nonpharmacological approaches are weight reduction, restriction of salt intake, stress reduction, lifestyle changes, exercise, discontinuance of alcohol intake, and cessation of smoking. Depending on ethnic background, other approaches, such as meditation and acupuncture, may be useful. For some individuals, these therapeutic approaches may be all that is needed. For example, it was shown during the 1980s that a relatively small reduction in weight (less than 10%) markedly lowered arterial pressure in essential hypertensives, resulting in a reduction or even total discontinuance of antihypertensive drugs in some patients. Weight reduction achieves two immediate goals: reduction of the workload placed on the heart and reduction in total blood volume. Another nondrug approach is stress relaxation and/or lifestyle modification. Stress reduction could lower sympathetic drive, thereby lowering activation of neurohormones as well as directly decreasing stimulation of the heart. Nondrug therapies such as stress reduction may take several months before evidence of their effectiveness is clear. For this reason, patients are usually placed on limited antihypertensive drug therapy to assist in lowering the blood pressure more quickly. Whether or not this is necessary depends on the individual patient.

The best source of currently recommended drugs is the JNC 7 report, which was discussed earlier and need not be repeated here. Pharmacotherapy of hypertension uses what is termed a stepped care approach: introduce one antihypertensive drug (selected by the severity of the hypertension and other risk factors present), and if reduction of arterial pressure is inappropriate, then complement the first agent with a second or shift to a different class of agents. In the United States, the most common drug initially prescribed (termed first-line therapy) is diuretics, which increase urine production by increasing the loss of sodium and water from the body. However, such agents have many secondary (and sometimes undesirable) actions such as increasing loss of potassium.

Some patients receive a type of diuretic that is potassium sparing; that is, it does not cause a loss of potassium. Excessive loss of potassium may lead to heart rhythm abnormalities. In the absence of a potassium-sparing diuretic agent, potassium supplementation, by drugs or diet, is often necessary. Diuretics may be used alone or in combination with other drugs, often being prescribed with agents that block sympathetic function because the latter tend to promote salt and water retention. An alternate first-line therapy is  $\beta$ -adrenergic receptor blockers (termed beta blockers). These drugs partially block the ability of the sympathetic nervous system to stimulate the heart and release renin from the kidney. Some beta blockers may also work in the brain to reduce sympathetic nervous system activity and the secretion of brain hormones that can raise arterial pressure. A relatively new class of agents, successful both in treating hypertension and in reducing adverse restructuring of the heart, are ACE inhibitors. ACE inhibitors not only block the formation of active angiotensin II, they also increase the lifetime and activity of another peptide, bradykinin, which relaxes vascular smooth muscle, releases vasodilator nitric oxide from endothelial cells, and releases vasodilator polyunsaturated fatty acids known as eicosanoids. If arterial pressure is still uncontrolled with these first-line drugs, calcium channel blockers or drugs that work primarily in the brain to reduce sympathetic activity would be added to one of the three first-line agents. If control is still not achieved, third-line agents, such as  $\alpha$ -1 adrenergic receptor blockers, would be tried. However, the latter have significant side effects, such as postural hypotension, and are generally reserved for unresponsive patients or medical emergencies. New agents are being developed continuously, so these categories will likely change with time.

It is generally considered that, with time (years), patients will progress to either combinations of drugs or the stronger third-line agents. However, many investigators believe that there has been inadequate attention to combining nondrug methods with pharmacotherapy. Reliance on pharmacotherapy without strong efforts to promote concurrent nondrug therapies has demonstrated that progression is inevitable.

## CONCLUSION

Hypertension is a disorder of regulation of systemic arterial pressure. Essential hypertension and many

forms of secondary hypertension have major genetic elements that dictate predisposition to the disorder. Early recognition of abnormal regulation of arterial pressure is essential if progress is to be arrested. All therapeutic approaches must focus on slowing progression of the disorder and returning, if possible, the arterial pressure to population-defined normative limits. However, what is normal for one individual may still be abnormal for another, or conversely, what is abnormal for one individual may be normal for another. Arterial pressure permits organisms to function and survive. It must be understood that hypertension may be the result of efforts by the body to satisfy the cardiovascular and autonomic requirements of the individual. Reversing hypertension without addressing the origins of the elevated arterial pressure may be therapeutically inadequate and potentially detrimental. Until more is known about the genes that regulate cardiovascular and autonomic coupling, the approach to the treatment of hypertension must be to slow progression of the disorder. Thus, the treatment of hypertension involves an ongoing series of decisions and evaluations by the physician based on the individual patient. Continued interaction and communication among patient, physician, and health professional are essential to the prevention of the damaging effects of hypertension. In the case of hypertension, as with many diseases or disorders, prevention remains the preferred first step in treatment.

## See Also the Following Articles

Aldosterone Receptors • Conn's Syndrome • Hypertension and Diabetes • Hypertension, Endocrine • Hypertension, Neurogenic • Hypertension, Renin and • Renal Vein Renin • Stress and Endocrine Physiology

## Further Reading

- Alemayehu, A., Breen, L., and Printz, M. P. (2002). A new inbred Wistar-Kyoto rat substrain apparently exhibiting salt sensitivity and borderline hypertension. *Am. J. Physiol. Heart Circ. Physiol.* **283**, H1181–H1190.
- Buijs, R. M., Van Eden, C. G., Goncharuk, V. D., and Kalsbeek, A. (2003). The biological clock tunes the organs of the body: Timing by hormones and the autonomic nervous system. *J. Endocrinol.* **188**, 17–26.
- Johnson, E. H., Gentry, W. D., and Julius, S. (1992). "Personality, Elevated Blood Pressure, and Essential Hypertension." Hemisphere Publishing, Washington, DC.
- Kaplan, N. M. (2002). "Kaplan's Clinical Hypertension," 8th ed. Lippincott Williams & Wilkins, Philadelphia.
- Page, I. H. (1987). "Hypertension Mechanisms." Grune Stratton/Harcourt Brace Jovanovich, Orlando, FL.
- Page, N. M. (2002). The endocrinology of pre-eclampsia. *Clin. Endocrinol.* **57**, 413–423.
- Printz, M. P., Jirout, M., Jaworski, R., Alemayehu, A., and Kren, V. (2003). The HXB/BXH rat recombinant inbred strain platform: A newly enhanced tool for cardiovascular, behavioral, and developmental genetics and genomics. *J. Appl. Physiol.* **94**, 2510–2522.
- National Heart Lung and Blood Institute (2003). "The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure" (NIH Publication No. 03-5233). National Institutes of Health, Bethesda, MD.

first 1301 bp of the 5'-flanking region, a classic silencer element within the first intron (intron A) of the gene, and a strong enhancer located far upstream (bp -5777 to -5552 in humans). There is accumulating evidence that renin expression is regulated according to the "variegation concept," meaning that transcription of a single gene is either switched on or switched off, but not gradually modified. A more or less pronounced expression of renin within a tissue would consequently be the result of more or less "switched-on," renin-transcribing cells.

Additional data show mechanisms for a posttranscriptional control of renin synthesis. These mechanisms seem to involve the stabilization of renin mRNA by the binding of specific proteins, whose expression is stimulated by cyclic AMP.

Molecular genetics studies did not find any evidence of an association between renin gene expression and hypertension.

## BIOSYNTHESIS, PROCESSING, AND EXPRESSION OF RENIN

Translation of renin genes yields a precursor molecule, preprorenin, which consists of 340 amino acids. Preprorenin is processed to enzymatically inactive prorenin (MW 57,000). Prorenin can be activated *in vitro* by acidification or prolonged cooling, as well as by exposure to neutral serine proteases (e.g., trypsin, plasmin, kallikrein) or acid proteases (e.g., pepsin, cathepsin D). More than 80% of the total circulating renin is prorenin.

Renin is expressed in a complex tissue-specific and developmentally specific pattern. Early in development, renin is abundant in smooth muscle cells in the intrarenal arteries. During development, however, renin expression is progressively more reduced to the "classical" place of synthesis, the juxtaglomerular apparatus, where it is produced and stored in granules.

Very strong, chronic stimuli, however, are able to reestablish a cell's ability to synthesize renin. In the heart, for example, where renin mRNA is normally absent, conditions known to increase renin synthesis in the kidney (such as sodium depletion for at least a week or chronic ACE inhibition) lead to a massive rise in cardiac renin mRNA expression. The same phenomenon (called "recruitment") was observed in afferent arteriolar cells extending further upstream of the juxtaglomerular apparatus, which usually show no renin mRNA expression.

## MECHANISMS CONTROLLING RENIN RELEASE

Renin is synthesized, stored, and secreted into the renal arterial circulation by the granular juxtaglomerular cells that lie in the walls of the afferent arterioles as they enter the glomeruli.

Effective stimuli for renin release and thereby for the induction of the renin-angiotensin-aldosterone cascade are sodium deficiency, hypovolemia, a significant decrease in arterial blood pressure, and an increased sodium concentration in the distal tubule (tubuloglomerular feedback).

The secretion of renin from juxtaglomerular cells is controlled by three pathways, as follows:

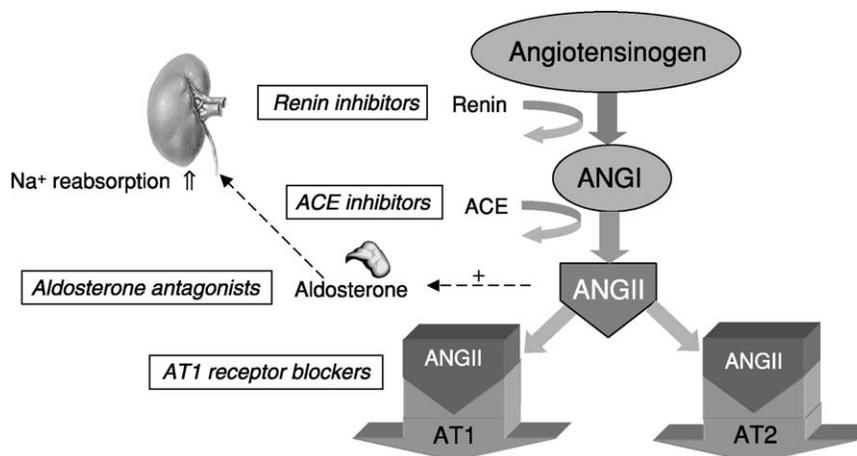
The first intrarenal mechanism controlling renin release is called the macula densa pathway. A specialized segment of the early distal tubule, the macula densa segment, comes into direct contact with the afferent and efferent arterioles of its own nephron. A change in NaCl reabsorption by the macula densa results in the transmission of chemical signals to nearby juxtaglomerular cells, which modify the release of renin; the increase in NaCl flux across the macula densa inhibits the release of renin and the decrease stimulates the release of renin. The other chemical signals mediating the macula densa pathway involve adenosine, prostaglandins, and nitric oxide.

The second mechanism controlling renin release is called the intrarenal baroreceptor pathway. In 1934, Goldblatt showed that it was possible to produce persistent hypertension in dogs by constricting the renal arteries. Later, the concept of a "baroreceptor mechanism for renin secretion" was formulated by Skinner. Increases and decreases in blood pressure in the preglomerular vessels inhibit and stimulate renin release, respectively. The underlying mechanism is believed to be a reduction in the tension within the wall of the afferent arteriole.

The third mechanism is called the  $\beta$ -adrenergic receptor pathway. It has long been known that an increase in sympathetic activity influences renin release. The studies of Kirchheim and related investigations have clearly established that an increase in sympathetic tone to the kidney stimulates renin release via the activation of  $\beta_1$ -adrenergic receptors on juxtaglomerular cells.

## RENIN ACTIONS

Renin acts on angiotensinogen, splitting off a decapeptide, angiotensin I, from the N-terminal end of the



**Figure 1** Drugs affecting the renin–angiotensin–aldosterone system. ACE, angiotensin-converting enzyme; ANG I, angiotensin I; ANG II, angiotensin II; AT1, angiotensin receptor type 1; AT2, angiotensin receptor type 2.

protein (Fig. 1). ANG I is acted on by a second proteolytic enzyme, ACE, that removes two more amino acids to form the highly active octapeptide, ANG II. The removal of one more amino acids yields the heptapeptide, angiotensin III (ANG III). ANG II and ANG III induce the release of the mineralocorticoid aldosterone, the most important hormone for sodium balance, from the adrenal cortex.

### THE SUBSTRATE FOR RENIN: ANGIOTENSINOGEN

Angiotensinogen, an  $\alpha_2$ -globulin, is the precursor molecule for angiotensin II and the substrate for renin. Angiotensinogen is synthesized and secreted mainly by the liver and is found in the  $\alpha$ -globulin fraction of plasma. Moreover, it is also found in diverse tissues expressing local RAASs. Its synthesis is stimulated by glucocorticoids, thyroid hormone, estrogens, and ANG II. Increases in angiotensinogen levels are associated with essential hypertension. Moreover, a linkage to essential hypertension was found in regions within or close to the angiotensinogen gene. Transgenic mice expressing the rat angiotensinogen gene are hypertensive and mice lacking the angiotensinogen gene are hypotensive.

### ANGIOTENSIN-CONVERTING ENZYME

ACE is the second enzyme (Fig. 1) in the cascade of ANG II synthesis, converting the inactive decapeptide, ANG I, to the potent octapeptide pressure hormone, ANG II. ACE is a nonspecific peptidase that

can cleave C-terminal dipeptides from various peptides (dipeptidyl carboxypeptidase). The enzyme kininase II is identical to ACE and contributes to the inactivation of kinins, such as bradykinin, and other potent vasodilator peptides. ACE is primarily localized on the luminal side of the vascular endothelium. The lung, which has a vast surface area of vascular endothelium, is rich in ACE. Additionally, ACE is present in other organs including kidney, heart, brain, and striated muscle skin, as it is a part of local RAASs.

### ANGIOTENSIN II AND ITS RECEPTORS

ANG II, an octapeptide hormone, is the main active component of the RAAS. It contributes to the regulation of blood pressure, plasma volume (via aldosterone-regulated sodium excretion), and sympathetic nervous activity. It is, moreover, involved in such diverse effects as proliferation, differentiation, regeneration, and apoptosis. The multiple actions of ANG II are mediated via specific, highly complex intracellular signaling pathways that are stimulated following an initial binding of the peptide to its cell surface receptors. The major actions of ANG II are mediated by two subtypes of G protein-coupled angiotensin receptors, the angiotensin receptor type 1 (AT1) and the angiotensin receptor type 2 (AT2) receptors, which are seven-transmembrane glycoproteins with only 32–34% sequence homology.

#### The AT1 Receptor

The AT1 receptor is involved in the classical physiological actions of ANG II: regulation of blood

pressure, electrolyte and water balance, thirst, hormone secretion, and renal function. The AT1 receptor belongs to the G protein-coupled receptor superfamily and typically activates phospholipase C through the heterotrimeric G<sub>q</sub> protein.

AT1 receptors are present in the human vasculature, lung, liver, brain, kidney, adrenal gland, skin, and endometrium.

The classical actions of AT1 receptors include the following: generalized vasoconstriction; increased release of noradrenaline from sympathetic nerve terminals, reinforcing vasoconstriction and increasing the rate and force of contraction of the heart; stimulation of proximal tubular reabsorption of sodium ions; secretion of aldosterone from the adrenal cortex; and cell growth in the cardiac left ventricle and in the arterial wall.

ANG II via the AT1 receptor furthermore controls cellular growth, adhesion, migration, and intercellular matrix deposition, influencing chronic adaptive changes in vascular and cardiac growth, remodeling, repair, and atherosclerosis. In vascular smooth muscle cells and in endothelial cells, ANG II via the AT1 receptor stimulates phospholipase A<sub>2</sub> activity, leading to the release of arachidonic acid and eicosanoids, which influence vascular and renal mechanisms important in the regulation of blood pressure and cell growth.

The expression of AT1 receptors is altered by various pathophysiological conditions: renovascular hypertension, myocardial infarction, ventricular hypertrophy, and bilateral nephrectomy. Overexpression of the vascular AT1 receptor can be observed in hypercholesterolemic men. These findings may help to explain why hypercholesterolemia is frequently associated with hypertension and why blockade of the RAAS attenuates the progression of atherosclerosis. The down-regulation of AT1 receptors in sepsis is the main reason for the attenuated responsiveness of blood pressure and of aldosterone formation to ANG II and, therefore, may contribute to the characteristic septic shock.

### The AT2 Receptor

The AT2 receptor is widely expressed in fetal tissues, where its expression is dramatically decreased after birth, being restricted to a few organs such as brain, adrenal, heart, kidney, myometrium, skin, and ovary. Although the AT1 receptor is dominant in the adult organism, an increase of AT2 receptor expression has been observed in pathological conditions, such as

vascular injury, myocardial infarction, congestive heart failure, renal failure, brain ischemia, and sciatic or optic nerve transection.

It seems that the AT2 receptor exerts a protective effect against an overstimulation of AT1 receptors by counteracting AT1 receptor-mediated actions; e.g., whereas the AT1 receptor stimulates cell proliferation, the AT2 receptor has an antiproliferative effect and promotes cell differentiation. Moreover, inhibition of AT2 receptors enhances the immediate left ventricular growth response to ANG II.

The signaling mechanisms of the AT2 receptor are diverse and only a few mechanisms have been characterized reasonably well. In some cases, they are coupled to G<sub>i</sub> proteins. One pathway in neurons (and perhaps other tissues) involves the activation of the protein serine/threonine phosphatase PP2A, which leads to the activation of the delayed rectifier K<sup>+</sup> channel, hyperpolarization of plasma membranes, and suppression of cellular activities stimulated by depolarization. Other pathways shown thus far involve the activation of protein phosphotyrosine phosphatases, mitogen-activated protein kinase phosphatase 1, phospholipase A<sub>2</sub>, and—very importantly—the release of bradykinin and nitric oxide (NO). The AT2 receptor stimulates the production of NO, which results in an increased formation of the second messenger cyclic GMP (cGMP). cGMP, in turn, mediates many of the biological actions of NO, such as vasodilation, natriuresis, and anti-growth by activating cGMP-dependent protein kinase.

A negative crosstalk between AT1 and AT2 receptors has been proposed not only on a functional level, but also on the level of intracellular signaling.

Further knowledge about the tissue-protective effects of the AT2 receptor might lead to new therapies in the future and should be confirmed by clinical studies.

### ALDOSTERONE

The mineralocorticoid aldosterone is produced in the zona glomerulosa of the adrenal medulla. Its main action is the reabsorption of Na<sup>+</sup> (Cl<sup>-</sup> and H<sub>2</sub>O follow) in exchange for K<sup>+</sup> in the distal tubules of the kidney. Low plasma sodium or high plasma potassium concentrations affect the zona glomerulosa cells of the adrenal directly, stimulating aldosterone release. Aldosterone secretion is also controlled indirectly by the juxtaglomerular apparatus, which is sensitive to the composition of the fluid in the distal tubule. A decrease in the sodium chloride

concentration of the filtrate is sensed by macula densa cells, which stimulate renin release. This leads—as illustrated above—to the formation of ANG II, which in turn, via AT1 receptors, stimulates the synthesis and release of aldosterone by the adrenal cortex. Moreover, in the process of aldosterone synthesis, angiotensins (ANG II and ANG III) regulate the corticosterone methyl oxidase I and II enzymes, which catalyze the hydroxylation and aldehyde formation at C-18 of corticosterone.

In the kidneys, in the late distal tubule and collecting duct, aldosterone binds to cytoplasmic receptors, which migrate to the nucleus and initiate DNA transcription, translation, and production of proteins, which activate Na<sup>+</sup> and K<sup>+</sup> channels and increase the synthesis of Na<sup>+</sup>/K<sup>+</sup> ATPase and production of ATP. As a result, the reabsorption of Na<sup>+</sup> (and the subsequent reabsorption of Cl<sup>-</sup> and H<sub>2</sub>O) and the secretion of K<sup>+</sup> and H<sup>+</sup> are increased. Na<sup>+</sup> reabsorption increases the osmolarity of extracellular fluids. This stimulates the hypothalamic osmoreceptor to release vasopressin from the posterior pituitary. Vasopressin leads to enhanced free-water reabsorption in the collecting duct, which expands extracellular volume and reduces plasma osmolarity. Vasopressin also acts on other tissues. In the vasculature, vasopressin increases intracellular Ca<sup>2+</sup> and potentiates vasopressor responses.

## DRUGS AFFECTING THE RAAS

### Renin Inhibitors

Since renin catalyzes the first and rate-limiting step of the RAAS cascade, interruption of the generation of ANG II by renin inhibitors at this highly specific initial step of the cascade has long been a therapeutic goal. The inhibition of renin would have the advantage of preventing ANG II synthesis without concomitant accumulation of other peptides, such as kinins or substance P, as in ACE inhibitor treatment, and it would have the advantage of the absence of ANG II overexpression and stimulation of other ANG II receptor subtypes as in AT1 receptor antagonist treatment. Despite these theoretical advantages, the lack of oral availability, low efficacy, and high costs of development have thus far prevented renin inhibitors from becoming successful drugs. However, a potent nonpeptidic inhibitor of renin, Aliskiren, with acceptable oral bioavailability, has been synthesized. Aliskiren has been shown to decrease in a dose-dependent manner plasma renin activity, ANG I and ANG II levels in healthy volunteers, and blood

pressure in hypertensive patients. Large clinical studies comparing renin inhibitors with the other blockers of the RAAS will be needed to establish the true clinical utility of renin inhibition.

### ACE Inhibitors

ACE inhibitors occupy the angiotensin-converting enzyme as false substrates, thus preventing the conversion of ANG I to ANG II and resulting in reduced levels of circulating and tissue ANG II. They also block the degradation of bradykinin and other vasodilatory peptides, which may have potential benefits in cardiovascular disease. ACE inhibitors affect capacitance and resistance vessels, reduce cardiac load as well as arterial pressure, improve endothelial function, and reduce left ventricular hypertrophy.

ACE inhibitors are well recognized as an important therapeutic step to control blood pressure in hypertensive patients and to reduce morbidity and mortality in patients with hypertension, congestive heart failure, or diabetes mellitus. They appear to possess unique cardioprotective benefits, even when used in patients without high blood pressure or left ventricular dysfunction. ACE inhibitors promote collateral vessel development and improve prognosis in patients who have had a coronary revascularization procedure. ACE inhibitors are also effective in the management of chronic renal diseases to delay the progression of renal failure.

### AT1 Receptor Antagonists

The “sartan” family comprises a rather new group of pharmaceuticals (losartan, valsartan, candesartan, irbesartan, telmisartan, eprosartan), which all act as antagonists on the AT1 receptor. AT1 receptor antagonists are specific for the RAAS, selective for the AT1 receptor, and act independently of the ANG II synthetic pathway, allowing a more selective blockade of the AT1-mediated effects of ANG II compared with ACE inhibitors. However, by inhibiting the AT1 receptor-mediated negative feedback of ANG II on renin release in the kidney, these drugs may evoke an overstimulation of the AT2 receptor by enhancing ANG II levels in the plasma. An AT2 receptor-mediated increase in the production of vasodilators (nitric oxide, cGMP, prostaglandins) as well as the anti-growth features of this receptor might contribute to a further decrease in blood pressure and prevent hypertrophy and remodeling. AT1 receptor

antagonists do not inhibit the degradation of kinins and cough is not a frequent side effect.

The AT1 receptor antagonists are widely used as antihypertensive agents, especially in patients with type 2 diabetes or ACE inhibitor intolerance.

### Aldosterone Antagonists

The mineralocorticoid hormone aldosterone is a product of the RAAS that contributes to the development of hypertension and myocardial hypertrophy and has the potential to cause edema through sodium and water retention. Spironolactone and its metabolite canrenone are antagonists of the aldosterone receptor and attenuate the effects of the hormone. However, spironolactone is associated with progestational and anti-androgenic side effects, such as gynecomastia and impotence, as a result of its binding to other steroid receptors. Eplerenone is the first agent of a new class of drugs known as selective aldosterone receptor antagonists, which provide effective and well-tolerated blood pressure reduction. Because eplerenone has little affinity for androgen and progesterone receptors, it produces fewer steroid-like effects (such as gynecomastia in men) than spironolactone. In addition to its action in lowering blood pressure, eplerenone appears to provide protection to the heart and kidney. Hyperkalemia is a dose-related adverse effect.

### Other Drugs Influencing Renin Release

Several other drugs, which do not directly interfere with the cascade of ANG II synthesis, indirectly modify the rate of renin synthesis or release. Loop diuretics stimulate renin release by blocking the reabsorption of NaCl at the macula densa. Antihypertensive drugs in general increase renin release by decreasing arterial blood pressure, which in turn

activates baroreceptors. Sympatholytic drugs and  $\beta$ -adrenoreceptor antagonists decrease renin release by blocking the  $\beta$ -adrenoreceptor pathway, leading to a diminished sympathetic tone in the kidneys.

### See Also the Following Articles

Angiotensin, Evolution of • Captopril • Conn's Syndrome • Conn's Syndrome, Diagnosis of • Hypertension, Endocrine • Renal Vein Renin • Renin • Tissue Renin-Angiotensin-Aldosterone System

### Further Reading

- de Gasparo, M., Catt, K. J., Inagami, T., Wright, J. W., and Unger, T. (2000). International Union of Pharmacology. XXIII. The angiotensin II receptors. *Pharmacol. Rev.* **52**, 415–472.
- Edling, O., Gohlke, P., Paul, M., Rettig, R., Stauss, H., Steckelings, U. M., and Unger, T. (1995). Physiology of the renin-angiotensin-system. In "ACE Inhibitors: Current Use and Future Prospects" (M. Schachter, ed.), pp. 3–19. Martin Dunitz Ltd., London, UK.
- Gilman, A. G., *et al.* (1996). "Goodman and Gilman's The Pharmacological Basis of Therapeutics," 9th ed. Pergamon, Elmsford, NY.
- Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (2003). "The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure," NIH Publication No.03–5233. National Institutes of Health, Bethesda, MD.
- Kaschina, E., and Unger, T. (2003). Angiotensin AT1/AT2 receptors: Regulation, signalling and function. *Blood Press.* **12**, 70–88.
- Morries, B. J. (2000). Renin. In "Handbook of Physiology, Section 7, The Endocrine System, Vol. III, Endocrine Regulation of Water and Electrolyte Balance" (J. C. S. Fray, ed.), pp. 3–58. Oxford University Press, New York.
- Page, C. P., Curtis, M. J., Sutter, M. C., Walker, M. J. A., and Hoffman, B. B. (1998). "Integrated Pharmacology." Mosby-Verlag, London, UK.
- Swales, J. D. (1994). "Textbook of Hypertension." Blackwell Scientific, Oxford, UK.

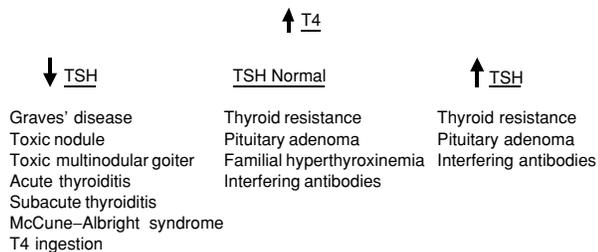


Figure 1 Causes of hyperthyroidism.

Biochemically, the cardinal feature of hyperthyroidism is a suppressed TSH. T4 and T3 levels are generally elevated, but in some cases only elevations in T3 are present (T3 toxicosis). Scanning using radionuclides reveals diffuse and increased uptake within the thyroid. In most patients with Graves' disease, elevated TRAB levels can be detected. Individuals with high TRAB levels and a large thyroid gland are likely to have a protracted course, with a very low chance of spontaneous remission. Individuals with a small gland and low TRAB levels have a moderate chance of spontaneous remission. Because Graves' disease does not commonly resolve spontaneously within a short period of time, treatment for hyperthyroidism is needed. Treatment approaches include antithyroid drugs, radioactive iodine ( $^{131}\text{-iodine}$ ), and surgery.

### Medical Therapy

Antithyroid drugs remain the first line of therapy for children with Graves' disease in many centers. Mainstays of antithyroid therapy include the thionamide derivatives propylthiouracil (PTU) and methimazole (MMI). MMI is 10-fold more potent than PTU and has a longer half-life. Recommended doses for initial therapy are 5 to 10 mg/kg/day for PTU and 0.5 to 1.0 mg/kg/day for MMI; even lower doses may be effective for induction or maintenance therapy. To control the hyperthyroid state, PTU and MMI are typically given every 8 h. However, once-a-day dosing may bring remission as rapidly as do divided doses and is especially well suited for maintenance therapy. Because MMI pills (5 or 10 mg) are smaller than PTU tablets (50 mg), and because fewer MMI pills are generally needed, MMI is more convenient to take than is PTU.

Although MMI and PTU promptly inhibit hormone formation, they do not inhibit thyroid hormone release. Thus, levels of circulating thyroid hormones may remain elevated for several weeks as stored hormone is released. Until circulating levels of thyroid

hormones normalize, the signs and symptoms of hyperthyroidism may be controlled with beta blockers such as atenolol (25 or 50 mg once or twice a day) or propranolol (2.5–10.0 mg two or three times a day). Iodine drops also help to control hyperthyroidism more rapidly than does PTU or MMI alone; high doses of iodine acutely reduce thyroid hormone synthesis and release, a phenomenon referred to as the Wolf–Chaikoff effect. Clinical responses to PTU or MMI are usually seen after 4 to 8 weeks, at which time biochemical hypothyroidism often develops and the thionamide dose can be reduced by 30 to 50%. If hypothyroidism develops, the dose of MMI or PTU can be reduced further or levo-thyroxine can be given to correct the hypothyroid state.

Approximately 25% of children treated with PTU or MMI will develop minor complications, and less than 1% of children will develop serious complications. Minor complications include rashes, urticaria, myalgias, elevations in liver enzymes, leukopenia, arthritis, and lymphadenopathy. Serious complications include agranulocytosis, hepatitis, and vasculitis. Severe hepatotoxicity is more of a problem with PTU than with MMI.

In children, remission rates after several years of drug therapy are usually less than 30%. Importantly, remission rates are less in prepubertal children than in pubertal children. Although many clinicians will treat children with antithyroid medication for years, the likelihood of long-term remission can be predicted by observing responses to short-term (4–12 months) treatment. If remission occurs, long-term follow-up is needed. If remission does not occur, drug treatment can be restarted or definitive therapy can be selected.

### Radioactive Iodine

Radioactive iodine is the most simple and cost-effective treatment for Graves' disease. Treatment is achieved when  $^{131}\text{-iodine}$  is trapped in thyroid cells, leading to thyroid cell destruction by internal radiation. It has been suggested that doses (administered activities) delivering 300 to 400 Gy (30,000–40,000 cGy or rads; 360–480  $\mu\text{Ci/g}$  of  $^{131}\text{-iodine}$ ) are required to achieve hypothyroidism in adults. However, the thyroid gland of children may be more sensitive to  $^{131}\text{-iodine}$  than that of adults, and it is possible to achieve hypothyroidism using lower doses in children. Treating our patients, we have observed good responses using between 180 and 300 Gy (200–300  $\mu\text{Ci/g}$ ) of  $^{131}\text{-iodine}$ .

Long-term cure rates are higher in patients treated with larger amounts of  $^{131}\text{I}$ -iodine than with smaller amounts. In children treated with a single dose of 150 to 200  $\mu\text{Ci/g}$ , hyperthyroidism persists in 5 to 20%, and 60 to 90% become hypothyroid. Hypothyroidism usually occurs within 2 to 6 months of treatment. If hyperthyroidism persists, additional courses of  $^{131}\text{I}$ -iodine are indicated. After repeated treatments, 6 to 12 months may pass before hypothyroidism develops.

Responses to  $^{131}\text{I}$ -iodine therapy are much less favorable in patients with large glands. Thus, we usually recommend surgery when the thyroid is more than 100 g. To determine thyroid volume in this setting, ultrasound is useful [volume = 0.52 (length  $\times$  depth  $\times$  width)].

Recent discussions have focused on the association of  $^{131}\text{I}$ -iodine therapy of Graves' disease with the development or progression of ophthalmopathy in adults. In contrast to adults, children rarely develop severe ophthalmopathy and proptosis is generally mild. In the unusual setting where there is severe eye disease, adjunctive prednisone therapy or surgery should be considered.

Radioactive iodine is being used at progressively younger ages, yet we do not know whether there is an age below which high-dose  $^{131}\text{I}$ -iodine therapy should be avoided. Risks of thyroid cancer after external irradiation are highest in children under 5 years of age and decline progressively with advancing age. Therefore, it may be prudent to avoid  $^{131}\text{I}$ -iodine therapy in young children. It is also important to note that when children are treated with  $^{131}\text{I}$ -iodine, higher doses that result in thyroid ablation and hypothyroidism should be used. Low doses of  $^{131}\text{I}$ -iodine (<80 Gy, <100  $\mu\text{Ci/g}$ ) are associated with an increased risk of thyroid neoplasms. Yet if higher doses are used and result in complete destruction of the thyroid gland, the risk of thyroid cancer will be much less.

### Surgery

Whereas subtotal (partial) thyroidectomy was advocated in the past for children and adults, total (near total) thyroidectomy is now recommended to reduce the risk of recurrent hyperthyroidism. In preparation for surgery, the child should be rendered euthyroid or hypothyroid. This is typically done with either PTU or MMI. One week before surgery, iodine is added to cause the gland to become firmer and less vascular, thereby facilitating surgery. The most frequent complications include pain and transient hypocalcemia.

Less common problems (1–4%) include hemorrhage, permanent hypoparathyroidism, and vocal cord paralysis.

Surgery is especially useful for patients with a large thyroid gland (>100 g) and for individuals who have not gone into remission with drug therapy and do not desire  $^{131}\text{I}$ -iodine. Because of the intricacies involved, surgery should be performed by surgeons with expertise in performing thyroidectomies in children.

### NEONATAL THYROTOXICOSIS

Thyrotoxicosis in neonates is a severe and life-threatening condition. If a mother has Graves' disease, there is a 1-in-80 chance that TRABs will be transferred to the fetus, resulting in intrauterine or neonatal hyperthyroidism. Rarely, neonatal thyrotoxicosis will persist.

If a mother with Graves' disease is taking antithyroid medications during pregnancy, fetal thyroid hormone synthesis will be inhibited, and this will prevent the development of intrauterine hyperthyroidism; however, the infant may be born with a goiter and hypothyroidism. At birth, circulating levels of T4 may be low and TSH may be elevated. In most cases, as the effects of antithyroid drugs wane, thyroid function normalizes within a week. Yet if there has been significant transplacental passage of TRABs, neonatal thyrotoxicosis will develop.

If a mother with a history of Graves' disease is not taking antithyroid drugs during pregnancy, the fetus may develop intrauterine hyperthyroidism. If not recognized, this may result in profound intrauterine thyrotoxicosis and growth retardation. Such infants may have prematurely fused cranial sutures, advanced skeletal ages, long-term learning problems, and/or mental retardation. If fetal hyperthyroidism is recognized by the presence of fetal tachycardia (heart rate >160 beats/min after 22 weeks), treatment of the mother with antithyroid drugs will reduce intrauterine thyrotoxicosis and improve fetal outcome.

Treatment of thyrotoxic newborns consists of antithyroid medications (PTU 5–10 mg/kg/day or MMI 0.5–1.0 mg/kg/day) and beta blockade (propranolol 1 mg/kg/day). Lugol's solution or saturated potassium iodide may be given (1–2 drops every 8 h) for 7 to 10 days to control biochemical hyperthyroidism more rapidly. After approximately 2 weeks of antithyroid drug therapy, thyroid hormone levels will decline. When thyroid hormone levels fall below normal, supplementary levo-thyroxine (37.5  $\mu\text{g/day}$  for term infants) is added to prevent hypothyroidism.

As TRABs are cleared from the infant's circulation, spontaneous recovery begins within 3 months and is usually complete by 6 months. Thus, treatment can be weaned after 3 months.

## HYPERFUNCTIONING THYROID NODULES

Warm or hot nodules lead to excessive production of thyroid hormone and can be associated with clinical and biochemical hyperthyroidism. Hyperfunctioning nodules are most commonly solitary lesions. However, some children may manifest hyperthyroidism in the setting of TMGs.

The possibility of a hyperfunctioning thyroid nodule should be considered in individuals with thyroid asymmetry, high T4 levels, suppressed TSH levels, and the absence of antithyroid antibodies. Ultrasonography, which is indicated whenever thyroid asymmetry is detected, is useful in detecting solitary nodules or multinodular goiters. Scanning using radionuclides reveals uptake by the hyperfunctioning nodule and decreased uptake by normal thyroid tissue.

Whereas there is a chance of spontaneous remission of the hyperthyroid state in children with Graves' disease, the chance of spontaneous remission of a solitary hyperfunctioning thyroid nodule or a TMG is rare. Thus, definitive treatment is indicated. In adults, ablation of hyperfunctioning nodules with <sup>131</sup>-iodine is often performed. With children, surgical excision of hyperfunctioning nodules is recommended because radiation-exposed normal thyroid tissue will remain after the hyperfunctioning nodule is ablated.

## INFECTIOUS THYROIDITIS

Occasionally a child will present with hyperthyroidism, tenderness over the thyroid gland, and fever due to bacterial infection on the thyroid, a condition referred to as acute thyroiditis. This can be associated with a fistula connecting a pyriform sinus in the pharynx to the thyroid. Fevers can be high, and erythrocyte sedimentation rates and white counts can be elevated. Ultrasonography may reveal a local abscess. When thyroid radionuclide scanning is performed, reduced uptake is observed.

The offending bacteria consist of Hemophyllus influenza, oral flora, group A Streptococcus, or Staphylococcus. Thus, treatment with a beta lactamase-resistant antibiotic is recommended. For severe cases, hospitalization and intravenous antibiotic

administration is indicated because there may be lymphatic drainage into the mediastinal region. Surgical drainage is needed if a localized abscess develops and the response to antibiotics is poor. Because the infectious process results in destruction of thyroid tissue, there may be release of preformed thyroid hormone and hyperthyroidism will occur. The hyperthyroid state is usually transient, and treatment with antithyroid drugs is not indicated. If the patient becomes symptomatic, beta blockers may be used. After the child has recovered, a pharyngeogram may be performed to test for a patent pyriform sinus tract. Occasionally, the tract may close as a result of the infection. However, if the tract persists and acute thyroiditis recurs, tract resection is needed.

Viral infections of the thyroid may occur and result in subacute thyroiditis. In comparison with acute thyroiditis, subacute thyroiditis may be less severe. However, there may be fever, thyroid tenderness, and hyperthyroidism that may last for several weeks.

## OTHER CAUSES OF HYPERTHYROIDISM

Hyperthyroidism can occur in Hashimoto's disease because autoimmune destruction of thyroid tissue can result in excessive release of thyroid hormone. Yet the hyperthyroid state is transient and is followed by hypothyroidism.

Thyrotoxicosis can occur when there is thyroid hormone ingestion. With severe ingestions, thyroid storm, heart failure, and seizure activity can occur. Treatment is supportive and involves the use of beta blockers. Epidemic thyrotoxicosis has also occurred following the ingestion of beef containing thyroid gland remnants.

Hyperthyroidism can result when individuals in areas of iodine insufficiency are given large doses of iodine via pharmaceuticals containing iodine or X-ray contrast material, a phenomenon referred to as the Jod-Basedow effect. Clinicians need to be aware that the commonly used, anti-arrhythmic drug Amiodarone contains large amounts of iodine and can result in hypo- or hyperthyroidism. Amiodarone-induced thyrotoxicosis can be due to iodine-induced thyroid hormone synthesis or to destructive thyroiditis.

Hyperthyroidism can develop in McCune-Albright syndrome, in which there are activating mutations of G proteins. Such individuals may have café-au-lait spots with irregular margins, bone lesions, and autonomous secretion of sex steroids and adrenal hormones.

TSH-producing pituitary adenomas are a rare cause of thyrotoxicosis in children. Affected individuals may present with clinical hyperthyroidism, goiter, and (occasionally) exophthalmos. Biochemical studies will reveal elevated T4 or T3 concentrations along with detectable or elevated TSH levels. Brain imaging studies are useful in visualizing pituitary adenomas. In contrast to Graves' disease or hyperfunctioning thyroid nodules, in which TSH levels remain suppressed, thyrotropin-releasing hormone testing reveals TSH and prolactin responsiveness when there is a TSH-producing pituitary adenoma. Treatment involves resection of the pituitary tumor.

### CONDITIONS ASSOCIATED WITH HYPERTHYROIDISM

Several autoimmune disorders are associated with hyperthyroidism, including diabetes mellitus, juvenile rheumatoid arthritis, nephritis, and inflammatory bowel disease. Thus, individuals with these conditions should have thyroid status assessed during routine office visits and whenever there is a change in the status of the underlying illness. Urticaria is also associated with hyperthyroidism and may occur before or after the hyperthyroid state is recognized and treated. Hyperthyroidism should be considered in patients with attention deficit/hyperactivity disorder (ADHD) because symptoms can be similar in both conditions.

### CONDITIONS THAT CAN BE CONFUSED WITH HYPERTHYROIDISM

#### Thyroid Hormone Resistance

When the thyroid hormone receptor is mutated, impaired tissue responsiveness results, leading to thyroid enlargement, elevated levels of T4 and T3 tachycardia, and behavioral problems. In contrast to primary disorders of the thyroid gland, TSH levels are normal or slightly elevated in this condition.

Most individuals with resistance to thyroid hormones have generalized thyroid hormone resistance. These individuals are eumetabolic and asymptomatic, with TSH levels in the normal range. In contrast, some individuals will have isolated pituitary thyroid hormone resistance. These individuals have symptoms of hyperthyroidism in that they are sensitive to the effects of increased thyroid hormone levels.

Because individuals compensate for thyroid hormone resistance by secreting more thyroid hormones, treatment is not necessary. However, patients with thyroid hormone resistance may be improperly diagnosed as having Graves' disease and ablation of the thyroid may be performed.

#### Familial Dysalbuminemic Hyperthyroxinemia

When free T4 values and TSH levels are normal and total T4 values are high, familial dysalbuminemic hyperthyroxinemia (FDH) needs to be considered. This autosomal dominant disorder is most commonly seen in Hispanic individuals and is due to increased affinity of albumin for T4, leading to increased circulating T4 levels and normal T3 values. In FDH and other conditions affecting thyroid hormone binding, the condition can be diagnosed by thyroid hormone-binding protein electrophoresis. Treatment is not needed, and the patient should be educated about the condition to avoid treatment by unsuspecting practitioners.

#### Thyroid Hormone and Heterophile Antibodies

Although T4- and TSH-level measurements usually accurately reflect circulating concentrations of the hormones, the serum of some individuals contains antibodies that interfere with the measurement of these hormones, resulting in falsely elevated determinations. These substances include thyroid hormone autoantibodies, heterophile antibodies (human anti-mouse antibodies), and rheumatoid factors. Interfering antibodies should be considered when clinical and laboratory findings do not agree.

### CONCLUSION

With current diagnostic methods, it is possible to identify causes of hyperthyroxinemia and thyrotoxicosis in children and adolescents with a high degree of fidelity. Because of the profound effects that the hyperthyroid state has on a child, prompt and proper treatment is essential. Vigilance is needed to detect hyperthyroidism in the child with or without known risk factors for autoimmune thyroid disease. Clinicians also need to distinguish states of hyperthy-

oxinemia that do not require treatment from those pathological conditions that do.

### See Also the Following Articles

Graves' Disease, Hyperthyroidism in • Hashimoto's Disease • Hyperthyroidism, Subclinical • McCune-Albright Syndrome • Pituitary Adenomas, TSH-Secreting • Thyroid Hormone Action • Thyrotoxicosis: Diagnosis • Thyrotoxicosis, Overview of Causes • Thyrotoxicosis, Treatment

### Further Reading

- Cooper, D. S. (1998). Antithyroid drugs for the treatment of hyperthyroidism caused by Graves' disease. *Endocrinol. Metab. Clin. North America* **27**, 225–247.
- David, E., Rosen, I. B., Bain, J., James, J., and Kirsh, J. C. (1995). Management of the hot thyroid nodule. *Am. J. Surg.* **170**, 481–483.
- Despres, N., and Grant, A. M. (1998). Antibody interference in thyroid assays: A potential for clinical misinformation. *Clin. Chem.* **44**, 440–454.
- Farror, C., Wellby, M. L., and Beng, C. (1987). Familial dysalbuminaemic hyperthyroxinaemia and other causes of euthyroid hyperthyroxinaemia. *J. Royal Soc. Med.* **80**, 750–752.
- Giuffrida, D., and Gharib, H. (1995). Controversies in the management of cold, hot, and occult thyroid nodules. *Am. J. Med.* **99**, 642–650.
- Hamburger, J. I. (1985). Management of hyperthyroidism in children and adolescents. *J. Clin. Endocrinol. Metab.* **60**, 1019–1024.
- Jeng, L. B., Lin, J. D., and Chen, M. F. (1994). Acute suppurative thyroiditis: A ten-year review in a Taiwanese hospital. *Scand. J. Infect. Dis.* **26**, 297–300.
- Levy, O. P. (1988). Treatment of childhood Graves' disease: A review with emphasis on radioiodine treatment. *Cleve. Clin. J. Med.* **55**, 373–382.
- Peters, H., Fischer, C., Bogner, U., Reiners, C., and Schleusener, H. (1995). Radioiodine therapy of Graves' hyperthyroidism: Standard vs. calculated <sup>131</sup>-iodine activity—Results from a prospective, randomized, multicentre study. *Eur. J. Clin. Invest.* **25**, 186–193.
- Refetoff, S. (1997). Resistance to thyroid hormone. *Curr. Ther. Endocrinol. Metab.* **6**, 132–134.
- Rivkees, S. A., and Cornelius, E. A. (2003). Influence of iodine-131 dose on the outcome of hyperthyroidism in children. *Pediatrics* **111**, 745–749.
- Rivkees, S. A., Sklar, C., and Freemark, M. (1998). Clinical review 99: The management of Graves' disease in children, with special emphasis on radioiodine treatment. *J. Clin. Endocrinol. Metab.* **83**, 3767–3776.
- Zimmerman, D., and Lteif, A. N. (1998). Thyrotoxicosis in children. *Endocrinol. Metab. Clin. North America* **27**, 109–126.

**Table I Clinical Implications of Subclinical Hyperthyroidism**

- 
- Impaired quality of life
  - Higher heart rate
  - Increased prevalence of supraventricular arrhythmias
  - Increased left ventricular mass
  - Impaired diastolic function
  - Impaired systolic function on effort
  - Increased cardiovascular morbidity and mortality
  - Reduced bone mineral density
  - Increased risk of fractures
- 

important alterations involving the morphology and function of the cardiovascular system and/or bone and mineral metabolism (Table I). Various authors have reported that SH patients scored higher on a symptom rating scale (specifically designed for clinically overt hyperthyroid patients) than did euthyroid controls. Patients complained mainly of symptoms and signs suggestive of adrenergic overactivity. These findings clearly indicate that the term “subclinical hyperthyroidism” is a misnomer. In fact, SH is not merely a laboratory oddity but rather a mild tissue hyperthyroidism that may be particularly dangerous in the elderly.

### Effects on the Cardiovascular System

Higher heart rate and increased prevalence of supraventricular arrhythmias have been reported in most of the clinical studies of patients with SH, with a three-fold increased risk of atrial fibrillation in elderly patients. SH has also been associated with relevant changes in left ventricular properties. Increased left ventricular mass, in some cases above the hypertrophic threshold, is a consistent finding in SH patients, variably associated with impaired diastolic function, systolic performance on effort, and exercise tolerance. Importantly, cardiovascular morbidity and mortality are significantly increased in patients with endogenous SH. Although the underlying mechanism(s) remains speculative, it is likely that increased left ventricular mass, heart rate, and supraventricular arrhythmias may play a major role in determining this increase.

### Effects on Bone and Mineral Metabolism

Although clinically overt hyperthyroidism is a well-recognized risk factor for osteoporosis and fractures,

it is still debatable whether this is also the case for persistent SH. Serum concentration of osteocalcin and urinary excretion of markers of collagen bone degradation and reabsorption (pyridinoline cross-links, hydroxyproline, and telopeptide type I) were found to be increased in patients with SH. Congruent with these findings, some (but not all) studies showed a significant reduction in bone mineral density and an increased risk of fractures in women with SH; both features become more apparent during the postmenopausal period.

### MANAGEMENT STRATEGIES

In general, the treatment of SH should be modeled according to the underlying etiology, the patient's age, and the presence of comorbidity. In patients with exogenous SH, in whom the L-T4 dose cannot be reduced (e.g., patients with high-risk thyroid cancer), beta-blocking drugs can attenuate the effect of adrenergic overactivity on the cardiovascular system, thereby mitigating cardiac abnormalities and improving quality of life. Conversely, in all cases where L-T4 may be tailored (i.e., in benign thyroid disease and in some cases of low-risk thyroid cancer), this is the mandatory treatment, eventually corroborated by beta blockade in the case of remnant adrenergic overactivity. The treatment of endogenous SH, similarly to that of overt hyperthyroidism, is aimed at redressing euthyroidism by means of antithyroid drugs, radioiodine, or surgery. In some selected patients, beta blockade may be an effective alternative. Similar considerations apply to the management of the effects of SH on bone and mineral metabolism. Bisphosphonates and/or estrogen may be used in postmenopausal women and in individuals who have a high risk of fractures.

### Acknowledgment

We are grateful to Jean Ann Gilder for editing the text.

### See Also the Following Articles

Bisphosphonates • Graves' Disease, Hyperthyroidism in • Hyperthyroidism, Childhood and Adolescence • Osteoporosis, Overview • TSH Function and Secretion

### Further Reading

Bauer, D. C., Ettinger, B., Nevitt, M. C., and Stone, K. L. for the study of Osteoporotic Fractures Research Group. (2001). Risk

- for fracture in women with low serum levels of thyroid-stimulating hormone. *Ann. Int. Med.* **134**, 561–568.
- Biondi, B., Fazio, S., Carella, C., Amato, G., Cittadini, A., Lupoli, G., Saccà, L., Bellastella, A., and Lombardi, G. (1993). Cardiac effects of long-term thyrotropin-suppressive therapy with levothyroxine. *J. Clin. Endocrinol. Metab.* **77**, 334–338.
- Biondi, B., Fazio, S., Carella, C., Sabatini, D., Amato, G., Cittadini, A., Bellastella, A., Lombardi, G., and Saccà, L. (1994). Control of adrenergic overactivity by  $\beta$ -blockade improves quality of life in patients receiving long term suppressive therapy with levothyroxine. *J. Clin. Endocrinol. Metab.* **78**, 1028–1033.
- Biondi, B., Fazio, S., Cuocolo, A., Sabatini, D., Nicolai, E., Lombardi, G., Salvatore, M., and Saccà, L. (1996). Impaired cardiac reserve and exercise capacity in patients receiving long-term thyrotropin suppressive therapy with levothyroxine. *J. Clin. Endocrinol. Metab.* **81**, 4224–4228.
- Biondi, B., Palmieri, E. A., Fazio, S., Cosco, C., Nocera, M., Saccà, L., Filetti, S., Lombardi, G., and Perticone, F. (2000). Endogenous subclinical hyperthyroidism affects quality of life and cardiac morphology and function in young and middle-aged patients. *J. Clin. Endocrinol. Metab.* **85**, 47021–4705.
- Fazio, S., Biondi, B., Carella, C., Sabatini, D., Cittadini, A., Panza, N., Lombardi, G., and Saccà, L. (1995). Diastolic dysfunction in patients on thyroid-stimulating-hormone suppressive therapy with levothyroxine: Beneficial effect of  $\beta$  blockade. *J. Clin. Endocrinol. Metab.* **80**, 2222–2226.
- Greenspan, S. L., and Greenspan, F. S. (1999). The effect of thyroid hormone on skeletal integrity. *Ann. Int. Med.* **130**, 750–758.
- Mercuro, G., Panzuto, M. G., Bina, A., Leo, M., Cabura, R., Petrini, L., Pigliaru, F., and Mariotti, S. (2000). Cardiac function, physical exercise capacity, and quality of life during long-term thyrotropin-suppressive therapy with levothyroxine: Effect of individual dose tailoring. *J. Clin. Endocrinol. Metab.* **85**, 159–164.
- Parle, J. V., Maisonneuve, P., Sheppard, M. C., Boyle, P., and Franklyn, J. A. (2001). Prediction of all-cause and cardiovascular mortality in elderly people from one low serum thyrotropin result: A 10-year cohort study. *Lancet* **358**, 861–865.
- Sawin, C. T., Geller, A., Wolf, P. A., Belanger, A. J., Baker, E., Bacharach, P., Wilson, P. W., Benjamin, E. J., and D'Agostino, R. B. (1994). Low serum thyrotropin concentrations as a risk factor for atrial fibrillation in older persons. *N. Engl. J. Med.* **331**, 1249–1252.

In adipose and muscle capillaries, TG in chylomicrons and VLDL are hydrolyzed into FFA by endothelial-bound lipoprotein lipase (LPL). ApoCII is an obligatory cofactor for LPL, whereas apoCIII may interfere with LPL. FFA is then reesterified and stored in adipocytes or is oxidized for energy in myocytes. Chylomicrons and VLDL are remodeled into the short-lived, smaller, denser, CE-enriched chylomicron remnants (CMR) and VLDL remnants (also called intermediate-density lipoprotein [IDL]), respectively. CMR and some IDL are cleared by apoE-mediated endocytosis through hepatic remnant receptors. IDL can also be hydrolyzed by hepatic

lipase, making smaller, CE-rich, low-density lipoprotein (LDL) particles. [Table I](#) summarizes the proteins involved in TG metabolism.

## CLINICAL IMPORTANCE OF HYPERTRIGLYCERIDEMIA

### Triglyceride and Atherosclerosis

Moderate hypertriglyceridemia is probably an independent coronary heart disease (CHD) risk factor. The Prospective Cardiovascular Munster (PROCAM) study followed approximately 5000 men, ages 40 to 65 years, for 8 years and showed an increased risk of myocardial infarction (MI) or sudden cardiac death as TG increased from less than 200 mg/dl (<2.3 mmol/L) to 799 mg/dl (9.0 mmol/L). This association remained significant after adjustment for LDL and high-density lipoprotein (HDL) cholesterol and for other CHD risk factors. Furthermore, meta-analyses of more than 46,000 men followed for an average of 8.4 years and of more than 10,000 women followed for an average of 11.4 years showed that for a 1-mmol/L TG elevation, CHD risk increased by 32 and 76%, respectively. Adjustment for HDL cholesterol attenuated this CHD risk increase to 14% in men and 37% in women.

Complex mechanisms underlie the association of hypertriglyceridemia with atherosclerosis, and the complexity obscures ascertainment of a direct causal relationship. Proatherogenic metabolic and/or biochemical abnormalities, such as obesity, diabetes, decreased HDL cholesterol, increased small dense LDL, increased FFA, dysglycemia, hyperinsulinemia, increased plasma viscosity, increased inflammatory molecules, impaired fibrinolysis, and prothrombosis, are often associated with elevated TG. In addition, triglyceride-rich lipoproteins (TGRLPs) or their remnants may act directly in atherogenesis, contributing to arterial wall foam cell formation. It has been considered axiomatic by experienced clinicians that chylomicrons are not directly atherogenic, and rare reports of atherosclerosis in genetic hyperchylomicronemia are the exceptions that prove this clinical rule. In contrast, CMR, VLDL, and IDL appear to be atherogenic. Because postprandial FFA released from lipolysis impairs physiological endothelial response, a newer concept is that postprandial lipemia may independently predict CHD and other vascular disease, although the definition of "postprandial hypertriglyceridemia" and standardized sample preparation and assay conditions have not yet been established.

**Table I** Proteins Involved in Triglyceride Metabolism

Protein	Function
Pancreatic lipase	Hydrolyzes dietary TG into monoglycerides and FFA
Fatty acid-binding protein	Transports FFA from intestinal lumen into enterocytes
Monoacylglycerol acyltransferase	Catalyzes conversion of acyl-CoA and 2-monoacylglycerol into diacylglycerol and CoA
Diacylglycerol acyltransferase	Catalyzes conversion of acyl-CoA and diacylglycerol into TG (the committed step in TG synthesis)
ApoB48	Structural protein required for chylomicron synthesis
ApoB100	Structural protein required for VLDL synthesis and ligand for LDL receptor
Microsomal TG transfer protein	Mediates assembly of TG with apoB48 and apoB100 to form chylomicrons and VLDL, respectively
ApoE	Ligand that mediates clearance of CMRs and some IDLs through hepatic remnant receptors
Lipoprotein lipase	Capillary endothelial-bound enzyme that liberates FFA from TG in chylomicrons and VLDL
ApoCII	Cofactor essential for LPL activity
ApoCIII	Inhibitor of LPL and hepatic lipase
Sterol regulatory element-binding protein-1	Membrane-bound transcription factor that regulates fatty acid synthesis
Fatty acid synthase	Enzyme that catalyzes all of the steps involved in the conversion of malonyl-CoA to palmitate

## Hypertriglyceridemia and Pancreatitis

Very high TG resulting from elevated chylomicrons is associated with increased risk of acute pancreatitis. The mechanism underlying this association is uncertain, but the relatively unique capacity of the pancreas to produce an exocrine lipase might play a role. Most patients with pancreatitis from high TG probably have preexisting abnormal lipoprotein metabolism, and milder hypertriglyceridemia will frequently persist after recovery. Pancreatitis risk is accentuated by diabetes, pregnancy, estrogen use, alcohol consumption, and/or any factor that can abruptly increase plasma TG to more than 1000 mg/dl ( $>11.3$  mmol/L). Hypertriglyceridemia-induced pancreatitis can be preceded by episodic nausea and epigastric pain, during which serum amylase might not exceed common diagnostic cut points. Specific exclusion of lipemic serum should be documented before accepting a report of "normal" amylase in a patient who presents with an acute abdomen. If this is not done, TG-related pancreatitis might be excluded in favor of another diagnosis. Associated clinical clues include eruptive xanthomata and/or lipemia retinalis. The onset of full-blown acute pancreatitis can be prevented by TG reduction through rigorous dietary fat restriction. Once pancreatitis has developed, treatment involves hemodynamic stabilization, cessation of oral intake (possibly with placement of a nasogastric tube), and control of any primary metabolic disturbances such as diabetes. With restriction of caloric intake, plasma TG can fall quite briskly. Plasmapheresis may be an option in extreme cases, although the benefit is very transient if the primary metabolic disturbance is not treated. The individual thresholds for developing pancreatitis can vary markedly, and some individuals can be symptom free and survive with apparently normal longevity despite very high plasma TG ( $>3500$  mg/dl [ $>40$  mmol/L]).

## CLASSIFICATION OF HYPERTRIGLYCERIDEMIA

### Primary Hypertriglyceridemia

It could be argued that during the third millennium, a classification system for TG disorders should be based on a molecular diagnosis. However, human genetics has uncovered the molecular basis of only a minority of TG disorders. Thus, the time-honored Fredrickson system of lipoprotein phenotypes is presented here because it remains an entrenched system

of diagnostic shorthand for lipidologists and clinical biochemists. Five of the six Fredrickson types contain elevated TG as an essential diagnostic feature.

### **Familial Chylomicronemia Syndrome (MIM 238600)**

Familial chylomicronemia (type I hyperlipoproteinemia) is characterized by excess chylomicrons. Overnight-refrigerated plasma develops a creamy supernatant and a clear infranatant. Patients are usually diagnosed during early infancy, although occasionally some may go unnoticed until later in life, when pancreatitis or lipemia is noted. Clinical features include eruptive xanthomata, lipemia retinalis, hepatosplenomegaly, focal neurological symptoms (e.g., memory loss, inability to concentrate), and recurrent epigastric pain. Modern laboratory methods have eliminated the artifactual hyponatremia that used to be a feature of this condition. Fasting TG exceeds 1000 mg/dl ( $>11.3$  mmol/L) and occurs together with low plasma LDL and HDL cholesterol. Causes of this syndrome include LPL deficiency, apoCII deficiency, and ill-defined heritable inhibitors of LPL.

LPL deficiency is the most common molecularly defined cause of familial chylomicronemia. LPL deficiency is an autosomal recessive condition, occurring with a frequency of 1 in 1 million and resulting from two defective alleles of the LPL gene. The biochemical diagnosis of LPL deficiency is established by finding markedly reduced LPL activity in postheparin plasma. Deficiency of the LPL cofactor apoCII is an autosomal recessive condition (MIM 207750) that is even less common than LPL deficiency. ApoCII deficiency is less clinically severe than LPL deficiency. ApoCII deficiency can be identified by isoelectric focusing, DNA sequencing of the APOC2 gene, or documenting the rescue of absent *ex vivo* postheparin lipolytic activity by the addition of normal serum.

### **Primary Mixed Hypertriglyceridemia (MIM 144650)**

Primary mixed hypertriglyceridemia (type V hyperlipoproteinemia) is characterized by elevated VLDL and chylomicrons. Some consider this condition to be a part of a chylomicronemia continuum that includes type I hyperlipoproteinemia. Type V hyperlipoproteinemia is typically an adult disease with a frequency up to 1 in 1000. Clinical features may include eruptive xanthomata, hepatosplenomegaly, lipemia retinalis, and recurrent epigastric pain with or without pancreatitis. Laboratory findings include fasting TG that exceeds 1000 mg/dl ( $>11.3$  mmol/L),

increased total cholesterol, and low LDL and HDL cholesterol concentrations. Although heterozygosity for the mutant LPL gene has on occasion been demonstrated, disease expression is usually associated with secondary factors such as alcohol, poor diet, obesity, diabetes, and hypothyroidism.

#### **Familial Hypertriglyceridemia (MIM 145750)**

Familial hypertriglyceridemia (type IV hyperlipoproteinemia) is probably polygenic or possibly autosomal dominant with variable penetrance. It seems to be the most common primary hypertriglyceridemia phenotype seen in clinical practice. It certainly is a frequent cause of mild to moderate hypertriglyceridemia. The main lipoprotein abnormality is increased VLDL. The molecular basis of the phenotype is unknown in most instances. Typically, patients have moderately elevated plasma TG concentration ranging from 200 to 500 mg/dl (2.3–5.7 mmol/L), usually with low HDL. Familial hypertriglyceridemia is associated with increased CHD risk and often with obesity, insulin resistance, hyperglycemia, hypertension, and hyperuricemia.

#### **Secondary Hypertriglyceridemia**

Some metabolic conditions are frequently, but not absolutely, associated with high TG. One interpretation of these associations is that patients who develop secondary hypertriglyceridemia might already have a subtle metabolic defect that is perhaps genetically determined and that creates susceptibility to clinical hypertriglyceridemia depending on the presence of a metabolic stress. Obesity is probably the most commonly associated clinical attribute in hypertriglyceridemic patients.

#### **Alcohol Ingestion**

Hypertriglyceridemia associated with alcohol intake is also due mainly to increased plasma VLDL, with or without chylomicronemia. Within the liver, ethanol is converted to acetate, and this has a sparing effect on fatty acid oxidation, resulting in increased hepatic TG production and enhanced VLDL secretion. In many individuals, plasma TG can remain within the normal range due to adaptive increase in lipolytic activity. However, ethanol can also impair lipolysis, leading to increased plasma TG.

#### **Renal Disease**

Nephrotic syndrome is characterized by an increase in apoB-containing lipoproteins, including VLDL.

The mechanism underlying this increase includes overproduction by the liver, which has also increased albumin synthesis to compensate for renal protein wasting. Also, with hypoalbuminemia, more FFA binds to lipoproteins, which might impair LPL. Uremia is associated with modest elevation in VLDL, reflecting impairment of lipolysis, possibly due to the toxic effect of uremic metabolites.

#### **Pregnancy**

Plasma TG normally rises during the third trimester of pregnancy by up to threefold due to increased hepatic secretion of VLDL and reduced LPL activity. The physiological increase in plasma TG has little clinical consequence. However, more pronounced TG increases have been reported in association with reduced or absent LPL activity, and with the apoE4/E2 genotype, with variable pregnancy outcome. Severe hypertriglyceridemia during pregnancy due to chylomicronemia is very rare but can be complicated by pancreatitis, which carries a significant risk of mortality for both the mother and the fetus.

#### **Medications**

Use of specific drugs has been associated with hypertriglyceridemia, which can be profound in susceptible individuals. Medications that commonly exacerbate hypertriglyceridemia include highly active antiretroviral combination therapies, oral estrogens, isotretinoin, beta blockers, thiazides, tamoxifen, and bile acid-binding resins. If a medication is considered to be an important determinant of hypertriglyceridemia, the indications for the particular treatment should be reviewed, particularly if the hypertriglyceridemia is marked. If dose reductions, changes in route of administration, and/or substitution with another class of medication are not possible, marked TG elevation should be treated with diet and/or pharmacological agents.

#### **Other Causes of Secondary Hypertriglyceridemia**

Hypothyroidism is usually associated with elevated LDL, but elevated TG may also be present, perhaps due to impaired lipolysis. Paraproteinemias, such as the hypergammaglobulinemia in macroglobulinemia, myeloma, lymphoma, and lymphocytic leukemias, and autoimmune disorders, such as systemic lupus erythematosus, may cause hypertriglyceridemia, possibly through immune-mediated interference of lipolysis.

## TRIGLYCERIDE-LOWERING THERAPIES

### Nonpharmacological Therapy

Patients with hypertriglyceridemia are frequently obese with insulin resistance, hypertension, and/or diabetes. Because these are also CHD risk factors, they should be identified and treated as part of a global risk factor reduction strategy. The marked sensitivity of plasma TG to energy balance means that treatment plans should include weight reduction, dietary modification, and exercise. Dietary modification should be aimed at weight loss, with decreased overall fat intake and a reduction in refined carbohydrates or so-called "high glycemic index" foods. In general, the severity of the hypertriglyceridemia dictates the severity of the fat restriction. For instance, in severe hyperchylomicronemia, recommendations are often restriction of fat to approximately 10 to 15% of total calories, with reductions in both saturated and unsaturated fat. In an adult, this represents 15 to 20 g/day of fat. The diet should include at least 5 g/day of polyunsaturated fat as a source of essential fatty acids, and fat-soluble vitamins must be provided. A specialized dietician can be very helpful in these circumstances.

For less severe hypertriglyceridemia, restriction of saturated fat together with increased aerobic activity may lead to substantial reductions in plasma TG. The National Cholesterol Education Program (NCEP) advises that carbohydrate and protein intake should be 55 to 60% and 15 to 20%, respectively, whereas total and saturated fat should be less than 30% and less than 7%, respectively, of daily calories.

Omega-3 fatty acids, such as eicosapentanoic and docosahexanoic acid, are components of both the Mediterranean diet and fish oils. Omega-3 fatty acids can reduce hepatic secretion of TGRLP. Ingestion of 3 to 4 g/day of omega-3 fatty acids, with caloric and saturated fat restriction, may reduce plasma TG by up to 20%. However, response to omega-3 fatty acids is not uniform when used as the sole TG-lowering therapy.

### Pharmacological Therapy

In general, monotherapy with pharmacological agents should be attempted first, together with diet. Combination treatment may be required for refractory severe hypertriglyceridemia but should be attempted only with caution and with a plan for frequent follow-up and monitoring of serum creatine kinase and transaminases.

#### **Fibric Acid Derivatives (fibrates)**

Fibrates, such as gemfibrozil, bezafibrate, and fenofibrate, are a mainstay of treatment of hypertriglyceridemia. Fibrates can reduce plasma TG by up to 50% and can raise plasma HDL cholesterol by up to 20%, although these percentages are variable. The mechanism of action of fibrates is complex but includes modulation of the activity of peroxisome proliferator-activated receptor- $\alpha$  in the liver, with reduced hepatic secretion of VLDL and increased lipolysis of plasma TG, possibly related to decreased secretion of apoCIII. Gemfibrozil reduced CHD events in the Veterans Affairs HDL Cholesterol Intervention Trial. Fibrates can sometimes increase plasma LDL cholesterol concentration, and this may require a change to another drug or the addition of a second agent. Fibrates are generally very well tolerated, with very rare reports of hepatitis and myositis.

#### **HMG-CoA Reductase Inhibitors (statins)**

In addition to lowering LDL, statins used at higher doses can produce clinically significant TG decreases, probably by lowering hepatic secretion of apoB-containing lipoproteins. However, statins should not be considered as firstline therapy when TG is much above 500 mg/dl (5.7 mmol/L). An advantage of statins is the preponderance of clinical trial results indicating marked reductions in CHD end points. Like fibrates, statins are well tolerated but very rarely may cause myopathy and/or hepatic toxicity, often with sentinel elevations in creatine kinase and/or transaminases.

#### **Niacin (nicotinic acid)**

Niacin has pleiotropic, incompletely defined effects on lipoprotein metabolism that may include inhibition of hepatic VLDL secretion and stimulation of lipolysis. Niacin, when administered 3 g daily in divided doses, can lower plasma TG by up to 45%, raise plasma HDL cholesterol by up to 25%, and reduce plasma LDL cholesterol by up to 20%. Older clinical trials suggested a reduction in CHD events related to niacin treatment. However, niacin can cause lightheadedness, cutaneous flushing, and/or pruritus. These adverse effects can be minimized by initiating therapy at low doses and then gradually increasing the daily dose, along with the concomitant use of aspirin. Less common adverse effects include elevation of liver enzymes, gastrointestinal distress, worsening of glucose tolerance, and elevation of uric acid. Longer acting preparations of niacin may reduce the frequency and intensity of these adverse effects, but the clinical impression is that their TG-lowering effect may be less than that seen with crystalline niacin.

## TARGET TRIGLYCERIDE GOALS

According to the recent NCEP guidelines, the primary aim of therapy when TG is more than 150 mg/dl (>1.7 mmol/L) is to attain the LDL cholesterol target for the CHD risk stratum. When TG is borderline high, emphasis should be placed on weight loss and exercise. For patients with moderately high TG, weight reduction, increased physical activity, and drug therapy may be considered in high-risk persons to achieve the NCEP targets. For patients with very high TG, the initial aim of therapy is prevention of acute pancreatitis through TG lowering by diet, medication, weight reduction, control of secondary metabolic factors, and increased physical activity. Once TG has been lowered to less than 500 mg/dl (5.7 mmol/L), attention may be directed to LDL lowering. The Canadian Working Group on Hypercholesterolemia and Other Dyslipidemias guidelines suggest TG of less than 266 mg/dl (3 mmol/L) in individuals with low CHD risk and TG of less than 177 mg/dl (2 mmol/L) in individuals with moderate to very high CHD risk, including those with diabetes mellitus and/or preexisting vascular disease.

## See Also the Following Articles

Abetalipoproteinemia • Anderson's Disease (Chylomicron Retention Disease) • Familial Low Cholesterol Syndromes, Hypobetalipoproteinemia • Lipoprotein(a)

## Further Reading

- Despres, J.-P., Lemieux, I., and Prud'homme, D. (2001). Treatment of obesity: Need to focus on high risk abdominally obese patients. *Br. Med. J.* **322**, 716–720.
- Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. (2001). Executive summary of the third report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). *J. Am. Med. Assoc.* **285**, 2486–2497.
- Fodor, J. G., Frohlich, J. J., Genest, J. J., Jr., and McPherson, P. R. (2000). Recommendations for the management and treatment of dyslipidemia: Report of the Working Group on Hypercholesterolemia and Other Dyslipidemias. *Can. Med. Assoc. J.* **162**, 1441–1447.
- Fojo, S. S. (1998). The familial chylomicronemia syndrome. *Endocrinol. Metab. Clinics North America* **27**, 551–567.
- Ginsberg, H. N. (2001). Hypertriglyceridemia: New insights and new approaches to pharmacologic therapy. *Am. J. Cardiol.* **87**, 1174–1180.
- Ginsberg, H. N., and Goldberg, I. J. (1998). Disorders of lipoprotein metabolism. In "Harrison's Principles of Internal Medicine" (A. S. Fauci, et al., eds.), 14th ed., pp. 2138–2149. McGraw-Hill, New York.
- Malloy, M. J., and Kane, J. P. (2001). Disorders of lipoprotein metabolism. In "Basic and Clinical Endocrinology" (F. S. Greenspan and D. G. Gardner, eds.), 6th ed., pp. 716–744. McGraw-Hill, New York.
- Miller, M. (2000). Current perspectives on the management of hypertriglyceridemia. *Am. Heart J.* **140**, 232–240.
- Rubins, H. B., Robins, S. J., Collins, D., Fye, C. L., Anderson, J. W., Elam, M. B., Faas, F. H., Linares, E., Schaefer, E. J., Schectman, G., Wilt, T. J., and Wittes, J. (2000). Gemfibrozil for the secondary prevention of coronary heart disease in men with low levels of high-density lipoprotein cholesterol: Veterans Affairs High-Density Lipoprotein Cholesterol Intervention Trial Study Group. *N. Engl. J. Med.* **341**, 410–418.

action. Therefore, calcium should be given in conjunction with magnesium for several days until normal PTH responsiveness is restored. Treatment of severe hypomagnesemia consists of giving 2 g magnesium sulfate intravenously over 5 to 10 min initially. If serum magnesium is still low, 6 g magnesium sulfate can be infused in 5% dextrose/0.5% saline over 24 h. For chronic hypomagnesemia, magnesium oxide (250–500 mg by mouth two to four times daily) should be adequate.

Hypermagnesemia is a much less common abnormality than hypomagnesemia and is usually observed as result of magnesium overdose, for example, in obstetric practice when high-dose magnesium infusions are used for treatment of toxemia of pregnancy and in patients with renal failure who are receiving an antacid or enema containing magnesium. Hypermagnesemia can result in reversible hypocalcemia by mechanisms similar to those operating in hypomagnesemia. Tetany is unusual in this entity of hypocalcemia due to the concomitant hypermagnesemia. Treatment of moderate hypermagnesemia simply requires magnesium withdrawal to allow magnesium excretion by the kidneys. In cases of severe hypermagnesemia, intravenous calcium (100–200 mg over 5–10 min) will antagonize the toxic effects of magnesium in addition to raising serum calcium.

## MANAGEMENT OF HYPOCALCEMIC EMERGENCY

Most authors agree that hypocalcemic emergency exists when corrected serum calcium is less than 7.5 mg/dl or in the presence of severe symptoms or

signs (e.g., tetany, seizures, severe hypotension, heart failure). The mainstay treatment in hypocalcemic emergency is parenteral calcium therapy. There are several forms of calcium available for intravenous administration (Table I). Calcium chloride has a much higher content of elemental calcium than does calcium gluconate (272 and 90 mg/10 ml, respectively). Therefore, administration of calcium chloride can raise serum calcium more rapidly than can calcium gluconate. However, calcium chloride has the disadvantage of being extremely irritating to the veins and soft tissues in case of extravasation.

As initial therapy, one or two ampoules, each containing 10 ml of 10% calcium chloride or 10 ml of 10% calcium gluconate, are diluted in 50 to 100 ml of 5% dextrose and given intravenously over 10 min. In a severe emergency situation (e.g., a patient with tetany), a 10-ml ampoule of 10% calcium gluconate can be infused directly over 4 min.

If hypomagnesemia is suspected (e.g., an alcoholic patient, a patient having diarrhea, a patient receiving diuretics), magnesium administration (e.g., magnesium sulfate [2 g intravenously]) is appropriate pending the results of serum magnesium. Oral calcium and a rapidly acting form of vitamin D, such as calcitriol, should be started simultaneously with intravenous calcium.

If the cause of hypocalcemia is likely to be longstanding or due to a permanent defect (e.g., hypoparathyroidism), continuous calcium infusion for a few days is indicated until the actions of oral calcium and vitamin D start. Different protocols exist regarding the concentration and rate of calcium drip. The recommended initial rate ranges from 0.3 to 2.0 mg elemental calcium/kg/h depending on the severity of the

Table I Parenteral Calcium Preparations

Calcium preparation	Elemental calcium	Initial treatment	Remarks
Calcium gluconate (10%)	90 mg per 10-ml ampoule	1–2 ampoules in 50 to 100 ml of 5% dextrose over 5 to 10 min; one ampoule may be infused directly over 4 min in emergency situations	Less irritating to the veins than calcium chloride; therefore, more convenient for prolonged infusion <sup>a</sup>
Calcium chloride (10%)	272 mg per 10-ml ampoule	1–2 ampoules in 50 to 100 ml of 5% dextrose over 5 to 10 min	Raises serum calcium more rapidly than does calcium gluconate, but more toxic to the veins
Calcium gluceptate (10%)	90 mg per 5-ml ampoule	1–2 ampoules in 50 to 100 ml of 5% dextrose over 5 to 10 min	Useful in patients who cannot tolerate large volumes of fluid

<sup>a</sup>Ten ampoules of calcium gluconate are added to 900 ml of 5% dextrose (i.e., 900 mg elemental calcium/L) to be infused at 50 ml/h (45 mg elemental calcium/h) and then rate titrated. If necessary, this solution can be infused as rapidly as over 4 to 6 h.

**Table II** Initial Management of Hypocalcemia

1. Confirm hypocalcemia by measurement of ionized calcium or total serum calcium corrected for the degree of hypoalbuminemia.
2. Measure serum magnesium, phosphorus, PTH, 25-hydroxy-vitamin D, and creatinine.
3. If corrected serum calcium is less than 7.5 mg/dl or in the presence of severe hypocalcemic symptoms, administer intravenous calcium (e.g., 1–2 ampoules of 10% calcium chloride or gluconate) in addition to oral calcium and calcitriol.
4. If serum calcium is still low and in long-term causes of hypocalcemia, start calcium drip for a few days until oral calcium and vitamin D actions start.
5. If hypocalcemia is mild with minimal symptoms and signs, if any start oral calcium and vitamin D.

case. The infusion rate is titrated according to serum calcium levels, which are monitored every 4 to 8 h to maintain serum calcium in the low-normal range. A simple protocol of calcium gluconate drip is shown in [Table I](#). A total infusion of 15 mg/kg calcium gluconate over 4 to 6 h will raise the serum calcium by approximately 2 to 3 mg/dl (0.5–0.75 mmol/L). Recommended calcium infusion rates should not be exceeded so as to avoid cardiac dysfunction. Cardiac monitoring is advisable in all patients, particularly those receiving digitalis due to their propensity for arrhythmias if the serum calcium level is raised excessively. Bicarbonate or phosphate should not be infused along with calcium because of possible intravenous precipitation of those calcium salts. Initial management of hypocalcemia is summarized in [Table II](#).

## LONG-TERM TREATMENT OF HYPOCALCEMIA

The most severe form of chronic hypocalcemia is hypoparathyroidism, which is due mostly to surgical removal or vascular compromise of the parathyroid glands. In hypoparathyroidism, the physiological calcium-retaining effect of PTH in renal tubules is lacking. Therefore, the goal of therapy in hypoparathyroid states is maintenance of serum calcium in the low-normal range rather than in the mid- or high-normal range so as to limit hypercalciuria,

nephrolithiasis, and nephrocalcinosis. Among the oral calcium preparations available ([Table III](#)), calcium carbonate is the most commonly used because it offers the highest amount of elemental calcium (40%) per unit tablet weight. In addition, it is the least expensive. Calcium carbonate tablets are preferentially taken with food because their absorption is enhanced in an acid environment. Gastrointestinal side effects include bloating and constipation. If these occur, switching to calcium citrate is recommended. The latter formula is better absorbed and tolerated, but is more expensive and has less elemental calcium, compared with calcium carbonate. Treatment usually starts with daily doses of 1000 to 4000 mg elemental calcium divided into three or four doses. In hypocalcemia of renal failure, hyperphosphatemia is reduced by limitation of dietary phosphate and the use of phosphate-binding agents such as calcium carbonate and calcium acetate.

## SIDE EFFECTS OF ORAL CALCIUM

The efficacy of calcium absorption from the gut declines as calcium intake increases, providing a protective mechanism against calcium toxicity. However, this adaptive mechanism can be overcome by high calcium intake (usually more than 4 g daily). The consequences of calcium toxicity are shown in [Table IV](#). Milk-alkali syndrome is a form of calcium toxicity initially described in patients consuming large

**Table III** Oral Calcium Preparations

Calcium salt	Elemental calcium percentage	Remarks
Carbonate	40	Inexpensive; may cause bloating and constipation
Citrate	30	Well tolerated and absorbed; expensive
Lactate	13	Poor in elemental calcium
Gluconate	9	Poor in elemental calcium
Glubionate	7	Available in palatable liquid form (252 mg/10 ml)
Acetate	25	Given as phosphate binders in renal failure

Table IV Complications of Calcium and Vitamin D Toxicity

<b>Renal</b>
Polyuria, polydipsia, hypercalciuria, nephrolithiasis, nephrocalcinosis
<b>Gastrointestinal</b>
Nausea, vomiting, constipation
<b>Cardiovascular</b>
Arrhythmias, short QT interval, hypertension
<b>Central nervous system</b>
Headache, altered mental status
<b>Others</b>
Metastatic calcification of soft tissues (e.g, kidney, lung) <sup>a</sup>
<b>Laboratory abnormalities</b>
Hypercalcemia, hyperphosphatemia <sup>a</sup> , high serum levels of 1,25-dihydroxy-vitamin D <sup>a</sup>

<sup>a</sup>Complications occurring primarily with vitamin D toxicity.

amounts of dairy calcium milk and absorbable antacids for treatment of peptic ulcer disease. Nowadays, the syndrome is more commonly seen with administration of excessive doses of calcium carbonate. The syndrome consists of nausea, vomiting, renal failure, metabolic alkalosis, and hypercalcemia.

## VITAMIN D THERAPY

The two precursor molecules of vitamin D are ergocalciferol (vitamin D<sub>2</sub>) and cholecalciferol (vitamin D<sub>3</sub>). In the liver, vitamin D is converted to 25-hydroxy-vitamin D, which requires further hydroxylation in the kidney to form the biologically active 1,25-dihydroxy-vitamin D (calcitriol). PTH activates the renal conversion of 25-hydroxy-vitamin D to 1,25-dihydroxy-vitamin D. Dihydroxycholesterol is an active synthetic analogue that does not require further metabolism and is rarely used nowadays.

Vitamin D is required for the treatment of most, if not all, cases of hypocalcemia for two reasons. First, it

improves the bioavailability of calcium present in food or calcium given as medication by providing optimal calcium absorption in the gut. Second, chronic hypocalcemic states are frequently associated with a decreased synthesis or function of one or more forms of vitamin D. For instance, in hypoparathyroidism, lack of PTH results in decreased synthesis of the active form of vitamin D, 1,25-dihydroxy-vitamin D. The latter is also decreased in severe renal insufficiency due to hyperphosphatemia and decreased functional renal mass. In end-stage liver disease, formation of 25-hydroxy-vitamin D is impaired. Selection of the vitamin D preparation depends on the specific abnormality in vitamin D synthesis and activation. Thus, in the hypoparathyroid states and renal disease, 1,25-dihydroxy-vitamin D is the most rational therapy. In severe hepatic disease, either 25-hydroxy-vitamin D or 1,25-dihydroxy-vitamin D could be used. In malabsorption, relatively large doses of any form of vitamin D are appropriate orally or intramuscularly (every 6 months).

Table V depicts the pharmacological properties of various vitamin D preparations. Vitamin D (either D<sub>2</sub> or D<sub>3</sub>) is extremely lipophilic, has virtually extensive storage sites in fatty tissue, and has a long half-life of about 30 days. On the contrary, 1,25-dihydroxy-vitamin D is a polar compound, not stored in large amounts of fat, and has a short half-life of approximately 6 h. The time periods required to normalize serum calcium can be shortened considerably by starting treatment with loading doses. This policy is particularly useful in severe hypocalcemia when rapid normalization of serum calcium is warranted. For instance, calcitriol can be given in a loading dose of 4 µg daily for a few days and then in a maintenance dose of 0.5 to 1.0 µg daily.

Cost also plays a role in choosing vitamin D formulation. In general, cost parallels the biological activity. Thus, the precursor compounds, ergocalciferol (vitamin D<sub>2</sub>) and cholecalciferol (vitamin D<sub>3</sub>), are

Table V Vitamin D Preparations

Preparation	Approximate daily dose (µg) for treatment of hypocalcemia	Onset of action (days)	Approximate time to normalize serum calcium	Time to reverse toxicity
Ergocalciferol (vitamin D <sub>2</sub> ), cholecalciferol (vitamin D <sub>3</sub> )	1000–3000 (40,000–120,000 U) <sup>a</sup>	10–14	4–8 weeks	4–12 weeks
Calcifediol (25-hydroxy D <sub>3</sub> )	50–225	7–10	2–4 weeks	2–6 weeks
Dihydroxycholesterol	300–1000	4–7	1–2 weeks	3–14 days
Calcitriol (1,25-dihydroxy D <sub>3</sub> )	0.25–2.5	1–2	3–7 days	2–10 days

<sup>a</sup>1 µg = 40 U.

**Table VI Treatment of Hypercalcemia Resulting from Calcium or Vitamin D Toxicity**


---

1. Aggressive hydration with normal saline
2. In patients who cannot tolerate fluids, hemodialysis against a low-dialysis calcium bath
3. Loop diuretics after adequate hydration
4. Corticosteroids, prednisone (30–40 mg daily orally) or hydrocortisone (50 mg every 8 h), particularly in vitamin D-induced hypercalcemia
5. Calcitonin (4–6 U/kg every 6–8 h subcutaneously)
6. In severe cases, biphosphonate infusion (e.g., pamidronate, 90 mg)

---

the least expensive, whereas 1,25-dihydroxy-vitamin D is the most expensive. 25-hydroxy-vitamin D<sub>3</sub> (calcifediol) is intermediate in cost.

### VITAMIN D TOXICITY

Vitamin D toxicity virtually occurs more easily with the most potent agent, 1,25-dihydroxy-vitamin D, than with the inactive vitamin D forms. However, because of the short duration of action of 1,25-dihydroxy-vitamin D, its toxicity is reversed within a few days (Table V). On the other hand, toxicity due to excessive administration of vitamin D<sub>2</sub> and vitamin D<sub>3</sub> can be protracted and last for several months (Table IV). Symptoms and signs of vitamin D overdose are related mainly to those of the prevailing hypercalcemia and hyperphosphatemia (Table IV). Treatment of calcium or vitamin D overdose is shown in Table VI.

### MONITORING OF THERAPY FOR CHRONIC HYPOCALCEMIA

Close monitoring of serum calcium is required initially every few days after starting oral therapy. Stable laboratory values are usually reached within 3 to 4 weeks. Subsequently, periodic measurement of serum calcium, phosphorus, and magnesium can be performed every 2 to 4 months. Urine calcium may be assessed shortly after initiation of therapy and then every 6 to 12 months to prevent the development of hypercalciuria and to keep urinary calcium at less than 350 mg/24 h. Frequently, hypoparathyroid patients whose serum calcium is maintained in the normal range have hypercalciuria due to the absence of PTH action on renal calcium reabsorption. Thiazide diuretics enhance calcium reabsorption in distal renal tubule and, therefore, limit the amount of urinary calcium. Thus, thiazides can be useful in many hypoparathyroid patients as adjunctive therapy to decrease hypercalciuria and calcium/vitamin requirements. In

advanced renal insufficiency, oral calcium, dialysate calcium, and calcitriol must be properly balanced to keep the calcium–phosphorus product within the normal range so as to avoid metastatic calcification. Acceptable targets are serum phosphorus of 4.5 mg/dl and total serum calcium of 10 mg/dl.

### CONCLUSION

Hypocalcemia is a common disorder characterized by decreased levels of ionized serum calcium, the only biologically active form of calcium. The serum magnesium level should be assessed in every case of hypocalcemia and should be corrected if it is abnormal. Hypocalcemic emergency is a life-threatening condition occurring when severe symptoms exist and/or corrected serum calcium is less than 7.5 mg/dl. Intravenous calcium is life-saving in hypocalcemic emergency, and in many cases calcium infusion for a few days is required until actions of oral calcium and vitamin D are in effect. Calcium carbonate is the oral preparation of choice due to its high content of elemental calcium and low cost. Calcitriol is the most potent and rapidly acting vitamin D form, but its widespread use is limited by its high cost. In hypoparathyroid states, the goal of therapy is to maintain serum calcium in the low-normal range to avoid hypercalciuria. The latter can be limited by the administration of thiazide diuretics.

### See Also the Following Articles

Hypercalcemia and Hypercalcemia Treatment • Hyperphosphatemia • Hypoparathyroidism • Kidney Stones • Magnesium Disorders • Vitamin D

### Further Reading

Bouillon, R. (2001). Vitamin D: From photosynthesis, metabolism, and action to clinical applications. *In* "Endocrinology" (L. J. DeGroot and J. L. Jameson, eds.), pp. 1009–1028. Saunders, Philadelphia.

- Bushinsky, D. A., and Monk, R. D. (1998). Calcium. *Lancet* **352**, 306–311.
- Heaney, R. P., and Weaver, C. M. (2003). Calcium and vitamin D. *Endocrinol. Metab. Clin. North America* **32**, 181–194.
- Levine, M. A. (2001). Hypoparathyroidism and pseudohypoparathyroidism. In “Endocrinology” (L. J. DeGroot and J. L. Jameson, eds.), pp. 1133–1153. Saunders, Philadelphia.
- Marx, S. J. (2000). Hyperparathyroid and hypoparathyroid disorders. *N. Engl. J. Med.* **343**, 1863–1875.
- Reber, P. M., and Heath, H. (1995). Hypocalcemic emergencies. *Med. Clin. North America* **79**, 93–107.
- Shane, E. (1999). Hypocalcemia: Pathogenesis, differential diagnosis, and management. In “Primer on the Metabolic Bone Diseases and Disorders of Mineral Metabolism” (M. J. Favus, ed.), pp. 223–226. Lippincott–Raven, Philadelphia.
- Thomas, M. K., and Demay, M. B. (2000). Vitamin D deficiency and disorders of vitamin D metabolism. *Endocrinol. Metab. North America* **29**, 611–627.



## Hypocortisolism

see Adrenal Insufficiency

**Table I** Effects of a 120-min Hyperinsulinemic Hypoglycemic (~50 mg/dl) Clamp on Neuroendocrine and Autonomic Responses in Overnight Fasted Men (means  $\pm$  SEs)

	Basal	Final 30 min of clamp
Norepinephrine (pg/ml)		
Euglycemic clamp	163 $\pm$ 15	225 $\pm$ 24 <sup>*</sup>
Hypoglycemic clamp	165 $\pm$ 10	328 $\pm$ 20 <sup>*,†</sup>
Epinephrine (pg/ml)		
Euglycemic clamp	38 $\pm$ 6	35 $\pm$ 5
Hypoglycemic clamp	36 $\pm$ 5	934 $\pm$ 110 <sup>*,†</sup>
Pancreatic polypeptide (pg/ml)		
Euglycemic clamp	168 $\pm$ 15	86 $\pm$ 15
Hypoglycemic clamp	132 $\pm$ 25	1209 $\pm$ 134 <sup>*,†</sup>
Glucagon (pg/ml)		
Euglycemic clamp	78 $\pm$ 10	59 $\pm$ 3
Hypoglycemic clamp	71 $\pm$ 12	253 $\pm$ 35 <sup>*,†</sup>
Cortisol ( $\mu$ g/dl)		
Euglycemic clamp	9 $\pm$ 1	10 $\pm$ 1
Hypoglycemic clamp	9 $\pm$ 1	25 $\pm$ 2 <sup>*,†</sup>
Growth Hormone (ng/ml)		
Euglycemic clamp	2 $\pm$ 1	1 $\pm$ 1
Hypoglycemic clamp	2 $\pm$ 1	53 $\pm$ 8 <sup>*,†</sup>
Muscle sympathetic nerve activity (bursts/min)		
Euglycemic clamp	28 $\pm$ 3	32 $\pm$ 3 <sup>*</sup>
Hypoglycemic clamp	25 $\pm$ 3	43 $\pm$ 4 <sup>†</sup>

Source. Data from Davis, S. N., Fowler, S., and Costa, F. (2000). Hypoglycemic counterregulatory responses differ between men and women with type 1 diabetes. *Diabetes* **49**, 65–72.

\*Values are significantly increased versus basal period.

†Values are significantly different versus euglycemic clamp.

endogenous glucose production that occur primarily from the liver but also the kidney. Glucagon is a powerful and quick-acting stimulus for hepatic glucose production. The increase in hepatic glucose production is primarily through glycogenolysis, but gluconeogenesis becomes more important as hypoglycemia is prolonged. Glucagon responses are absent within 5 years' duration of T1DM, causing an increased reliance on epinephrine for defense against hypoglycemia. Growth hormone from the pituitary gland, cortisol from the adrenal cortex, and norepinephrine spillover from the sympathetic nervous system (SNS) also increase in response to moderate hypoglycemia but have modest metabolic effects. Pancreatic polypeptide (an index of parasympathetic nervous system activity), oxytocin, and vasopressin are also released during hypoglycemia but have no discernible metabolic effects.

**Table II** Effects of a 120-min Hyperinsulinemic Hypoglycemic (~50 mg/dl) Clamp on Metabolic and Cardiovascular Responses in Overnight Fasted Men (means  $\pm$  SEs)

	Basal	Final 30 min of clamp
Endogenous glucose production (mg/kg/min)		
Euglycemic clamp	2 $\pm$ 1	0 $\pm$ 1 <sup>*</sup>
Hypoglycemic clamp	2 $\pm$ 1	2 $\pm$ 1 <sup>†</sup>
Glucose infusion rate (mg/kg/min)		
Euglycemic clamp	0 $\pm$ 0	8 $\pm$ 1 <sup>*</sup>
Hypoglycemic clamp	0 $\pm$ 0	0.3 $\pm$ 0.1 <sup>†</sup>
Glycerol ( $\mu$ mol/L)		
Euglycemic clamp	33 $\pm$ 3	14 $\pm$ 2 <sup>*</sup>
Hypoglycemic clamp	38 $\pm$ 2	37 $\pm$ 4 <sup>†</sup>
Plasma free fatty acids ( $\mu$ mol/L)		
Euglycemic clamp	440 $\pm$ 70	119 $\pm$ 24 <sup>*</sup>
Hypoglycemic clamp	590 $\pm$ 70	228 $\pm$ 30 <sup>*,†</sup>
Lactate ( $\mu$ mol/L)		
Euglycemic clamp	705 $\pm$ 128	1059 $\pm$ 58 <sup>*</sup>
Hypoglycemic clamp	661 $\pm$ 50	1195 $\pm$ 115 <sup>*,†</sup>
Heart rate (beats/min)		
Euglycemic clamp	60 $\pm$ 3	62 $\pm$ 3
Hypoglycemic clamp	59 $\pm$ 3	70 $\pm$ 4 <sup>*,†</sup>
Mean arterial pressure (mm Hg)		
Euglycemic clamp	83 $\pm$ 3	81 $\pm$ 4
Hypoglycemic clamp	81 $\pm$ 3	81 $\pm$ 3

Source. Data from Davis, S. N., Fowler, S., and Costa, F. (2000). Hypoglycemic counterregulatory responses differ between men and women with type 1 diabetes. *Diabetes* **49**, 65–72.

\*Values are significantly different versus basal period.

†Values are significantly different versus euglycemic clamp.

## Symptoms of Hypoglycemia

The symptoms of hypoglycemia can be separated into those that are neurogenic (autonomic) and those that are neuroglycopenic. Examples of neurogenic symptoms experienced when glucose levels fall below 60 mg/dl are tremulousness, palpitations, anxiety, sweating, and paresthesias. These symptoms are autonomic in origin and stem from increased sympathetic drive. For example, sweating and paresthesias are cholinergically mediated via sympathetic nerve fibers, and tremor is correlated with increased circulating levels of epinephrine. Symptoms that are neuroglycopenic generally occur at glucose levels of 50 mg/dl or less, result directly from brain glucose deprivation, and include difficulty in thinking, confusion, weakness, fatigue, seizures, coma, and death. Impairment of cognitive function occurs at glucose levels of about

45 mg/dl and can be responsible for accidents that are sometimes fatal.

## HYPOGLYCEMIA AND T1DM

Persistent hyperglycemia has been determined to be the cause of long-term microvascular complications associated with T1DM. The Diabetes Control and Complications Trial, a landmark multicenter randomized clinical trial, was developed to determine the risks and benefits of tight glucose control. The results showed that tight glucose control, defined as hemoglobin A<sub>1C</sub> (HbA<sub>1C</sub>)  $\leq 7.2\%$ , obtained by using either multiple insulin injections or continuous subcutaneous insulin infusion prevented or delayed the progression of diabetes complications as compared with conventional treatment (HbA<sub>1C</sub>  $\geq 9.0\%$ ). However, patients with tight glucose control also suffered a threefold increase in the incidence of severe hypoglycemia (requiring outside assistance to recover) and coma. Ninety percent of all patients with T1DM experience symptoms of hypoglycemia. In T1DM, hypoglycemia is initially caused by insulin excess but can be potentiated by lack of food intake, physical activity, and autonomic dysfunction. Hypoglycemia-associated autonomic dysfunction is an acute failure of the autonomic response to hypoglycemia that is induced by prior episodes of hypoglycemia. This creates a situation where T1DM patients have reduced capability to defend against impending hypoglycemia. Unfortunately, hypoglycemia is the complication of diabetes most feared by diabetics. Consequently, the increased prevalence of hypoglycemia is the major roadblock preventing patients from realizing the benefits of tight glucose control.

## HYPOGLYCEMIC UNAWARENESS

One component of hypoglycemia-induced autonomic dysfunction is hypoglycemic unawareness. Hypoglycemic unawareness occurs when an individual has reduced or a complete loss of symptoms that indicate the presence of hypoglycemia. The pathogenesis of hypoglycemic unawareness is multifactorial but includes a shifting of the glycemic threshold. That is, symptoms occur only at progressively lower glucose levels. If this threshold occurs below a critical level, hypoglycemic symptoms might not be activated and a patient might slip into a coma with little or no warning. Importantly, hypoglycemia unawareness can be reversed by meticulous avoidance of iatrogenic hypoglycemia.

## Role of Antecedent Stress in Incidence of Hypoglycemia

Many studies have demonstrated the importance of antecedent hypoglycemia in the pathogenesis of hypoglycemic-associated autonomic dysfunction. Patients with tight glucose control typically have low glycemic thresholds due to the repeated exposure to prior hypoglycemia. Prior episodes of hypoglycemia cause a shift in the glycemic threshold that result in a progressive diminution of autonomic and neuroendocrine responses to subsequent episodes of hypoglycemia. As stated previously, the autonomic and neuroendocrine responses to hypoglycemia are necessary to stimulate the liver to release more glucose and to inhibit peripheral uptake of glucose in an attempt to defend against the falling glycemia. With a reduction in these defenses, combined with the absence of glucagon, T1DM patients are left particularly vulnerable to repeated hypoglycemia. [Tables III and IV](#) contain data from healthy individuals exposed to either day 1 clamped euglycemia or hypoglycemia followed by a subsequent bout of hypoglycemia on day 2. These data clearly demonstrate the blunted day 2 neuroendocrine and ANS responses and reveal the considerably greater amounts of glucose that had to be infused during the day 2 experiments to maintain the desired level of hypoglycemia.

Exercise is an important adjunct treatment of T1DM. Physical activity improves insulin sensitivity, helps in body weight maintenance, and can reduce postprandial hyperglycemia. Unfortunately, exercise is also associated with increased hypoglycemia in T1DM patients. This may be due in part to the fact that prior episodes of hypoglycemia also reduce the autonomic and neuroendocrine response to a subsequent bout of prolonged exercise. The reverse is also true, as prior prolonged exercise will blunt autonomic and neuroendocrine responses to next-day hypoglycemia. Also, one bout of exercise in the morning can even cause a blunted neuroendocrine and metabolic counterregulatory response to a second bout of exercise in the afternoon. Thus, insulin delivery (reduced) and carbohydrate intake (increased) during exercise following a hypoglycemic episode may have to be modified to prevent further hypoglycemia.

The mechanism(s) responsible for the blunting effect of prior episodes of hypoglycemia is unknown at this time. Cortisol is known to blunt SNS responses to stress in both animal and human experimental models. Therefore, prior increases in cortisol during hypoglycemia or exercise could be one factor that

**Table III** Effects of Day 1 Hypoglycemia (~50 mg/dl) on Neuroendocrine and Autonomic Responses to Day 2 Hypoglycemia (~50 mg/dl) in Overnight Fasted Men and Women (means  $\pm$  SEs)

	Final 30 min of day 2	
	Basal	hypoglycemic clamp
Norepinephrine (pg/ml)		
Day 1 euglycemia	178 $\pm$ 12	346 $\pm$ 37 <sup>*</sup>
Day 1 hypoglycemic	172 $\pm$ 18	293 $\pm$ 21 <sup>*,†</sup>
Epinephrine (pg/ml)		
Day 1 euglycemia	42 $\pm$ 3	950 $\pm$ 4 <sup>*</sup>
Day 1 hypoglycemic	33 $\pm$ 3	421 $\pm$ 83 <sup>*,†</sup>
Pancreatic polypeptide (pg/ml)		
Day 1 euglycemia	163 $\pm$ 38	1174 $\pm$ 129 <sup>*</sup>
Day 1 hypoglycemic	125 $\pm$ 17	862 $\pm$ 151 <sup>*,†</sup>
Glucagon (pg/ml)		
Day 1 euglycemia	102 $\pm$ 12	375 $\pm$ 28 <sup>*</sup>
Day 1 hypoglycemic	72 $\pm$ 12	171 $\pm$ 25 <sup>*,†</sup>
Cortisol ( $\mu$ g/dl)		
Day 1 euglycemia	8 $\pm$ 1	22 $\pm$ 1 <sup>*</sup>
Day 1 hypoglycemic	5 $\pm$ 1	17 $\pm$ 2 <sup>*,†</sup>
Growth hormone (ng/ml)		
Day 1 euglycemia	2 $\pm$ 1	46 $\pm$ 6 <sup>*</sup>
Day 1 hypoglycemic	2 $\pm$ 1	32 $\pm$ 5 <sup>*,†</sup>
Muscle sympathetic nerve activity (bursts/min)		
Day 1 euglycemia	30 $\pm$ 7	44 $\pm$ 3 <sup>*</sup>
Day 1 hypoglycemic	28 $\pm$ 9	33 $\pm$ 8 <sup>*,†</sup>

Source. Data from Davis, S. N., Shavers, C., Mosqueda-Garcia, R., and Costa, F. (1997). Effects of differing antecedent hypoglycemia on subsequent counterregulation in normal humans. *Diabetes* **46**, 1328–1335.

\*Values are significantly increased versus basal period ( $P < 0.05$ ).

†Values are significantly reduced versus day 1 euglycemia ( $P < 0.05$ ).

causes blunted SNS responses to subsequent episodes of hypoglycemia or bouts of exercise.

## CONCLUSION

A vicious cycle is generated for T1DM patients when intensive control results in iatrogenic hypoglycemia. An increased frequency of hypoglycemia leads to lower glycemic thresholds, blunted neuroendocrine and ANS responses, hypoglycemia unawareness, and (unfortunately) further episodes of hypoglycemia. Alterations in treatment that focus on preventing hypoglycemia but allow maintenance of tight control of blood glucose levels can add significantly to the quality of life for T1DM patients.

**Table IV** Effects of Day 1 Hypoglycemic (~50 mg/dl) on Metabolic and Cardiovascular Responses to Day 2 Hypoglycemia (~50 mg/dl) in Overnight Fasted Men and Women (means  $\pm$  SEs)

	Final 30 min of day 2	
	Basal	hypoglycemic clamp
Endogenous glucose production (mg/kg/min)		
Day 1 euglycemia	2 $\pm$ 1	2 $\pm$ 1
Day 1 hypoglycemic	2 $\pm$ 1	1 $\pm$ 2 <sup>*,†</sup>
Glucose infusion rate (mg/kg/min)		
Day 1 euglycemia	0 $\pm$ 0	0 $\pm$ 0
Day 1 hypoglycemic	0 $\pm$ 0	4.3 $\pm$ 2.2 <sup>*,†</sup>
Glycerol ( $\mu$ mol/L)		
Day 1 euglycemia	32 $\pm$ 4	29 $\pm$ 4
Day 1 hypoglycemic	35 $\pm$ 2	24 $\pm$ 2 <sup>*,†</sup>
Plasma free fatty acids ( $\mu$ mol/L)		
Day 1 euglycemia	384 $\pm$ 100	178 $\pm$ 32 <sup>*</sup>
Day 1 hypoglycemic	492 $\pm$ 43	199 $\pm$ 20 <sup>*,†</sup>
Lactate ( $\mu$ mol/L)		
Day 1 euglycemia	924 $\pm$ 101	1747 $\pm$ 14 <sup>*</sup>
Day 1 hypoglycemic	945 $\pm$ 96	1535 $\pm$ 73 <sup>*,†</sup>
Heart rate (beats/min)		
Day 1 euglycemia	63 $\pm$ 4	78 $\pm$ 5 <sup>*</sup>
Day 1 hypoglycemic	61 $\pm$ 3	72 $\pm$ 5 <sup>*</sup>
Mean arterial pressure (mm Hg)		
Day 1 euglycemia	83 $\pm$ 3	84 $\pm$ 4
Day 1 hypoglycemic	82 $\pm$ 2	82 $\pm$ 3

Source. Data from Davis, S. N., Shavers, C., Mosqueda-Garcia, R., and Costa, F. (1997). Effects of differing antecedent hypoglycemia on subsequent counterregulation in normal humans. *Diabetes* **46**, 1328–1335.

\*Values are significantly different versus basal period ( $P < 0.05$ ).

†Values are significantly different versus day 1 euglycemia ( $P < 0.05$ ).

## See Also the Following Articles

Beckwith-Wiedemann Syndrome (BWS) • Diabetes, Type 1 • Glucose Physiology, Normal • Hypoglycemic State, Non-Diabetic

## Further Reading

- Cryer, P. E. (1981). Glucose counterregulation in man. *Diabetes* **30**, 261–264.
- Cryer, P. E. (2001). Hypoglycemia-associated autonomic failure in diabetes. *Am. J. Physiol. Endocrinol. Metab.* **281**, E1115–E1121.
- Davis, S. N., and Cherrington, A. D. (1993). The hormonal and metabolic responses to prolonged hypoglycemia. *J. Lab. Clin. Med.* **121**, 21–31.

- Davis, S. N., Shavers, C., Costa, F., and Mosqueda-Garcia, R. (1996). Role of cortisol in the pathogenesis of deficient counterregulation following antecedent hypoglycemia in normal man. *J. Clin. Invest.* **98**, 680–691.
- Diabetes Control and Complications Trial Research Group. (1991). Epidemiology of severe hypoglycemia in the Diabetes Control and Complications Trial. *Am. J. Med.* **90**, 450–459.
- Diabetes Control and Complications Trial Research Group. (1993). The effect of intensive treatment of diabetes on the development and progression of long term complications in insulin-dependent diabetes mellitus. *N. Engl. J. Med.* **329**, 977–986.
- Galassetti, P., Mann, S., Tate, D., Neil, T., Costa, F., Wasserman, D., and Davis, S. N. (2001a). Effect of antecedent prolonged exercise on subsequent counterregulatory responses to hypoglycemia. *Am. J. Physiol.* **280**, E908–E917.
- Galassetti, P., Tate, D., Neill, R. A., Morrey, S., and Davis, S. N. (2002). Effect of gender on counterregulatory responses to euglycemic exercise in type 1 diabetes. *J. Clin. Endocrinol. Metab.* **87**, 5144–5150.
- Galassetti, P., Tate, D., Neill, R. A., Morris, P. G., and Davis, S. N. (2001b). Effect of antecedent hypoglycemia on neuroendocrine responses to subsequent exercise in type 1 diabetes. *Diabetes* **50**, A54.
- Hepburn, D. A., Deary, I. J., Frier, B. M., Patrick, A. W., Quinn, J. D., and Fisher, B. M. (1991). Symptoms of acute insulin-induced hypoglycemia in humans with and without IDDM: Factor-analysis approach. *Diab. Care* **14**, 949–957.
- McGregor, V. P., Banarer, S., and Cryer, P. E. (2001). Limited impact of vigorous exercise on defenses against hypoglycemia: Relevance to hypoglycemia-associated autonomic failure. *Diabetes* **40**, A138.
- McGregor, V. P., Greiwe, J. S., Banarer, S., and Cryer, P. E. (2002). Elevated endogenous cortisol reduces autonomic neuroendocrine and symptom responses to subsequent hypoglycemia. *Am. J. Physiol. Endocrinol. Metab.* **282**, E770–E777.

of cognitive function (2.5–3.0 mmol/L) are sometimes observed in everyday life conditions in some individuals.

## VARIOUS FORMS OF NON-DIABETIC HYPOGLYCEMIA, DIAGNOSIS, AND TREATMENT

As reviewed in details elsewhere by Lefèbvre and Scheen, there are two principal forms of hypoglycemia: (1) exogenous hypoglycemia attributable to the administration (injection or ingestion) of a hypoglycemic compound and (2) endogenous hypoglycemia.

### Exogenous Hypoglycemia

Insulin is by far the most frequent cause of hypoglycemia. In non-diabetic as well as diabetic patients, insulin has been used for homicidal or suicidal purposes. Severe unexplained hypoglycemia in a non-diabetic individual should always raise the possibility of an exogenous insulin administration, either suicidal or criminal. Such cases are encountered more frequently in the medical milieu or in the families or neighborhoods of diabetic patients. Inadvertent insulin administration to a hospitalized non-diabetic patient has also been reported. In psychiatric patients, purposely induced insulin shock therapy sometimes leads to prolonged hypoglycemia and irreversible brain damage. Factitious hypoglycemia due to clandestine self-administration of insulin must always be considered in the differential diagnosis of hypoglycemia; again, this situation is encountered more frequently in the relatives of diabetic patients, in the medical or paramedical profession, or in diabetic patients themselves. Similarly, oral antidiabetic agents, mainly sulfonylureas, can be involved in the pathogenesis of hypoglycemia in non-diabetic individuals: inadvertent administration, accidental ingestion (mainly in children), suicide attempt, and clandestine ingestion (a variety of factitious hypoglycemia). Alcohol ingestion may lead to hypoglycemia if it occurs in fasting conditions given that accidental ingestion of alcohol in children can induce severe hypoglycemia. Sometimes, alcohol favors the “reactive hypoglycemia” following sugar ingestion. Numerous other agents or drugs may induce hypoglycemia, including salicylates, quinine,  $\beta$ -receptor blocking agents, pentamidine, hypoglycins, ouabain, mebendazole, isoproterenol, disopyramide, tranlycypamine, haloperidol, clofibrate, and angiotensin-converting enzyme inhibitors.

### Endogenous Hypoglycemia

Endogenous hypoglycemia may be organic or functional.

#### **Organic Hypoglycemia**

##### *Insulinomas*

Insulinomas are uncommon neoplasms, most often benign, that derive from the B cells of the islets of Langerhans of the pancreas. An insulinoma should be suspected in any patient presenting with the triad described by Whipple: symptoms precipitated by fasting or exercise, proven hypoglycemia associated with symptoms, and relief of symptoms by glucose. The demonstration of endogenous plasma insulin (and C-peptide) levels inappropriate to the prevailing blood glucose level is the cornerstone of the diagnosis. Simultaneous determination of blood glucose and plasma insulin (and C-peptide) levels after an overnight fast and, mainly, during a 24- to 48-h fast is probably the best procedure to demonstrate relative hyperinsulinism. When one is convinced of the diagnosis, one attempts to localize the tumor before sending the patient to surgery. Preoperative localization procedures include tomodesitometry, conventional ultrasonography, selective arteriography, magnetic resonance imaging, effluent vein catheterization, and (the most useful) endoscopic transduodenal ultrasonography. Surgical removal of the tumor is the first and obvious choice of treatment. It must always be accompanied by intraoperative ultrasensitive pancreas echotomography, particularly because there may be more than one tumor.

Medical management, mainly using diazoxide, is reserved for patients who do not accept surgery or in whom major contraindications for the operation exist. Streptozotocin, in association with fluorouracil or doxorubicin, is considered to be the most effective antitumor agent for treating the rare metastatic malignant insulinomas, possibly after surgical reduction of the tumor mass and/or removal of liver metastasis. Other therapeutic options have been reviewed by Lefèbvre and Scheen.

##### *Extrapaneatic Neoplasms*

Extrapaneatic neoplasms secreting a large form of insulin-like growth factor-II, known as “big IGF-II,” are usually large tumors present as masses in the mediastinum or the retroperitoneal space. They often have a mesenchymal origin but can originate from the liver, gastrointestinal tract, or pancreas or can be associated with lymphomas and leukemias. In these patients, hypoglycemia coexists with low or

undetectable insulin and C-peptide levels, whereas high circulating levels of big IGF-II are found. Surgery is the treatment of choice.

#### *Neonatal and Infancy Hypoglycemia*

Numerous inborn errors of metabolism can induce hypoglycemia in neonates and young infants. They include hereditary fructose intolerance, fructose-1,6-diphosphatase deficiency, phosphoenolpyruvate carboxykinase deficiency, some cases of galactosemia, and some of the 11 varieties of glycogen storage disease. Nesidioblastosis, now called "persistent hyperinsulinemic hypoglycemia of infancy" (PHHI), is a rare disease leading to persistent hypoglycemia of infancy. It is basically histologically characterized by the budding off from duct epithelium of endocrine cells and by the presence of microadenomas in the pancreas. The onset of symptoms of beta cell hyperplasia may occur during the first days of life but most commonly within the first 6 months. A few cases beginning with symptoms beyond 1 year of age have been reported. The group of Saudubray in Paris reported the features of 52 neonates with hyperinsulinism. Of these, 30 had diffuse B-cell hyperfunction and 22 had focal adenomatous islet cell hyperplasia. Among the latter, the lesions were in the head of the pancreas in 9, in the isthmus in 3, in the body in 8, and in the tail in 2 neonates. Partial pancreatectomy has been successful in curing 19 of the 22 neonates in whom this procedure has been proposed. Recent studies have shown that congenital hyperinsulinism with focal or diffuse nesidioblastosis can be associated with several mutations affecting the beta cell such as the genes encoding for the sulfonylurea receptor, the glucokinase enzyme, and glutamate dehydrogenase.

Other causes of hypoglycemia during infancy (more functional in nature) include erythroblastosis fetalis, infants of diabetic mothers, leucine-induced hypoglycemia, ketotic or ketogenic hypoglycemia, maple sugar urine disease, and adrenal hyporesponsiveness.

#### **Functional Hypoglycemia**

##### *Alimentary Hypoglycemia*

Alimentary hypoglycemia can occur 1 to 2 h after a carbohydrate-rich meal in individuals who have had a gastrectomy or who, for other reasons, have rapid gastric emptying. It is believed that the rapid dumping of carbohydrates in the upper small intestine elicits an excessive insulin release mediated by both the release of intestinal gut factors (e.g., GLP-1, GIP) and a rapid rise in blood glucose. The treatment is identical to that of spontaneous reactive hypoglycemia.

#### *Spontaneous Reactive Hypoglycemia*

Spontaneous reactive hypoglycemia is a poorly defined entity. The term is usually applied to a syndrome with the following features: (1) symptoms that resemble those seen in insulin-induced hypoglycemia (e.g., diaphoresis, tachycardia, tremulousness, headache) but that often are accompanied by other symptoms less typical of hypoglycemia (e.g., fatigue, drowsiness, feelings of incipient syncope, depersonalization, irritability, lack of motivation); (2) symptoms that may be episodic, sometimes aggravated by carbohydrate-rich meals; and (3) plasma glucose concentrations that drop to 45 mg/dl (2.5 mmol/L) or less at one or more of the half-hourly samples taken in a 5- to 6-h glucose tolerance test. Abnormal insulin secretory patterns have been reported in certain patients.

This entity has had a widespread vogue, particularly in the United States, over the past 30 years but has been said to be diagnosed more rarely elsewhere in the world. The American Diabetes Association and the Endocrine Society issued a joint statement to the effect that this entity is probably overdiagnosed. Indeed, the very existence of this condition has been called into question following several studies demonstrating that 25 to 30% of apparently healthy individuals without any hypoglycemic symptoms may exhibit low plasma glucose values on being given a glucose load. Furthermore, the similarity of the symptoms to those of hyperventilation, and indeed to those of other functional syndromes, emphasizes the need to reevaluate the whole matter of so-called functional or reactive hypoglycemia. The question of cause and effect has not been settled. It would be reasonable at the current time to restrict the diagnosis of reactive hypoglycemia to individuals in whom hypoglycemic blood glucose levels are demonstrated in samples taken after the sort of meals that are said to induce their symptoms. Furthermore, and as has been discussed by Lefèbvre, some patients have adrenergic responses after a meal or during oral glucose tolerance test (OGTT) without hypoglycemia. Such "adrenergic hormone postprandial syndrome" probably results from an altered glycemic threshold (a higher glucose level) for generating an adrenergic response. This results in confusion. A critical analysis of the reactive hypoglycemia syndrome can be found in the proceedings of an international symposium held in Rome in September 1986.

Diet is the first treatment of alimentary and reactive hypoglycemia. Simple sugars should be omitted and replaced by complex carbohydrates. If symptoms persist, small but frequent high-protein, low-carbohydrate meals should be tried. The pharmacological

treatment of choice is  $\alpha$ -glucosidase inhibitors such as acarbose and miglitol.

#### *Alcohol-Promoted Reactive Hypoglycemia*

Alcohol-promoted reactive hypoglycemia can occur when insulinotropic sugars (e.g., glucose, saccharose) are ingested together with alcohol (e.g., beer, gin and tonic, rum and cola, whisky and ginger ale). Such mixtures should be avoided in susceptible individuals.

#### *Other Causes of Functional Hypoglycemia*

Other causes of functional hypoglycemia include discontinuation of total parenteral nutrition, an endocrine deficiency state (glucocorticoid, growth hormone, or glucagon deficiency), severe liver disease, profound malnutrition, prolonged muscular exercise, the autoimmune insulin syndrome (where hypoglycemia is considered to be the consequence of inappropriate release of insulin from insulin-antibody complexes), and the rare syndrome of antibodies directed against the insulin receptor (where hypoglycemia is attributed to an insulinomimetic action of the antibody).

## CONCLUSION

Hypoglycemia in non-diabetic individuals is not a rare condition. It is diagnosed when the blood glucose level is lower than the lowest limit of normal, that is, lower than about 3 mmol/L (or 54 mg/dl), a value also corresponding to the threshold for symptoms in various experimental studies performed in healthy volunteers in whom mild hypoglycemia was induced using graded insulin infusion. Hypoglycemia can result from the administration (injection or ingestion) of a hypoglycemic compound (e.g., insulin, oral antidiabetic agents, alcohol, various drugs). It can also be endogenous in nature. Organic endogenous hypoglycemia can be due to an insulin-producing tumor (insulinoma) or an extrapancreatic neoplasm. In neonates and young infants, hypoglycemia results mainly from various inborn errors of metabolism or of nesidioblastosis, now known as the syndrome of persistent

hyperinsulinemic hypoglycemia of infancy. Functional hypoglycemia is called alimentary if it is due to gastrectomy or a too rapid gastric emptying. It is recognized as spontaneous reactive hypoglycemia if it occurs without any identified cause. Caution must be exerted in the diagnosis of this type of hypoglycemia. However, the diagnosis of reactive hypoglycemia can be made on the basis of a careful clinical and biochemical strategy. In such a case, simple therapeutic measures can be applied and the patient's quality of life can potentially be markedly improved.

## See Also the Following Articles

Glucose Physiology, Normal • Hypoglycemia • Insulin Secretion: Functional and Biochemical Aspects

## Further Reading

- Andreani, D., Marks, V., and Lefèbvre, P. J. (eds.) (1984). "Hypoglycemia." Raven, New York.
- De Lonlay-Debeney, P., Poggi-Travert, F., Fournet, J. C., Sempoux, C., Dionisi, C., Brunelle, F., Touati, G., Rahier, J., Junien, C., Nihoul-Fékété, C., Robert, J. J., and Saudubray, J.-M. (1999). Clinical features of 52 neonates with hyperinsulinism. *N. Engl. J. Med.* **340**, 1169–1175.
- Lefèbvre, P. J. (1991). Hypoglycemia or non-hypoglycemia. *In* "Diabetes" (H. Rifkin, J. A. Colwell, and S. I. Taylor, eds.), pp. 757–761. Excerpta Medica, Amsterdam.
- Lefèbvre, P. J., Andreani, D., and Marks, V. (1988). Statement on "post-prandial" or "reactive" hypoglycemia. *Diabetologia* **31**, 68.
- Lefèbvre, P. J., and Scheen, A. J. (2002). Hypoglycemia. *In* "Diabetes Mellitus" (D. J. Porte, R. S. Sherwin, and A. Baron, eds.), pp. 959–972. McGraw-Hill, New York.
- Lteif, A. N., and Schwenk, W. F. (1999). Hypoglycemia in infants and children. *In* "Endocrinology and Metabolism Clinics of North America: Hypoglycemic Disorders" (F. J. Service, ed.), pp. 619–646. W. B. Saunders, Philadelphia.
- Marks, V. (1987). Glycemic stability in healthy subjects: Fluctuations in blood glucose concentration during the day. *In* "Hypoglycemia" (D. Andreani, V. Marks, and P. J. Lefèbvre, eds.), pp. 19–24. Raven, New York.
- Marks, V., and Rose, F. C. (1981). "Hypoglycemia," 2nd ed. Blackwell Scientific, Oxford, UK.
- Mitrakou, A., Ryan, C., Veneman, T., Mokan, M., Jensen, T., Kiss, I., Durant, J., Cryer, P., and Gerich, J. (1991). Hierarchy of glycemic thresholds for counterregulatory hormone secretion, symptoms, and cerebral dysfunction. *Am. J. Physiol.* **260**, E67–E74.

**Table 1** Causes of Hypoparathyroidism

<b>Development abnormalities of the parathyroid gland</b>
X-linked
Autosomal recessive hypoparathyroidism (GCMB)
DiGeorge syndrome
Kenney–Caffey syndrome
Barakat syndrome
Mitochondrial neuromyopathies
Kearns–Sayre syndrome
MELAS syndrome
Mitochondrial trifunctional protein deficiency
<b>Damage to the parathyroid glands</b>
Surgical
Infiltrative disorders
Granulomatous diseases
Wilson's disease
Metastases
Hemochromatosis
Radiation
Autoimmune polyglandular syndrome type 1
<b>Reduced parathyroid gland function due to altered regulation</b>
Primary
Gain-of-function mutations in the calcium-sensing receptor gene
PTH gene mutations (autosomal recessive/autosomal dominant)
Secondary
Maternal hyperparathyroidism
Hypomagnesemia
<b>Impaired PTH secretion</b>
Pseudohypoparathyroidism

are disrupted. A deletion of 11 nucleotides of the *rxn40* gene occurs at the translocation junction, and loss of function of this gene is responsible for at least part of the DiGeorge phenotype. Another partial transcript, called *nex2.2–nex3*, was also identified from this breakpoint. Both *rxn40* and *nex2.2–nex3* are deleted in all DGS patients with 22q11 deletions, and studies aimed to demonstrate hemizygoty and mutations in these genes in patients without deletions on 22q11 are required to prove their role in DGS. Deletion of the *UDF1L* gene (located on 22q11), encoding a protein involved in the degradation of ubiquitinated proteins, is present in all patients with the 22q11 deletion syndrome, which includes patients with DGS, velo–cardio–facial syndrome (VCFS), and the conotruncal anomaly face syndrome (CAFS). A smaller deletion removing exons 1 to 3 of the *UDF1L* gene was found in one patient with hypocalcemia, cleft palate, small mouth, low-set ears and interrupted aortic arch, syndactyly of the toes, and deficiency of

T lymphocytes. Patients with late-onset DGS have microdeletions in the 22q11 region and develop symptomatic hypocalcemia during childhood or adolescence with only mild phenotype.

Hypoparathyroidism is also a component of the neuromyopathies caused by mitochondrial gene defects: Kearns–Sayre syndrome (KSS), MELAS syndrome, and a mitochondrial, trifunctional, protein deficiency syndrome. KSS is characterized by pigmentary retinopathy and progressive external ophthalmoplegia before 20 years of age and is often associated with heart block or cardiomyopathy. MELAS syndrome consists of childhood-onset mitochondrial encephalopathy, lactic acidosis, and stroke-like episodes. Both KSS and MELAS syndrome have been reported to occur with insulin-dependent diabetes mellitus and hypoparathyroidism. The molecular defects range from large deletions, rearrangement, and duplication of the mitochondrial genomes in many tissues (in KSS) to single-base mutations in one of the tRNA genes found only in a restricted range of cell types (in MELAS syndrome). Mitochondrial trifunctional protein deficiency is a disorder of fatty acid oxidation that is associated with pigmentary retinopathy, peripheral neuropathy, and acute fatty liver degeneration occurring in pregnant women carrying affected children.

Hypoparathyroidism has been reported to occur in 50% of patients with Kenney–Caffey syndrome. It is a congenital anomaly and is associated with parathyroid agenesis, growth retardation, and medullary stenosis of tubular bones. Genetic analysis of the PTH gene did not reveal abnormalities. Both dominant and recessive modes of inheritance have been described. The occurrence of hypoparathyroidism, nerve deafness, and a steroid-resistant nephrosis leading to renal failure, which has been referred to as Barakat syndrome, has been reported in four brothers of one family. An association of hypoparathyroidism with congenital lymphedema, nephropathy, mitral valve prolapse, and brachytelephalangy has been found in two brothers from another family. Molecular genetic studies have not been performed in these two families.

In most cases, the molecular defect responsible for isolated autosomal recessive hypoparathyroidism remains unknown. A homozygous deletion of the human homologue of the *Drosophila* glial cell missing gene *b* (GCMB located on chromosome 6p23) has been identified in a proband with parathyroid agenesis. *Gcm2*, the murine homologue of GCMB, encodes a transcription factor expressed exclusively in the parathyroid cells. Animal models have shown that mice deficient in *Gcm2* had mild hypocalcemia and

serum PTH levels similar to those of the wild type, with the thymus serving as auxiliary source for the peptide hormone.

### Damage to the Parathyroid Glands

The most common cause of hypoparathyroidism is surgical damage to the parathyroids as a result of total thyroidectomy for thyroid cancer, radical neck dissection for laryngeal or esophageal cancer, or repeated parathyroidectomies for multigland disease. The incidence of permanent hypoparathyroidism is approximately 1 to 4%. Permanent hypoparathyroidism can be caused by damage of the blood supply to the parathyroid glands but also by inadvertent removal of parathyroid tissue (a rare event). A permanent hypoparathyroidism is suggested by prolonged hypocalcemia, which may develop immediately or weeks to years after neck surgery. In patients at a higher risk for developing permanent hypoparathyroidism, parathyroid tissue may be autotransplanted into the brachioradialis or sternocleidomastoid muscle at the time of parathyroidectomy or may be cryopreserved for subsequent transplantation if necessary. Transient hypoparathyroidism occurs when the damage to the parathyroid glands is reversible and other mechanisms of hypocalcemia after thyroid surgery also intervene. After surgery for primary hyperparathyroidism, transient hypocalcemia can occur for the suppression of normal parathyroids by hypercalcemia. These glands recover quickly, and the serum calcium level returns to normal within days. More prolonged hypocalcemia can occur in "hungry bone syndrome" due to severe hyperparathyroidism, where there is an increased uptake of calcium into remineralizing bones.

Hypoparathyroidism (rarely) may occur in patients with emochromatosis and iron overload, Wilson's disease, and neoplastic or granulomatous involvement of the parathyroid glands. Hypoparathyroidism has been described in a small number of patients receiving extensive radiation to the neck and mediastinum.

Autoimmune disease of the parathyroid glands can occur as an isolated condition or can be associated with other disorders in autoimmune polyendocrinopathy–candidiasis–ectodermal dystrophy (APECED) syndrome. APECED, an autosomal recessive disorder that is also known as autoimmune polyglandular syndrome type 1 (APS1), is more frequent in Finns and Iranian Jews. It is characterized by a variable combination of destructive autoimmune phenomena leading to a failure of the parathyroid glands, adrenal cortex, gonads, pancreatic beta cells, gastric parietal cells and

thyroid gland, chronic mucocutaneous candidiasis, alopecia, vitiligo, keratopathy, and dystrophy of dental enamel and nails. In most cases, candidiasis is the first clinical manifestation to appear, usually before 5 years of age, followed by hypoparathyroidism before 10 years of age and later by primary adrenal insufficiency before 15 years of age. The etiology of APECED has been clarified and is due to loss-of-function mutations of the AIRE (autoimmune regulator) gene. It is located on chromosome 21q22 and encodes a 57-kDa protein with characteristics of a transcription factor.

### Reduced Parathyroid Gland Function Due to Altered Regulation

Altered regulation of parathyroid gland function may be primary or secondary. Primary alterations now have an established genetic basis. Three rare genetic defects have been defined. The largest group involves the calcium-sensing receptor (CaR). The CaR, which belongs to a family of G protein-coupled receptors and cell surface receptors, regulates the secretion of PTH from the parathyroid glands and the reabsorption of calcium by the renal tubules in response to alterations in serum calcium concentration. Gain-of-function mutations in the CaR gene, leading to a constitutive activated protein (i.e., a receptor with a decreased set point for extracellular calcium concentrations), cause a functional hypoparathyroid state with hypocalcemia and hypercalciuria. The disorder is inherited as an autosomal dominant trait. Sporadic cases due to *de novo* activating mutations of the CaR have also been described. The consequence of the activated parathyroid glands is chronic suppression of PTH secretion, whereas the activated CaR in the kidney induces hypercalciuria. Although some patients have symptoms of hypocalcemia, these tend to be more mild and intermittent than might be expected. Many affected individuals are asymptomatic and are detected only on family screening after an affected individual has been identified. Clinically, it is important to distinguish these patients from those with other forms of hypoparathyroidism because treatment with vitamin D to correct hypocalcemia could lead to marked hypercalciuria, nephrocalcinosis, and renal impairment.

Isolated hypoparathyroidism has also been associated with abnormalities of the PTH gene. A heterozygous base substitution (T to C) in exon 2, leading to the substitution of arginine for cysteine in the signal peptide, has been described in a family with dominant isolated hypoparathyroidism. Functional studies have

shown that this mutation impairs the interaction between the nascent protein and the translocation machinery, and the cleavage of the mutant signal sequence by solubilized signal peptidase is ineffective. A homozygous mutation of the donor splice site of exon 2 and intron 2 of the PTH gene has been identified in one family with autosomal recessive isolated hypoparathyroidism. This mutation, involving a single base transition at position 1 of intron 2, results in loss of exon 2, which encodes the initiation codon and the signal peptide, and causes PTH deficiency. In another patient with autosomal recessive isolated hypoparathyroidism born to consanguineous marriage, a T-to-C mutation was found in the first nucleotide of codon 3 of the 25-amino acid signal peptide sequence, resulting in the substitution of serine with proline.

Secondary causes of impaired PTH secretion include maternal hyperparathyroidism. In this condition, the high concentrations of calcium in the maternal serum cross the placenta and inhibit PTH secretion by the infant's parathyroid glands. Hypocalcemia in the newborn usually develops by the third week of life and is often self-limited.

Hypomagnesemia due to defective renal tubular reabsorption of magnesium or intestinal absorption may impair secretion of PTH and contribute to hypoparathyroidism. This defect is corrected by magnesium treatment.

### Impaired PTH Action

A bioinactive PTH able to cause hypoparathyroidism has not yet been described. More commonly, ineffective PTH action appears to be due to peripheral hormone resistance (pseudohypoparathyroidism).

## CLINICAL MANIFESTATIONS

The predominant clinical manifestations of the disease are those related to hypocalcemia (Table II). In the acute setting, neuromuscular irritability, including perioral paresthesias, tingling of the fingers and toes, and spontaneous or latent tetany with grand mal seizures and laryngeal spasm, can be evident. Chronic hypocalcemia can be asymptomatic and usually recognized by routine blood tests. Patients with chronic hypocalcemia may have coarse hair, dry skin, muscle cramps, extrapyramidal signs, cataracts, alopecia, mental retardation, and personality disorders. Abnormal dentition, enamel hypoplasia, and absence of adult teeth may suggest that hypocalcemia has

**Table II Signs and Symptoms of Hypocalcemia**

<b>Physical signs</b>
Trousseau's sign
Chvostek's sign
<b>Manifestations of increased neuromuscular excitability</b>
Muscle cramps
Paresthesias of distal extremities
Tetany
Mental disorders
Grand mal seizures
Laryngeal spasm
<b>Calcium precipitation</b>
Basal ganglia calcifications
Cataract

been present since childhood. On ECG, a prolonged QT interval is often observed, and in patients with severe hypocalcemia, reversible congestive heart failure may occur. Nonspecific electroencephalographic changes may be observed.

The classic physical findings of hypocalcemia are also manifestations of increased neuromuscular excitability. Trousseau's sign is elicited after inflation of a blood pressure cuff above systolic pressure for 3 to 5 min. A positive response consists of the development of typical carpal spasm, with relaxation occurring 5 to 10 s after the cuff is deflated. Chvostek's sign is elicited by tapping over the facial nerve just anterior to the ear. The response ranges from twitching of the lip at the corner of the mouth to twitching of all the facial muscles on the stimulated side. Simple twitching at the corner of the mouth occurs in some normal persons, but more extensive muscle contraction is a reliable sign of latent tetany.

Calcification of basal ganglia or more widespread calcium deposits in intracranial structures may be detected on routine radiographs, computed tomography (CT) scans, or magnetic resonance.

## LABORATORY INVESTIGATIONS

The biochemical hallmarks of hypoparathyroidism are low serum calcium and high serum phosphorus in the presence of normal renal function. Serum calcium levels are often approximately 6 to 7 mg/dl, whereas serum phosphorus levels are 6 to 9 mg/dl. Ionized calcium concentration is less than 1 mmol/L. Serum concentration of PTH is low or undetectable except in cases of PTH resistance, in which it is high-normal or elevated. Serum levels of 1,25-dihydroxy-vitamin D are usually low normal or low. The 24-h urinary

excretion of calcium is decreased except in cases of gain-of-function mutations in the CaR, where it is elevated. Nephrogenous cyclic AMP (cAMP) excretion is low, and renal tubular reabsorption of phosphorus is increased. Phosphorus excretion and urinary cAMP increase after administration of exogenous PTH with the exception of PTH resistance.

## TREATMENT

The aim of treatment is to normalize extracellular ionized calcium concentration so as to abolish symptoms due to hypocalcemia and to prevent its long-term complications.

### Chronic Treatment

Despite wide differences around the world in treatment modalities due to various availability of vitamin D and its analogues, the general idea is to supplement calcium and vitamin D metabolites so as to achieve stable normocalcemia.

Because of great variability in intestinal calcium absorption in adults, it is usually necessary to supply 1 to 2 g of elemental calcium daily. Calcium must be administered in divided doses with meals to minimize epigastric bloating and irritation. Calcium can be administered in several forms: 1 g of elemental calcium is present in 2.5 g calcium carbonate, 5 g calcium citrate, 8 g calcium lactate, or 10 g calcium gluconate. Calcium carbonate and calcium citrate are equally effective in obtaining appropriate serum calcium levels.

Ergocalciferol (vitamin D<sub>2</sub>) and cholecalciferol (vitamin D<sub>3</sub>) have been widely used at doses ranging from 1.25 to 10 mg daily. Their duration of action is very long due to their wide distribution in body fat; thus, responses to changes in these doses are very slow to manifest, and when intoxication is present, it persists for a long time. Therefore, during long-term treatment, it is necessary to measure periodically (every 3–4 months) serum calcium and phosphorus concentrations. In selected cases (e.g., previous history of kidney stones), urinary calcium and creatinine clearance should also be evaluated. If mild hypercalcemia develops, a reduction in the dose of vitamin D (and/or of calcium) is sufficient to restore serum calcium levels to normal, usually within 1 month. In the case of more severe hypercalcemia, treatment should be stopped and serum calcium should be monitored weekly; hydration and glucocorticoids can occasionally be started. Nowadays, calcitriol

[1,25(OH)<sub>2</sub>D<sub>3</sub>] is considered the treatment of choice. It is initially administered orally at a dose of 0.5 to 1 µg daily in divided doses. The dose should be carefully titrated to match the individual's need. In fact, during the first month of treatment, the required dose, when changed according to serum calcium levels as monitored every 3 to 4 days, may increase to as much as 1.5 to 2 µg per day. During chronic treatment, serum calcium and phosphorus should be checked every 3 to 4 months, keeping in mind that, because of the lack of PTH, renal tubular calcium reabsorption is decreased. Thus, hypercalciuria may occur with increased risk of nephrolithiasis. If hypercalcemia develops, treatment must be temporarily discontinued, with usual prompt restoration of normal serum levels.

Careful monitoring of serum calcium concentration is particularly needed when other drugs are administered to the patient. In fact, many different preparations may interfere with calcium and vitamin D absorption and metabolism, and these effects are amplified by the absence of counterregulatory action exerted by PTH. Thiazide diuretics enhance renal tubular calcium reabsorption and may induce hypercalcemia, whereas loop diuretics cause hypocalcemia by increasing calcium excretion. Moreover, glucocorticoids reduce intestinal calcium absorption by antagonizing vitamin D effects, cholestyramine inhibits vitamin D absorption, and agents such as aluminum and magnesium hydroxide reduce calcium absorption by promoting its precipitation in the gastrointestinal tract.

In pseudohypoparathyroidism, the risk of hypercalciuria is less present because the action of PTH on distal renal tubular calcium reabsorption is preserved. However, at the beginning of treatment, high doses are usually needed for the high prevalence of osteomalacia. During follow-up, while bone lesions are being repaired (as shown by a decline in serum alkaline phosphatase and a rise in serum calcium concentration), dose requirements progressively decrease.

Occasionally, some patients may experience unsatisfactory control, with ample fluctuations of serum calcium, despite the absence of dose variations. Poor compliance should be suspected, but variations in calcium intake and intestinal absorption problems should also be considered. Particular care must be taken in the management of hypoparathyroidism during the third trimester of pregnancy and puerperium. The dose of vitamin D will usually have to be reduced because of the actions of placenta, which can synthesize 1,25-dihydroxy-vitamin D<sub>3</sub>, and of prolactin, which can stimulate 1α-hydroxylation.

In young patients, when hypoparathyroidism is diagnosed, the therapy regimen is similar to that described for adults. However, doses must be tailored to consider body weight. Calcium is usually initiated at the level of 30 to 50 mg/kg body weight and is gradually increased according to serum calcium levels. Also in children, the most used vitamin D preparation is calcitriol due to its effectiveness at nontoxic doses, shorter half-life, and more rapid correction of hypercalcemia when it occurs.

It must be remembered that vitamin D, in all its forms, can be a difficult drug to use. Unpredictable changes in therapeutic regimens may occur even after a long period of stable treatment; therefore, continued monitoring is mandatory.

### Treatment after Surgery for Hyperparathyroidism

After surgery for primary hyperparathyroidism due to parathyroid adenoma, serum calcium usually falls to normal levels quickly, sometimes evolving to transient hypocalcemia. In general, no treatment is needed except for oral calcium supplementation if symptoms are present. Too aggressive supplemental calcium during the early postoperative phase may delay restoration of normal parathyroid function; this seems to be triggered by mild hypocalcemia. However, if preoperatively hyperparathyroid bone disease was present, so-called hungry bone disease usually develops. Serious hypocalcemia can be observed, requiring high doses of calcitriol (1–3 µg daily) and large calcium supplements for a long period of time.

Symptomatic patients should also be treated by intravenous calcium (2 mg/kg of elemental calcium over 10 min) and, in the case of recurrence of symptoms, by a more prolonged infusion (15 mg/kg of elemental calcium over 24 h, with half of the total amount administered during the initial 6 h). In general, 15 mg/kg of elemental calcium infused during 4 to 6 h is capable of increasing serum calcium concentration by 2 to 3 mg/dl. To reduce the amount of fluids, the concentration of the solution can be increased to as much as 200 mg of elemental calcium per 100 ml without the risk of irritating the vein. Particular care must be taken if the patient is taking medications such as digitalis (calcium may increase sensitivity to adverse effects of this drug such as arrhythmias); additives, such as phosphate and bicarbonate, must be avoided. A dramatic fall in vitamin D and calcium requirements will be observed after bone disease healing. In the case of multiple parathyroid excision,

long-lasting hypoparathyroidism will ultimately result.

Bone remodeling after parathyroidectomy may cause hypomagnesemia, which contributes to the risk of tetany by inhibiting PTH secretion by the remaining glands and by exacerbating hypocalcemic symptoms. In the instance of magnesium deficiency, the administration of 2.4 mg/kg body weight of the element over a period of 10 min may be required. Once magnesium has been repleted, stores are usually maintained through a regular diet. An oral supplementation of 300 to 600 mg daily may be necessary in the presence of diarrhea, fistulous drainage, or other abnormalities.

### Future Trends

Studies are in progress to verify whether parenteral administration of the active form of PTH or oral administration of modified molecules could be used in the treatment of hypoparathyroidism. In patients with parathyroid hyperplasia or in those who underwent repeated neck explorations to identify the adenomas, parathyroid tissue can be autotransplanted into the brachioradialis or sternocleidomastoid muscle at the time of parathyroidectomy, or it can be cryopreserved for later transplantation if required.

### See Also the Following Articles

Autoimmune Polyglandular Syndrome • Hyperparathyroidism, Primary • Hyperphosphatemia • Hypocalcemia, Therapy • Parathyroid Glands, Pathology • Parathyroid Hormone (PTH) • Parathyroid Surgery • Pseudohypoparathyroid States

### Further Reading

- Betterle, C., Dal Pra, C., Mantero, F., and Zanchetta, R. (2002). Autoimmune adrenal insufficiency and autoimmune polyendocrine syndromes: Autoantibodies, autoantigens, and their applicability in diagnosis and disease prediction. *Endocrine Rev.* **23**, 327–364.
- Garfield, N., and Karaplis, A. (2001). Genetics and animal models of hypoparathyroidism. *Trends Endocrinol. Metab.* **12**, 288–294.
- Golzman, D. G., and Cole, D. E. C. (1999). Hypoparathyroidism. In "Primer on Metabolic Bone Diseases and Disorders of Mineral Metabolism" (M. J. Favus, ed.), pp. 226–230. Lippincott Williams & Wilkins, Philadelphia.
- Gunther, T., Chen, Z., Kim, J., Priemel, M., Rueger, J. M., Amling, M., Moseley, J. M., Martin, T. J., Anderson, D. J., and Karsenty, G. (2000). Genetic ablation of parathyroid glands reveals another source of parathyroid hormone. *Nature* **406**, 199–203.
- Marx, S. J. (2001). Hyperparathyroid and hypoparathyroid disorders. *N. Engl. J. Med.* **343**, 1863–1875.

- O'Riordan, J. L. H. (1994). Treatment of hypoparathyroidism. In "The Parathyroids" (J. P. Bilezikian, M. A. Levine, and R. Marcus, eds.), pp. 801–804. Raven, New York.
- Okano, K., Furukawa, Y., Morii, H., and Fujita, T. (1982). Comparative efficacy of various vitamin D metabolites in the treatment of various types of hypoparathyroidism. *J. Clin. Endocrinol. Metab.* **55**, 238–243.
- Parkinson, D. B., and Thakker, R. V. (1992). A donor splice mutation in the parathyroid hormone gene is associated with autosomal recessive hypoparathyroidism. *Nat. Genet.* **1**, 149–152.
- Schinke, M., and Izumo, S. (2001). Deconstructing DiGeorge syndrome. *Nat. Genet.* **27**, 238–240.
- Sunthornthepvarakul, T., Churesigaew, S., and Ngowngarmratana, S. (1999). A novel mutation of the signal peptide of the preproparathyroid hormone gene is associated with autosomal recessive hypoparathyroidism. *J. Clin. Endocrinol. Metab.* **84**, 3792–3796.
- Thakker, R. V. (2001). The molecular genetics of hypoparathyroidism. In "The Parathyroids: Basic and Clinical Concepts" (J. P. Bilezikian, R. Marcus, and M. A. Levine, eds.), pp. 779–790. Academic Press, London.

Table I Clinically Relevant Hormones Secreted by the Pituitary Gland

Anterior pituitary gland	Posterior pituitary gland
Luteinizing hormone and follicle-stimulating hormone (gonadotropins)	Vasopressin/Antidiuretic hormone
Growth hormone	
Thyrotropin	Oxytocin
Adrenocorticotrophic hormone	
Prolactin	

100,000 population, but this excludes those clinically insignificant lesions that have been reported at autopsy in up to 27% of individuals without an antemortem diagnosis of pituitary pathology. Primary pituitary carcinomas are extremely rare, but metastatic deposits at this site do occur, usually arising from malignancies of the lung or breast. Pituitary adenomas are divided into microadenomas (<1 cm) and macroadenomas (>1 cm) on the basis of size. In addition, there are a variety of classifications that determine the degree of invasiveness. Figure 1 shows the magnetic resonance imaging (MRI) appearance of a pituitary macroadenoma with significant suprasellar extension.

Pituitary adenomas may also be defined as functioning (70%) or nonfunctioning (30%). The former group produces systemic effects by secreting hormones into the bloodstream. The most common functioning pituitary tumors are prolactinomas, followed by those causing acromegaly and Cushing's disease, with gonadotropinomas and thyrotropinomas occurring only rarely. Both functioning and nonfunctioning tumors may result in hypopituitarism when there is insufficient normal pituitary tissue remaining. The typical sequence of hyposecretion starts with GH deficiency, followed by gonadotropin loss before progressing to loss of thyrotropin-stimulating hormone (TSH) and then adrenocorticotrophic hormone (ACTH) reserves. Mild hyperprolactinemia is often

present when the tumor itself does not secrete prolactin due to interruption of the influence of dopamine from the hypothalamus that normally inhibits prolactin release. Vasopressin deficiency in association with pituitary adenomas is extremely rare.

### Pituitary Surgery/Radiotherapy

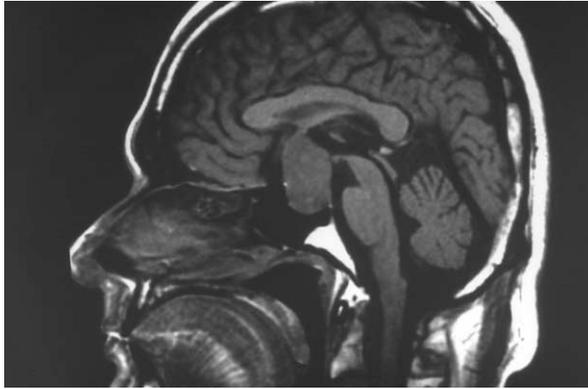
Hypopituitarism may be iatrogenic, and the incidence is approximately 10 to 15% posttranssphenoidal surgery. The effects of conventional radiotherapy may take many years to manifest fully; consequently, it is vital that those patients who are at risk for developing new endocrine deficiencies as a result of irradiation be reassessed every 1 to 2 years. Approximately 50% of patients will have evidence of hypopituitarism 5 years postradiotherapy.

### Parapituitary Tumors

Parapituitary tumors may be intradural (e.g., craniopharyngiomas, gliomas, hamartomas, meningiomas) or extradural (e.g., chordomas, metastases, chondrosarcomas, Brown tumors), but all are less common than primary pituitary adenomas. Craniopharyngiomas account for up to 4% of all intracranial tumors, and although they are often thought to be childhood tumors, 50% of all craniopharyngiomas present in

Table II Causes of Hypopituitarism

<ul style="list-style-type: none"> <li>• Pituitary tumors</li> <li>• Pituitary surgery/radiotherapy</li> <li>• Parapituitary tumors (e.g., craniopharyngiomas, meningiomas, gliomas, leukemic infiltrates)</li> <li>• Pituitary apoplexy</li> <li>• Inflammatory/infiltrative conditions (e.g., lymphocytic hypophysitis, sarcoidosis, hemochromatosis, histiocytosis X, amyloidosis)</li> <li>• Infection (e.g., tuberculosis, pituitary abscess)</li> <li>• Head injuries</li> <li>• Congenital malformations/empty sella syndrome (e.g., agenesis, hypoplasia, primary empty sella)</li> <li>• Genetic and idiopathic disorders (e.g., isolated ACTH deficiency, Kallman's syndrome, PIT1 mutations)</li> </ul>
--



**Figure 1** Visual perimetry showing typical bitemporal hemianopia associated with optic chiasm compression.

patients over 16 years of age. They appear to follow a more aggressive course, particularly if onset is during childhood, with increased morbidity and mortality compared with patients with other pituitary or parapituitary lesions.

### Pituitary Apoplexy

Pituitary apoplexy describes destruction of functioning pituitary tissue as a result of infarction and may arise due to ischemia or hemorrhage. The former usually arises in patients with preexisting macroadenomas, whereas the latter may occur as a complication of anticoagulant therapy or pregnancy (Sheehan's syndrome). Sheehan's syndrome is rare in the developed world. Early surgery has been advocated for this condition, but each individual case must be assessed separately taking into account the neurological deficit, visual loss, and medical comorbidity.

### Inflammatory/Infiltrative Conditions

Lymphocytic hypophysitis is an autoimmune condition characterized by diffuse infiltration of the gland by inflammatory cells that predominantly, but not exclusively, occurs in women toward the end of pregnancy. The condition may present very acutely over a few days or more insidiously, and histological confirmation is required to be certain of the diagnosis. It may result in isolated anterior pituitary hormone deficiencies.

Infiltrative (e.g., hemochromatosis) and granulomatous conditions (e.g., sarcoidosis, histiocytosis X) can cause hypopituitarism; therefore, a high index of suspicion is needed for these conditions.

### Infection

Communicable diseases such as tuberculosis must remain part of the differential diagnosis of hypopituitarism, and focal abscesses of the pituitary gland may occur rarely.

### Head Injuries

Cranial trauma can cause disruption of the pituitary stalk, which results most commonly in diabetes insipidus; however, hyposcretion of any pituitary hormone is possible. Monitoring and frequent re-assessment of these patients are essential because the deficiencies may resolve over time.

### Congenital Malformations/Empty Sella Syndrome

The so-called empty sella syndrome may be congenital secondary to arachnoid herniation through a diaphragmatic defect or may be acquired postsurgery, postradiotherapy, or post-pituitary infarction. Imaging in this condition reveals an enlarged empty pituitary fossa, but this finding does not by definition exclude the presence of a functioning tumor.

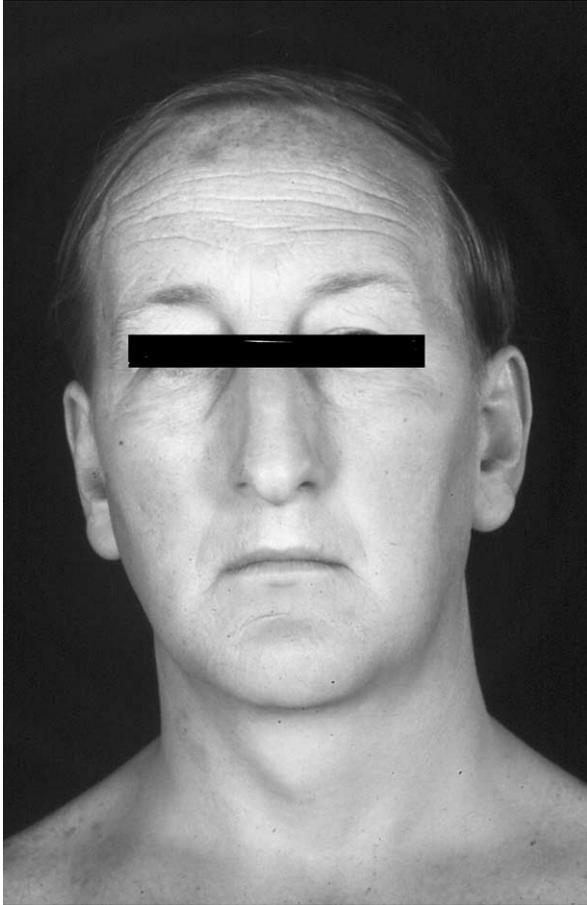
### Genetic and Idiopathic Disorders

Isolated defects of single pituitary hormones have been described but are rare and usually do not progress to involve hyposcretion of other components of the gland. Examples include isolated ACTH deficiency and gonadotropin deficiency (Kallman's syndrome). Mutations in the PIT1 gene are an important genetic cause of hypopituitarism.

## CLINICAL FEATURES

The clinical presentation of hypopituitarism varies according to exactly which endocrine deficiencies are present and also depends on the age at which pituitary insufficiency developed, that is, pre- or postpuberty. The clinical picture may also be influenced by the underlying pathology and whether the presentation is acute or chronic.

Patients with panhypopituitarism are classically pale due to lack of ACTH, have thin hair, display premature ageing of the skin with increased wrinkles, and lack secondary sexual characteristics. They frequently experience reduced muscle strength and increased fatigue, and they may exhibit psychological problems such as depression. This typical appearance is shown in [Fig. 2](#). The symptoms and signs may be

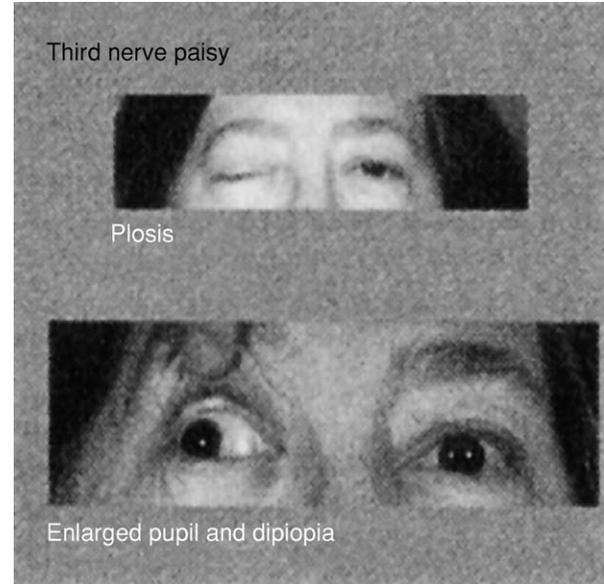


**Figure 2** Sagittal MRI scan showing pituitary macroadenoma with extensive suprasellar extension.

divided into those due to direct mass effects of the tumor (if present) and those due to the endocrine deficiencies.

### Tumor Mass Effects

Expansion of the tumor may be associated with visual field loss. Classically, a bitemporal upper quadrantinopia occurs first, progressing to a bitemporal hemianopia as a result of the tumor compressing the optic chiasm; however, a variety of defects can occur depending on which parts of the visual pathways are affected. Ongoing visual recovery can occur up to 6 months postdecompression, but established optic atrophy secondary to long-standing pressure on the optic nerves is a bad prognostic sign. Headaches may occur and are thought to be the result of the enlarging tumor stretching the dura; more rarely, obstructive hydrocephalus may develop. Cerebrospinal fluid rhinorrhea or cranial nerve palsies affecting the III, IV, or VI nerve may be seen if the tumor invades the sphenoid



**Figure 3** Patient with right third cranial nerve palsy secondary to infiltration of cavernous sinus by pituitary macroadenoma.

or cavernous sinuses, respectively. [Figure 3](#) shows a patient who developed a third cranial nerve palsy from a pituitary macroadenoma extending into the cavernous sinus. More extensive tumor infiltration into surrounding tissues can result in seizures and personality changes very rarely.

### Symptoms and Signs Related to Endocrine Deficiencies

The typical presentation associated with deficiencies of each of the pituitary hormones is outlined in [Table III](#).

### DIAGNOSIS

The diagnosis of hypopituitarism is dependent on the appropriate biochemical investigations being carried out in a patient where the clinical findings have raised the suspicion of pituitary insufficiency. The presentation in adult life is often subtle and insidious, with symptoms that overlap with both other medical conditions and the normal population; consequently, the diagnosis is frequently delayed unless the investigating physician retains a high index of suspicion for a possible underlying endocrine etiology.

In any patient with suspected pituitary or hypothalamic pathology, imaging of the area, ideally with MRI scanning and formal visual field assessment, is mandatory. These investigations help to assess for the presence of damage to the optic pathways, may assist in elucidating the underlying cause, and will

Table III Signs and Symptoms of Hypopituitarism

Hormone deficiency	Clinical features
Growth hormone	Increased fatigue, reduced muscle strength and exercise capacity, impaired psychological well-being, reduced lean body mass, and increased cardiovascular risk
Luteinizing hormone and follicle-stimulating hormone	Women: oligomenorrhea/amenorrhea, infertility, dyspareunia, breast atrophy, and loss of libido Men: loss of libido, erectile dysfunction, testicular atrophy, infertility, and loss of secondary sexual characteristics
Thyrotrophin	Increased fatigue, weight gain, cold intolerance, dry skin and hair, constipation and muscle weakness
Adrenocorticotrophic hormone	Increased fatigue, anorexia, nausea and vomiting, weight loss, generalised weakness, hypoglycemia, and hyponatremia
Prolactin	Failure of lactation
Vasopressin	Polydipsia, polyuria and nocturia

also help in planning ongoing management of the condition. Additional tests that may help to elucidate the likely etiology should be carried out as appropriate (e.g., serum ferritin, angiotensin-converting enzyme from serum or cerebrospinal fluid, pituitary biopsy).

Specific endocrine tests can be divided into those that assess either anterior or posterior pituitary function, and tests of anterior pituitary function can be further subdivided according to whether they are static measurements at a single point in time or dynamic tests that assess pituitary reserve more formally.

### Basal Anterior Pituitary Hormone Assays

To gain the most information about each axis from a single blood test, the basal concentrations of both the pituitary hormone and the target hormone should be assayed:

- Luteinizing hormone (LH), follicle-stimulating hormone (FSH), and estradiol or testosterone (9:00 AM)
- TSH and free thyroxine
- ACTH and cortisol (9:00 AM)
- Prolactin
- Insulin-like growth factor-1 (IGF-1)

Testosterone and cortisol should be measured at 9:00 AM as a baseline because this is the time of peak secretion within the circadian rhythm. Most reference ranges have been devised for 9:00 AM samples, so it is easier to interpret and compare results taken at this time.

### Dynamic Tests of Anterior Pituitary Function

The preceding basal pituitary assays can provide vital information, particularly when the pituitary component is appropriately elevated with a low target hormone indicating primary target organ failure. The situation is frequently more complex when either a low target hormone is found in conjunction with an inappropriately low pituitary hormone or ongoing monitoring is required for patients with known partial hypopituitarism. In these cases, dynamic testing may help to quantify pituitary reserve more accurately.

#### Assessing the Hypothalamic–Pituitary–Adrenal Axis

A 9:00 AM cortisol level of less than 3.6 µg/dl (100 nmol/L) is generally sufficient to diagnose adrenal insufficiency with confidence, whereas a level of 18.1 µg/dl (500 nmol/L) or more is indicative of adequate ACTH reserve. Patients with 9:00 AM values between these limits generally warrant further investigation.

It is important to be aware that oral estrogen therapy artificially raises serum cortisol levels by increasing cortisol-binding globulin, so it is recommended that any such medication be stopped for at least 6 weeks (or a transdermal preparation be substituted) before carrying out any biochemical assessment that relies on serum cortisol measurements.

The main tests employed in assessing ACTH reserve are the insulin tolerance test (ITT) and the short synacthen test (SST), both originally described during the 1960s.

### *Insulin Tolerance Test*

Cortisol and GH are counterregulatory hormones; consequently, artificially inducing hypoglycemia is a useful stimulus by which to provoke their secretion. In the ITT, 0.15 IU/kg of soluble insulin is usually administered intravenously, and if the required peak cortisol level (usually 18.1–21.0  $\mu\text{g}/\text{dl}$  [500–580 nmol/L]) is achieved during the test, then that patient is determined to have sufficient ACTH reserve to mount an adequate response to acute stress. The test is contraindicated in those with epilepsy, ischemic heart disease, or postcraniotomy, and the majority of endocrinologists would not recommend its use beyond 70 years of age. Despite the potential hazards, the complication rate of the test remains very low when conducted by experienced personnel.

### *Short Synacthen Test*

The SST is useful in diagnosing adrenal insufficiency but does not help any further with the differential diagnosis unless a plasma ACTH is measured simultaneously. Patients with established primary or secondary hypoadrenalism should not reach the required peak cortisol level (usually 21.0  $\mu\text{g}/\text{dl}$  [580 nmol/L]) 30 min following parenteral administration of 250  $\mu\text{g}$  of synthetic ACTH. In patients who develop ACTH deficiency acutely (e.g., postsurgery), this test is not appropriate because it can take several weeks for the adrenals to involute following the loss of normal circulating ACTH levels. Given the excellent safety profile of the SST, most endocrinologists reserve the ITT for patients with borderline SST results or for cases where simultaneous assessment of GH is required.

The low-dose SST, using 1  $\mu\text{g}$  of synthetic ACTH, has become increasingly popular during recent years as the 250- $\mu\text{g}$  dose is seen as being so suprphysiological that false-negative results may occur. This dose is currently not routinely available, and further assessment of the test is required.

### *Additional Tests*

The glucagon test has the advantage, as does the ITT, of being able to assess ACTH and GH reserves simultaneously, but the glucagon test provides a weaker stimulus for secretion of ACTH and GH than does the ITT and, consequently, is more often associated with false-negative results. Glucagon transiently increases the plasma glucose, and the test relies on the subsequent fall to create relative hypoglycemia as a stimulus for ACTH and GH release.

The caveat to the tests outlined previously is that in critically ill patients with suspected adrenal insufficiency, treatment should not be delayed while

investigations are carried out. In this setting, random ACTH and plasma cortisol measurements can be taken and parenteral high-dose hydrocortisone therapy can be quickly instituted. The results can then be reviewed retrospectively and steroid replacement can be continued if indicated, with any further assessment delayed until the acute illness has resolved.

### **Assessing the Pituitary–Thyroid Axis**

Baseline assessment of TSH and thyroxine (free or total) is usually sufficient for the diagnosis of primary and secondary hypothyroidism. Measuring free thyroxine is essential if the patient is pregnant or taking oral estrogen therapy due to the increase in thyroid hormone-binding globulin that occurs. The TSH level in secondary hypopituitarism is rarely undetectable and in fact may be low, normal, or even mildly elevated. The main important differential diagnosis is from the so-called “sick euthyroid” syndrome, which is associated with critical systemic illness. Measurement of thyroxine-binding globulin may be helpful in complex cases because congenital deficiencies of this protein can occur, giving a similar result to secondary hypothyroidism.

The thyrotropin-releasing hormone (TRH) test (assessment of TSH response to administration of intravenous TRH) has been largely abandoned, due to its lack of diagnostic sensitivity and specificity, in favor of newer and more sensitive assays for TSH and thyroxine.

### **Assessing Growth Hormone Reserve**

GH is secreted in a pulsatile fashion and is released in response to stress or exercise and at night during sleep. IGF-1 mediates many of the actions of GH, is released in response to GH secretion, and itself inhibits GH release at the level of the pituitary and the hypothalamus. Therefore, isolated GH measurements are generally meaningless, and approximately 30% of patients with GH deficiency would be missed if a normal IGF-1 were the sole criterion for diagnosis.

Tests to further assess GH may be subdivided into those using physiological methods and those relying on pharmacological agents to stimulate GH release.

#### *Physiological Methods*

These methods include urinary GH assessment, 24-h profiles of serum GH, and measurement of GH response during sleep or exercise. These techniques are very labor intensive, and because the results show significant overlap with normal controls, they are now seldom used clinically.

### *Pharmacological Methods*

The ITT is still regarded as the gold standard for assessing GH reserve. The main second-line test until recently has been the glucagon test. Newer agents include arginine, growth hormone-releasing hormone (GHRH), and growth hormone-releasing peptides (GHRP) such as hexarelin. Clonidine testing has been largely discredited for diagnosing GH deficiency in adults. Although research is ongoing, there is increasing evidence that combinations of these newer tests may provide accurate results with the advantage of a more acceptable safety profile as compared with the ITT.

### *Assessing the Pituitary–Gonadal Axis*

Usually, measurement of baseline LH, FSH, and testosterone or estradiol is sufficient to confirm hypogonadotropic hypogonadism. Estradiol is best measured during the follicular phase of the menstrual cycle, and testosterone is ideally measured on a 9:00 AM sample. Both sex steroids bind to sex hormone-binding globulin (SHBG); therefore, this may need to be assayed to accurately determine the amount of biologically active hormone present.

Women who are menstruating regularly are not, by definition, gonadotropin deficient, although the cycles may clearly be anovulatory. Ovulation can be assessed separately by measurement of a progesterone level on day 21 of the menstrual cycle. Management of subfertility due to hypopituitarism is discussed later in this article.

Dynamic tests (e.g., clomiphene test) are only indicated to distinguish hypothalamic disease from pituitary disease, but these are generally for academic interest and rarely alter clinical management.

### **Posterior Pituitary Assessment**

It is vital that ACTH reserve has been fully assessed, and that glucocorticoid replacement has been instituted if appropriate, prior to formally assessing posterior pituitary function because cortisol is essential for the kidneys to handle and excrete a water load normally and so diabetes insipidus may be concealed in patients who are steroid deficient.

Paired plasma and urine osmolalities in combination with urea and electrolytes and accurate fluid balance may be sufficient to confirm a diagnosis of cranial diabetes insipidus (vasopressin deficiency); however, in less clear-cut cases, or where the differential diagnosis of nephrogenic diabetes insipidus or

primary polydipsia needs to be excluded, a formal water deprivation test may be required.

A variety of protocols for water deprivation tests have been described, but essentially they assess the response of urine output and osmolality to a period of fluid restriction followed by a dose of exogenous vasopressin. Plasma osmolalities and serial weights are also recorded. The tests must be carefully supervised to prevent patients from becoming dangerously dehydrated and to ensure that the fluid restriction is maintained.

Newer assays that offer more accurate measurements of endogenous vasopressin, normally present at very low concentrations in the circulation, may provide an alternative means of diagnosing diabetes insipidus, but currently these are not widely used.

## **TREATMENT**

Treatment of pituitary insufficiency must involve addressing the underlying cause as well as commencing and optimizing appropriate hormone replacement therapy.

### **Treatment of Underlying Cause**

Determining the etiology of the hypopituitarism will help to guide management of the underlying condition. Options may include the use of medical treatments or surgical intervention and/or radiotherapy.

### **Hormone Replacement Therapy**

Treatment regimens generally involve replacing the target hormone rather than the pituitary or hypothalamic component of the pathway. The main exceptions to this are the use of recombinant GH and the treatment of infertility caused by hypogonadotropic hypogonadism.

### **GH Deficiency**

During the final few decades of the past century, evidence accumulated that, contrary to the accepted doctrine, GH may have an important role to play in adulthood. During the 1990s, with the availability of recombinant GH eliminating the risks associated with GH derived from human cadavers, many adult patients with severe GH deficiency are now on appropriate replacement. GH is administered as a daily subcutaneous injection using a pen device that most patients master quickly. Treatment is commenced at low dose and titrated up slowly according to clinical

and biochemical (IGF-1) parameters. There is convincing evidence that GH replacement in GH-deficient adults is associated with significant improvements in mood, energy levels, and overall quality of life. An increase in lean body mass, with a reduction in fat mass and a resultant increase in exercise tolerance, is seen in patients receiving treatment. GH therapy is also associated with an improvement in the lipid profile and with long-term treatment results in improved bone mineral density in these patients at risk for osteoporosis.

### **LH/FSH Deficiency**

The possible complications of estrogen therapy in postmenopausal women are well publicized and may lead women to refuse treatment. It is vital that each case is considered individually and that the patient is helped to make an informed decision because the severe long-term consequences of estrogen deficiency clearly outweigh the risks of such therapy, particularly in young women. Essentially any combined oral contraceptive pill, or any form of estrogen licensed for postmenopausal use (in combination with a cyclical progestagen to avoid endometrial hyperplasia and subsequent malignancy in women with intact uteri), is suitable, and the decision comes down to patient choice and physician familiarity. Many younger women prefer the combined oral contraceptive pill because it is more acceptable among their peers. Transdermal HRT preparations are also available. These have the advantage, compared with oral estrogen preparations, of not affecting cortisol-binding globulin and, consequently, not affecting the monitoring of hydrocortisone replacement.

Androgen replacement in men is usually initiated with intramuscular depot testosterone injections. Trough and/or peak levels are monitored, and the therapy is optimized by adjusting the dose or dose interval. Subcutaneous testosterone implants provide a more physiological profile and avoid the sharp peaks and troughs associated with intramuscular injections, but this must be balanced against the risk of infection and scarring at the site of implantation. Transdermal testosterone patches are available but can cause local skin irritation and are less popular than their estrogen counterparts, whereas oral testosterone has a very short half-life and is rarely recommended as the therapy of choice.

Restoration of fertility is usually possible in these patients by using gonadotropin or gonadotropin-releasing hormone (GnRH) to stimulate ovulation or spermatogenesis. If gonadotropin deficiency occurs prior to puberty, careful and gradual dose titration is

required to ensure that development occurs at an appropriate rate so that final height is achieved, secondary sexual characteristics develop normally, and there is sufficient time to cope with the usual psychological changes occurring during these years.

### **TSH Deficiency**

Thyroxine replacement is the treatment of choice for secondary hypothyroidism, as it is for primary hypothyroidism, and the average adult replacement dose for a patient with total deficiency is approximately 100 to 150  $\mu\text{g}/\text{day}$ . Treatment can usually be initiated at a dose of 75 to 100  $\mu\text{g}/\text{day}$ , with the dose altered in 25- $\mu\text{g}$  increments based on six weekly reassessments of the free thyroxine. In patients with severe hypothyroidism or coexisting ischemic heart disease, lower starting doses of 25 to 50  $\mu\text{g}$  can be used. Triiodothyronine rarely is superior due to its more rapid onset of action and shorter half-life, but thyroxine is appropriate routinely. In primary hypothyroidism, TSH is a sensitive marker that can guide accurate replacement, but this is not the case in secondary hypothyroidism; the serum TSH may be low, normal, or even mildly elevated in this condition. Most endocrinologists aim for a replacement dose of thyroxine that results in asymptomatic patients and free thyroxine on the upper end of the normal reference range.

It is essential that patients receive adequate steroid replacement, if required, prior to commencing with thyroxine; otherwise, a potentially dangerous hypoadrenal crisis may be precipitated.

### **ACTH Deficiency**

Hydrocortisone, the generic pharmaceutical name for cortisol, is generally accepted to be the glucocorticoid of choice for replacement therapy because it directly replaces the deficient hormone and can be easily measured in the plasma or urine. Alternatives include cortisone acetate or the synthetic glucocorticoids prednisolone and dexamethasone. Cortisone acetate must first be metabolized to cortisol to achieve its glucocorticoid effect; therefore, its onset of action is delayed compared with that of hydrocortisone, which has largely replaced it as the treatment of choice. Prednisolone and dexamethasone have a longer duration of action than does hydrocortisone and can be administered twice and once daily, respectively, but the inability to monitor steroid replacement with these drugs outweighs any advantages and makes them second-choice agents.

Traditional steroid replacement regimens have been reevaluated during recent years in light of increasing

evidence that they provide significantly supraphysiological doses and also poorly mimic endogenous cortisol secretion. The currently recommended starting replacement regimen is 10 mg of hydrocortisone on waking, 5 mg at lunchtime, and 5 mg during the early evening. Clearly, very few patients are average, and hydrocortisone replacement might need significant adjustment in individual patients depending on symptoms and the results of biochemical monitoring. Patients with secondary hypoadrenalism do not normally require mineralocorticoid supplementation because the renin–angiotensin–aldosterone pathway should be intact.

### **Prolactin Deficiency**

Prolactin deficiency is rare because hypopituitarism is generally associated with mild hyperprolactinemia. When present, as in Sheehan's syndrome, it is associated with failure of lactation, but it is otherwise clinically silent.

### **Vasopressin Deficiency**

Hyposecretion of vasopressin from the posterior pituitary results in partial or complete cranial diabetes insipidus. Desmopressin, a synthetic analogue of arginine vasopressin, is the replacement drug of choice and is available in oral, intranasal, and parenteral forms. Compared with endogenous vasopressin, it has a significantly longer half-life and is generally administered between one and three times per day. Desmopressin should be started at low dose and gradually titrated up until the polyuria is controlled. Monitoring paired urine and plasma osmolality, urea and electrolytes, and fluid balance is essential, particularly during the early stages and with dose adjustments.

## **MONITORING AND LONG-TERM REPLACEMENT**

The aim of endocrine therapy in hypopituitarism is to achieve adequate replacement, ensuring that the patient is asymptomatic and able to lead a normal life, but also to avoid the potential complications caused by long-term exposure to supraphysiological doses of hormones.

The adverse effects of excess glucocorticoid replacement are well known; consequently, the aim of therapy increasingly is to provide an individual with the lowest dose of hydrocortisone possible to alleviate symptoms. Use of 24-h collections for urinary-free cortisol in combination with hydrocortisone day curves (varying from 3 points up to 10 points) can

help to guide exact replacement doses for a particular individual.

Supraphysiological thyroxine replacement is also increasingly being recognized as a significant problem due to the increased risk of atrial fibrillation, osteoporosis, and possible associations with dementia, but it is less easy to define in secondary hypothyroidism than in primary hypothyroidism, where TSH monitoring provides an accurate guide.

Excess androgen replacement can lead to clinically significant polycythemia, and monitoring of the hematocrit is recommended. Concerns also remain that testosterone replacement may exacerbate prostatic hypertrophy and malignancy. In view of this, although the current policy in the United Kingdom is not to carry out population screening for prostate cancer, most endocrinologists currently recommend biannual prostate-specific antigen (PSA) monitoring for men receiving androgen therapy.

Overreplacement with GH can, via sodium and water retention, lead to weight gain, peripheral edema, and carpal tunnel syndrome. Arthralgia and myalgia are also well recognized side effects of treatment. GH replacement leads to enhanced insulin resistance that may unmask previously undiagnosed diabetes and contribute to increased cardiovascular risk. The concern that GH therapy would be associated with an increased risk of subsequent malignancy has not been borne out by the evidence.

In patients with cranial diabetes insipidus and intact thirst responses, severe dehydration is usually avoided because adequate fluid intake is maintained. If patients complain of excessive thirst, 24-h records of fluid balance can help to guide a dose increase of desmopressin. Hyponatremia as a consequence of lack of free water excretion due to overtreatment with desmopressin is a potentially more serious complication of treatment. This can be avoided by regular biochemical monitoring when therapy is commenced and after any dose adjustment (paired urine and serum osmolality in addition to urea and electrolytes) in addition to advising patients that they should be aware of the effect of the drug wearing off (i.e., polyuria) 1 to 2 h before their next dose.

In patients with partial hypopituitarism, the potential for new deficiencies must always be borne in mind, particularly if the patients have received pituitary radiotherapy given that the lag time between irradiation and developing clinically significant endocrine deficiencies may be many years. In these patients, ongoing monitoring is essential.

Patients now take on average lower replacement doses of steroid than ever before, leaving them

potentially more vulnerable to hypoadrenalism when intercurrent illness occurs. Consequently, these patients must receive clear verbal and written advice on when and how to adjust their glucocorticoid replacement. It is also vital that patients' general practitioners and family members be made aware of the diagnosis, and most patients are advised to carry some form of identification, such as a medic alert bracelet, to assist medical staff if the patients lose consciousness. Mild intercurrent illnesses that are not associated with significant diarrhea, vomiting, or a fever generally do not warrant an alteration in steroid replacement. Patients should be advised to double their standard regimen in the event of a pyrexial illness and to seek medical assistance early if profuse vomiting or diarrhea is preventing the administration or absorption of oral hydrocortisone. Many patients have an emergency steroid pack consisting of a single vial of 100 mg of hydrocortisone that can be administered intramuscularly by a trained relative or visiting general practitioner. Severe illness or surgery usually necessitates hospital admission for regular parenteral hydrocortisone (50–100 mg intramuscularly every 6 h).

Hormone replacement regimens can be complex and require patients to actively participate in their treatment to achieve the best quality of life possible. Patient education is fundamental to achieving this aim, and charities such as the Pituitary Foundation (United Kingdom) and the Pituitary Tumor Network Association (United States), which offer advice and support to patients and relatives, are invaluable.

## THE FUTURE

There is good evidence that patients with acromegaly and Cushing's disease are at increased risk for premature death, but there is increasing evidence to support the hypothesis that other patients with hypopituitarism have a mortality exceeding that of the background population. Studying this population is clearly

complex because it includes patients with a variety of endocrine deficiencies due to differing underlying pathologies who have received a selection of treatment modalities, but it appears that the premature mortality might not be entirely attributable to GH and other endocrine deficiencies but rather might reflect increased mortality from respiratory and vascular disease.

There is significant evidence that replacement therapy with dehydroepiandrosterone is beneficial in Addison's disease, but replacement therapy is also postulated to be of benefit in improving well-being in secondary hypoadrenalism. Further research is required to clarify these and other contentious issues.

## See Also the Following Articles

Hypopituitarism, Hormonal Therapy for • Hypothalamic Disease • Hypothalamus–Pituitary–Thyroid Axis • Pituitary Gland Anatomy and Embryology • Pituitary Tumors, Molecular Pathogenesis • Pituitary Tumors, Surgery

## Further Reading

- DeGroot, L. J., and Jameson, J. L. (2001). "Endocrinology" (3 vols.). W. B. Saunders, Philadelphia.
- Howlett, T. A. (2002). Hypopituitarism *In* "The Oxford Textbook of Endocrinology" (H. E. Turner and J. A. H. Wass, eds.), pp. 145–151. Oxford University Press.
- Lamberts, S. W. J. (1997). "The Diagnosis and Treatment of Pituitary Insufficiency." BioScientifica, Bristol, UK.
- Regal, M., Páramo, C., Sierra, J. M., and García-Mayor, R. V. (2001). Prevalence and incidence of hypopituitarism in an adult Caucasian population in northwestern Spain. *Clin. Endocrinol.* **55**, 735–740.
- Tomlinson, J. W., Holden, N., Hills, R. K., Wheatley, K., Clayton, R. N., Bates, A. S., Sheppard, M. C., and Stewart, P. M. (2001). Association between premature mortality and hypopituitarism: West Midlands Prospective Hypopituitary Study Group. *Lancet* **357**, 425–431.
- Turner, H.E., and Wass, J. A. H. (eds.) (2002). "The Oxford Handbook of Endocrinology." Oxford University Press.
- Wass, J. A. H., and Thakker, R. V. (2001). "Endocrine Disorders (Medicine)." Medicine Publishing, Abingdon, UK.

Table 1 Causes of Hypopituitarism

Common causes of hypopituitarism	Less common causes of hypopituitarism	Rare causes of hypopituitarism
Pituitary tumor	Other cranial tumors	Empty sella syndrome
Pituitary surgery	Rathke's cleft cyst	Trauma
Pituitary radiotherapy	Pituitary apoplexy including Sheehan's syndrome	Granulomatous hypophysitis
	Idiopathic isolated hormone deficiencies	Hemachromatosis
		Pituitary metastases
		Internal carotid artery
		Aneurysm

## HORMONE REPLACEMENT THERAPY IN THE PATIENT WITH HYPOPITUITARISM

### Adrenocorticotropin Deficiency

Hydrocortisone is the drug of choice and is the firstline treatment in most countries. Hydrocortisone has traditionally been given in two doses: a larger dose of 20 mg on waking and a smaller dose of 10 mg taken with the evening meal. However, Esteban and colleagues convincingly demonstrated daily cortisol production rates in healthy individuals to be less than half of the previously accepted values. Other work has highlighted potentially detrimental effects on bone density and glucose tolerance of high tissue exposure to glucocorticoids, so a more tailored approach to therapy is now the norm. Many patients on twice-daily replacement regimens report fatigue or headache in the afternoon that is improved by taking their hydrocortisone in three divided doses, although quality-of-life assessments were unchanged with lowered total doses of hydrocortisone in one study. Therefore, doses are fine-tuned according to patient well-being and serum cortisol levels by many centers. Urine-free cortisol measurements are not useful because saturation of corticosteroid-binding globulin (CBG) following oral hydrocortisone results in supraphysiological urine-free cortisol excretion. The hydrocortisone day curve used by our institution consists of serum sampling for cortisol at time 0 before the first dose is given and then at 30 min and 1, 2, 3, 5, 7, 9, 9.5, 10, and 11 h, with hydrocortisone administered at the usual times, typically at 0, 5, and 9 h. The aim is to achieve adequate circulating levels of cortisol throughout the day without excessive peaks (e.g., >1000 nmol/L) or troughs (e.g., <100 nmol/L) before or after doses. Timing and size of dose may then be adjusted, and the hydrocortisone day curve may be repeated, to confirm appropriate cortisol levels. Therefore, a typical modern treatment regimen may provide a lower total dose of

hydrocortisone, typically 15 to 20 mg divided across three time points (e.g., 10 mg on waking, 5 mg at noon, and 5 mg at 5:00 to 6:00 PM).

If a hypopituitary patient is acutely unwell or requires surgery, hydrocortisone should be given via intermittent intramuscular injections. In the event of an emergency, 100 mg will provide full glucocorticoid cover for at least 6 h. Alternatively, hydrocortisone may be given as a continuous intravenous infusion of 1 to 3 mg/h.

The most commonly used alternative to hydrocortisone is cortisone acetate, which is widely used in Scandinavia and the Netherlands. Cortisone acetate has an approximate dose equivalence of 25 mg to 20 mg hydrocortisone. The absorption of cortisone acetate may be slower than that of hydrocortisone, and cortisone acetate is biologically inactive until conversion to cortisol by the enzyme 11 $\beta$ -hydroxysteroid dehydrogenase type 1 (11 $\beta$ HSD1). Furthermore, the net dehydrogenase activity of 11 $\beta$ HSD varies among individuals, being increased in normal females compared with males, decreased in the presence of increased body fat in hypopituitary patients, and decreased in patients with type 1 diabetes mellitus. In addition, 11 $\beta$ HSD1 is inhibited by growth hormone (GH) and insulin-like growth factor-1 (IGF-1) in hypopituitary adults and in acromegaly, resulting in a measurable reduction in circulating cortisol as well as in a fall in the urinary cortisol-to-cortisone metabolite ratio that reflects the relative dehydrogenase-to-reductase activity of 11 $\beta$ HSD. In practical terms, patients receiving cortisone acetate replacement therapy may require further dose titration after the implementation of GH therapy. Other steroid preparations such as prednisolone may also be used in the place of hydrocortisone (dose equivalence 5 mg prednisolone to 20 mg hydrocortisone). However, in view of the inability to monitor plasma glucocorticoid levels in this drug, it should be used only in exceptional circumstances.

Table II Clinical Features of Pituitary Insufficiency

Pituitary hormone	Adrenocorticotropin	Thyrotropin	Growth hormone	Luteinizing hormone/Follicle-stimulating hormone	Arginine vasopressin	Prolactin and oxytocin
Effector hormone	Cortisol	Thyroxine	Insulin-like growth factor-1	Estrogen / Progesterone / Testosterone		
Clinical/ Laboratory findings	Fatigue, weakness	Fatigue, weakness	Fatigue, weakness	Men	Thirst	Failure of lactation
	Mental depression	Mental depression	Mental depression	Loss of libido	Polyuria	
	Anorexia, weight loss	Weight gain	Increased visceral adiposity	Erectile dysfunction	Polidipsia	
		Constipation		Hot flushes	Hypernatremia	
	Dizziness, Orthostatic hypotension	Cold Intolerance	Osteoporosis	Diminished facial and body hair	Increased serum osmolality with Inappropriately low urine osmolality	
		Menorrhagia	Hyperlipidemia			Disinhibition of prolactin
			Reduced muscle strength	Fine facial wrinkles		
	Nausea, vomiting			Gynecomastia		
	Diarrhea	Hyponatremia		Testicular atrophy		Men
		Hyperlipidemia		Infertility		
	Hyponatremia	Normocytic anemia				Loss of libido
	Hypoglycemia	Bradycardia		Women		
	Hyperkalemia	Hyporeflexia		Menstrual disturbance		Impotence
Normocytic anemia					Women	
Eosinophilia			Hot flushes			
			Loss of libido			
			Dyspareunia		Menstrual disturbance	
			Infertility			
					Galactorrhea	

### Patient Education

Education is important for patients with ACTH deficiency. It is vital that these patients understand the rationale for treatment to ensure long-term compliance and also that they know how to adjust their treatment in the case of an emergency. The first dose should be taken on waking, and the last dose should be taken several hours before going to bed to ensure safe glucocorticoid cover for the waking hours but to minimize insomnia at night. Patients should also carry documentation, such as a steroid card or a medic alert bracelet, to alert others of their steroid requirement in the case of an emergency. Patients should understand that during intercurrent illness, their steroid requirements will increase; therefore, they should be confident about doubling their hydrocortisone dose during

pyrexial illness and should carry a vial of 100 mg hydrocortisone for intramuscular administration in the event of severe illness, vomiting, or diarrhea.

### Antidiuretic Hormone Deficiency

Synthetic vasopressin (DDAVP) is available to be taken orally, nasally, or via subcutaneous injection. Dose and route of administration are altered according to the clinical setting as well as to objective fluid status and serum and urine osmolality and biochemistry. For example, in the acute clinical setting, where diabetes insipidus follows pituitary surgery, DDAVP is best administered via the subcutaneous route to provide a reliable predictable response. Very small doses (0.5–1 µg) will effectively control polyuria

**Table III Basic Investigations for Diagnosis of Pituitary Hormone Deficiencies**

<b>ACTH</b>
9 AM cortisol
Insulin tolerance test
<b>TSH</b>
Free (or total) T <sub>4</sub>
TSH
<b>LH/FSH</b>
Basal LH and FSH
9 AM testosterone/estradiol
<b>GH</b>
Basal IGF-1
Insulin tolerance test
<b>Arginine vasopressin</b>
Synchronous serum and urine osmolality
Water deprivation test

for up to 16 h. This is traditionally given at night to prevent nocturia and allow undisturbed sleep, with polyuria expected to restart late during the following afternoon if at all. In the nonacute setting, subcutaneous administration is inappropriate, so the oral or nasal route is chosen according to patient preference. DDAVP is best absorbed via the nasal route, allowing smaller doses to be used (typically 10–20 µg once or twice daily). However, this requires some experience and expertise in the technique of self-administration, and some patients will prefer the oral route. Oral DDAVP is less effective, such that larger doses are required and must be taken more frequently (100–300 µg taken two or three times daily). Doses of more than 200 µg have not been found to increase the antidiuretic effect significantly, although they may allow less frequent doses. Dose titration and patient education are required to achieve optimal symptom control, particularly at night, and to maintain a normal serum osmolality and sodium concentration. Slight undertreatment, with normal water homeostasis being maintained by thirst mechanisms, is the preferred approach. In those situations where ADH deficiency is accompanied by a reduced thirst threshold, it is necessary to carefully monitor body weight and urine output on a fixed dose of DDAVP and to adjust fluid intake accordingly.

### Thyroid-Stimulating Hormone Deficiency

Thyroid hormone replacement has been practiced for more than a century. Animal-derived thyroid extract

was first given by mouth in 1892 and was widely used until the 1970s. More recently, synthetic levothyroxine sodium (L-thyroxine [L-T<sub>4</sub>]) has become established as the preferred replacement drug. Recent trials have examined potential benefits of simultaneous administration of L-T<sub>4</sub> and triiodothyronine (T<sub>3</sub>).

Healthy human thyroid secretes predominantly T<sub>4</sub> but also T<sub>3</sub>, which accounts for approximately 25% of daily whole-body T<sub>3</sub> production. T<sub>3</sub> is responsible for most of the biological activity of thyroid hormone, and the majority is derived by deiodination of circulating T<sub>4</sub>. The original desiccated thyroid extracts contained both L-T<sub>4</sub> and T<sub>3</sub> (approximately 121 and 20 µg, respectively), resulting in supraphysiological T<sub>3</sub> levels and a high incidence of adverse effects. Synthetic L-T<sub>4</sub> became available during the 1950s, and early studies showed it to be well absorbed and effective at maintaining serum T<sub>4</sub> levels. T<sub>3</sub> displays superior gastrointestinal absorption, but its short half-life results in a need for multiple daily dosing. As a result use of T<sub>3</sub> is largely restricted to thyroid cancer patients undergoing frequent isotopic imaging or treatment and to those patients with ischemic heart disease where gentle incremental increases in thyroid replacement are indicated. L-T<sub>4</sub> has a longer half-life, enabling once-daily dosing without major adverse consequences of occasional tablet omission.

Most authorities advocate L-T<sub>4</sub> as the preferred agent for thyroid replacement. Improvements in the sensitivity of TSH assays led to the recognition that previous L-T<sub>4</sub> doses induced suppression of TSH, with the average L-T<sub>4</sub> dose required to normalize TSH in the area of 1.6 µg/kg. In hypopituitarism, TSH is low or absent and is not a useful marker of adequacy of L-T<sub>4</sub> replacement. In these patients, measurement of circulating thyroid hormones supported by clinical assessment are the best guides to dose titration. Experience from patients with primary hypothyroidism indicates that doses of L-T<sub>4</sub> that normalize serum TSH commonly result in high normal or elevated free thyroxine (FT<sub>4</sub>) levels, whereas T<sub>3</sub> levels are normal. This may be explained by the need for higher FT<sub>4</sub> levels to compensate for the absence of thyroidal T<sub>3</sub> secretion.

When starting L-T<sub>4</sub> replacement, consideration of the duration of hypopituitarism and the presence of comorbidities, particularly ischemic heart disease, is necessary. In young patients with short-duration TSH deficiency, a full replacement dose of L-T<sub>4</sub> may be initiated (typically a starting dose of 100 µg). To avoid precipitating cardiac events, a low dose (e.g., 25–50 µg) should be commenced and gradually

titrated against serum FT4 levels and clinical response in the elderly and in those with possible long-standing hypopituitarism. Occasionally, initiation of T3 followed by a crossover to L-T4 once safety is established is appropriate.

The serum FT4 level is the best method to assess status of L-T4 replacement in hypopituitarism. Changes in values following a dose are modest, but it is best to monitor immediately prior to the next dose. Consideration of increases in L-T4 replacement should be made for patients using oral estrogen therapy. Such treatment increases levels of circulating thyroid-binding globulin, resulting in a decrease in bioavailable FT4, and this phenomenon will not be evident if serum total thyroxine measurements are made. Dose requirements may fall with increasing age, perhaps related to reduced clearance in the elderly. In hypopituitary adults with growth hormone deficiency (GHD), conversion of FT4 to T3 may be reduced, and it is not uncommon for patients to require replacement L-T4 doses that result in elevated FT4 levels to generate normal T3 concentrations.

#### **Other Issues**

Generic and brand name L-T4 preparations are generally bioequivalent and interchangeable. However, prescribing practice does differ; in the United Kingdom, most authorities favor generic L-T4, whereas in the United States and some European countries, brand names are favored.

Excessive L-T4 replacement may be associated with adverse effects. Studies have indicated that replacement therapy results in lumbar spine bone loss in premenopausal women; however, it remains unclear whether fracture risk is increased. Doses of L-T4 that result in suppressed TSH in those with primary hypothyroidism are associated with a risk of atrial fibrillation and increased left ventricular mass.

It is recognized that a small subset of patients requiring L-T4 replacement continue to feel symptomatic despite normalization of FT4 levels. Because exposure to T3 may differ between tissues in such patients, the potential benefits of combined T4 and T3 replacement has been the focus of recent study. Early evidence in humans indicates that mood and psychological status may be improved in hypothyroid individuals receiving a combination replacement regimen. However, this approach must still be considered experimental, and until slow-release T3 preparations are available, use of T3 in routine replacement in hypopituitary patients will be limited.

## **Gonadotropin Deficiency**

### **Males**

The gonadotropins luteinizing hormone (LH) and follicle-stimulating hormone (FSH) are responsible for gonadal sex steroid production and gamete production and maturation in both genders. Deficiency of LH and FSH results in absence of sperm (azospermia) and testosterone (the main circulating androgen) in the male. The goals of treatment of gonadotropin deficiency are to (1) induce secondary sexual characteristics in patients that have not had spontaneous puberty, (2) maintain or restore normal sexual function, and (3) restore fertility.

### *Effects of Testosterone Deficiency*

Testosterone is an important regulator of body composition in postpubertal adults, and testosterone deficiency results in reduced lean body mass (predominantly skeletal muscle) and increased fat mass. Testosterone replacement in the male with hypogonadism results in correction of these abnormalities and normalization of body composition.

Evidence indicates an important role for testosterone in the regulation of bone mineralization. Cross-sectional data indicate an association between serum testosterone levels and bone mineral density (BMD), and induced hypogonadism results in decreases in BMD. Testosterone replacement in the male with hypogonadism results in substantial increases in BMD. However, the precise mechanisms of testosterone-mediated bone anabolism remain poorly understood. It has been suggested that the majority of effects are estrogen mediated following aromatization of testosterone.

Testosterone has important effects on behavior, and hypogonadism results in decreases in sexual interest, arousal, and the achievement of erections. Improved sexual function has been clearly demonstrated following testosterone replacement. In some studies, improvements in mood have been recorded following testosterone replacement in hypogonadism. Data regarding androgens and lipids are somewhat contradictory. Cross-sectional studies indicate a positive correlation between high-density lipoprotein (HDL) cholesterol and testosterone but indicate inconsistent relationships between testosterone and total and low-density lipoprotein (LDL) cholesterol. Similarly, both increased and decreased total, LDL, and HDL cholesterol have been reported following testosterone replacement in hypogonadal men. Other risk factors for cardiovascular disease may also be altered by androgen status. These include PAI-1,

fibrinogen, and antithrombin III as well as proteins C and S.

Growth of both facial hair and body hair is reduced in hypogonadism. Prostate cancer rarely occurs in testosterone-deficient men, and antiandrogen treatment can lead to regression of prostate carcinoma. Although testosterone replacement should be avoided in patients known to have carcinoma of the prostate, there is no evidence that appropriate replacement in the deficient male is associated with a greater risk than in the healthy population.

#### *Replacement Treatment in Males with Gonadotropin Deficiency*

Testosterone replacement should be considered in all males with gonadotropin deficiency resulting in low serum testosterone. The goals of treatment are to improve quality of life through restoration of normal circulating levels with a minimum of adverse effects. A number of routes of administration are available, and the choice should be tailored to the individual. Monitoring of testosterone therapy is based on measurement of serum testosterone levels and assessment of the clinical response. Supraphysiological circulating levels of total testosterone should be avoided because they may contribute to the development of adverse effects. These include changes in personality, acne vulgaris, gynecomastia, lipid disorders, polycythemia, and possibly prostate malignancy.

#### *Modes of Testosterone Replacement*

Oral preparations are subject to a high first-pass hepatic metabolism. A number of modified testosterone products have been tried with variable success, and large doses are necessary. Testosterone undecanoate is the only available oral preparation, and when it is taken it is absorbed through intestinal lymphatics, avoiding complete first-pass effects. The half-life is short, requiring more than a single daily dose, and individual pharmacodynamics vary considerably. The most reliable monitoring method is measurement of serum dihydrotestosterone. In most centers, oral testosterone undecanoate is reserved for those individuals in whom other modes of replacement have not been successful.

Intramuscular administration of testosterone esters is the most widely used method of testosterone replacement in the hypopituitary male. Esterification increases the half-life, with various esters resulting in different dosing interval requirements ranging from days to several months. Many preparations result in supraphysiological peaks immediately following administration, with falls to low levels prior to

subsequent dosing. The commonly used preparations are "Sustanon" (a mixture of propionate, phenylpropionate, and deconate esters), testosterone isocaproate, and testosterone enanthate. Usually, these preparations are administered every 2, 3, or 4 weeks and often require visits to a family practice or hospital clinic. The potentially longer acting testosterone buciclate is under development and may result in less frequent injections with a smoother serum profile.

Typically, subcutaneous implants are inserted on a 4- to 6-month basis, with 600 to 800 mg of testosterone the common dose range. Peak serum testosterone levels are seen 2 to 4 weeks postplacement, with a gradual decline thereafter. The main disadvantage is need for a minor surgical procedure. A variety of transdermal preparations have become available. Nonscrotal patches deliver testosterone via a permeation-enhancing vehicle, and a new patch is applied each day. Treatment with 2.5 to 5.0 mg of testosterone daily provides total serum testosterone levels within the normal range in the majority of hypopituitary adults. The patch should be applied at night because the majority of testosterone is delivered within the first 12 h, mimicking the normal circadian rhythm of testosterone levels. The major restriction relates to skin irritation that affects approximately 50% of people and has limited the use of this modality. A scrotal patch that delivers testosterone with less skin irritation is also available. Shaving of the site of application is required, and the patch must be changed daily.

More recently, a testosterone gel preparation that results in less skin irritation and more satisfactory serum levels has become available. It is likely that this preparation will be widely used when it becomes licensed throughout Europe and as availability increases. Buccal and sublingual preparations are also undergoing investigation.

#### *Fertility Treatment in Gonadotropin Deficiency*

The production of testosterone is governed by LH, and this axis is responsible for the induction of puberty and maintenance of secondary sexual characteristics. FSH, in addition to being needed for an adequate level of testosterone, is required for complete spermatogenesis. In patients with gonadotropin deficiency, use of FSH in addition to human chorionic gonadotropin (hCG, which has LH activity) will both induce puberty and restore fertility. Although hCG is exclusively available as urinary extracts, increasingly recombinant preparations of FSH are being used.

In prepubertal adults, a number of different protocols of gonadotropin administration have been

suggested, but a commonly used algorithm consists of the following:

1. hCG (2000 IU intramuscularly) three times weekly for 6 to 8 weeks to elevate the serum testosterone levels
2. FSH (37.5 IU intramuscularly) three times weekly to facilitate spermatogenesis

The semen should be analyzed monthly to assess response. If the response is suboptimal, doses may be doubled. In most cases, satisfactory serum testosterone levels and quality and number of spermatozoa are seen, but treatment may need to continue for up to 1 year. In postpubertal adults, treatment to restore fertility should be commenced with hCG (2000 IU) three times weekly. This may be sufficient to induce spermatogenesis alone, but addition of FSH is necessary in most patients. Our preferred practice is to administer high-dose FSH (300 IU) three times weekly and hCG (2000 IU) two times weekly to eliminate the need to increase the dose if increased spermatogenesis fails to occur after 4 months of therapy. Regardless of the dose used, an interval of at least 4 months is required to achieve an increase in sperm count and treatment should then be continued for a further 6 months or until conception is achieved. Because these agents are expensive efforts to cryopreserve, sperm should be made once levels are adequate.

#### *Gonadotropin-Releasing Hormone Therapy*

Pulsatile administration of gonadotropin-releasing hormone (GnRH) results in production and secretion of LH and FSH in patients with hypothalamic disorders provided that anterior pituitary function is preserved. Typically, a small infusion pump is worn and doses of GnRH (~25 ng/kg body weight) are delivered subcutaneously every 90 min. Adequate serum levels of the gonadotropins and testosterone are measurable in most patients after 7 days. Fertility should be assessed using seminal analysis after 4 months of treatment. Because pulsatile therapy is expensive and requires considerable effort, cryopreservation of sperm should be considered.

#### **Females**

The main circulating estrogen in the healthy premenopausal female is ovarian-derived 17 $\beta$ -estradiol, and production rates are in the area of 100 to 300  $\mu$ g/day. In addition, estrogen production occurs in peripheral tissues from conversion of testosterone and androstenedione. Ovarian estrogen production is regulated by the gonadotropins, and in healthy women the

ovarian failure seen in the menopause results in a 10- to 20-fold increase in circulating FSH levels and a threefold increase in LH. In the hypopituitary female, absence of gonadotropins results in decreased circulating estrogen and progesterone levels and anovulation.

#### *Effects of Estrogen Deficiency*

Estrogen deficiency results in both short-term and long-term effects, and individuals vary in the pattern and severity of complications encountered. The majority of patients with hypopituitarism have had long-term gonadotropin deficiency, but those with new onset of hypothalamo-pituitary disease or surgery may experience variable degrees of vasomotor instability, fatigue, mood disturbance, and sexual dysfunction. The latter is principally related to atrophic change in vaginal epithelium.

Long-term consequences of estrogen deficiency are frequently seen in women with early onset of hypopituitarism who have not undergone estrogen replacement. The most important consequence is reduced bone mineralization resulting in increased risk of osteoporotic fracture. The development of osteoporosis in an individual depends on the peak adult bone mass achieved and the subsequent rate of bone loss. Genetic factors are the most important determinants of peak bone mass; however, estrogen deficiency throughout adolescence and early adulthood contributes to reduced bone mineralization. Thus, onset of estrogen deficiency both early in life and later (following achievement of peak bone mass) can affect bone health. BMD can be assessed using a variety of techniques, including quantitative computed tomography (CT), photon absorptiometry, and (more recently) dual-energy X-ray absorptiometry (DXA). The wide clinical availability and relative inexpense of DXA means that increasingly endocrinologists are using this tool to assess bone health and to monitor the effect of treatments in hypopituitary patients.

Dyslipidemia is the other major long-term consequence of estrogen deficiency. Circulating estradiol alters lipoprotein metabolism, resulting in reduced LDL cholesterol and increased HDL cholesterol. Use of estrogen replacement results in favorable effects on the lipid profile; however, androgenic progestins may counteract this in part.

#### *Replacement Therapy in Females With Gonadotropin Deficiency*

Hormone replacement therapy was first introduced for the treatment of menopausal symptoms some 30 years ago. Early preparations used unopposed

estrogen preparations that were typically taken for 21 days followed by 7 tablet-free days. This regimen resulted in endometrial hyperplasia and increased risk of endometrial carcinoma. Since the 1970s, estrogen replacement has been combined with cyclical progesterone, which normalizes or even reduces the risk of endometrial cancer. However, these treatments result in restoration of a menstrual bleed in the majority of women with intact uteri.

The aims of treatment of gonadotropin deficiency in the hypopituitary female are to counteract the features of estrogen deficiency, specifically maintaining bone and sexual health, and to assist fertility in some patients. In those not seeking fertility, estrogen replacement treatment should be tailored to the individual patient given that a number of different routes of administration are available, attesting to the fact that none is perfect.

Estrogen may be administered via oral, subcutaneous, and transdermal routes. Oral estrogens are in the form of natural estrogens (conjugated equine, estradiol valerate, and 17 $\beta$ -estradiol) and typically are taken for 21 days followed by 7 tablet-free days. For the younger patient, the oral contraceptive preparations are slightly unorthodox, but we see no reason why they should not be used, especially given that they may allow the patient to feel more “normal” in relation to her peers. Oral preparations are subject to extensive first-pass hepatic metabolism, and large doses are required (typically 1–2 mg/day). To avoid endometrial hyperplasia, these should be combined with a progestin or the “Mirena” intrauterine device.

Transdermal preparations are usually applied twice weekly and contain 50 to 100  $\mu$ g of estrogen in cyclical combination with a progestin. Subcutaneous implants are inserted as a minor surgical procedure 6 times monthly and necessitate cyclical progestin treatment in those with intact uteri. Recently, selective estrogen receptor modulators (SERMs), including tibolone and raloxifene, have become available. These agents provide bone mineral effects with reduced effect on endometrial tissue, resulting in decreased or absent menstrual bleeding. However, raloxifene has little effect in treating the vasomotor symptoms of acute estrogen deficiency. In those intolerant of conventional estrogen treatment, consideration should be given to the use of a SERM to preserve bone health. The nonestrogenic compound tibolone provides an alternative treatment for postmenopausal symptoms and also exerts favorable effects on bone.

Monitoring of estrogen replacement is based largely on clinical response. In females with intact

uteri, it may be anticipated that a cyclical menstrual period will occur. Synthetic estrogens and conjugated equine estrogens do not result in predictable levels of circulating estradiol, and this commonly used measurement is of little practical benefit in patients using these preparations. Typically, estrogen replacement should be continued provided that there are no contraindications until the age of the natural menopause—approximately 51 years. Thereafter, the use of these compounds must be balanced against the potential adverse effects related to thromboembolism, breast cancer, and cardiovascular morbidity. For many years, estrogen replacement was considered safe and relatively free from adverse effects. Little information is available regarding specific use of estrogen compounds in hypopituitary females, and much of the safety data are derived from the postmenopausal population taking HRT. A recent study provided clear evidence that combined estrogen/progestin is associated with increased incidence of invasive breast cancer, thromboembolism, and coronary events in women ages 50 to 75 years, highlighting that these preparations should be tailored to each individual rather than used as a matter of routine.

Fertility treatment in the hypopituitary female requires ovulation induction using hCG and hMG/recombinant FSH. Typically, this is supervised through fertility clinics and may involve ovum storage and *in vitro* fertilization.

### Growth Hormone Deficiency

Unlike glucocorticoids, thyroid hormone, and sex steroids, the use of GH replacement therapy in adult hypopituitarism is largely restricted to patients affected by a characteristic clinical syndrome. This syndrome includes reduced well-being with low energy levels, poor self-esteem, and social isolation; abnormal body composition with increased android (abdominal and visceral) fat and decreased lean body mass and total body water; osteopenia with reduced bone remodeling; and an adverse cardiovascular risk profile that includes dyslipidemia, hyperfibrinogenemia, and elevated levels of plasminogen activator inhibitor (PAI). As with other forms of endocrine replacement therapy, the goal of therapy is to maximize the clinical benefits while minimizing the risks of excess exposure. Early studies of GH replacement used weight and/or surface area GH doses, but it is now accepted that the dose of GH needs to be individually tailored to each patient. Several strategies exist for monitoring, although each strategy has important limitations.

### Tolerability

In the early trials of GH replacement, symptoms related to sodium and water retention (edema, arthralgia, and myalgia) were common and required dose reductions in up to 40% of patients. However, even when GH doses are lowered due to adverse symptoms, biochemical markers of GH action, such as serum IGF-1, remain elevated in up to one-third of patients, suggesting that the absence of symptoms of GH overtreatment is a relatively crude method of judging excess GH exposure.

### Clinical Monitoring

In theory, each of the features of the GH deficiency syndrome could serve as a marker of efficacy for dose monitoring, but correction of abnormal body composition has received most attention. Although these studies illustrate the value of body composition changes as a marker of GH efficacy, they also suggest that increasing the dose of GH due to persistent abnormal body composition may result in overtreatment, judged by biochemical markers of GH action. Similar conclusions may be drawn from studies of changes in well-being in response to GH replacement. Therefore, the question arises as to whether supraphysiological levels of serum IGF-1 have an adverse effect on adult hypopituitarism. Clinical experience with acromegaly, a condition known to be associated with excess morbidity and mortality and epidemiological data, regarding the risk of developing carcinoma of the prostate or breast would suggest that biochemical overtreatment should be avoided until more data are available.

### Biochemical Monitoring

Potential candidates for the biochemical monitoring of GH dose include serum IGF-1, IGF-binding protein 3 (IGFBP-3), and acid labile subunit (ALS). Patients reporting symptoms of GH excess are more likely to have an elevated serum IGF-1 level than to have elevated IGFBP-3 or ALS, suggesting that IGF-1 is a more sensitive marker of GH action during GH replacement. Importantly, treatment protocols that specifically aim to avoid elevated serum IGF-1 levels are associated with similar clinical benefits to those observed in the original placebo-controlled studies. However, serum IGF-1 levels may not reflect "true" GH status; up to 30% of patients with severe GH deficiency have serum IGF-1 levels in the lower part of the age-adjusted normal range. Furthermore, in most studies, blood samples for serum IGF-1 are drawn during the early morning. The diurnal variation in serum IGF-1 levels seen with a single

nightly GH injection (a morning peak with a nadir at night) may lead to an artifactually high incidence of supraphysiological levels of serum IGF-1.

### Gender Differences in Growth

#### Hormone Responsiveness

Various studies indicate that clinical and biochemical responses to a given dose of GH are greater in men than in women, a discrepancy that is likely due, at least in part, to modulation of GH action by estrogen. Furthermore, there are data to suggest that the route of estrogen administration may affect the GH maintenance dose, with lower GH doses required in women taking transdermal, as opposed to oral, estrogen preparations.

### Summary

There is no biological marker in adults that is the equivalent of linear growth in children by which to judge the efficacy of GH replacement. Most clinicians agree that GH is most appropriately commenced at low doses, building up slowly to the final maintenance dose. This, in turn, is best determined by a combination of clinical response and measurement of serum IGF-1, avoiding supraphysiological levels of this GH-dependent peptide.

### See Also the Following Articles

ACTH (Adrenocorticotrophic Hormone) • Estrogen Replacement, Oral • FSH (Follicle-Stimulating Hormone) • Growth Hormone (GH) • Hormone Replacement Therapy, Male • Hormone Replacement, Transdermal • Hypopituitarism • LH (Luteinizing Hormone) • Pituitary Tumors, Molecular Pathogenesis • Pituitary Tumors, Surgery • TSH Function and Secretion

### Further Reading

- Bunevicius, R., Kazanavicius, G., Zalinkevicius, R., and Prange, A. J., Jr. (1999). Effects of thyroxine as compared with thyroxine plus triiodothyronine in patients with hypothyroidism. *N. Engl. J. Med.* **11**, 424-429.
- Burman, P., Johansson, A. G., Siegbahn, A., Vessby, B., and Karlsson, F. A. (1997). Growth hormone (GH)-deficient men are more responsive to GH replacement therapy than women. *J. Clin. Endocrinol. Metab.* **82**, 550-555.
- Carroll, P. V., Christ, E. R., Bengtsson, B. A., Carlsson, L., Christiansen, J. S., Clemmons, D., Hintz, R., Ho, K., Laron, Z., Sizonenko, P., Sonksen, P. H., Tanaka, T., and Thorner, M. (1998). Growth hormone deficiency in adulthood and the effects of growth hormone replacement: A review. *J. Clin. Endocrinol. Metab.* **83**, 382-395.
- Cook, D. M., Ludlam, W. H., and Cook, M. B. (1999). Route of estrogen administration helps to determine growth hormone (GH) replacement dose in GH-deficient adults. *J. Clin. Endocrinol. Metab.* **84**, 3956-3960.

- Drake, W. M., Coyte, D., Camacho-Hubner, C., Jivanji, N. M., Kaltsas, G., Wood, D. F., Trainer, P. J., Grossman, A. B., Besser, G. M., and Monson, J. P. (1998). Optimizing growth hormone replacement by dose titration in hypopituitary adults. *J. Clin. Endocrinol. Metab.* **83**, 3913–3919.
- Lam, K. S., Wat, M. S., Choi, K. L., Ip, T. P., Pang, R. W., and Kumana, C. R. (1996). Pharmacokinetics, pharmacodynamics, long term efficacy, and safety of oral 1-deamino-8-D-arginine vasopressin in adult patients with central diabetes insipidus. *Br. J. Clin. Pharmacol.* **42**, 379–385.
- Peacey, S. R., Guo, C., Robinson, A. M., Price, A., Giles, M. A., Eastell, R., and Weetman, A. P. (1997). Glucocorticoid replacement therapy: Are patients overtreated and does it matter? *Clin. Endocrinol.* **46**, 255–261.
- Samsioe, G. (2002). The role of ERT/HRT. *Best Pract. Res. Clin. Obstetr. Gynaecol.* **16**, 371–381.
- Tchernof, A., Labrie, F., Belanger, A., Prud'homme, D., Bouchard, C., Tremblay, A., Nadeau, A., and Despres, J. P. (1997). Relationships between endogenous steroid hormone, sex-hormone binding globulin, and lipoprotein levels in men: Contribution of visceral obesity, insulin levels, and other metabolic variables. *Atherosclerosis* **133**, 235–244.
- Toft, A. D. (1994). Thyroxine therapy. *N. Engl. J. Med.* **331**, 174–180.
- Women's Health Initiative Investigators. (2002). Risks and benefits of estrogen plus progestin in healthy postmenopausal women: Principal results from the Women's Health Initiative Randomized Controlled Trial. *J. Am. Med. Assoc.* **84**, 3956–3960.

**Table I Causes of the Syndrome of Hyporeninemic Hypoaldosteronism**

Diabetes mellitus
Gout
Interstitial nephritis
Glomerulonephritis
Nephrolithiasis
Urinary tract obstruction
Systemic lupus erythematosus
Myeloma
Amyloid
AIDS
Use of nonsteroidal anti-inflammatory drugs

release; the administration of prostaglandin synthesis inhibitors could impair renin release.

Several lines of evidence suggest that the primary defect responsible for the hyperkalemia in many patients resides in the adrenal gland. Angiotensin II and ACTH directly stimulate aldosterone release by the adrenal zona glomerulosa. Because the aldosterone response to angiotensin II and ACTH is impaired in many patients, the syndrome of hyporeninemic hypoaldosteronism (HH) may be of mixed causes.

## DIAGNOSIS

The diagnosis of HH is considered for patients in whom hyperkalemia and metabolic acidosis are out of proportion to the degree of renal impairment (i.e., glomerular filtration rate >20 ml/min). "Pseudohyperkalemia" (e.g., hemolysis, thrombocytosis) should be excluded.

The next step is to demonstrate a normal cortisol response to ACTH stimulation. Then, the response of renin and aldosterone levels to stimulation (e.g., upright posture, sodium restriction) should be measured. Low renin and aldosterone levels establish the diagnosis of HH.

## TREATMENT

Several different therapeutic regimens can be employed to lower the plasma potassium concentration. However, the physician should keep in mind that the majority of patients suffering from this syndrome are elderly and have associated conditions such

as diabetes mellitus, atherosclerotic cardiovascular disease, and hypertension.

Mineralocorticoid replacement, usually given as fludrocortisone acetate, is the mainstay of therapy. Large doses of the steroid, often as high as 0.4 to 1.0 mg daily, are frequently required to reduce the plasma level to normal. This observation suggests that the renal tubular cell is resistant to the action of mineralocorticoids, at least with respect to their effect on potassium secretion. Sensitivity to the sodium-retentive effects of fludrocortisone appears to be less affected given that marked sodium retention with edema, exacerbation of hypertension, and congestive heart failure are common consequences. In many patients, the administration of a loop diuretic may be required to prevent or treat these complications.

If the hyperkalemia fails to respond to fludrocortisone acetate, or if excessive sodium retention is encountered, a trial of potassium-wasting diuretics is desirable. In many patients, particularly those with a primary tubular potassium secretory defect, thiazide diuretics are effective. In some patients, sodium bicarbonate may improve potassium excretion and correct the hyperkalemia. If these measures fail, one can always employ a sodium-potassium exchange resin such as sodium polystyrene sulfonate.

In many patients, particularly those with only mild to moderate hyperkalemia, it is best not to institute any therapy but to check the plasma potassium concentration periodically to ensure that the hyperkalemia has not reached dangerous levels.

In all patients, drugs known to impair renal potassium excretion should be avoided. Volume contraction, which decreases distal sodium delivery, also should be prevented.

## See Also the Following Articles

Aldosterone Receptors • Primary Aldosteronism (PAL) • Renal Vein Renin • Renin • Tissue Renin-Angiotensin-Aldosterone System

## Further Reading

- DeFronzo, R. A. (1980). Hyperkalemia and hyporeninemic hypoaldosteronism. *Kidney Intl.* 17, 118-134.
- DeLeiva, A., Christlieb, A. R., Melby, J. C., Graham, C. A., Day, R. P., Luetscher, J. A., and Zager, P. G. (1976). Big renin and biosynthetic defect of aldosterone in diabetes mellitus. *N. Engl. J. Med.* 295, 639-643.

- Oh, M. S., Carroll, H. J., Clemmons, J. E., Vagnucci, A. H., Levison, S. P., and Whang, E. S. (1974). A mechanism for hyporeninemic hypoaldosteronism in chronic renal disease. *Metabolism* **23**, 1157-1166.
- Perez, G. O., Lespier, L., Jacobi, J., Oster, J. R., Katz, F. H., Vaamonde, C. A., and Fishman, L. M. (1977). Hyporeninemia and hypoaldosteronism in diabetes mellitus. *Arch. Int. Med.* **137**, 852-855.
- Romero, J. C., Dunlap, C. L., and Strong, C. G. (1976). The effect of indomethacin and other anti-inflammatory drugs on the renin-angiotensin system. *J. Clin. Invest.* **58**, 282-288.
- Schindler, A. M., and Sommers, S. C. (1966). Diabetic sclerosis of the renal juxtaglomerular apparatus. *Lab. Invest.* **15**, 877-884.
- Tuck, M. L., and Mayes, D. M. (1980). Mineralocorticoid biosynthesis in patients with hyporeninemic hypoaldosteronism. *J. Clin. Endocrinol. Metab.* **50**, 341-347.

Table I Classification of Anovulation Caused by the Central Nervous–Hypothalamic–Pituitary System

---

<b>Physiologic anovulation</b>
Prepubertal period
Postmenopausal period
Pregnancy and postpartum
<b>Functional hypothalamic anovulation</b>
Psychogenic or stress factors
Nutritional factors
Exercise-related factors
<b>Pharmacologic anovulation</b>
Opiate agonists
Dopaminergic antagonists
<b>Psychiatric-associated anovulation</b>
Anorexia nervosa
Pseudocyesis
<b>Organic defects</b>
Isolated GnRH deficiency
Pituitary tumors
Sheehan's syndrome
Empty-sella syndrome
Head trauma
Inappropriate prolactin secretion

---

hypoestrogenic milieu might be seen in the genital tract. However, complains of hot flashes are uncommon. Hypothalamic anovulation may be related to some psychiatric conditions, especially anorexia nervosa. This is an uncommon psychiatric disorder that is usually diagnosed in adolescent and college-aged females. Anorexia nervosa is much more common in women with amenorrhea and therefore should be considered and ruled out in young patients undergoing evaluation due to anovulation. Several organic defects are related to a small but significant number of hypothalamic anovulatory females. Such disorders include pituitary tumors, Sheehan's syndrome (due to circulatory collapse), empty sella syndrome, and head trauma. Isolated gonadotropin-releasing hormone (GnRH) deficiency, which shares many biochemical features with functional hypothalamic anovulation, is characterized by a decrease in the secretion of endogenous GnRH. The disturbed secretion of GnRH leads to hypogonadotropic hypogonadism, eunuchoid features, incomplete development of secondary sexual characteristics, primary (as opposed to secondary in functional anovulation) amenorrhea, and, in some cases, anosmia. This disorder is the result of developmental or migratory-failure of GnRH neurons. Hypoplasia of the olfactory bulb can be identified in some patients.

## CHARACTERISTIC OF FUNCTIONAL HYPOTHALAMIC ANOVULATION

Most women with functional hypothalamic anovulation have a history of normal onset of menarche with regular menstrual cycles between 26 and 35 days. These patients tend to be highly motivated and involved in stressful occupations. Changing lifestyles and an increasing emphasis on exercise, nutrition, and slenderness are expected to result in an increased incidence of hypothalamic anovulation. In some women, emotional or stressful events precede the onset of anovulatory amenorrhea. Other environmental and interpersonal factors (e.g., social maladjustment and psychosexual difficulties) can be identified during the first interview. Women with functional hypothalamic anovulation usually have low to normal body weight. Several studies have shown that these women have more psychological characteristics associated with eating disorders, such as a feeling of general inadequacy, insecurity, lack of control over life, and confusion in identifying and responding to food-related body sensations and emotional states. A thorough physical examination usually reveals normal secondary sexual development. Signs of other hormonal imbalances, such as thyroid enlargement, galactorrhea, and androgenic excess, should be searched for during the examination. A finding of such an endocrinological disorder usually challenges the diagnosis of functional hypothalamic anovulation.

## PATHOPHYSIOLOGY OF FUNCTIONAL HYPOTHALAMIC ANOVULATION

The common underlying defect of these various disorders is the deficiency (or slowing in frequency) in the pulsated secretion of GnRH. When exogenous GnRH is administered to women with hypothalamic anovulation, luteinizing hormone (LH) and follicle-stimulating hormone responses are either normal or exaggerated, suggesting that the primary defect is at a central level with preserved function of the pituitary. A growing body of evidence indicates that an increase in endogenous opiate activity plays a significant role in reducing the pulsated GnRH secretion (and consequently LH) in functional hypothalamic anovulation. Blockage of endogenous opiate receptors by naloxone may increase LH pulse and amplitude in women with hypothalamic anovulation. Long-term treatment with opiate receptor antagonists may result in a spontaneous return of gonadotropin secretion and ovulatory function. Studies have demonstrated that blockage of the dopamine receptor with metaclopramide can also induce an increase in LH secretion.

Disruption of the reproductive function during chronic exposure to stress has been reported in both animals and humans. In humans, there is a functional association between chronic stress and the onset of ovulatory dysfunction and amenorrhea. Evidence links stress, chronic activation of the hypothalamic–pituitary–adrenal axis, and reproductive dysfunction with anovulation. In women with hypothalamic anovulation, there is a significant increase in daytime cortisol levels, a delayed or absent response in ACTH and cortisol secretion during noon meals, and a blunted response to corticotropin-releasing hormone.

Anovulation associated with nutritional factors and weight loss is a well-characterized clinical syndrome. The incidence of weight loss-related anovulation varies widely from series to series, depending on the population studied. It is well-known that the proper amount of fat is crucial for the timely development of menarche (total fat should comprise  $\geq 17\%$  of body weight, corresponding to a body mass index of approximately  $19 \text{ kg/m}^2$ ). Regular ovulation and menstrual cycles necessitate the maintenance of at least 22% of body weight as fat. Weight loss by self-enforced abstinence, starvation, chronic illness, or exercise leads to impaired GnRH secretion and peripheral impaired estrogen status.

Regular strenuous physical activity in women is often associated with anovulation and menstrual disturbances (e.g., delayed menarche, oligomenorrhea, amenorrhea, and luteal phase defect). Endocrinological status is affected by exercise, with some similarity to starvation-associated modifications. Two stress-related mechanisms have been implicated in amenorrheic athletes. Exercise-induced increases in  $\beta$ -endorphin levels influence the frequency and amplitude of LH pulses. Endurance exercise may also cause the release of corticotropin-releasing factor, inhibiting gonadotropin secretion and activating the adrenal release of corticosteroids and androgenic steroids. Although weight loss plays a crucial role in exercise-related anovulation, menstrual function may resume spontaneously during training, even if there is no change in body weight or composition.

## MANAGEMENT

Because of the functional nature of hypothalamic anovulation, major emphasis should be placed on carefully conducted initial interviews, focusing on lifestyle and interpersonal relationships. Since spontaneous recovery is observed in some cases, expectant management with periodic reassessment is a viable option. For women with persistent anovulation,

a major concern is the long-term effect of hypoestrogenism on bone density. When significant osteopenia is observed, estrogen replacement therapy should be initiated. For women who are interested in fertility, a trial of clomiphene citrate is indicated. For those who fail to respond to clomiphene, the recommended and logical approach is ovulation induction by physiologic replacement therapy with pulsate GnRH.

## CONCLUSION

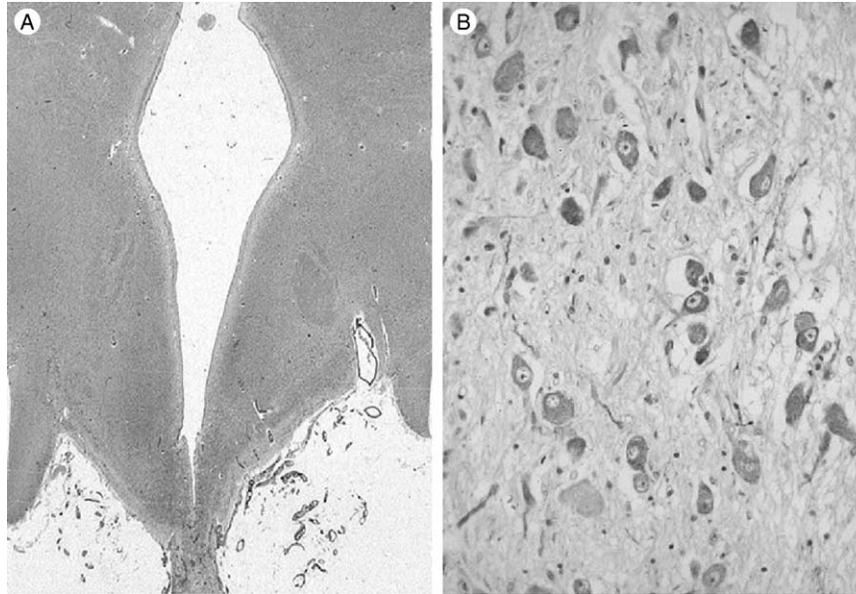
The development of functional hypothalamic amenorrhea is linked to lifestyle and environmental stressors. Psychogenic, nutritional, or exercised-related factors cause a temporary shift into a “resting mode” of the reproductive cycle at the central level. Neuroendocrine factors, which appear to modulate reduced GnRH activity, include the opiate and dopamine neuronal system. In most women, reactivation of the hypothalamic unit occurs after lifestyle modification and accommodation of stress factors. Therapy is aimed at ensuring adequate bone density and resolving consequent infertility.

## See Also the Following Articles

Eating Disorders and the Reproductive Axis • Gonadotropin-Induced Ovulation • Gonadotropin-Releasing Hormone (GnRH) Actions • Osteoporosis, Overview • Ovarian Stimulation: Clomiphene Citrate • Polycystic Ovary Syndrome (PCOS) • Superovulation and Intrauterine Insemination

## Further Reading

- Berga, S. L., Daniels, T. L., and Giles, D. E. (1997). Women with functional hypothalamic amenorrhea but not other forms of anovulation display amplified cortisol concentrations. *Fertil. Steril.* **67**, 1024–1030.
- Grana-Barcia, M., Liz, J., Jimenez, E., Novo, A., and Aguilar, J. (1998). Ovulation induction with pulsatile GnRH in a patient with anovulation of hypothalamic origin and central diabetes insipidus. *Gynecol. Endocrinol.* **12**, 203–207.
- Liu, J. H. (1990). Hypothalamic amenorrhea: Clinical perspectives, pathophysiology, and management. *Am. J. Obstet. Gynecol.* **163**, 1732–1736.
- Marcus, M. D., Loucks, T. L., and Berga, S. L. (2001). Psychological correlates of functional hypothalamic amenorrhea. *Fertil. Steril.* **76**, 310–316.
- Marshall, J. C., Eagleson, C. A., and McCartney, C. R. (2001). Hypothalamic dysfunction. *Mol. Cell. Endocrinol.* **183**, 29–32.
- Mendes, M. C., Ferriani, R. A., Sala, M. M., Moura, M. D., and de Sa, M. F. (1999). Induction of ovulation with clomiphene citrate with metaclopramide in patients with amenorrhea of hypothalamic origin. *Gynecol. Endocrinol.* **13**, 149–154.
- Schachter, M., Shoham, Z. (1994) Amenorrhea during the reproductive years—Is it safe? *Fertil. Steril.* **62**, pp. 1–16.



**Figure 1** Normal hypothalamus. (A) The third ventricle is lined by the lateral hypothalamus that harbors the paraventricular nuclei. Inferiorly, the infundibulum carries nerve fibers to the pituitary stalk. (B) The nuclei of the hypothalamus are collections of ganglion cells; in this supraoptic nucleus, the neurons contain vasopressin immunoreactivity.

adult human they are ill defined. From animal studies, it is known that individual nuclei have important physiological functions; for example, there are specific nuclei implicated in hunger or satiety control, temperature regulation, olfaction, circadian rhythms, sexual drive, ovarian regulation, and parenting behaviors. However, functionally, a given hormone is often produced in more than one nucleus, and in many instances a single nucleus produces more than one hormone. These data raise doubts about the concept of individual nuclei as designated functional entities.

### CLINICAL MANIFESTATIONS OF HYPOTHALAMIC DISEASE

Lesions of the hypothalamus cause headache, nausea, vomiting, somnolence, behavioral alterations, psychosis, and dementia. Hypothalamic destruction can result in bulimia or anorexia. Visual disturbance can result from oculomotor alterations or optic nerve damage. Hypopituitarism and diabetes insipidus are common manifestations. In severe cases, patients can develop hydrocephalus. Inflammation can result in meningitis.

Diabetes insipidus is the most common and often the initial manifestation. The presentation of hypopituitarism varies with age. In children, hypothalamic dysfunction may present with dwarfism. In adults, sexual dysfunction is the most common endocrine

complaint, with impotence in males and primary or secondary amenorrhea in females. If the disease is predominantly hypothalamic or causes interruption of the pituitary stalk, pituitary hypofunction is associated with hyperprolactinemia due to destruction of the dopaminergic neurons that maintain tonic inhibition of that pituitary hormone, and stimulation confirms an intact pituitary response. If the lesion causes destruction of hypophysial tissue as well as hypothalamic disease, there is a reduction of all basal pituitary hormones or more subtle changes with reduced response to stimulation.

Some tumors are associated with pituitary hormone excess. Occasionally, this is due to the production of hormones stimulating pituitary function, such as in hypothalamic gangliocytomas that secrete adenyhypophysiotrophic hormones. Alternatively, the clinical manifestations may be due to the production of substances that simulate or mimic pituitary hormones. For example, germ cell tumors associated with precocious puberty produce  $\beta$ -chorionic gonadotropin. Occasionally, patients manifest excessive secretion of vasopressin, resulting in the syndrome of inappropriate antidiuretic hormone.

### DEVELOPMENTAL DISORDERS

The failure of development of areas of the hypothalamus can lead to variable clinical manifestations. Major defects are incompatible with life.

Septo-optic dysplasia or de Morsier's syndrome is a complex developmental disorder with variable manifestations of aplasia of the septum pellucidum, hypoplasia of the optic nerves, and endocrine dysfunction as well as vegetative alterations. The disorder results from mutations of a homeobox transcription factor, *Hesx1*, which is required for normal development of the affected regions of the central brain.

Kallman's syndrome is an X-linked developmental defect due to mutation of the *KAL-1* gene that encodes an extracellular glycoprotein, anosmin-1, that is expressed during the period of human organogenesis in the early olfactory system and is required for normal migration of gonadotropin-releasing hormone (GnRH)-containing neurons. This disorder results in gonadal insufficiency in males due to hypothalamic GnRH deficiency associated with anosmia.

Lawrence–Moon–Biedl syndrome is a highly polymorphic disorder that includes pigmentary retinopathy, mental retardation, spinal paraplegia and hypogonadism associated with obesity, and digital anomalies. The genetic basis is unknown and the variable clinical manifestations suggest incomplete penetrance with differential expression of the anomaly.

Prader–Willi syndrome is characterized by obesity, short stature, delayed puberty, infertility, mental retardation, muscle hypotonia, hypopigmentation, and seizure disorder. It has been linked with deletions of chromosome 15q11–q13. This region contains an imprinting center that regulates paternally expressed genes whose expression is lost; some cases are attributed to maternal uniparental disomy. However, the causative gene is unknown.

Wolfram's syndrome (WS) is characterized by diabetes insipidus, diabetes mellitus, optic atrophy, and deafness (hence the acronym DIDMOAD). This rare neurodegenerative disorder exhibits autosomal recessive inheritance due to compound heterozygous mutations of the *WFS1* gene, a member of a novel gene family that encodes wolframin, an endoglycosidase H-sensitive membrane glycoprotein that localizes in the endoplasmic reticulum. *WFS1* is expressed predominantly in selected neurons in the hippocampus, amygdala, olfactory tubercle, and superficial layer of the allocortex—components of the limbic system or structures closely associated with this system that account for the psychiatric, behavioral, and emotional abnormalities of this syndrome. Heterozygous mutations are associated with nonsyndromic, low-frequency sensorineural hearing loss affecting only the 2000-Hz and lower range. This is an unusual disorder that worsens over time without progressing to profound deafness.

Idiopathic growth hormone (GH) insufficiency is likely due to a defect in the synthesis or secretion of growth hormone-releasing hormone (GHRH) since most patients have normal pituitary GH somatotroph structure and hormone content, and they respond to GHRH administration.

## HYPOTHALAMIC INFLAMMATION

### Infectious Lesions

Acute and chronic infections of the hypothalamus are rare but they do occur, usually in association with sphenoid sinus infection, cavernous sinus thrombosis, otitis media mastoiditis, or peritonsillar abscess. Pituitary tumors have been associated with the development of pituitary abscess that can spread to the hypothalamus. It has been suggested that bony erosion by the tumor predisposes such patients to the spread of sinonasal infection. Rarely, infection results from vascular seeding of distant or systemic infection.

### Noninfectious Inflammatory Lesions

Sarcoidosis is a multisystem granulomatous disease of unknown etiology. It has long been attributed to an infectious agent; however, none have been identified. Disease onset usually occurs in adults and there is a predilection for blacks and females. There are usually systemic manifestations, and neural involvement is rare; however, it can occur in the hypothalamus, usually involving the meninges at the infundibulum and floor of the third ventricle. The granulomatous inflammation has a subacute or protracted course of tissue destruction that may respond to steroid suppression, and there are rare reports of spontaneous resolution.

Neuroinfundibulohypophysitis is a rare inflammatory condition that affects the infundibulum, the pituitary stalk, and the neurohypophysis and may be part of a range of autoimmune disorders, including lymphocytic hypophysitis. Lymphocytic hypophysitis occurs mainly in women and most often presents in the later stages of pregnancy. Infundibulohypophysitis shows no sexual predilection and usually presents with diabetes insipidus. The cause is unclear.

A lesion similar to orbital pseudotumor characterized by chronic inflammation and fibrosis has been reported to involve the parasellar tissues associated with other sclerosing lesions, such as Riedel's thyroiditis, retroperitoneal fibrosis, and sclerosing cholangitis. The etiology and appropriate management of these disorders are uncertain.

## METABOLIC LESIONS

### Neurodegenerative Processes

Alzheimer's disease, Parkinson's disease, Huntington's chorea, and other neurodegenerative diseases can involve the hypothalamus, resulting in variable endocrine, behavioral, and vegetative abnormalities.

### Systemic Processes

Wernicke's encephalopathy can alter hypothalamic function, but usually the manifestations are due to mamillary body degeneration. Hemochromatosis results in hypogonadotropic hypogonadism, but this is thought to be due to iron deposition in the pituitary gonadotrophs rather than a primary hypothalamic lesion.

## CYSTIC LESIONS

### Rathke's Cleft Cysts

These cysts originate in the remnants of Rathke's pouch of the pituitary. Rathke's cleft arises from the oropharynx and migrates upward with an anterior and posterior limb that ultimately give rise to the anterior and intermediate lobes of the adenohypophysis, respectively. In the human, the intermediate lobe is vestigial, and its remnants line small cystic cavities that are remnants of the cleft and usually <5 mm in diameter. When they enlarge and become detectable, they can cause symptoms. Initially the symptoms are hypopituitarism and diabetes insipidus, due to sellar compression. When they develop suprasellar extension, symptoms of hypothalamic involvement ensue. There are rare cases of purely suprasellar Rathke's cleft cysts, and occasionally abscess formation develops within these lesions.

They usually present in adults but do occur in infants and young children. Computed tomography (CT) scans reveal low-density cystic areas with capsular enhancement in most cases; the magnetic resonance imaging (MRI) appearance is variable. Morphologic examination reveals a cyst lining characterized by ciliated cuboidal or columnar epithelium resembling respiratory epithelium, with occasional goblet cells and squamous elements. The degree of ciliation and the propensity for squamous metaplasia distinguish these cysts from the neuroepithelial-derived colloid cysts of the third ventricle.

The management of these lesions involves surgical drainage with or without partial excision. The

recurrence rate is low. Most symptoms and signs are relieved postoperatively, but permanent hypopituitarism and diabetes insipidus require hormone replacement therapy.

### Arachnoid Cysts

These lesions may be congenital anomalies or acquired cysts in the arachnoid of the sellar and suprasellar region. The cystic nature of these lesions evident on CT or MRI scans may make it difficult to distinguish them from other cysts that occur in this area. These cysts are filled with clear, colorless fluid and are lined by arachnoid laminar connective tissue with a single layer of flattened epithelium. These lesions are also managed by drainage, with partial cyst wall excision.

### Dermoid and Epidermoid Cysts

Dermoid and epidermoid cysts arise from epithelial cells that are misplaced during embryologic development or, rarely, traumatically. Epidermoid cysts are also known as cholesteatoma. These epithelial cysts can occur intracranially, most commonly at the cerebellopontine angle but often in the suprasellar area.

The cystic nature of these lesions evident on CT or MRI scans may make it difficult to distinguish them from other cystic lesions in this area. Epidermoid cysts are lined by keratinizing squamous epithelium; dermoid cysts are distinguished by the additional presence of skin appendages, including hair follicles and sweat glands.

Management usually involves surgical resection. Complications include rupture with chemical meningitis due to keratin debris or the development of squamous carcinoma.

### Aneurysms

Aneurysms of the carotid arteries can expand to give rise to masses in the suprasellar region that compress hypothalamic structures and result in endocrine and other hypothalamic manifestations. It is important to distinguish these from other cystic lesions prior to surgery.

### Meningoencephalocele

Encephaloceles involving the hypothalamus have been reported to cause symptoms of a mass lesion with variable endocrine manifestations.

## PRIMARY HYPOTHALAMIC NEOPLASMS

### Craniopharyngioma

Craniopharyngiomas are thought to derive from the remnants of Rathke's pouch. They represent up to 13% of intracranial neoplasms and are the most common sellar tumor of childhood. They can occur at any age from infancy to old age, but the peak incidence is from 5 to 20 years. A second small peak occurs in the sixth decade. In some series, males are more often affected than females.

The majority of tumors occur in the suprasellar region; 15% have an intrasellar component. Craniopharyngiomas are usually cystic or may be cystic and solid, and approximately half have radiologically detectable calcification. Although they may be as small as 1 cm, the majority are much larger at the time of diagnosis. Grossly, they are well-circumscribed tumors but there may be little or no capsule at the interface with brain parenchyma. The lesions usually contain a thick oil-like fluid that is described as "black sludge." Cholesterol crystals and calcification may be seen grossly. Rarely, these tumors may contain bone and/or teeth. Microscopically, they are composed of cords or islands of epithelial cells in a loose fibrous stroma and with intervening cysts (Fig. 2). The epithelium usually has an outer palisaded layer, a midzone of stellate epithelial cells, and a superficial keratinizing layer. Masses of keratin often form the nidus for calcification. Occasionally, there is an inflammatory component. Microscopically, the borders are frequently irregular and may be associated with gliosis in the adjacent brain tissue.

Two subtypes of craniopharyngioma have been recognized; however, there is frequent overlap. The adamantinomatous type has a predominance of stellate components, which yields a pattern resembling the dental ameloblastic organ and is similar to that seen in adamantinomas. The papillary variant is less common, rare in children, and believed to have a better prognosis. It is characterized by solid or cystic growth of pseudopapillary squamous epithelium and lacks the palisading, fibrosis, and cholesterol accumulation that characterizes the typical craniopharyngioma.

Craniopharyngiomas are extremely infiltrative. They may cause extensive tissue damage, extending into the hypothalamus and as high as the third ventricle; obstruction of the ventricle may result in hydrocephalus. Rarely, craniopharyngiomas may



**Figure 2** Craniopharyngiomas are composed of squamoid epithelium that derives from remnants of Rathke's pouch. The infiltrative lesions destroy hypothalamic parenchyma.

spontaneously rupture or form abscesses. Because of their highly infiltrative nature, they are often incompletely excised surgically. There is a 10–40% recurrence rate, particularly in younger patients. Many surgeons advocate postoperative radiation. There is a single report of malignant transformation of a craniopharyngioma. This occurred on the fifth recurrence after a 35-year history of disease and 8 years following radiotherapy.

### Neuronal Tumors

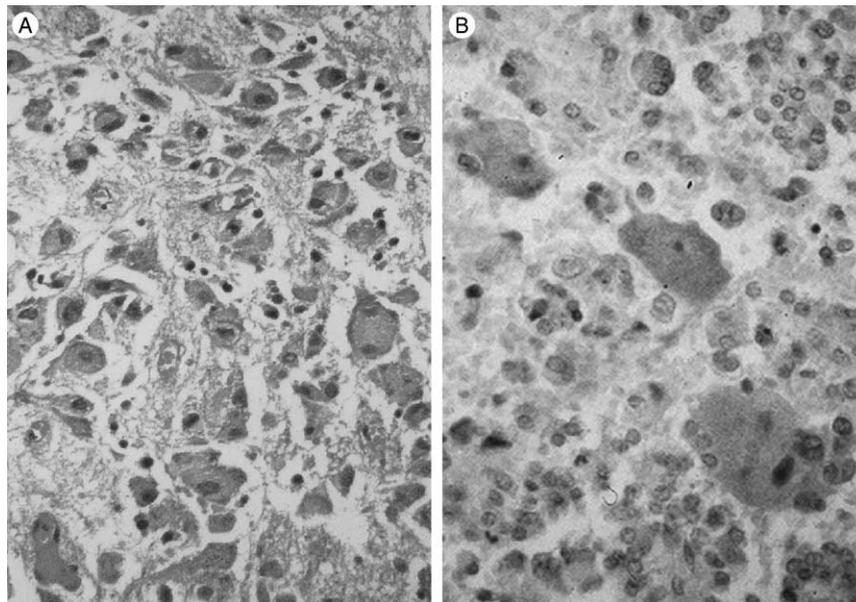
Hypothalamic neuronal tumors are very rare neoplasms that have been reported in the literature as "gangliocytomas" or "ganglioneuromas." These lesions are composed of mature neurons resembling hypothalamic ganglion cells and are capable of producing hypothalamic peptides (Fig. 3). Some researchers prefer to consider these hamartomas because they occur in children and it is thought that they represent developmental anomalies. They are commonly called

ganglioneuromas or gangliocytomas because there is evidence that they represent neoplasms that have their onset in adulthood and are capable of continued growth. Such tumors have also been described in other parts of the brain as extrahypothalamic gangliocytomas. The term choristoma has been applied to describe collections of hypothalamic neurons within the adenohypophysis.

In addition to the usual symptoms of a hypothalamic mass lesion, some of these tumors are hormonally active and cause endocrinopathies that are usually mediated by the pituitary. They have been associated with acromegaly, precocious puberty, Cushing's disease, and amenorrhea–galactorrhea. The most common localization of these peculiar neoplasms is in the hypothalamus or tuber cinereum, with variable involvement of the third ventricle. They may be nodular, pedunculated, sessile, solid, or cystic. They may be within, attached to, or completely detached from the hypothalamus. These tumors may also be found within the parenchyma of the adenohypophysis. They vary from microscopic lesions composed of only a few cells and measuring a few millimeters to masses that measure more than 5 cm. In particular, hypothalamic hamartomas associated with acromegaly often present as a solitary intrasellar mass that mimics a simple pituitary GH-producing pituitary adenoma; microscopic examination reveals a gangliocytoma associated with a pituitary adenoma in such cases. They are composed

of randomly oriented large mature ganglion cells with large nuclei and very prominent nucleoli; binucleated and even multinucleated cells are seen. Most of the ganglion cells contain abundant cytoplasm, and Nissl granules are conspicuous at the periphery of the cell body. Some tumors have a glial stroma.

In patients with acromegaly, the ganglion cells may contain GHRH (Fig. 3B), glucagon, somatostatin, vasoactive intestinal peptide (VIP), corticotropin-releasing hormone (CRH), GnRH, or gastrin; associated pituitary adenomas generally contain GH but some patients have somatotroph hyperplasia and, rarely, no pituitary abnormality is identified. In patients with amenorrhea–galactorrhea, prolactin and VIP have been localized in the cytoplasm of ganglion cells. One patient had an associated lactotroph adenoma; another had a mixed GH- and prolactin-producing pituitary adenoma. Tumors associated with Cushing's disease contain CRH and may also express ACTH,  $\beta$ -lipotropin, and somatostatin; they have been associated with corticotroph hyperplasia or corticotroph adenoma. Children with precocious puberty have neuronal tumors that contain GnRH but may also contain immunoreactivity for CRH,  $\beta$ -endorphin, and oxytocin. Some gangliocytomas not associated with clinical evidence of hormone excess show multiple immunoreactivities, including positivity for VIP, galanin,  $\alpha$ -subunit of glycoprotein hormones, somatostatin, and serotonin. In a few cases, there is



**Figure 3** Hypothalamic gangliocytoma. (A) The tumor is composed of ganglion cells resembling those of the normal nuclei of the hypothalamus but lacking organization and containing abnormal binucleate neurons. (B) The tumor cells of this gangliocytoma associated with acromegaly contain GHRH immunoreactivity.

ultrastructural evidence of an intimate association between neurons and adenohipophysial cells, both adenomatous and nontumorous, supporting the theory of an endocrine relationship.

The pathogenesis of these lesions is not known. Their association with pituitary adenomas in some patients suggests that there may be a common causative mechanism. It has been suggested that sellar gangliocytomas may arise by abnormal differentiation during neoplastic proliferation of adenohipophysial cells. However, it is more likely that these tumors have a true hypothalamic origin. It may be that the same tumorigenic stimulus results in tumors of both hypothalamus and pituitary.

These tumors are diagnosed at the time of surgery and cure has resulted from complete surgical resection in several published cases. However, despite the apparent cure of the endocrine disturbance, there may be radiographic evidence of residual tumor. When surgery cannot accomplish total resection of the lesion, the prognosis is varied. Most of these neuronal tumors have an extremely low proliferative potential and the overall prognosis is good. A few have caused severe hypothalamic destruction and have resulted in the demise of the patient. Radiotherapy appears to ameliorate hormone excess in these patients, but it does not appear to be effective in preventing regrowth of the neuronal tumor.

## Gliomas

Gliomas are neoplasms derived from and composed of neuroglial cells. They include astrocytomas and glioblastoma multiforme, oligodendrogliomas, and ependymomas. These tumors show a wide range of biologic behaviors.

Aggressive, rapidly lethal malignant gliomas or glioblastoma multiforme in the hypothalamus occur following radiotherapy for pituitary adenoma, craniopharyngioma, or suprasellar germinoma. These tumors can occur 5–25 years after conventional radiation doses of 42.5–66 Gy. Glioma of the optic nerve is rare and usually occurs in children or adolescents who are predisposed due to neurofibromatosis or Beckwith–Wiedemann syndrome.

The radiologic features vary with the tumor type. On CT scan, low-grade astrocytomas are low-density lesions that do not enhance with contrast; enhancement suggests malignant transformation, particularly if it is peripheral to a central hypodense area of necrosis, as seen in glioblastoma multiforme. On MRI, astrocytomas usually have high T2-weighted signal intensity.

The most common glioma in the sellar region is the pilocytic astrocytoma. Most common in children, hence the term juvenile type, this is a relatively discrete, often cystic mass with prominent enhancement on CT scan. These tumors usually arise in the wall of the third ventricle, where they displace and compress the optic chiasm; they may also arise in the optic nerve, usually at the chiasm. The low-grade lesions in children have a relatively good prognosis because they grow slowly, but chiasmal lesions can be more aggressive. In contrast, optic gliomas that occur sporadically in adults are usually rapidly fatal tumors.

## Meningiomas

Meningiomas are tumors derived from the meninges and their derivatives in the meningeal spaces. They may arise from dural fibroblasts or pial cells, but the most common are of arachnoid origin. Approximately 20% of tumors of arachnoid and meningotheial cells occur in the sellar and parasellar area; they are most common in the sphenoid ridge and tuberculum sellae but also occur in the clivus. Occasionally, meningiomas have arisen in the sellar region following irradiation of the area for pituitary adenoma or craniopharyngioma. These tumors are more common in women than in men, possibly due to their expression of progesterone receptors and estrogen receptors.

Meningiomas are histologically diverse tumors. The more familiar types and the most common in the suprasellar hypothalamic area are the meningotheial, fibrous or fibroblastic, and transitional variants. Papillary patterns and anaplastic changes portend a more aggressive behavior and the possibility of metastases.

The management of these lesions is generally surgical resection. These highly vascular lesions can create problems with intraoperative bleeding.

## Chordomas

Chordomas are rare lesions thought to derive from remnants of the notochord. They occur in the midline, most often in the sacral region, but also in the region of the clivus and occasionally in vertebrae and within the sella turcica. These slowly growing but locally aggressive neoplasms usually occur in patients older than 30 years of age, but patients with the unusual sphenoid lesions tend to be younger.

Chordomas are usually lobulated, calcified, and expansile osteolytic lesions that may cause characteristic elevation of the periosteum, in which case they may be suspected on the basis of the radiologic

findings. They have a characteristic gelatinous gross appearance and may have areas of calcification. Histologically, they are composed of large polyhedral cells with a characteristic bubbly or “physaliphorous” vacuolation. The tumor cells have immunohistochemical markers of epithelial differentiation, keratins and epithelial membrane antigen (EMA), as well as S100 protein and vimentin. Some exhibit positivity for carcinoembryonic antigen. High-grade sarcomatous areas may herald dedifferentiation and are sometimes seen in metastatic deposits. These aggressive variants are rare in intracranial chordomas. In contrast, the sphenoid-occipital region is the preferred site of the “chondroid” variant, which exhibits areas of cartilaginous differentiation that may dominate the histologic picture. These chondroid chordomas lack keratin and EMA positivity and are considered by some to be chondrosarcomas. This variant has a better prognosis than the usual clival chordomas.

Surgery is the preferred initial therapeutic approach; however, the location and extent of tumor may make complete extirpation impossible. Radiotherapy is indicated for incompletely resected lesions. Mean survival is approximately 4 or 5 years, and metastases to lung, liver, bone, and lymph nodes occur rarely.

### Schwannomas

Tumors composed of spindle-shaped Schwann cells derive from cranial nerves; they are also known as neurilemmoma or neurinoma. Hypothalamic and parasellar schwannomas occur rarely. On CT scan, these lesions are visualized as sellar or parasellar masses that enhance with contrast and can mimic pituitary adenoma. They are hypovascular on angiography.

The morphology of these lesions in the pituitary region is identical to that of schwannomas elsewhere. These tumors are usually encapsulated and sometimes cystic lesions composed of spindle-shaped cells arranged in compact Antoni type A and loose Antoni type B areas. The former may exhibit palisading that leads to the formation of Verocay bodies. Immunohistochemical staining for S100 protein is strong in these lesions. Using electron microscopy, these tumors are recognized by the prominent basal lamina surrounding individual cells and by the characteristic “long-spacing” collagen of the stroma.

Similar to Schwannomas elsewhere, these lesions are usually benign and are amenable to surgical resection unless they involve critical structures in the parasellar region, in which case conservative surgical resection is indicated.

### Germ Cell Tumors

These tumors are derived from germ cells that are residual along the midline; they are identical to germ cell tumors of the gonads and mediastinum. Intracranial germ cell tumors represent less than 1% of intracranial neoplasms, but in children they constitute up to 6.5% of such lesions. After the pineal, the suprasellar region is the second most common site of involvement. These tumors occur most often before the age of 20 years and occur more often in males than in females.

The most common type, the germinoma, is usually a well-demarcated tumor that has high density on CT scan and enhances with contrast. Teratomas may exhibit fat densities and calcifications that are recognized radiologically. Whereas the CT appearance of some of the other germ cell tumors may not be distinctive, MRI is more sensitive and can show features that are obscured on CT scan.

The classification of parasellar and intrasellar germ cell tumors includes teratomas, germinomas, embryonal carcinomas, endodermal sinus tumors, and choriocarcinomas. Germinomas and teratomas predominate, and mixed germ cell tumors that show features of more than one tumor subtype are frequent, usually germinoma combined with one of the other tumor types or a combination of embryonal carcinoma and immature teratoma, known as teratocarcinoma. Recognition of even minor components that are characterized by more aggressive elements can predict a worse prognosis and may alter management.

If the diagnosis is suspected preoperatively, biopsy can be performed to type the lesion and determine whether surgery or other modalities, such as radiation or chemotherapy, are indicated. Germinomas are uniquely radiosensitive and long-term remission is achieved in approximately 70% of patients. The recognition of more aggressive elements in mixed tumors can predict failure of radiotherapy. Other tumors are more aggressive, and despite surgery, radiation, and chemotherapy, they may recur or metastasize both within and beyond the central nervous system. Seeding via ventriculoperitoneal shunt may also result in dissemination.

### Vascular and Mesenchymal Tumors

These tumors arise from vessel walls, fibrous connective tissue and fat, bone, and cartilage. They are rare in this area and can be benign or malignant.

The development of sarcoma in this region may be sporadic but more commonly is the result of previous

ionizing irradiation. There are several reports in the literature of osteosarcoma and fibrosarcoma of the sella developing 4–21 years after irradiation for pituitary adenoma or craniopharyngioma. These aggressive neoplasms result in the rapid demise of the patient.

## SECONDARY NEOPLASMS

### Pituitary Adenomas

Pituitary adenomas are common lesions, occurring in 20–27% of the population, and at least one-third of these give rise to clinical manifestations. Many of these tumors exhibit suprasellar or parasellar invasion involving the hypothalamus.

### Metastatic Carcinoma

Cranial metastasis is a manifestation of a systemic and usually terminal malignancy (Fig. 4). The most common primary sites of these lesions are the lung, breast, and gastrointestinal tract. Most hypothalamic metastases are not associated with specific clinical symptomatology and are therefore found incidentally at autopsy. Patients may present with diabetes insipidus, and large tumors may cause headache, visual field defects, ophthalmoplegia, and ptosis. Rarely is there an endocrine manifestation, usually resulting in hypopituitarism. A few reported cases have identified hormone excess, and usually the culprit is ectopic ACTH production in a location that mimics primary pituitary Cushing's disease. A fascinating case of CRH production by a hypothalamic metastasis of prostate small cell carcinoma resulted in pituitary corticotroph hyperplasia.

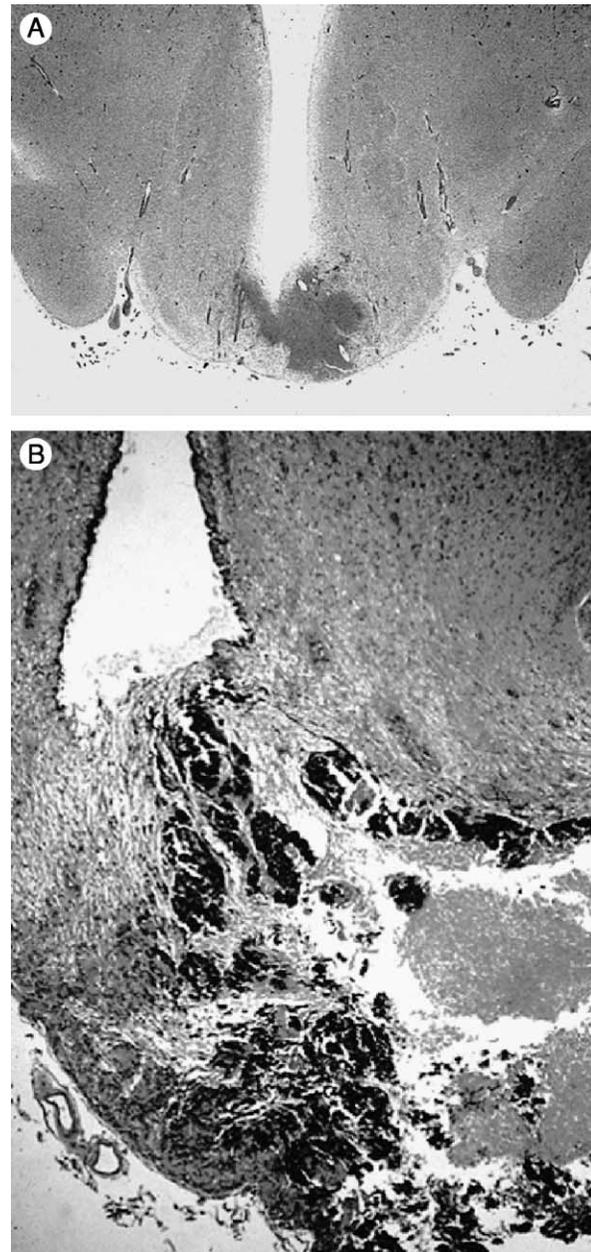
Since these patients have disseminated malignancy, therapy is aimed at palliation. Surgical decompression with or without radiotherapy can relieve symptoms.

### Lymphoma and Leukemia

Tumors of leukocytic, lymphocytic, or plasmacytic differentiation are usually systemic disorders. Very rarely, lesions in the region of the hypothalamus appear to be solitary and primary at that site. As in the central nervous system in general, lymphoma or leukemia are most often meningeal and extradural.

Lymphomas are seen on MRI as masses with signals of similar intensity to those of their surroundings on T1-weighted images, but they are hyperintense on T2-weighted views. They enhance with contrast media and usually in a uniform pattern.

Histologically, the lymphomas of the central nervous system resemble systemic lymphomas. They are almost always non-Hodgkin lymphomas with a diffuse rather than follicular pattern; the majority are B cell tumors. Plasmacytomas are composed of well-differentiated plasma cells.



**Figure 4** Metastatic tumor. (A) In this case of Langerhans cell histiocytosis, the lesion infiltrates and destroys normal hypothalamus. (B) This metastatic small cell lung carcinoma has replaced the arcuate nucleus and invades the infundibulum and superior pituitary stalk.

## Langerhans Cell Histiocytosis

The disorder formerly known as histiocytosis X, encompassing eosinophilic granuloma, Hand–Schüller–Christian disease, Letterer–Siwe disease, Langerhans cell granuloma, and several other eponymic variants, is now classified as Langerhans cell histiocytosis. This localized, multifocal, or disseminated proliferation of epithelioid histiocyte-like Langerhans cells is thought to be a reactive rather than neoplastic process with an immunologic aberration underlying its etiology. The classical presentation of Hand–Schüller–Christian disease involves the hypothalamus, and the disorder may involve the pituitary gland.

CT scans of Langerhans cell histiocytosis are characterized by ill-defined contrast-enhancing hypodense masses with areas of edema. MRI can detect the small multifocal lesions of this disorder more readily than CT scan; these lesions may be hypointense but can have slightly increased T1 contrast and intense T2-weighted images. Involvement of the pituitary stalk is characteristic of this disorder and may precede clinical manifestations of the disease.

Langerhans cells are characterized by an epithelioid, histiocyte-like appearance. They have kidney-shaped nuclei and abundant cytoplasm and stain for CD1 and S100 proteins. They are admixed with chronic inflammatory cells, foamy macrophages, and eosinophils. By electron microscopy, they contain pathognomonic Birbeck granules that allow a definitive diagnosis.

The prognosis of this disorder is variable. When systemic involvement is manifest, it is frequently lethal and apparently isolated lesions can progress rapidly to widely disseminated disease that is unresponsive to any form of therapeutic intervention. Surgery has been reported to be curative of localized lesions, and radiotherapy has been used postoperatively with success in cases of isolated involvement of this region.

## See Also the Following Articles

Craniopharyngiomas • Diabetes Insipidus, Nephrogenic • Diabetes Insipidus, Neurogenic • Hypothalamic Regulation of Appetite and Obesity • Hypothalamic-Pituitary Unit • Hypothalamus, Anatomy of • Kallmann's Syndrome and Idiopathic Hypogonadotropic Hypogonadism • Pituitary Tumors, Molecular Pathogenesis • Prader-Willi Syndrome

## Further Reading

- Asa, S. L. (1998). "Tumors of the Pituitary Gland. Atlas of Tumor Pathology," Third Series, Fascicle 22. Armed Forces Institute of Pathology, Washington, DC.
- Braunstein, G. D. (2001). Hypothalamic syndromes. In "Endocrinology" (L. J. DeGroot and J. L. Jameson, eds.), 4th ed., pp. 269–281. Saunders, Philadelphia.
- Puchner, M. J. A., Lüdecke, D. K., Saeger, W., Riedel, M., and Asa, S. L. (1995). Gangliocytomas of the sellar region—A review. *Exp. Clin. Endocrinol.* **103**, 129–149.
- Scheithauer, B. W. (1998). The hypothalamus and neurohypophysis. In "Functional Endocrine Pathology" (K. Kovacs and S. L. Asa, eds.), 2nd ed., pp. 171–246. Blackwell, Boston.

**Table 1** Causes of Hypothalamic Hypogonadism

Idiopathic
Idiopathic hypogonadotropic hypogonadism
Kallmann syndrome
Adult-onset idiopathic hypogonadotropic hypogonadism
Fertile eunuch syndrome
Adrenal hypoplasia congenita
Delayed puberty
Functional
Stress
Exercise
Dieting
Hyperprolactinemia
Organic
Tumors
Infiltrative disorders
Head trauma
Radiation therapy
Miscellaneous
Prader–Willi syndrome
Laurence Moon syndrome
Moebius syndrome
Gordon Holmes spinocerebellar ataxia

increases a genetic transmission, and X-linked and autosomal KS pedigrees have been reported. The gene implicated in Kallmann syndrome, *Kal*, is located at Xp22.3. The protein encoded by the *Kal* gene, anosmin, shows structural homologies to proteins involved in the neural development. Anosmin is thought to function as a substrate adhesion molecule necessary for establishing normal connections between olfactory axons and the mitral cells of the olfactory bulb. Xp22.3 deletions and a number of different *Kal* point mutations have been described. The role of these mutations is unclear, especially when considering discordances in the phenotype of brothers bearing the same mutation.

In X-linked Kallmann syndrome, there is a well-demonstrated link between GnRH deficiency and the dysgenesis of the olfactory structure that can be anatomically evidenced by nuclear magnetic resonance in most patients. *Kal* mutations have not been found in normosomic patients; however, it is still unclear whether Kallmann syndrome and HH with normosomia are distinct entities. In addition to anosmia, a variety of abnormalities have been described in Kallmann syndrome. These include synkinesias, renal agenesis, sensorineural defects not observed in normosomic patients, and rarer defects such as cryptorchidism and eye movement disorders. Craniofacial palate defects can be present in both situations.

Although X-linked Kallmann syndrome is the best characterized form, autosomal mutations in genes not yet identified have been found to account for the majority of cases of this condition. In 106 patients studied at Massachusetts General Hospital, only 11% of the familial cases were X linked and less than 5% of the sporadic cases were found to bear mutations in the *Kal* gene. The most obvious candidate gene for GnRH deficiency is the GnRH gene; however, to date, alterations at 8p21–p12, where this gene is located, have not been demonstrated.

Neonatal adrenal insufficiency, caused by *DAX1* missense or nonsense mutations, is also associated with idiopathic hypogonadotropic hypogonadism (IHH). Because earlier diagnosis and adequate treatment of the adrenal insufficiency have allowed these patients to reach adult age, secondary hypogonadism has become a component of this syndrome.

### Late Onset

An acquired form of isolated GnRH deficiency, defined as adult-onset IHH, has been described. These patients have a normal pubertal development followed by a clinical picture of hypogonadism. The endocrine profile is indistinguishable from that of patients with congenital GnRH deficiency, being characterized by low circulating levels of testosterone and an absence of the normal pulsatile pattern of gonadotropin. Most of these patients normalize their pituitary gonadal axis during pulsatile GnRH administration. This form differs from hypothalamic amenorrhea in that it lacks any precipitating factor such as exercise, stress, or weight loss and also is irreversible. However, in a small subset of patients, the gonadal function remains normal even after the cessation of GnRH therapy. These patients may represent an incomplete form of GnRH deficiency that is relieved by the priming effect of treatment or, alternatively, by forms of very delayed puberty.

The fertile eunuch syndrome is also thought to represent an incomplete form of GnRH deficiency. The pathophysiology of this disorder may be due to a secretion of endogenous GnRH sufficient to support spermatogenesis and testicular growth but not testosterone levels adequate for a normal virilization. Delayed puberty may also be a mild form of partial GnRH deficiency.

### Functional

In these circumstances, the failure of GnRH to be released is transient and frequently leads to so-called

hypothalamic amenorrhea. A syndrome of functional GnRH deficiency is less recognizable in men mostly due to the lack of a clear clinical marker. There are some data to suggest that caloric restriction or strenuous physical exercise lowers testosterone levels.

### **Hypothalamic Amenorrhea**

Idiopathic amenorrhea is a diagnosis of exclusion and generally refers to the cessation of menstrual cycles in young women without demonstrable organic abnormalities of the pituitary gonadal axis. Stressful factors are associated with the onset of hypothalamic amenorrhea such as emotional events, exercise, and weight loss. From a hormonal point of view, hypothalamic amenorrhea is characterized by hypoestrogenism with low-normal gonadotropin and normal prolactin (PRL) levels. The frequency and amplitude of the gonadotropin pulses are diminished, and in some cases a pubertal pattern with amplification of pulsatility of luteinizing hormone (LH) during sleep can be demonstrated. The pattern of ovarian steroid secretion varies in relation to the clinical situation. In severe cases, when only minimal changes in gonadotropin levels can be detected, hormonal ovarian activity is virtually absent. When gonadotropin pulses are still detectable, albeit diminished in number and amplitude, estrogens and androgens are measurable and spontaneous, although erratic menses may occur. The initial phases of hypothalamic amenorrhea are often characterized by an inadequate luteal phase.

There is increasing evidence that idiopathic amenorrhea can be the result of endocrine and metabolic responses to a restriction in food intake. Evidence of subclinical eating disorders has been shown in hypothalamic amenorrhea, particularly as reduction in fat intake. This circumstance also occurs in women with stable weight who are nonathletic, harboring levels of emotional distress or depression not different from those of menstruating controls. It is currently thought that stress may contribute to a disordered eating behavior but that it is not directly responsible for the menstrual alterations. Actually, weight loss observed in patients with eating disorders, particularly in those with anorexia nervosa, is the most obvious cause of hypothalamic amenorrhea, and a similar condition is also observed in women with low body weight and reduced fat stores such as female athletes.

The disruption of the physiological pulsatile secretion of GnRH neurons is probably due to changes in the activity of neuroendocrine circuits involving neuropeptides such as neuropeptide Y (NPY), opioids, corticotropin-releasing factor (CRF), and neurotransmitters (e.g., dopamine, noradrenaline,

$\gamma$ -aminobutyric acid (GABA), serotonin). However, reliable evaluations of these factors cannot be made, so their role in the pathophysiology of hypothalamic amenorrhea is undefined.

Recently, the role of leptin, a peptidic hormone produced by adipocytes, has gained relevant importance in the regulation of the reproductive axis. Leptin is an indicator of nutritional status; thus, it may trigger changes in GnRH pulsatility in response to reduced fat stores. Whether the subtle alterations of adrenal function and the increased levels of  $\beta$ -endorphins found in patients with hypothalamic amenorrhea are of clinical relevance is still to be established.

### **Pathological Hyperprolactinemia**

Patients with macroprolactinoma (i.e., adenoma diameter > 1 cm) suffer from a disorder of gonadal function due to both tumor mass effect and hyperprolactinemia, but in microadenoma patients or in drug-induced hyperprolactinemia, hypogonadism recognizes a functional origin due to neuroendocrine effects of elevated PRL levels that impairs pulsatile release of GnRH and ultimately that of gonadotropin. According to the hypothesis of Scanlon, it is possible that hyperprolactinemia enhances dopaminergic tone at the hypothalamic levels with a consequent inhibitory effect on gonadotropin release.

### **Organic**

Structural lesions of the hypothalamus can interfere with the neuroendocrine regulation of GnRH. In these conditions, the impairment of gonadal function is almost invariably associated with deficiency of other pituitary functions. In children, the most common form of hypothalamic hypogonadism is craniopharyngioma presenting with visual field defects, growth failure, and/or diabetes insipidus. Other common hypothalamic tumors such as meningiomas, dysgerminoma, and gliomas are accompanied by hypogonadism. Secondary hypothalamic localizations of breast tumors in women and lung tumors in men may also determine secondary gonadal impairment generally associated with diabetes insipidus.

Infiltrative disorders of the hypothalamus, such as sarcoidosis, histiocytosis, and hemochromatosis, may also cause hypothalamic hypogonadism as well as radiotherapy for central nervous system (CNS) tumors or leukemia.

### **Miscellaneous Disorders**

Secondary hypogonadism with the characteristics of HH is found in genetic conditions such as

Prader-Willi syndrome, Bardet-Biedl syndrome, and Moebius syndrome (i.e., the association of congenital ophthalmoplegia and facial paresis). In a single patient with this latter syndrome, the association with hypogonadism and anosmia has been reported. HH resistant to treatment with GnRH has been reported in a patient with Gordon Holmes spinocerebellar ataxia.

## DIAGNOSIS

As a form of secondary hypogonadism, hypothalamic hypogonadism is characterized by signs and symptoms of gonadal failure, low levels of gonadal hormones, and low levels of gonadotropin. This latter finding is not absolute given that a single evaluation of follicle-stimulating hormone (FSH) and LH, particularly in the acquired forms, may display levels in the normal range. Actually, the marker of the hypothalamic disturbance is represented by alterations in gonadotropin pulsatility.

In general, the most severe forms of congenital hypogonadism in patients who show no evidence of puberty are associated with no detectable pulses of gonadotropin when examined by serial samples at 10- to 20-min intervals for 12 to 24 h. Pulses of normal frequency but of reduced amplitude attributed to impaired responsiveness of the GnRH receptors may also be observed.

In patients with acquired hypogonadism, the unpulsatile pattern of gonadotropin strongly suggests an organic cause or is associated with severe nutritional disturbances as in anorexia nervosa. In patients with the functional form, there is generally only a decrease in pulse frequency with a pattern similar to that observed during the early follicular phase of the cycle.

In patients with hypogonadotropic hypogonadism, the possibility of an organic disease of the hypothalamic-pituitary region must always be considered. Nuclear magnetic resonance imaging is always indicated in male patients and in females without a clear-cut history of functional amenorrhea. A determination of basal PRL levels must always be performed together with a gonadotropin assessment. Supranormal PRL levels, in the absence of other causes of hyperprolactinemia such as primary hypothyroidism, renal or liver failure, and treatment with antidopaminergic agents, render nuclear magnetic resonance imaging mandatory. In particular, marginally elevated PRL levels (i.e., in the range of 40–120 ng/ml) should be viewed as strongly suspected for a space-occupying lesion at the hypothalamic level that impairs the dopamine delivery to the lactotrophs. PRL levels above 120 ng/ml are diagnostic for a PRL-secreting adenoma.

When basal PRL levels have been carefully evaluated by means of at least three samples taken at bed rest and during saline infusion to avoid stress-related elevation of the hormone levels, dynamic tests such as thyrotropin-releasing hormone (TRH) injection are definitely of no value. When neuroradiology has identified an organic disease, an evaluation of the thyroid, adrenal function, and growth hormone (GH) secretion should be performed.

## TREATMENT

The treatment of HH depends on the nature of the disease and on the goal to be achieved in relation to pubertal stage and desire of fertility. In prepubertal patients, gonadal hormones should be administered at low initial doses that are increased gradually to obtain the development of secondary sexual characters. In men, long-acting preparations of testosterone are effective, starting with 50 mg monthly and increased by 50 mg every 3 to 6 months. Alternatively, low doses of human chorionic gonadotropin (hCG) have the advantage of stimulating testicular growth.

In females, the therapy can be started with low doses of estrogens, for example, 5 µg of ethynyl estradiol to obtain breast development, followed by a cyclical therapy in combination with a progestin. During adulthood, different approaches should be followed depending on whether fertility is desired. If fertility is not desired, the goal of the treatment in men is to maintain virilization, normal sexual function, and adequate muscle and bone mass. Usually, 100 to 200 mg of long-acting testosterone esters are given every 2 to 4 weeks, although the dosage can be adjusted to obtain a testosterone level in the lower end of the normal range before the injection of the next dose. The disadvantage of this route of administration is an unfavorable pharmacokinetic profile often characterized by supranormal levels during the first week after the injection. Testosterone undecanoate, an orally administrable preparation, has an absorption that is influenced by diet composition.

The transdermal testosterone preparations applied to scrotal skin or, more recently, to nonscrotal skin provide an effective method of substitutive treatment with the drawbacks of high cost and frequent skin irritations. Testosterone implants are also an effective, albeit not readily available, way in which to treat hypogonadism.

In female patients, substitutive treatment is carried out with low-dose pills or with conjugated estrogens or transdermal estradiol associated with cyclic progestin to achieve endometrial protection. If fertility is desired,

the treatment has to be carried out with pulsatile GnRH, gonadotropins, or clomiphene citrate.

In men, pulsatile GnRH therapy is practically performed by subcutaneous route of injection through a portable pump, although intravenous administration produces the most physiological pattern of gonadotropin secretion. The dose per bolus is highly variable, ranging from 25 to 600 ng/kg delivered every 120 min. The dose is individually tailored to achieve normalization of testosterone and gonadotropin levels. This route of therapy is effective in obtaining spermatogenesis in the majority of patients, with treatments lasting up to 2 years or even longer in patients with very small basal testicular volume. When testicular volume reaches 8 ml, semen analyses are obtained.

Gonadotropin treatment is carried out by a combination of hCG and recombinant FSH. The starting schedule is 500 to 2000 IU of hCG two or three times weekly, followed after 6 months by FSH (75 IU three times weekly) in relation with testicular growth and induction of spermatogenesis. In patients with adult-onset HH, spermatogenesis may be induced by the use of hCG alone.

The main side effect of gonadotropin treatment is gynecomastia, which can be minimized by reducing hCG doses to keep testosterone levels at the lowest limit of the normal range.

The therapeutic approaches to achieving fertility in women are similar to those in men, but the treatment with gonadotropin in women is slightly less effective in terms of cumulative rate of contraception and is more tolerated in terms of multiple folliculogenesis than that with GnRH. The better results obtained by GnRH are probably due to the fact that it is possible to mimic the hormone dynamics of a normal

menstrual cycle and to achieve the maturation of a single follicle.

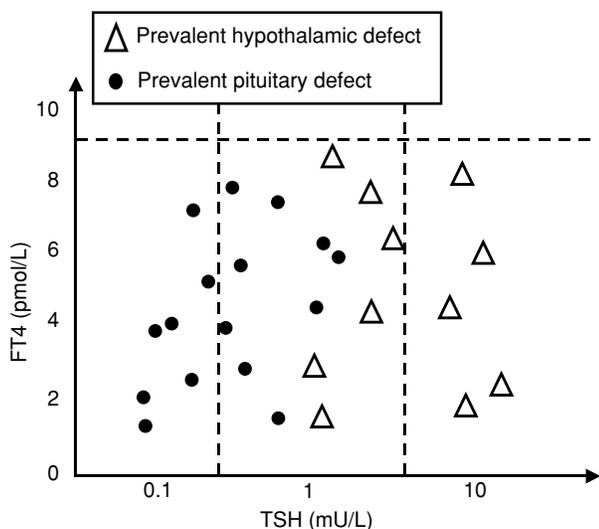
When ovulation by gonadotropin therapy is needed, treatment is started with 150 IU of FSH and the dose is increased until the levels of estradiol begin to rise. Follicular growth is monitored by ultrasound, and follicular rupture is induced by 10,000 IU of hCG. In patients with the acquired form of hypothalamic amenorrhea, induction of ovulatory cycles can also be attempted by clomiphene citrate (50 mg/day for 5 days).

### See Also the Following Articles

Anorexia Nervosa • Delayed Puberty and Hypogonadism, Female • Delayed Puberty and Hypogonadism, Male • Diabetes Insipidus, Neurogenic • Gonadotropin-Releasing Hormone (GnRH) Actions • Gynecomastia • Hypothalamus, Anatomy of • Kallmann's Syndrome and Idiopathic Hypogonadotropic Hypogonadism • Prader-Willi Syndrome • Undescended Testes

### Further Reading

- Hayes, F., Seminara, S., and Crowley, F., Jr. (1998). Hypogonadotropic hypogonadism. *Endocrinol. Metab. Clinics North America* **27**, 739–759.
- Quinton, R., Cheow, H., Bouloux, R., and Jacobs, H. (1999). Kallmann's syndrome: Is it always for life? *Clin. Endocrinol.* **50**, 481–487.
- Quinton, R., Duke, M., Robertson, A., Maffin, G., de Zoysa, P., Azcona, G., MacColl, G., Jacobs, H., Conway, G., Besser, M., Stanhope, R., and Bouloux, P. (2001). Idiopathic gonadotropin deficiency: Genetic questions addressed through phenotypic characterization. *Clin. Endocrinol.* **55**, 163–174.
- Warren, P., and Fried, L. (2001). Hypothalamic amenorrhea. *Endocrinol. Metab. Clinics North America* **30**, 611–623.
- Yen, S. (1993). Female hypogonadotropic amenorrhea. *Endocrinol. Metab. Clinics North America* **22**, 29–58.



**Figure 1** Serum levels of FT4 and TSH in patients with central hypothyroidism. Dashed lines indicate the lower and upper limits of the normal ranges. The evident lack of correlation between TSH and FT4 demonstrates that quantitative impairment of TSH secretion does not completely explain the pathogenesis of CH and indicates the involvement of qualitative alterations of hypothalamic regulation in a large portion of CH patients, particularly those with prevalent hypothalamic defects.

role played by hypothalamic factors in the generation of circadian TSH rhythm.

The differential diagnosis of CH is indicated by circulating levels of immunoreactive TSH: High TSH is typical of patients with prevalent hypothalamic defects (HH), whereas low TSH indicates the almost complete lack of functional thyrotropes (pituitary or secondary hypothyroidism). In CH patients with normal TSH, important information on the entity of hypothalamic impairment can be obtained by TRH testing. Indeed, differential diagnosis of CH is one of the residual indications for TRH tests. Pure HH is characterized by delayed/exaggerated/prolonged responses of immunoreactive TSH after intravenous injections of TRH (0.2 mg), but net TSH responses  $>4.0$  mU/liter can be considered indicative of a prevalent hypothalamic defect. Conversely, blunted TSH responses ( $<4.0$  mU/liter) are typical of a prevalent pituitary defect.

Before giving the diagnosis of HH (or CH), one should exclude the presence of “nonthyroidal illnesses” leading to the biochemical picture of the low T3 syndromes. These conditions include acute (e.g., myocardial infarction, severe infections, and extended burns) and chronic (e.g., liver or kidney insufficiency) diseases, fasting (as in anorexia nervosa), or recovery from general surgery. They also occur in patients with

an intact hypothalamic–pituitary–thyroid axis. In contrast to HH, these conditions are generally characterized by low T3 but normal TSH and FT4. Reduction of circulating TSH and FT4 occurs only in prolonged and manifest disease states and is likely due to suppression of hypothalamic TRH gene expression.

## CLINICAL PRESENTATION

Depending on the cause of CH, some characteristics of the disease may vary greatly. In general, CH of genetic origin is found in familial settings, the defect is congenital, and the onset generally occurs in early infancy with severe hypothyroid manifestations. In these cases, the defect is classically located in the pituitary, although a typical form of hypothalamic defect with mild CH manifestations is seen in patients with resistance to leptin. Genetic CH can be either isolated or combined with other pituitary hormone deficiencies (CPHD) and can be associated with growth disorders, delayed puberty, and/or neurological defects. Neonatal forms of CH are not recognized in the screening programs for congenital hypothyroidism based on TSH evaluation in dried blood spot, and neurological defects may become more severe if the clinical diagnosis and start of treatment are delayed. Nevertheless, several genetic forms of CPHD have delayed and progressive onset of hypothyroidism beyond the vulnerable period of central nervous system development during infancy and childhood.

Nongenetic CH is typically sporadic and acquired in most cases as the consequence of large sella tumors. Among the lesions potentially leading to CH, those more frequently associated with hypothalamic defects include large pituitary macroadenomas with suprasellar extension, craniopharyngiomas and suprasellar tumors, and those due to cranial irradiation, vascular accidents, and head traumas. In these situations, defects of posterior pituitary function with diabetes insipidus as well as visual field defects due to compression of the optic nerves may be concomitant. These lesions generally affect both pituitary and hypothalamic functions to various extents, leading to mild hypothyroidism, always combined with defects of other anterior pituitary hormones and frequently associated with high PRL levels due to stalk resection or compression. The association with signs and symptoms secondary to the altered secretion of other pituitary hormones (e.g., menstrual disorders, decreased libido, hair loss, galactorrhea, pallor, and altered lipid metabolism) or to local compression (visual defects,

headache, etc.) may cover the specific hypothyroid manifestations in several cases. Therefore, TSH and FT4 should always be included in the biochemical evaluation of patients with hypothalamic–pituitary diseases.

## PATHOGENESIS

The pathogenesis of typical HH is due to defects affecting the function of hypothalamic TRH neurons. These defects occur in various pathological conditions, including large pituitary macroadenomas with suprasellar extension, craniopharyngiomas and suprasellar tumors (such as meningiomas or gliomas), cranial irradiation, vascular accidents (including Sheehan's syndrome), head traumas (including traumatic delivery), and infiltrative diseases (such as sarcoidosis). As shown in Fig. 1, quantitative impairment of TSH secretion does not completely explain the pathogenesis of CH, and it indicates the involvement of qualitative alterations of hypothalamic regulation in a large number of CH patients, particularly those with prevalent hypothalamic defects. Some patients with CH of hypothalamic origin have normal or even elevated circulating TSH levels. The lack of any sign of primary thyroid insufficiency (absent anti-thyroid autoantibodies and normal response to exogenous TSH injection) suggested that hypothyroidism in these patients may result from the secretion of biologically inactive TSH molecules. This hypothesis was indirectly supported by the blunted T3 response to endogenous TRH-stimulated TSH, despite exaggerated and prolonged responses of endogenous TSH to the tripeptide. The very low bioactivity of TSH molecules circulating in several patients with hypothalamic CH of various origins (idiopathic, secondary to pituitary or hypothalamic tumors, or due to cranial irradiation or Sheehan's syndrome) was obtained by different bioassays. The reduced bioactivity was due to an impaired binding of circulating TSH to its specific thyroid receptor. Chronic TRH treatment restored both TSH receptor binding and bioactivity, indicating that this hypothalamic factor is necessary for the secretion of TSH molecules with structural features essential for appropriate thyroid stimulation. Lectin affinity chromatography showed that the oligosaccharide structure of TSH molecules circulating in hypothalamic CH is altered. The description of TRH knockout mice with the same biochemical picture as that of patients with hypothalamic CH has given definitive support for the pathogenic role played by TRH deficiency in this syndrome.

Genetic alterations associated with HH in humans are limited to defects in the gene encoding leptin receptor (leptin-R), a typical cytokine receptor involved in the complex network regulating food intake and body weight. Homozygous or compound heterozygous mutations leading to the production of inactive leptin-R mutants have been shown to be associated with a complex phenotype that includes familial obesity and hyperphagia, short stature with growth hormone deficiency, and puberty defects with low gonadotropins and mild hypothyroidism. Hypothyroidism of these patients is typically of hypothalamic origin because they have normal/elevated immunoreactive TSH and delayed/prolonged responses of endogenous TSH to TRH. A possible explanation for the pathogenesis of CH in patients with leptin-R mutations may derive from *in vitro* experiments showing that expression and stimulation of leptin-R result in strong activation of the TRH promoter. Moreover, leptin stimulates release of TRH from primary culture of hypothalamic neurons in a dose-dependent manner.

Another gene possibly involved in human cases of isolated CH is the TRH gene. Indeed, TRH KO mice are viable and are considered the animal model of HH. Similar to what was previously seen in humans with HH, subnormal T3 increases occur in these mice after stimulation of endogenous TSH by TRH, despite an exaggerated response. This finding represents indirect evidence that TSH circulating in these mice is devoid of biological activity. Relative hyperglycemia is apparently the only associated abnormality in TRH KO mice and may represent a distinctive element useful for determining the possible involvement of the TRH gene in human cases.

## TREATMENT

As for primary hypothyroidism, treatment of HH is based on the administration of L-thyroxine. Before starting L-thyroxine therapy, particular attention should be paid to possible concomitant corticotropin (ACTH) defects leading to partial adrenal insufficiency. Failure to recognize this concomitant defect may lead to acute adrenal insufficiency during the initial days of therapy. Therefore, screening for adrenal insufficiency, with evaluation of urinary free cortisol and/or low-dose ACTH testing (1 µg), is recommended for patients with HH. Alternatively, treatment with steroids may be advised before starting L-thyroxine.

L-Thyroxine replacement therapy should be tailored for each patient, as in primary hypothyroidism. Nevertheless, this goal is not easily reached for

HH patients because TSH measurement loses any value as a monitoring index. Therapy should be tailored in order to maintain FT<sub>4</sub> (and FT<sub>3</sub>) concentrations in the mid-normal range, provided that blood for hormone measurement is withdrawn before the morning of L-thyroxine administration. Evaluation of parameters of thyroid hormone action (such as cholesterol, bone markers, or sex hormone-binding globulin), particularly those not affected by defects in other pituitary functions (e.g., soluble receptor of interleukin-2), could be helpful in case of doubts. Replacement doses range from 1.1 to 1.7  $\mu\text{g}/\text{kg}/\text{day}$  in more than two-thirds of patients and are relatively lower in those older than age 60.

### See Also the Following Articles

Hypothalamic Disease • Hypothyroidism, Causes of • Hypothyroidism, Congenital, Long-Term Follow-Up • Hypothyroidism, Congenital, Screening Programs • Hypothyroidism, Diagnosis of • Hypothyroidism, Subclinical •

Hypothyroidism, Systemic Manifestations of • Hypothyroidism, Treatment of • Thyroid Hormone Metabolism • TSH Function and Secretion

### Further Reading

- Ferretti, E., Persani, L., Jaffrain-Rea, M. L., Giambona, S., Tamburrano, G., and Beck-Peccoz, P. (1999). Evaluation of the adequacy of l-T<sub>4</sub> replacement therapy in patients with central hypothyroidism. *J. Clin. Endocrinol. Metab.* **84**, 924–929.
- Persani, L. (1998). Hypothalamic thyrotropin-releasing hormone and thyrotropin biological activity. *Thyroid* **8**(10), 941–946.
- Persani, L., Ferretti, E., Borgato, S., Faglia, G., and Beck-Peccoz, P. (2000). Circulating TSH bioactivity in sporadic central hypothyroidism. *J. Clin. Endocrinol. Metab.* **85**, 3631–3635.
- Scalon, M. F. (2001). Thyrotropin-releasing hormone and thyroid-stimulating hormone. In "Endocrinology," (L. J. DeGroot and J. L. Jameson, eds.), 4th ed., pp. 1279–1289. Saunders, Philadelphia.
- Weintraub, B. D., Kazlauskaitė, R., Grossmann, M., and Szkudlinski, M. Z. (2001). Thyroid-stimulating hormone and regulation of the thyroid axis. In "Endocrinology," (L. J. DeGroot and J. L. Jameson, eds.), 4th ed., pp. 1345–1360. Saunders, Philadelphia.

by the ARN and EEN span overlapping neuroanatomical sites in the hypothalamus. Fourth, the onset and termination of appetite or hunger is triggered by discrete neural timing devices in the hypothalamus. Fifth, both the ARN and EEN are subject to moment-to-moment modulation by the internal milieu of hormones from adipocytes, gastrointestinal (GI) tract and endocrine glands, and neural signals from the GI tract.

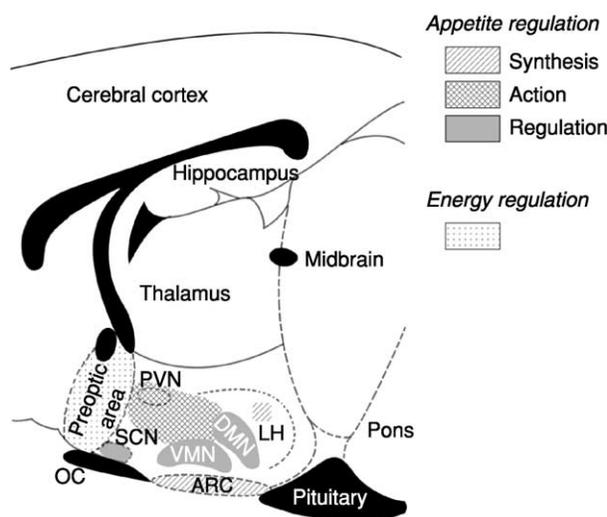
## NEUROANATOMICAL SUBSTRATE FOR ENERGY HOMEOSTASIS

The topography of the ARN and EEN circuitries and the clocks in the hypothalamus is described in this section (Figs. 1–3).

### Neuroanatomical Sites Regulating Appetite

#### Arcuate Nucleus

The arcuate nucleus (ARC), located at the base of the hypothalamus on either side of the third cerebroventricle, is the most important site in the hypothalamic integration of energy balance. Distinct clusters of neurons in the ARC produce orexigenic peptides such as neuropeptide Y (NPY) coexpressed with agouti-related peptide (AgRP) and the amino acid  $\gamma$ -aminobutyric acid (GABA), galanin (GAL) and the opioid dynorphin, and pro-opiomelanocortin (POMC)-derived  $\beta$ -endorphin and  $\alpha$ -melanocyte-stimulating hormone ( $\alpha$ -MSH) coexpressed with cocaine- and amphetamine-regulating transcript (CART) (Fig. 3). One of the terminal fields of these orexigenic and anorexigenic signal-producing neurons is the paraventricular nucleus (PVN). In addition, NPY-, AgRP-, and GABA-coexpressing neurons communicate directly with POMC neurons in the ARC via their receptors:  $Y_1$ , MC3, and  $GABA^A$ , respectively. NPY neurons also communicate with the orexigenic GAL- and dynorphin-producing neurons in the ARC. Thus, an interconnected orexigenic network of NPY and cohorts participates in the daily patterning of energy intake by two modalities: direct activation of  $Y_1/Y_5$  receptors (R) in the PVN and indirect activation by coacting with AgRP and GABA to restrain release of the anorexigen,  $\alpha$ -MSH, in the PVN. Furthermore, the ARC is strategically positioned outside the blood–brain barrier to process modulatory impulses from hormones secreted by adipocytes, the GI tract, and endocrine glands.



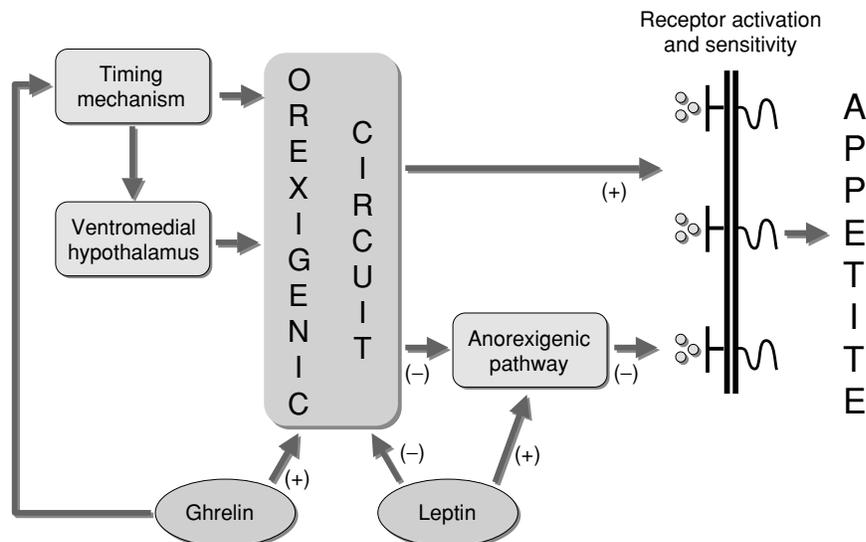
**Figure 1** Diagrammatic representation of hypothalamic sites engaged in regulation of appetite and energy expenditure on a sagittal section near the midline of the rat brain. AC, anterior commissure; DMN, dorsomedial nucleus; LH, lateral hypothalamus; OC, optic chiasm; PC, posterior commissure; PVN, paraventricular nucleus; SCN, supra-chiasmatic nuclei; VMN, ventromedial nucleus; synthesis = hypothalamic sites where orexigens and anorexigens are synthesized; action = hypothalamic areas where orexigens and anorexigens act; regulation = hypothalamic sites involved in regulation of synthesis and release of orexigens and anorexigen; energy regulation = hypothalamic site involved in nonshivering thermogenesis to expend energy. Adapted with permission from Kalra, S. P., Dube, M. G., Pu, S., Xu, B., Horvath, T. L., and Kalra, P. S. (1999). Interacting appetite-regulating pathways in the hypothalamic regulation of body weight. *Endocrine Rev.* 20, 68–100. Copyright 1999, The Endocrine Society.

#### Lateral Hypothalamus

Contrary to previous views, the lateral hypothalamus (LH) has been shown to relay excitatory influences to the orexigenic NPY system in the ARC–PVN axis. A distinct subpopulation of neurons producing melanin-concentrating hormone (MCH) and orexin (ORX) stimulate appetite by enhancing NPY release in the PVN. Apparently, these two orexigenic tracts also relay afferent hormonal and neural information on energy status in the body to the NPY ARC–PVN axis.

#### Ventromedial Hypothalamus

The ventromedial hypothalamus (VMH), including the ventromedial nucleus (VMN), is linked bidirectionally with the ARN in both the ARC and PVN. Because localized disruption of signaling in the VMN produces hyperphagia and leptin insensitivity, it is believed that the VMN restrains the ARN in the daily management of energy balance and may also be an integral component of the timing mechanism in the hypothalamus.



**Figure 2** Schematic representation of the basic elements of the hypothalamic circuitry that regulates the daily pattern of feeding. The information from the brain-timing mechanism, entrained to external photoperiodic signals, and the internal clock, driven by hormonal signals (ghrelin), stimulates the orexigenic circuit either directly or indirectly through multisynaptic routes in the medial basal hypothalamus. This results in stimulation of appetite mediated by postsynaptic targets and curbs the tonic restraint of the anorexigenic circuit. Afferent hormonal signals from the periphery, gastric ghrelin and adipocytes' leptin, exert opposing affects on operation of the orexigenic and anorexigenic circuits to assist in timely initiation and termination of appetite. (+) = stimulatory; (-) = inhibitory. Adapted with permission from Kalra, S. P., Dube, M. G., Pu, S., Xu, B., Horvath, T. L., and Kalra, P. S. (1999). Interacting appetite-regulating pathways in the hypothalamic regulation of body weight. *Endocrine Rev.* **20**, 68–100. Copyright 1999, The Endocrine Society.

### Paraventricular Nucleus

PVN and its two subdivisions, parvocellular PVN (pPVN) and magnocellular PVN (mPVN), are the terminal fields of the orexigenic and anorexigenic signal-producing neurons. The PVN consists of a heterogeneous population of neurons that fan out on either side of the roof of the third cerebroventricle in the hypothalamus. Target neurons that express postsynaptic receptors for propagation and transmission of signals to either initiate or terminate appetite are resident in the PVN.

### Neuroanatomical Sites Regulating Energy Expenditure

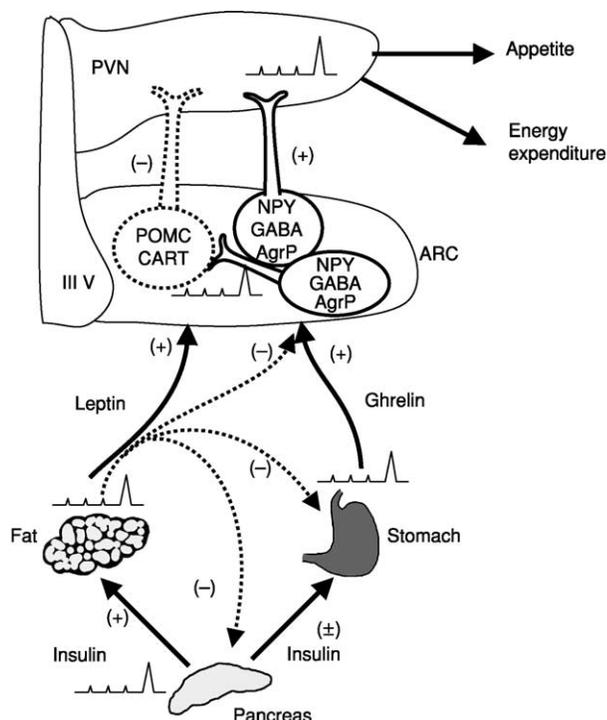
#### Medial Preoptic Area

The medial preoptic area (MPOA) located rostral to the PVN is the only site shown to augment non-shivering thermogenic energy expenditure without affecting appetite. The MPOA is a component of the sympathetic nervous system (SNS), and neurons expressing leptin R transduce and relay energy expenditure information to the brown adipose tissue (BAT) in the periphery.

### Neuroanatomical Sites for the Timing Mechanism

#### Suprachiasmatic Nucleus and the Basal Hypothalamus

Daily meal patterning is a highly regulated phenomenon. Although entrained to external cues of the daily photoperiod and seasonality, it is highly vulnerable to availability of food supply. Animals in the wild or laboratory rodents maintained on an *ad libitum* diet under a controlled light–dark cycle exhibit multiple high-amplitude feeding episodes in the arousal phase during the dark phase in nocturnal animals or during the light phase in diurnal animals. In contrast, humans and subhuman primates maintained in the laboratory consume a few timed meals during the daytime and early hours of the night. Two distinct neural timing mechanisms in the hypothalamus operate upstream from the ARN in the daily trigger of ingestive behavior. Two small round nuclei dorsal to the optic chiasm on either side of the third cerebroventricle, the suprachiasmatic nuclei (SCN), are major coordinators of the internal circadian organization for ingestive behavior. Destruction of the SCN results in unregulated feeding, apparently due to the loss of a “CLOCK”



**Figure 3** Model showing the dynamics in the feedback circuitry involved in hypothalamic integration of appetite and energy expenditure. The primary components of the ARN, the orexigenic NPY and coexpressed GABA and AgrP in an interplay with the anorexigenic pathway emanating from POMC and CART coexpressing neurons, stimulate and terminate appetite in a timely fashion, as directed by the timing mechanism. NPY produced and released in the ARC–PVN axis displays two rhythmic efflux patterns (circadian and ultradian), as directed by two functionally opposing rhythmic afferent hormonal signals: leptin from adipocytes and ghrelin from the stomach. Also depicted is the rhythmic feedback relationship in the fat–stomach–pancreas axis in the periphery. See text for details. (+) = stimulatory; (–) = inhibitory; (±) = unresolved. Adapted with permission from Kalra, S., and Kalra, P. (2003). Rhythmic, reciprocal ghrelin and leptin signaling: New insight in the development of obesity. *Reg. Pept.* 111, 1–11.

transcription factor. However, uncoupling meal timing from the circadian clock in primates is regulated by neuronal PAS domain protein 2 (NPAS2) transcription factor in the forebrain sites, VMH, dorsomedial nucleus, subparaventricular zone, and LH. Apparently, a timely interplay between CLOCK and NPAS2 transcription factors, in response to incoming internal and external environmental information, initiates first the anticipatory synthesis and then the timely release of orexigenic peptides in the PVN (Figs. 1 and 2).

## OPERATION OF THE HYPOTHALAMIC ARN

### NPY and Cohorts

NPY, the most prominent, physiologically relevant, appetite-stimulating neurotransmitter in vertebrates, is produced by two discrete subpopulations of neurons in the hypothalamic ARC and in the brainstem (BS). It is released in the ARC–PVN axis to stimulate feeding by activating the  $Y_1/Y_5$  R primarily in the mPVN. Increases in production and release of NPY, as seen in genetic models of obesity such as obese *fa/fa* Zucker rats and *ob/ob* mice, underlie hyperphagia accompanied by hyperinsulinemia, glucose intolerance, and hyperglycemia, a phenotype resembling the clinical metabolic syndrome. Infusion of NPY centrally stimulates an episodic pattern of feeding, rapid weight gain, and obesity. To replenish the energy depleted by fasting or diet restriction, episodic NPY secretion in the PVN is accelerated to sustain the appetitive drive. NPY is a physiological orexigenic signal because increased NPY synthesis, storage, and release in the ARC–PVN axis are triggered by the clock mechanism in anticipation of mealtime and NPY secretion subsides as food consumption is initiated. Blockade of the anticipatory rise in either synthesis or release suppresses feeding. Whereas feeding in the complete absence of NPY in null mutant mice is unchanged, the obesity in leptin-deficient *ob/ob* mice is suppressed and the magnitude of fasting-induced feeding is drastically curtailed, thereby implying a crucial role of NPY in induction of hyperphagia and obesity. Paradoxically, experimentally induced diminution in NPY release in the PVN produces unregulated hyperphagia and obesity. In fact, a low abundance of NPY up-regulates  $Y_1$  R to render animals hypersensitive to NPY such that extremely low doses of NPY that are normally ineffective evoke near maximal stimulation of feeding. Germ line deletion of any one of the NPY receptors,  $Y_1$ ,  $Y_2$ , or  $Y_5$ , also results in unregulated feeding and late-onset obesity. Overall, an imbalance in NPY signaling locally in the ARC–PVN axis, elicited by either high or low NPY abundance, leads to altered eating patterns and obesity.

Although NPY is a primary player in the ARN, several other chemically distinct cohorts of disparate origin also assist in propagation of the innate appetitive drive. NPY neurons coexpress AgrP and GABA in the ARC and coexpress the adrenergic neurotransmitters, norepinephrine and epinephrine, in the BS. Whereas GABA coreleased with NPY in the PVN augments feeding synergistically by activation of

GABA<sup>A</sup> and Y<sub>1</sub>/Y<sub>5</sub> R, respectively, GABA and AgrP also attenuate the  $\alpha$ -MSH-induced restraint on feeding via GABA<sup>A</sup> R and MC3 R expressed locally in the ARC POMC neurons. Moreover, AgrP released in the mPVN antagonizes the melanocortin restraint on feeding by competitive antagonism at MC4 receptors.

Galanin,  $\beta$ -endorphin, MCH, and ORX are additional neuropeptides that either coact with NPY or stimulate NPY release to enhance feeding. Thus, the intricate relationship of NPY with a host of orexigenic cohorts represents a failproof system with ample backup to meet varied environmental challenges and sustain the instinctual appetitive drive for survival. Seemingly, a three-pronged action of NPY—direct stimulation of Y<sub>1</sub>/Y<sub>5</sub> R in the mPVN, stimulation of the release of orexigenic opioid,  $\beta$ -endorphin, via elicitation of GAL release and GAL action on its own, and a curb on the restraint exercised by anorexigenic  $\alpha$ -MSH and CART—initiates and sustains a robust appetite at mealtime. ORX and MCH stimulate feeding by eliciting NPY release in the PVN. This line of communication between ORX and MCH neurons in the LH, along with the synergistic effect of adrenergic transmitters coreleased with NPY from BS projections to the ARC–PVN axis, is apparently important in relaying the combined hormonal and neural information from peripheral visceral organs (Figs. 2 and 3).

## REGULATION OF THE ARN

An interplay of afferent hormonal and neural pathways, under the direction of central clocks and energy demands of the organism, is fundamental for integration of various effector pathways in the ARN, as discussed in this section (Figs. 2 and 3).

### Regulation by Afferent Hormonal Signals

The current view advocates timely reciprocal interaction between the afferent hormonal signals, the anorexigenic leptin and orexigenic ghrelin, in driving the ARN to enforce the daily patterns of food intake.

#### **Adipocyte Hormone Leptin: An Anorexigen**

It is now well established that leptin, produced by adipocytes and the stomach in the periphery and by a subset of neurons in the hypothalamus, is the predominant signal that restrains appetite and stimulates nonshivering thermogenic energy expenditure. Leptin mutant *ob/ob* mice and humans display unabated

phagia, abnormal weight gain, and obesity. Replacement of leptin either by giving daily injections or by increasing expression of leptin selectively in the hypothalamus with leptin gene therapy reinstates the tonic restraint on feeding and concurrently augments energy expenditure to normalize weight. Also, lack of leptin R, as in germ line mutant rodents or by conditional hypothalamic leptin R deletion, reproduces the obese phenotype and metabolic abnormalities. Reinstallation of leptin R by leptin R gene therapy in the ARC of these rodents normalizes energy homeostasis.

Leptin secretion from adipocytes in ultradian and circadian patterns assists in expression of the tonic restraint on the ARN. The daily peaks and valleys in secretion are abolished by increased adiposity with age and by lifestyle changes. Consumption of a calorie-rich diet results in increased ultradian leptin secretion and loss of the circadian pattern. These abnormal hormonal patterns are associated with pathophysiological sequelae, leading to abnormal rate of weight gain, obesity, and metabolic disorders. Current evidence demonstrates that hyperleptinemia generates central leptin insufficiency, resulting in loss of tonic restraint on hypothalamic effector pathways and dysregulation of weight homeostasis. An increase in leptin availability selectively in hypothalamic sites by leptin gene therapy normalizes food consumption and weight homeostasis.

The tonic restraint on the ARN by leptin is mediated by the long form of leptin R expressed by orexigenic and anorexigenic signal-producing neurons of the ARN. Leptin inhibits the synthesis and secretion of the orexigenic NPY, AgrP, MCH, and ORX and enhances anorexigenic melanocortin signaling by increasing  $\alpha$ -MSH release in the PVN. This concurrent dual mode of central action by postprandial leptin hypersecretion is pivotal in terminating the meal. Conversely, curtailed restraint resulting from low leptin levels during the pre-meal interval or in response to energy depletion by food restriction or fasting sets these neurochemical sequelae in motion in reverse order, that is, increased orexigenic and decreased anorexigenic signaling to propagate hunger and initiate the drive to the source of food. In addition, leptin restrains feeding by opposing ghrelin-induced stimulation of appetite and by synergizing with the gut peptide cholecystokinin (CCK).

#### **Gut Hormones**

##### *Ghrelin: An Orexigen*

Ghrelin is as effective an orexigen as is NPY. Increased circulating levels readily stimulate feeding,

and sustained high levels promote hyperphagia and adiposity. Ghrelin, produced in the stomach and in a discrete subpopulation of neurons located in the sub-PVN zone, is a putative physiological signal for generating hunger. Ultradian secretion of ghrelin from the stomach is accelerated during the pre-meal interval and by fasting. Unlike multiple leptin targets, central ghrelin targets are the ghrelin R expressing ARC NPY neurons only. Increasing ghrelin levels accelerate rhythmic NPY discharge in the PVN to stimulate appetite.

The orexigenic effects of ghrelin are regulated by leptin in two ways. Centrally, leptin restrains ghrelin-induced NPY release, and peripherally, it inhibits ghrelin secretion. Consequently, ghrelin hypersecretion during the pre-meal interval and fasting is due, in part, to attenuated leptin restraint on ghrelin secretion, and the postprandial rise in leptin secretion restores inhibition of ghrelin secretion peripherally and ghrelin-induced appetite centrally. It is highly likely that this dual action of leptin is crucial for suppression of ingestive behavior during the inter-meal interval.

#### *Peptide YY 3-36*

Peptide YY (PYY) is a member of the NPY family and is produced by endocrine cells of the gut. Peptide YY 3–36 (PYY3–36) levels rise in the circulation after a meal. Microinjection of PYY3–36 into the ARC inhibited feeding by reducing NPY release in the PVN through activation of inhibitory presynaptic  $Y_2$  auto-receptors. Despite the contrary evidence that PYY3–36 levels are reduced in obese patients, it is possible that the postprandial rise of PYY3–36 may constitute one of the satiety signals to hypothalamic effector pathways.

#### *Cholecystokinin and Glucagon-like Peptides 1 and 2*

CCK has long been implicated as a putative satiety signal from the duodenum. CCK receptors on afferent sensory fibers of the vagus terminating in the nucleus of the tractus solitarius (NTS) in the brainstem mediate termination of a meal by CCK. Thus, a postprandial increase in CCK may reduce meal duration by relaying information via the gut–BS axis to the hypothalamus. However, the precise impact of this part hormonal, part neural route of communication on the motivational drive emanating in the hypothalamic ARN remains to be ascertained.

Like CCK, glucagon-like peptides (GLP-1, GLP-2, and oxyntomodulin), derivatives of posttranslational cleavage of prepro-glucagon in the intestinal mucosa and neurons in the NTS, are putative candidates for

a neurohormonal role in mediating postprandial satiety. GLP-1 and GLP-2 inhibit food intake after central administration, possibly mediated by feeding relevant hypothalamic sites, including the dorsomedial nucleus. Interestingly, postprandial gastric loading activated GLP1/GLP2-expressing neurons in the NTS that express leptin receptors, thereby implying that this GLP pathway represents a central leptin-responsive system that exerts a tonic restraint on appetite.

#### **Pancreatic Hormones: Insulin**

Insulin, secreted by beta cells of the pancreas, regulates glucose homeostasis and has been suggested to also act centrally to inhibit food intake. In the periphery, insulin stimulates leptin secretion from adipocytes. The observations that leptin and insulin secretion rise postprandially, hyperinsulinemia precedes hyperleptinemia in response to consumption of a high-fat diet, and leptin replacement alleviates hyperglycemia and hyperphagia that ensue after destruction of pancreatic beta cells are consistent with the view that it is leptin and not insulin that normally evokes the previously advocated central effects of insulin on energy homeostasis.

On the other hand, cross-talk between white adipose tissue (WAT) and the pancreas affects energy homeostasis indirectly. Normally, insulin stimulates leptin secretion and storage of excess energy as fat. Leptin (as an autocrine signal), in turn, restrains insulin secretion. However, consumption of a calorie-rich diet promotes hyperinsulinemia, hyperleptinemia, and obesity. This metabolic imbalance, especially hyperleptinemia, for extended periods engenders various disorders of metabolic syndrome. Reduction in adiposity generally reinstates the normal cross-talk between WAT and the pancreas and alleviates metabolic disorders.

#### **Steroids**

##### *Glucocorticoids*

The lipophilic steroids readily cross the blood–brain barrier to access target neurons in the hypothalamus. The glucocorticoids (corticosterone in rodents and cortisol in humans) secreted by the adrenal cortex stimulate appetite. Glucocorticoid therapy invariably promotes weight gain and adiposity. Stimulation of food intake is affected by increased hypothalamic NPY signaling facilitated by the ARC NPY neurons expressing glucocorticoid receptors. Furthermore, the observations that mRNA expression in the ARC NPY neurons and peptide levels in the PVN nerve terminals rise after glucocorticoid treatment, and that the daily increase in corticosterone secretion precedes the

onset of a meal, support a facilitatory role of glucocorticoids in the NPY-induced daily patterning of ingestive behavior.

### *Estrogens*

The primary ovarian estrogen, estradiol 17 $\beta$ , suppresses appetite and fat deposition. Whereas ovariectomy rapidly increases phagia and body weight, estrogen replacement reverses these effects. Estradiol 17 $\beta$  decreases NPY release in the PVN, an action mediated by estrogen R expressed on ARC NPY neurons.

In summary, the interplay of the leptin–ghrelin–NPY loop, in conjunction with the modulatory influence of an array of signal molecules of hypothalamic and peripheral origin, represents the hardcore wiring of the basic dynamic machinery that initiates and terminates episodic feeding in vertebrates (Fig. 3). Normally, leptin tonically restrains feeding by restricting efflux of orexigenic NPY, GABA, and AgrP and increases (either directly or indirectly through NPY links) the release of anorexigenic  $\alpha$ -MSH and CART in the PVN. Likewise, in the periphery, leptin tonically restrains ghrelin secretion from the stomach and blocks its central orexigenic action by opposing its excitatory effects on NPY release in the PVN. Consequently, meal initiation is a sequential neurohumoral event initiated first by removal of leptin restraint to allow increased synthesis and storage of NPY in the ARC–PVN axis and of ghrelin in the stomach preceding the impending hypersecretion. Thereafter, a trigger from the timing mechanism elicits episodic ghrelin release from the stomach. This peripheral orexigenic signal and the neural clock together accelerate rhythmic NPY discharge in the PVN to initiate feeding. Leptin hypersecretion postprandially reasserts the tonic restraint on NPY and ghrelin release and curtails feeding behavior. Thus, the ebb and flow in the tonic restraint imposed by leptin on information flow from the peripheral afferent and hypothalamic ARN generate the daily meal pattern.

## ENERGY EXPENDITURE NETWORK

Hypothalamic SNS innervations relay signals that regulate general activity and nonshivering thermogenesis mediated by uncoupling protein-1 (UCP-1) in the BAT in the periphery. Obesity induced by ablation of the VMH decreases sympathetic tone, general activity, and thermogenic energy loss. Several sites in the anterior hypothalamus, including the

MPOA, are intimately involved in regulation of body temperature. It is now apparent that these sites are also involved in regulating nonshivering thermogenesis, an essential component of energy homeostasis. Injections of leptin readily increase general activity and nonthermogenic energy expenditure by activation of BAT UCP-1 signaling. Increased expression of leptin after microinjection of viral vectors encoding the leptin gene in various hypothalamic leptin R-expressing sites—the ARC, PVN, VMN, and MPOA—decreased adiposity and increased nonshivering thermogenic energy expenditure. However, unlike that observed in other hypothalamic sites, increased leptin transgene expression in the MPOA decreased adiposity by augmenting BAT-mediated thermogenesis without decreasing energy intake (Figs. 1 and 3). Thus, a leptin-responsive hypothalamic EEN network distinct from the ARN, originating in the MPOA and traversing caudally through several hypothalamic nuclei en route to BAT, is an integral component in the hypothalamic integration of energy homeostasis. The existence of distinct central pathways, the ARN and EEN, demonstrates that it is now possible to design therapies to curb weight gain and adiposity by increasing energy expenditure without interrupting normal food intake.

## CONCLUSION

Multidisciplinary research has uncovered various aspects of the spatiotemporal patterning of the hypothalamic ARN driven by a host of internal factors (Figs. 1–3). Within the ARN, the orexigenic NPY and its cohorts in the ARC–PVN axis constitute the primary neural pathway for initiation and termination of a meal. To accomplish this complex task efficiently in a strict temporal fashion and in synchrony with external driving forces (food availability, photoperiod, and seasonality), the NPY system employs a distinct anorexigenic network composed of  $\alpha$ -MSH- and CART-producing POMC neurons in the ARC. To monitor energy expenditure and metabolic efficiency in the body, the ARN is also intimately linked with the EEN and SNS in the MPOA and other hypothalamic sites for relay of neural efferents to the BAT, WAT, and GI.

At least three basic elements coordinate the operation of the ARN under the direction of external and internal environmental factors: (1) the intrahypothalamic timing mechanism, (2) innervations of MCH and ORX neurons from the LH for relay of afferent messages ascending the multisynaptic neural

pathway from the periphery, and (3) afferent hormonal signals from the periphery. Reciprocal circadian and ultradian rhythmic secretion patterns of leptin and ghrelin encode a corresponding pattern of NPY release in the PVN to regulate feeding. In addition, leptin restrains gastric ghrelin and pancreatic insulin secretion. Insulin, in turn, promotes leptin and possibly modulates ghrelin efflux. Thus, a dynamic interplay at two levels, one peripherally among adipocytes, stomach, and pancreas and the other among various components of the hypothalamic ARN and peripheral visceral organs, underlies the daily management of feeding behavior. Environmental challenges sustained over extended periods, particularly excess consumption of an energy-rich diet and a sedentary lifestyle, produce subtle and progressive derangements in feedback communication among peripheral visceral organs and the ARN. A major consequence of this disruption is positive energy balance promoting a gradual increase in adiposity, morbid obesity, and metabolic syndrome.

This new fundamental knowledge of the hypothalamic integrative processes at the cellular and molecular levels has elucidated a cohesive framework and identified vulnerable loci for designing therapeutic remedies to stem the onslaught of environmental obesity and metabolic diseases and, thereby, to enhance the quality of life and extend the life span.

### See Also the Following Articles

Anorexia Nervosa • CCK (Cholecystokinin) • Circadian Rhythms: Hormonal Facets • Eating Disorders and the Reproductive Axis • Ghrelin • Hunger and Satiation • Hypothalamus, Anatomy of • Neurohypophysial Hormone Regulatory Systems • Obesity, Childhood and Adolescence • Obesity, Treatment of • Peptide YY (PYY)

### Further Reading

Bagnasco, M., Dube, M. G., Kalra, P. S., and Kalra, S. P. (2002). Evidence for the existence of distinct central appetite and energy expenditure pathways and stimulation of ghrelin as revealed by hypothalamic site-specific leptin gene therapy. *Endocrinology* **143**, 4409–4421.

- Baratta, M. (2002). Leptin: From a signal of adiposity to a hormonal mediator in peripheral tissues. *Med. Sci. Monit.* **8**, RA282–RA292.
- Clark, J. T., Kalra, P. S., Crowley, W. R., and Kalra, S. P. (1984). Neuropeptide Y and human pancreatic polypeptide stimulate feeding behavior in rats. *Endocrinology* **115**, 427–429.
- Cowley, M. A., Pronchuk, N., Fan, W., Dinulescu, D. M., Colmers, W. F., and Cone, R. D. (1999). Integration of NPY, AGRP, and melanocortin signals in the hypothalamic paraventricular nucleus: Evidence of a cellular basis for the adipostat. *Neuron* **24**, 155–163.
- Cowley, M. A., Smith, R. G., Diano, S., Tschop, M., Pronchuk, N., Grove, K. L., Strasburger, C. J., Bidlingmaier, M., Esterman, M., Heiman, M. L., Garcia-Segura, L. M., Nillni, E. A., Mendez, P., Low, M. J., Sotonyi, P., Friedman, J. M., Liu, H., Pinto, S., Colmers, W. F., Cone, R. D., and Horvath, T. L. (2003). The distribution and mechanism of action of ghrelin in the CNS demonstrates a novel hypothalamic circuit regulating energy homeostasis. *Neuron* **37**, 649–661.
- Cummings, D. E., Purnell, J. Q., Frayo, R. S., Schmidova, K., Wisse, B. E., and Weigle, D. S. (2001). A preprandial rise in plasma ghrelin levels suggests a role in meal initiation in humans. *Diabetes* **50**, 1714–1719.
- Dhillon, H., Kalra, S. P., and Kalra, P. S. (2001). Dose-dependent effects of central leptin gene therapy on genes that regulate body weight and appetite in the hypothalamus. *Mol. Ther.* **4**, 139–145.
- Dube, M. G., Beretta, E., Dhillon, H., Ueno, N., Kalra, P. S., and Kalra, S. P. (2002). Central leptin gene therapy blocks high fat diet-induced weight gain, hyperleptinemia, and hyperinsulinemia: Effects on serum ghrelin levels. *Diabetes* **51**, 1729–1736.
- Green, C. B., and Menaker, M. (2003). Clocks on the brain. *Science* **301**, 319–320.
- Horvath, T. L., Diano, S., Sotonyi, P., Heiman, M., and Tschop, M. (2001). Minireview: Ghrelin and the regulation of energy balance: A hypothalamic perspective. *Endocrinology* **142**, 4163–4169.
- Kalra, S. P., Bagnasco, M., Otukonyong, E. E., Dube, M. G., and Kalra, P. S. (2003). Rhythmic, reciprocal ghrelin and leptin signaling: New insight in the development of obesity. *Regul. Pept.* **111**, 1–11.
- Kalra, S. P., Dube, M. G., Pu, S., Xu, B., Horvath, T. L., and Kalra, P. S. (1999). Interacting appetite-regulating pathways in the hypothalamic regulation of body weight. *Endocrine Rev.* **20**, 68–100.
- Kalra, S. P., and Kalra, P. S. (2003). Keeping obesity and metabolic syndrome at bay with central leptin and cytokine gene therapy. *Curr. Med. Chem. Cent. Nerv. System Agent* **3**, 189–199.
- Kalra, S. P., and Kalra, P. S. (2004). NPY: A novel on/off switch for control of appetite and reproduction. In “Handbook of Experimental Pharmacology” (M. C. Michel, ed.), in press. Springer-Verlag, Berlin.

**Table II** Hormones of the Hypothalamic–Pituitary Unit<sup>a</sup>

Hypothalamic hormone	Pituitary hormone	Target organ	Feedback hormone
GnRH	FSH	Ovary	Estradiol and inhibin
	LH	Testis	Testosterone and inhibin
TRH	TSH	Thyroid	T3 and T4
GHRH	GH	Multiple	IGF-1
Somatostatin	GH	Multiple	IGF-1
CRH	ACTH	Adrenal	Cortisol
Dopamine	Prolactin	Breast	?
AVP	ACTH	Adrenal	Cortisol

<sup>a</sup>Abbreviations used: ACTH, corticotropin; AVP, arginine vasopressin; CRH, corticotropin-releasing hormone; FSH, follicle-stimulating hormone; GH, growth hormone; GHRH, growth hormone-releasing hormone; IGF-1, insulin-like growth factor-1; LH, luteinizing hormone; GnRH, gonadotropin-releasing hormone; T3, triiodothyronine; T4, thyroxine; TRH, thyrotropin-releasing hormone; TSH, thyrotropin.

Anterior lobe or adenohypophysis: No physical connection exists between the adenohypophysis and the hypothalamus. Therefore, the interaction between these two organs is achieved through hypothalamic neurohormones that reach the anterior lobe through the portal circulatory plexus, the hypophyseal–portal circulation.

Intermediate lobe: Its significance is unknown. It appears especially prominent in the fetus and in pregnant women.

## FUNCTIONS

The hypothalamic–pituitary unit has been shown to regulate a pleiotopic set of functions in the organism

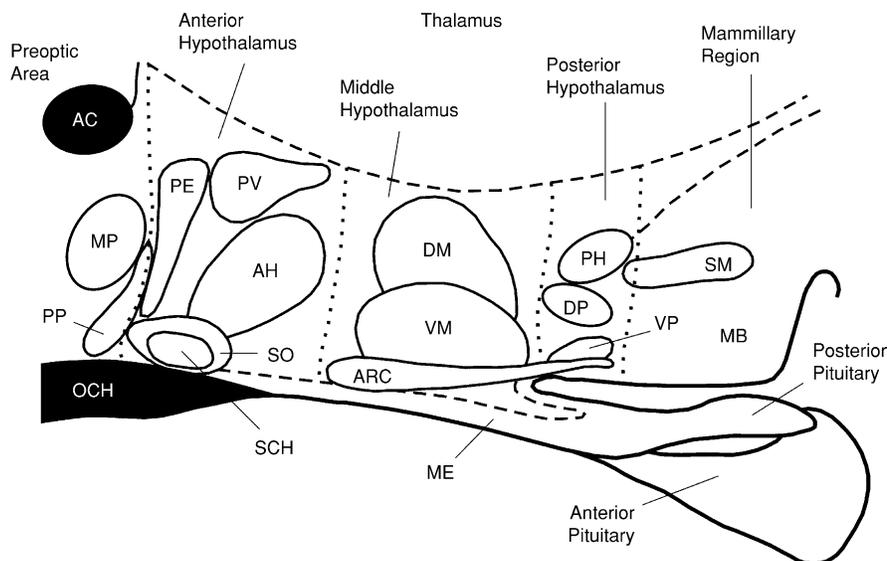
(Table I). The majority of the hypothalamic–pituitary functions are endocrine in nature. To exert their actions, hormones must first be recognized by specific receptors on the surface of cells of target organs or tissues (Table II). Target organs, in turn, are capable of influencing the hypothalamic–pituitary unit through feedback loops, thereby promoting integration and coordination between peripheral organs and tissues and the hypothalamic–pituitary unit, which are essential for the maintenance of the finely tuned physiologic homeostasis.

## See Also the Following Articles

Hypothalamic Disease • Hypothalamus, Anatomy of • Hypothalamus–Pituitary–Thyroid Axis • Pituitary Gland Anatomy and Embryology • Pituitary Gland, Evolution of

## Further Reading

- Challis, J. R. G., Matthews, S. G., Gibb, W., and Lye, S. J. (2000). Endocrine and paracrine regulation of birth at term and preterm. *Endocr. Rev.* **21**, 514–550.
- Doufas, A. G., and Mastorakos, G. (2000). The hypothalamic–pituitary–thyroid axis and the female reproductive system. *Ann. N. Y. Acad. Sci.* **900**, 65–76.
- Mantozoros, C. S. (2000). Role of leptin in reproduction. *Ann. N. Y. Acad. Sci.* **900**, 174–183.
- Morand, E. F., and Leech, M. (2001). Hypothalamic–pituitary–adrenal axis regulation of inflammation in rheumatoid arthritis. *Immunol. Cell. Biol.* **79**, 395–399.
- Reichlin, S. (1992). Neuroendocrinology. In “Williams Textbook of Endocrinology” (J. D. Wilson and D. W. Foster, eds.), pp. 135–219. Saunders, Philadelphia.
- Thormer, M. O., Vance, M. L., Horvath, E., and Kovacs, K. (1992). The anterior hypophysis. In “Williams Textbook of Endocrinology” (J. D. Wilson and D. W. Foster, eds.), pp. 221–310. Saunders, Philadelphia.



**Figure 1** Hypothalamic regions and cell groups within regions projected into the midsagittal section of the rat hypothalamus. AC, anterior commissure; AH, anterior hypothalamic nucleus; ARC, arcuate nucleus; DM, dorsomedial nucleus; DP, dorsal pre-mammillary nucleus; MB, mammillary body; ME, median eminence; MP, medial preoptic nucleus; OCH, optic chiasm; PE, periventricular nucleus; PP, preoptic periventricular nucleus; PV, paraventricular nucleus; SCH, supra-chiasmatic nucleus; SM, supra-mammillary nucleus; SO, supraoptic nucleus; PH, posterior hypothalamic nucleus; VM, ventromedial nucleus; VP, ventral pre-mammillary nucleus.

### Cell Groups and Areas

#### Preoptic Region

The preoptic region extends from the basal forebrain and lamina terminalis to the level of the rostral half of the optic chiasm and continues into the anterior hypothalamus. Dorsally, it is bordered by the anterior commissure and the bed nucleus of the stria terminalis.

The cell-rich medial part of the preoptic region contains several cell groups, including the medial preoptic and preoptic periventricular nuclei, and contains gonadotropin-releasing hormone (GnRH) neurons. The lateral part is occupied by the medial forebrain bundle. In the midline, a circumventricular organ, the organum vasculosum laminae terminalis (also called the supraoptic crest) can be found around the tip of the third ventricle. The organ is outside of the blood-brain barrier.

#### Anterior Hypothalamus

The anterior hypothalamus is a continuation of the preoptic region and extends caudally to the level of the arising of the median eminence. Its medial part comprises the anterior periventricular, supra-chiasmatic, anterior hypothalamic, and paraventricular nuclei. The anterior periventricular nucleus contains somatostatin immunoreactive neurons projecting to the median eminence and dopaminergic neurons terminating in the posterior and intermediate lobe of the

pituitary gland (periventriculo-hypophyseal dopaminergic system). The supra-chiasmatic nucleus (a key structure in the control of biological rhythms) is situated just above the optic chiasm on the two sides of the third ventricle. It receives direct input from the retina and indirect neuronal input from the retina via the lateral geniculate body and from brainstem serotonergic neurons. The supraoptic nucleus is located along and around the lateral edge of the optic chiasm and is composed of magnocellular neurons. In humans, the supraoptic nucleus is situated above the rostral end of the optic tract. Neurons of this cell group synthesize oxytocin and vasopressin and project to the posterior pituitary. The paraventricular nucleus lies close to the two sides of the third ventricle, above the anterior hypothalamic nucleus. Magnocellular and parvocellular divisions can be distinguished. Magnocellular neurons synthesize oxytocin and vasopressin and project to the posterior pituitary. The cell group contains corticotropin-releasing hormone (CRH)- and thyrotropin-releasing hormone (TRH)-producing neurons that terminate in the median eminence as well as medium-sized neurons that give rise to long descending projections to the brainstem and spinal cord. The lateral part of this hypothalamic region contains the supraoptic nucleus and fibers and cells of the medial forebrain bundle. Caudally, the so-called retrochiasmatic area containing loosely

packed cells and fibers occupies the ventral part of the anterior hypothalamus. In both the medial and lateral parts of the anterior hypothalamus, magnocellular neurons form small groups (called accessory nuclei).

#### *Middle Hypothalamus (Tuberal Region)*

The middle hypothalamus, starting behind the retrochiasmatic area, ends at the level of the separation of the pituitary stalk. The medial basal part of the region contains the arcuate and ventromedial nuclei and the median eminence, whereas the medial dorsal part contains the dorsomedial nucleus. The lateral part is called the lateral hypothalamic area and is largely occupied by the medial forebrain bundle.

The arcuate (also called the infundibular) nucleus, made up of five subdivisions, is an elongated cell group in the most ventromedial part of the middle hypothalamus and posterior hypothalamus, mostly on the two sides of the third ventricle (infundibular recess). The cell group contains, in addition to several other neuropeptides, pro-opiomelanocortin-synthesizing neurons and growth hormone-releasing hormone (GHRH) as well as dopaminergic neurons. These latter two project to the median eminence, forming the tubero–infundibular pathway. The ventromedial nucleus is a major cell group composed of five subdivisions that contain morphologically and functionally distinct neurons with neuronal connections to various components of the limbic system. The dorsomedial nucleus divided into three subdivisions is caudal to the paraventricular nucleus.

#### *Posterior Hypothalamus (Premammillary Region)*

The main cell groups of the posterior hypothalamus, a relatively small part of the hypothalamus, are the ventral and dorsal premammillary, tuberomammillary, supramammillary, and posterior hypothalamic nuclei. The ventral premammillary nucleus is a small cell group in continuation with the ventromedial and partly the arcuate nuclei. The dorsal premammillary nucleus is located caudal to the ventromedial nucleus at the sides of the inframammillary recess of the third ventricle. The tuberomammillary nucleus is located in the most ventral and medial part of the posterior hypothalamus. Some neurons of the cell group synthesize histamine, providing the only source of neuronal histamine in the brain. The posterior hypothalamic nucleus is a fairly large group of cells occupying the dorsal part of the periaqueductal central gray. The unpaired supramammillary nucleus is dorsal to the mammillary body and projects to the

hippocampus, dentate gyrus, and medial septum diagonal band nuclei.

#### **Characteristics of Hypothalamic Cell Groups**

The hypothalamic cell groups are not at all homogeneous, and in most cases subgroups can be distinguished. For example, in the arcuate nucleus, there are dopamine, GHRH, enkephalin, galanin, substance P,  $\gamma$ -aminobutyric acid (GABA), atrial natriuretic peptide, neurotensin, gastrin-releasing peptide (GRP), and glutamate-containing neurons. The suprachiasmatic nucleus, representing a key structure of the biological clock, is composed of vasopressin, vasoactive intestinal peptide, somatotropin release-inhibiting hormone (SRIH), substance P, GABA, and GRP-containing nerve cells. The situation is further complicated by the colocalization of various neuropeptides in the same neurons. The functional significance of colocalization of substances in the nerve cells needs to be clarified.

The paraventricular nucleus is unique among hypothalamic cell groups in housing substantial populations of cells that participate in the control of anterior and posterior pituitary secretions and also is associated with the autonomic nervous system. This nucleus is the predominant source of CRH and TRH. Neurons of the cell group synthesize oxytocin and vasopressin for release into the general circulation from terminals in the posterior pituitary. A third major cell type contains neurons that give rise to long descending projections to the brainstem and spinal cord that include sensory and motor structures associated with the autonomic nervous system. These three visceromotor populations are essentially separate and exhibit a high degree of topographic organization. On this high degree of anatomical organization is imposed a somewhat imprecise manner of chemical coding.

The neurons of a cell group display rich intrinsic connections. A local network of fibers seems to be a common feature of hypothalamic nuclei. Local circuit neurons may synchronize the activities of peptidergic neurons in a hypothalamic nucleus to integrate or coordinate them as a functional unit.

## Hypothalamic Connections

### **Intrahypothalamic Connections**

A large number of intrahypothalamic connections exist between hypothalamic cell groups, and in most cases these connections are reciprocal. For example, arcuate neurons, in addition to projecting to the outer layer of the median eminence, project to several other

regions such as the ventromedial nucleus, lateral hypothalamic area, anterior hypothalamic area, and preoptic area and also receive afferents from the preoptic area, the ventromedial nucleus, and several other hypothalamic regions.

The abundant intranuclear connections, as well as the rich connections among the various hypothalamic cell groups, support the general impression that the hypothalamus should be considered a neuronal network of quasirandom internal connections. In this network, where the impulses leave the hypothalamus through the main axons, excitation can spread from a given focus in any direction and can establish an infinite number of closed, self-reexciting chains.

### **Extrahypothalamic Connections**

#### *Connections with Telencephalic and Diencephalic Structures*

A large number of regions project to the hypothalamus, including various parts of the limbic system (e.g., amygdaloid complex, hippocampus, septum), thalamus, basal ganglia, and cortex. Although the hypothalamus receives a large amount of sensory input through the regions mentioned, there are also some more or less direct pathways. There is a direct projection from the retina to the hypothalamus, primarily to the suprachiasmatic nucleus. The effect of light on the hypothalamus, particularly on its control of the anterior pituitary, may be mediated by this pathway. Inputs from the olfactory bulb have relatively free access to the hypothalamus, although direct connections from the olfactory bulb to the hypothalamus are not known. The piriform cortex, which receives fibers from the olfactory bulbs, sends projections directly to the hypothalamus. Other olfactory pathways reach the hypothalamus mainly via the amygdala.

The efferent pathways of the hypothalamus appear to reciprocate several of the major afferent hypothalamic connections. Many such reciprocating connections are contained in the medial forebrain bundle, the dorsal longitudinal fasciculus, and the stria terminalis. These pathways appear to close neuronal circuits between the hypothalamus and several of the limbic forebrain structures. The major brain structures receiving hypothalamic efferents are the amygdala, hippocampus, and septum.

#### *Hypothalamic Connections with the Brainstem*

The hypothalamus has extremely rich connections with the brainstem. A significant part of the ascending fibers are aminergic, and they terminate in various hypothalamic cell groups. Noradrenalin-containing fibers arise from the medulla oblongata (from the

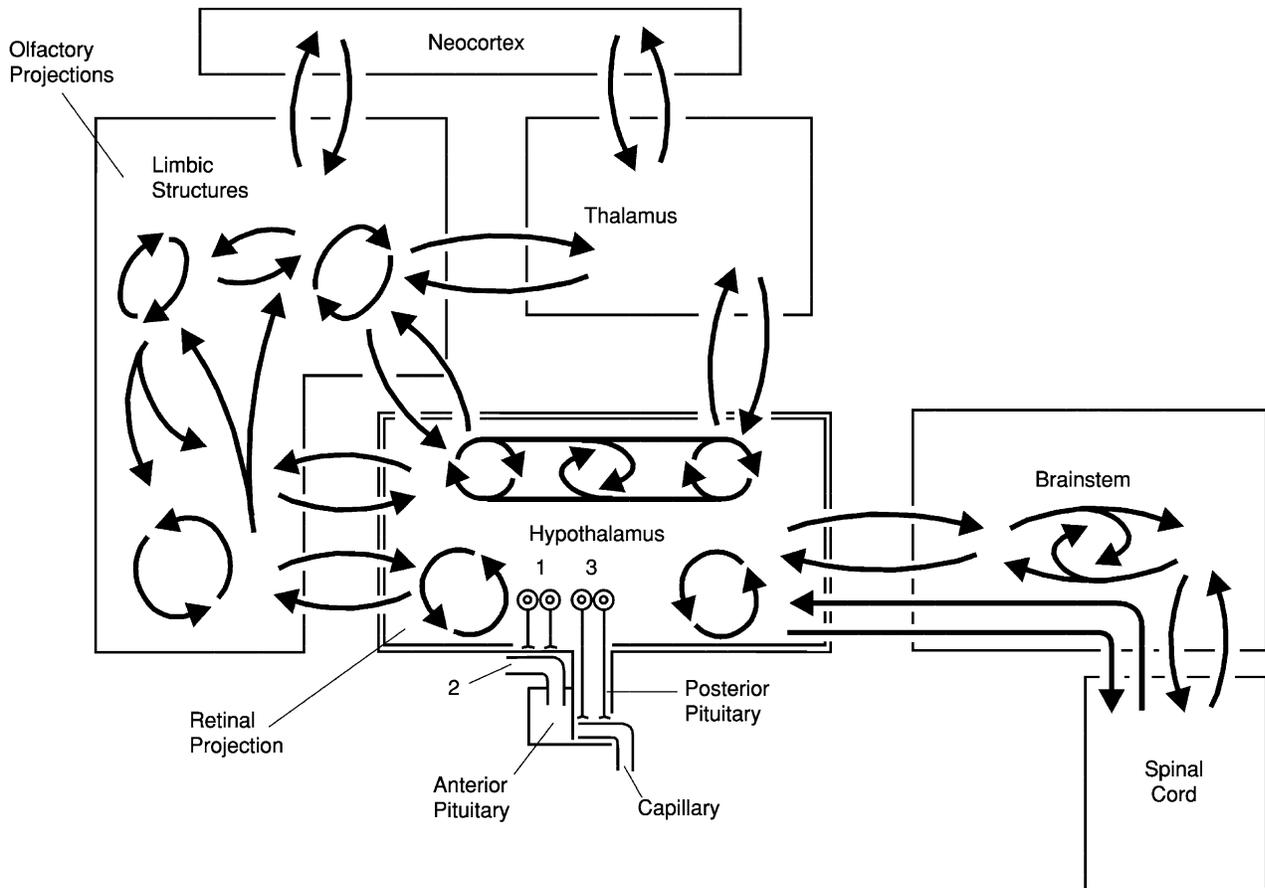
so-called A1 and A2 catecholaminergic cell groups) and pons (from the locus coeruleus). Adrenalin-synthesizing neurons are in the medulla oblongata (C1 and C2 adrenergic cell groups). Ascending serotonergic fibers arise from the pons and midbrain raphe nuclei.

Descending fibers from the hypothalamus, primarily from the paraventricular, arcuate, and medial preoptic nuclei, terminate in the brainstem and spinal cord.

Recent findings indicate that there are polysynaptic neuronal connections between hypothalamic structures and endocrine glands such as the gonads, adrenal, and pancreas. The neural connections of the hypothalamus are summarized in [Fig. 2](#).

The reciprocal connections of the hypothalamus with limbic forebrain structures and the brainstem are of such magnitude that it appears possible to interpret the hypothalamus, at least partly, as a way station in both the ascending and descending limbs of a polysynaptic neural circuit that extends between the limbic forebrain, on the one hand, and the primarily paramedian mesencephalic region, on the other. It may be assumed that the functional state of the hypothalamus is determined, to a significant extent, by the neural events that take place in the limbic structures and the lower brainstem, with both having a very integrated structural organization, including several reciprocal interconnections and neural circuits. In addition, they receive a vast amount of information from both the external and internal environments flowing in along neural and humoral pathways (there are hormone receptors in the hippocampus and amygdala). It should be mentioned that the hypothalamus itself also contains hormone receptors as well as several other types of receptors.

The extremely complex neuronal network of the hypothalamus and its reciprocal connections suggests that, apart from a few exceptions such as the supraoptico- and paraventriculohypophyseal system, there are not well-defined regions and pathways that are specifically and exclusively concerned with a discrete hypothalamic function. Of course, this does not exclude the predominance of one or the other hypothalamic area in the involvement of a particular hypothalamic function. Instead of a mosaic-type pattern, the hypothalamus can rather be envisaged as some kind of computer. This computer has a number of built-in programs, and its elements are involved in several processes. It elaborates the solution for each actual situation on the basis of a wealth of information that is partly stored and partly streaming in continuously. The results are then



**Figure 2** A simplified drawing of the bidirectional neural connections within the hypothalamus (intranuclear and internuclear) and between hypothalamic and extrahypothalamic structures. The relations of the hypothalamus to the anterior and posterior pituitary are also indicated. (1) Neurons synthesizing hypophysiotropic neurohormones. (2) Portal capillary. (3) Neurons of the supraoptic and paraventricular nuclei terminating in the posterior pituitary.

distributed over a number of neural and humoral output channels.

### **Major Neuroanatomical Pathways Connecting the Hypothalamus with Extrahypothalamic Structures**

#### *Medial Forebrain Bundle*

The medial forebrain bundle running longitudinally and occupying the majority of the lateral hypothalamus contains a high number of fibers with different destinations from hypothalamic nuclei and from limbic, cortical, and brainstem structures directed to hypothalamic cell groups. It also contains many fibers that just pass through the hypothalamus.

#### *Stria Terminalis*

The stria terminalis is a major pathway between the amygdala and the hypothalamus that provides reciprocal connections between the two structures.

#### *Fornix*

The fornix is a main link of the limbic system connecting the hippocampus, the septum, and the mammillary body. Some of its fibers deviate from the main bundle and terminate in the preoptic area as well as around the ventromedial nucleus of the hypothalamus.

#### *Medial Corticohypothalamic Tract*

The medial corticohypothalamic tract connects the hippocampus with the arcuate, ventromedial, and ventral premammillary nuclei.

#### *Dorsal Longitudinal Fasciculus*

The dorsal longitudinal fasciculus contains ascending and descending fibers connecting the dorsal hypothalamus and posterior hypothalamus with the periaqueductal central gray of the mesencephalon.

## HYPOTHALAMUS–ANTERIOR–PITUITARY SYSTEM

### Neurovascular Contact between the Hypothalamus and Anterior Pituitary

The connections between the hypothalamus and anterior pituitary are neurovascular (Fig. 3). Hypophysiotropic substances, called tropic hormone-releasing hormones (factors) or release-inhibiting hormones (factors), are produced by the hypothalamus. The chemically identified hypophysiotropic substances are CRH or corticotropin-releasing factor (CRF); GnRH, also called luteinizing hormone-releasing hormone (LHRH) or gonadoliberin; GHRH; SRIH, somatostatin, or growth hormone-inhibiting factor (GIF); TRH; and prolactin-inhibiting factor (PIF), at least one of which is dopamine (DA). These substances are transported by the hypophyseal portal vascular system to the anterior pituitary cells. The median eminence and the proximal part of the pituitary stalk are the site where the axons of the neurons synthesizing the hypophysiotropic neurohormones are very close to the portal vessels.

### Location of Neurons Synthesizing the Hypophysiotropic Neurohormones and Projecting to the Median Eminence

#### Corticotropin-Releasing Hormone

The most prominent CRH-containing cell group of the hypothalamus is the paraventricular nucleus, mainly its medial parvocellular part. The vast majority of the CRH terminals in the median eminence and pituitary stalk arise from here.

#### Gonadotropin-Releasing Hormone

In general, there are two areas that contain a significant amount of immunoreactive GnRH nerve cell bodies

projecting to the median eminence: (1) the septal-preoptic–suprachiasmatic region and (2) the mediobasal area of the middle hypothalamus and posterior hypothalamus, especially the arcuate and premammillary nuclei. But there are great variations in the number of such cells in these two regions of various species. In humans and primates, they are concentrated mainly in the second region. In the rat, many GnRH cells are in the medial preoptic area, the diagonal band of Broca, and the septal nuclei, whereas some cells are in the anterior hypothalamic area.

GnRH neurons originate in the medial olfactory placodal epithelium of the developing nose, migrate across the nasal septum, and enter the forebrain with the nervus terminalis, a cranial nerve that is a part of the accessory olfactory system and projects directly from the nose to the septal-preoptic area and hypothalamus. This migratory route for GnRH-expressing neurons could explain the deficiency of gonadotropins seen in hypogonadotropic hypogonadism with anosmia.

#### Growth Hormone-Releasing Hormone

The majority of the GHRH immunoreactive neurons projecting to the median eminence are in the arcuate nucleus.

#### Somatotropin Release-Inhibiting Hormone

SRIH immunoreactive neurons terminating in the median eminence are located mainly in the medial preoptic and anterior periventricular areas.

#### Thyrotropin-Releasing Hormone

Although TRH immunoreactive neurons are widely distributed in the central and peripheral nervous systems, those projecting to the median eminence are gathered mostly in the medial parvocellular division of the paraventricular nucleus.

#### Prolactin-Inhibiting Factor

At least one PIF is dopamine and is produced by the tuberoinfundibular dopaminergic neurons. These neurons are situated in the arcuate nucleus and in the ventral part of the anterior periventricular nucleus.

The location of neurons synthesizing the hypophysiotropic neurohormones and projecting to the median eminence is summarized in Fig. 4.

The hypophysiotropic neurons receive very significant neural input mediated by several chemical messengers. Various peptidergic, monoaminergic, and GABAergic axons terminate on perikarya and dendrites of these nerve cells, providing the structural basis for the assumption that the actions of the neurotransmitters or neuromodulators are at least partly

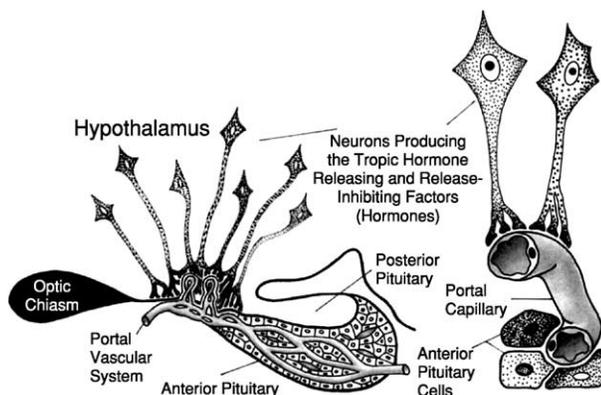
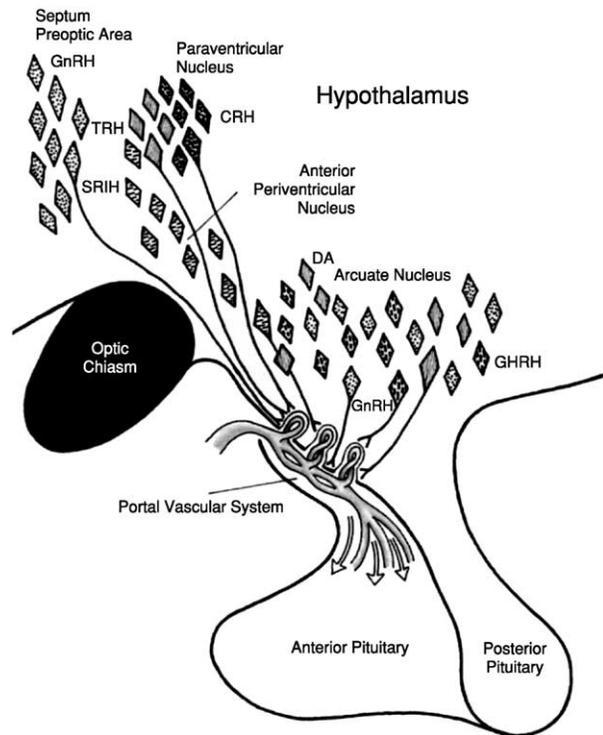


Figure 3 Schematic illustration of the neurovascular contact between the hypothalamus and the anterior pituitary.



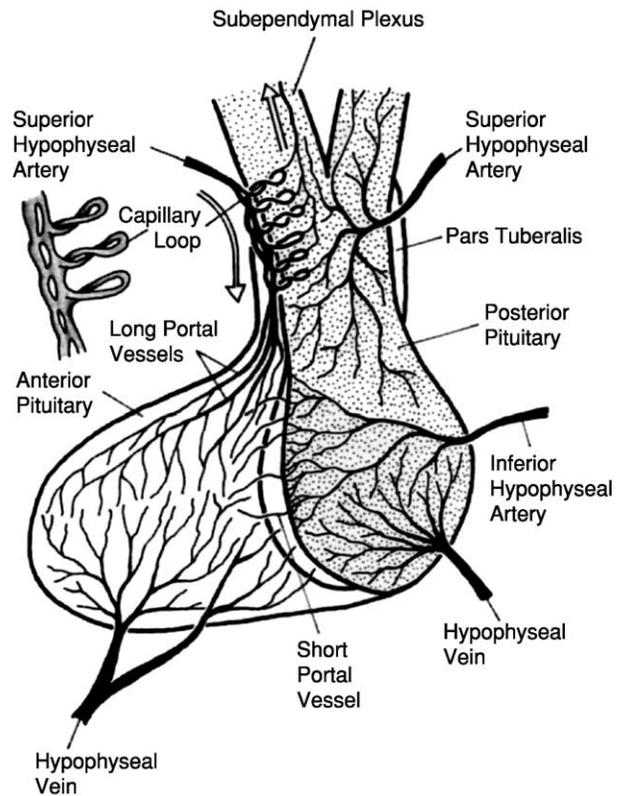
**Figure 4** Schematic drawing of midsagittal section of the hypothalamo-pituitary unit showing the location of neurons producing hypophysiotropic neurohormones and projecting to the outer layer of the median eminence. CRH, corticotropin-releasing hormone; DA, dopamine; GHRH, growth hormone-releasing hormone; GnRH, gonadotropin-releasing hormone; SRIH, somatostatin release-inhibiting hormone; TRH, thyrotropin-releasing hormone.

exerted directly on the structures producing the hypophysiotropic neurohormones. In addition, axons containing one or the other hypophysiotropic neurohormone form synaptic connections with neurons synthesizing the same peptide. This may be the morphological basis for an ultrashort feedback mechanism or may indicate an intrinsic circuit.

However, it should be kept in mind that neurons containing the trophic hormone-releasing or release-inhibiting neurohormones are widely distributed in the central nervous system; some of them are even present in other tissues. Not all of these neurons in the brain terminate in the hypothalamic median eminence and pituitary stalk; instead, some project to other brain structures.

### Hypophysial Portal Vascular System

Besides the trophic hormone-releasing and release-inhibiting hormones, the portal vascular system represents the key structure required for the operation of



**Figure 5** Schematic illustration of the hypophysial portal vascular system.

the neurohumoral (neurovascular) mechanism controlling pituitary trophic functions. It transports the substances released from the nerve terminals in the median eminence to the pituitary.

The main features of the hypophysial portal vascular system are the following (Fig. 5). The so-called superior hypophysial arteries form a dense plexus, largely of precapillary character, within the so-called pars tuberalis, a small part of the pituitary gland. This plexus is especially dense on the contact surface between the median eminence and the pars tuberalis (mantleplexus). From this plexus arise the capillary loops that penetrate into the tissue of the median eminence and infundibular stem. The mantleplexus and capillary loops drain toward the portal vessels (some of the capillary loops drain toward the subependymal plexus of the third ventricle) that lie on the ventral surface of the pituitary stalk. These vessels are called the long portal vessels. Part of the blood from the posterior pituitary reaches the anterior pituitary by way of vessels known as short portal vessels. Each portal vessel supplies a certain part of the pituitary. Anastomoses between these vessels are rare. The majority of the portal blood is directed from

the median eminence toward the pituitary, but some blood may flow in the reverse direction, toward the hypothalamus.

The presence of tropic hormone-releasing and release-inhibiting substances in the portal blood is well documented, as is the fact that the concentration of these substances is much higher in the portal blood than in the peripheral plasma and that changes occur under certain experimental conditions.

### Structure of the Median Eminence

The median eminence arising from the ventral surface of the tuberal region of the hypothalamus is a slight midline prominence. It continues into the pituitary stalk and represents the contact area between the nerve terminals of the neurons synthesizing the hypophysiotropic neurohormones and the precapillaries and capillaries of the portal vascular system. The inner surface of the median eminence is covered by ependymal cells. Two layers can be distinguished in the median eminence. The inner layer contains the fibers of the supraoptico- and paraventriculo-hypophysial system terminating in the posterior pituitary. The outer layer contains the tropic hormone-releasing and release-inhibiting hormones and the vessels of the portal vascular system. It must be mentioned that in addition to the terminals of the neurons synthesizing the hypophysiotropic neurohormones, there are many other neurons containing chemical messengers different from these compounds (most of them are also peptides, but “classic” neurotransmitters are also present), which also terminate in the median eminence. The functional significance of these other chemical messengers is not known. The possibility of interactions of the various substances at the median eminence level exists.

### HYPOTHALAMUS–POSTERIOR PITUITARY SYSTEM

Magnocellular neurons of the supraoptic and paraventricular nuclei and parvocellular neurons of the paraventricular nucleus project to the posterior pituitary, forming the so-called supraoptico- and paraventriculo-hypophysial tract (Fig. 6). The neurons synthesize oxytocin and vasopressin. The hormones are synthesized in the cell bodies of the nerve cells and are transported down the axons of these neurons to their endings in the posterior pituitary. Some of the neurons make oxytocin and others synthesize vasopressin. Oxytocin- and vasopressin-containing cells are evident in both cell groups. Vasopressin- and

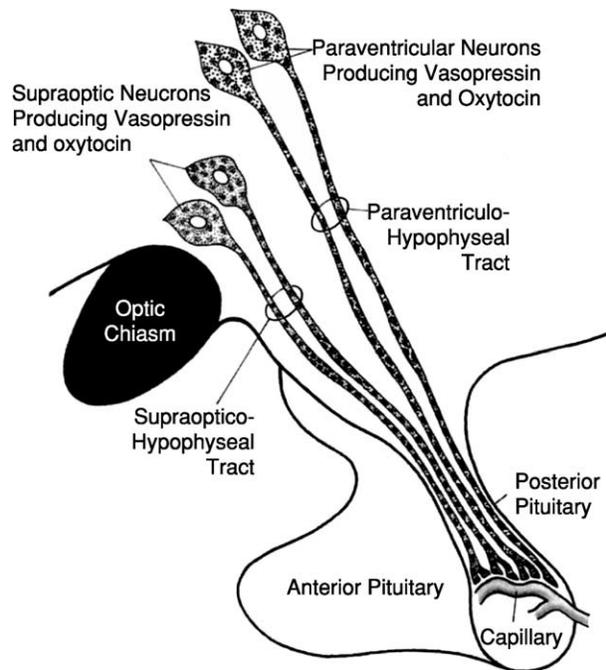


Figure 6 The hypothalamus–posterior–pituitary system.

oxytocin-producing neurons of the paraventricular nucleus project not only to the posterior pituitary but also to the brainstem and spinal cord and may be involved in cardiovascular control.

### Neurosecretion

The term “neurosecretion” was originally coined to describe the secretion of hormones by neurons, but the term is now somewhat misleading because it became evident that nearly all neurons secrete chemical messengers.

### BLOOD SUPPLY

The hypothalamic arteries arise directly from the circle of Willis (circulus arteriosus Willisii). Six major groups of arteries supply the hypothalamus. The arteries cover each other like shells in a mediolateral direction. The vessels entering at the midline supply the medial and basal parts of the hypothalamus, whereas those entering laterally supply the lateral and dorsal parts of the hypothalamus. None of the hypothalamic nuclei is supplied by a single artery.

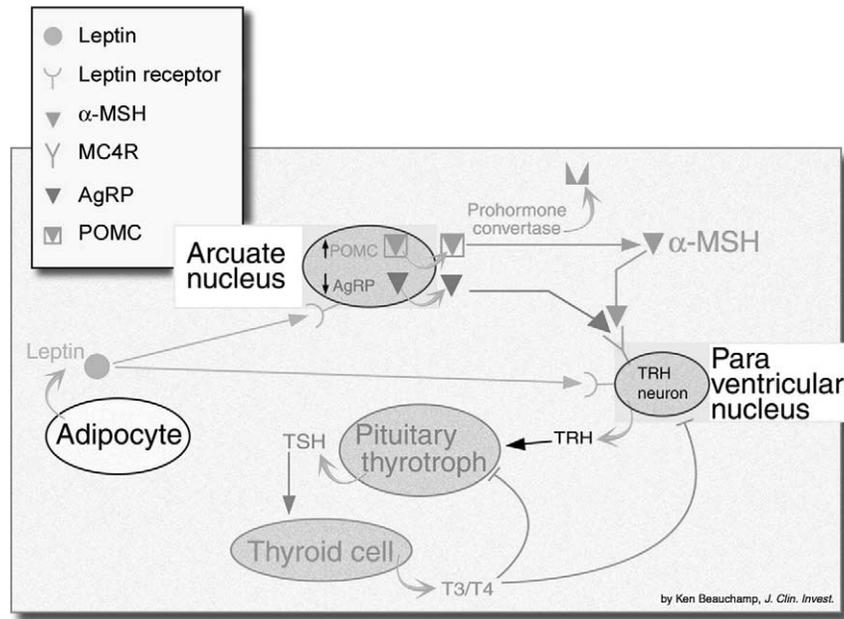
Most of the hypothalamic veins enter into the anterior cerebral, basilar, and interpeduncular veins. The anterior cerebral vein drains into the basilar vein. The venous blood is drained by the basal vein, which enters the great cerebral vein of Gallen.

### See Also the Following Articles

Hypothalamic Disease • Hypothalamic Hypogonadism • Hypothalamic Hypothyroidism • Hypothalamic Regulation of Appetite and Obesity • Hypothalamic-Pituitary Unit • Hypothalamus–Pituitary–Thyroid Axis • Pituitary Gland Anatomy and Embryology • Pituitary Tumors, Molecular Pathogenesis

### Further Reading

- Halász, B. (1994). Hypothalamo–anterior pituitary system. *In* “The Pituitary Gland” (H. Imura, ed.), 2nd ed., pp. 1–28. Raven Press, New York.
- Haymaker, W., Anderson, E., and Nauta, W. J. H. (1969). “The Hypothalamus.” Charles C Thomas, Springfield, IL.
- Morgane, P. J., and Panksepp, J. (eds.) (1979). “Handbook of the Hypothalamus,” vol. 1: “Anatomy of the Hypothalamus.” Marcel Dekker, New York.
- Palkovits, M. (2000). Micro- and macroscopic structure, innervation, and vasculature of the hypothalamus. *In* “Neuroendocrinology in Physiology and Medicine” (P. M. Conn, and M. E. Freeman, eds.), pp. 23–40. Humana Press, Totowa, NJ.
- Swanson, L. W. (1987). The hypothalamus. *In* “Handbook of Chemical Neuroanatomy” (A. Björklund, T. Hökfelt, and L. W. Swanson eds.), “Integrated Systems of the CNS,” vol. 5 pp. 1–124. Elsevier, Amsterdam.
- Szentágothai, J., Flerkó, B., Mess, B., and Halász, B. (1972). “Hypothalamic Control of the Anterior Pituitary,” 3rd ed. Akadémiai Kiadó, Budapest, Hungary.



**Figure 1** Interactions between the fat-derived hormone leptin and TRH neurons in the PVN. A sufficient level of leptin signaling is needed to maintain TRH expression in the hypothalamic PVN, which is necessary for normal production of TSH and production of thyroid hormone by the thyroid. Two mechanisms may be involved. In one, leptin regulates arcuate nucleus neurons expressing pro-opiomelanocortin (POMC) (induced by leptin) and AgRP (suppressed by leptin). These arcuate neurons project to TRH neurons, where they influence TRH expression by  $\alpha$ -MSH (stimulatory) and AgRP (inhibitory) input on MC4Rs. Leptin may also act directly on TRH neurons through leptin receptors on these cells. In the absence of leptin signaling, the feedback action of T<sub>4</sub>/T<sub>3</sub> at the hypothalamic level is impaired. Hence, although levels of T<sub>4</sub>/T<sub>3</sub> may be low, TRH and TSH levels remain suppressed. Reprinted with permission from Flier *et al.*, *J. Clin. Invest.* **105**, 859–861 (2000).

hormone (TSH) secretion via specific high-affinity membrane receptors present on thyrotropes [TRH receptor type 1 (TRH-R1)]. The termination of TRH signals is probably mediated by a membrane-bound ectoenzyme termed TRH-degrading enzyme (TRH-DE) for its remarkable specificity to hydrolyze only TRH and closely related peptides. It is localized on lactotrophs, on which TRH-DE activity and mRNA levels are tightly regulated by thyroid hormone. In addition to TRH, thyroid hormone regulates TSH secretion via negative feedback action on TSH synthesis and release. T<sub>4</sub> is locally deiodinated by D2 to the biologically active thyroid hormone T<sub>3</sub>. The negative feedback effect of T<sub>3</sub> is exerted via TRs expressed in the anterior pituitary. In addition to TRH-R1 and TRs, the TSH receptor appears to be present in the human anterior pituitary as well. It is expressed in a subpopulation of folliculostellate cells and is assumed to allow for ultra-short loop regulation of TSH secretion. Putative recognition of the TSH-R by TSH-R antibodies may have clinical relevance in Graves' disease.

## Thyroid Gland

TSH acts on the thyroid gland by binding to the TSH-R, which is located on the plasma membranes of thyroid cells. It stimulates thyroglobulin gene expression, iodide uptake and organification, and hydrolysis of iodinated thyroglobulin stored in the luminal colloid. The endpoint of TSH action is the production and release of T<sub>4</sub> and some T<sub>3</sub> by the thyroid gland. In addition, the thyroid gland enlarges as a result of long-term stimulation by TSH. The molecular pathways involved in this trophic effect of TSH have been largely uncovered in the past decade. In addition to endocrine regulation, neural regulation of the thyroid gland is probably a modality. The thyroid gland is richly innervated by sympathetic, parasympathetic, and sensory nerve fibers that contain various neuropeptides, such as neuropeptide Y. Based on the distribution pattern of these fibers, a role for neural control in the regulation of local blood flow and in thyroid hormone synthesis and secretion has been postulated. Studies in rats have

shown multisynaptic contacts between hypothalamic nuclei and the thyroid gland. Thus, the hypothalamus may be involved in both neuroendocrine and neural regulation of the thyroid gland.

## CONCLUSION

The HPT axis is a neuroendocrine feedback system allowing the organism to keep serum thyroid hormone concentrations within a narrow range by negative feedback of thyroid hormone at the level of the hypothalamus and anterior pituitary. In addition, it responds to various internal and external conditions, such as food deprivation and illness, with major changes in serum thyroid hormone concentrations. These changes can be seen as a useful adaptation of the organism to its environment. At the level of the hypothalamus, studies have demonstrated the importance of the thyroid hormone receptor isoform TR $\beta$ 2 for negative feedback of thyroid hormone on TRH neurons in the PVN. The SCN generates the circadian rhythm in serum TSH, whereas the arcuate nucleus appears to be pivotal in the down-regulation of the HPT axis in response to food deprivation in rodents. Monosynaptic neural connections between the arcuate nucleus and TRH cells in the PVN are involved in this response. At the level of the pituitary, TRH signals are mediated by TRH-R1 and terminated by TRH-DE. The TSH receptor is expressed in folliculostellate cells of the human anterior pituitary,

which offers a substrate for ultra-short loop regulation of TSH secretion. At the level of the thyroid, TSH stimulates not only synthesis and release of thyroid hormone but also growth. The functional meaning of multisynaptic neural connections between the hypothalamus and the thyroid is unclear at present.

## See Also the Following Articles

Hypothalamic-Pituitary Unit • Hypothalamus, Anatomy of  
• Thyrotropin-Releasing Hormone (TRH) • TSH Function and Secretion

## Further Reading

- Abel, E. D., Ahima, R. S., Boers, M., Elmquist, J. K., and Wondisford, F. E. (2001). Critical role for thyroid hormone receptor  $\beta$ 2 in the regulation of paraventricular thyrotropin-releasing hormone neurons. *J. Clin. Invest.* **107**, 1017–1023.
- Ahima, R. S., Prabakaran, D., Mantzoros, C., Qu, D., Lowell, B., Maratos-Flier, E., and Flier, J. S. (1996). Role of leptin in the neuroendocrine response to fasting. *Nature* **382**, 250–252.
- Flier, J. S., Harris, M., and Hollenberg, A. N. (2000). Leptin, nutrition, and the thyroid: The why, the wherefore, and the wiring. *J. Clin. Invest.* **105**, 859–861.
- Kalsbeek, A., Fliers, E., Franke, A. N., Wortel, J., and Buijs, R. M. (2000). Functional connections between the suprachiasmatic nucleus and the thyroid gland as revealed by lesioning and viral tracing techniques in the rat. *Endocrinology* **141**, 3832–3841.
- Prummel, M. F., Brokken, L. J. S., Meduri, G., Misrahi, M., Bakker, O., and Wiersinga, W. M. (2000). Expression of the thyroid-stimulating hormone receptor in the folliculo-stellate cells of the human anterior pituitary. *J. Clin. Endocrinol. Metab* **85**, 4347–4353.

**Table 1 Causes of Hypothyroidism**

1. Central (hypothalamic/pituitary) hypothyroidism
a. Loss of functional tissue
Tumors (pituitary adenoma, craniopharyngioma, meningioma, germinoma, glioma, metastases)
Trauma (surgery, irradiation, head injury)
Vascular (ischemic necrosis, hemorrhage, stalk interruption, aneurysm of internal carotid artery)
Infections (abscess, tuberculosis, syphilis, toxoplasmosis)
Infiltrative (sarcoidosis, histiocytosis, hemochromatosis)
Congenital (pituitary hypoplasia, septo-optic dysplasia, basal encephalocele)
b. Functional defect in TSH biosynthesis and release
Mutations in genes encoding for TRH receptor, TSH- $\beta$ , or Pit-1
Drugs: dopamine, glucocorticoids, L-thyroxine withdrawal
2. Primary (thyroidal) hypothyroidism
a. Loss of functional thyroid tissue
Chronic autoimmune thyroiditis
Reversible autoimmune hypothyroidism (silent and postpartum thyroiditis, cytokine-induced thyroiditis)
Surgery and irradiation ( $^{131}\text{I}$ or external irradiation)
Infiltrative and infectious diseases, subacute thyroiditis
Thyroid dysgenesis
b. Functional defects in thyroid hormone biosynthesis and release
Congenital defects in thyroid hormone biosynthesis
Iodine deficiency and iodine excess
Drugs: antithyroid agents, lithium, natural and synthetic goitrogenic chemicals
3. "Peripheral" (extrathyroidal) hypothyroidism
Thyroid hormone resistance
Infantile hemangioma

childhood cases are caused by craniopharyngiomas (TSH deficiency in 53% of patients) or cranial irradiation (e.g., for dysgerminoma or hematological malignancies).

### Adulthood

Adult cases of central hypothyroidism are most frequently due to pituitary macroadenomas (TSH deficiency in 10–25% of patients), pituitary surgery, or irradiation. TSH deficiency sometimes disappears after selective removal of a pituitary adenoma. Cranial radiotherapy for brain tumors causes hypothyroidism in 65% of patients, depending on the radiation dose, and its onset varies between 1 and 26 years after the irradiation. Radiotherapy for pituitary tumors is followed by hypothyroidism in at least 15% of patients (up to 55% when combined with surgery). Less

common causes are severe head trauma (TSH deficiency in 85% of patients), ischemic necrosis of the pituitary from postpartum hemorrhage (syndrome of Sheehan) or severe shock, pituitary apoplexy (hemorrhage in a pituitary adenoma), and lymphocytic hypophysitis. Lymphocytic hypophysitis is likely an autoimmune disease, presenting as a pituitary mass with hypopituitarism predominantly in women during pregnancy and the postpartum period.

Dopamine infusions in critically ill patients may cause central hypothyroidism by inhibition of pituitary TSH secretion. A huge excess of glucocorticoids (e.g., in Cushing's syndrome) also dampens TSH secretion. Removal of the dopamine or steroid excess restores TSH release. A transient inhibition of TSH release is observed after withdrawal of long-term  $T_4$  treatment in TSH-suppressive doses, lasting for about 6 weeks.

### PRIMARY HYPOTHYROIDISM

The hallmark of primary hypothyroidism is an elevated serum TSH. The pituitary senses the lower plasma concentrations of  $T_4$  and responds by increasing the release of TSH. Primary hypothyroidism is a very prevalent disease worldwide, especially in iodine-deficient regions. It is also very common in iodine-sufficient regions. The incidence in the adult population is 4.1 cases per 1000 women per year and 0.6 cases per 1000 men per year. The most frequent causes are chronic autoimmune thyroiditis, followed by thyroidectomy and  $^{131}\text{I}$  treatment. The incidence of congenital hypothyroidism is approximately 1 in 3500 newborns.

#### Childhood

Congenital primary hypothyroidism can be caused by structural loss of thyroid tissue due to thyroid agenesis or an ectopic thyroid gland or by functional defects in thyroid hormone biosynthesis due to loss-of-function mutations in genes encoding for the TSH receptor, the sodium iodide symporter, thyroglobulin, or thyroid peroxidase. Neonatal mass screening programs will detect nearly all cases of congenital primary hypothyroidism. Chronic autoimmune thyroiditis is relatively rare during childhood.

#### Adulthood

Hypothyroidism secondary to chronic autoimmune thyroiditis is caused mainly by destruction of

thyrocytes. Its atrophic variant is much more common than the goitrous variant. The incidence is higher in women than in men and increases with advancing age. Still other variants are silent or painless thyroiditis and postpartum thyroiditis (occurring in 4–6% of women during the first year after delivery and in up to 25% of patients with type 1 diabetes mellitus), which are mostly self-limiting diseases with spontaneous recovery. Another type of reversible autoimmune hypothyroidism is induced by treatment with the cytokines interleukin-2 and interferon- $\alpha$ , associated with the occurrence of thyroid peroxidase antibodies.

Hypothyroidism after subtotal thyroidectomy is less common after surgery for nontoxic or toxic nodular goiter (15% of patients) than after surgery for Graves' hyperthyroidism (40% of patients after 10 years). Most patients become hypothyroid during the first year after surgery. Likewise, hypothyroidism after  $^{131}\text{I}$  therapy is less common in patients with toxic nodular goiter (6–13%) than in patients with Graves' hyperthyroidism (up to 70% after 10 years). External radiotherapy of the neck for lymphomas or head and neck cancer causes hypothyroidism in 25 to 50% of patients, depending on pretreatment with iodine-containing radiographic contrast agents, radiation dose, and duration of follow-up.

Hypothyroidism is rarely caused by thyroid infections or thyroidal infiltration during the course of invasive fibrosis, cystinosis, or amyloidosis. In contrast, hypothyroidism during the recovery phase of subacute thyroiditis is not uncommon. Iodine excess may induce hypothyroidism by failure to escape from the Wolff–Chaikoff effect. Sources of iodine excess include an iodine-rich diet (e.g., seaweed) and iodine-containing medications such as potassium iodide, kelp, topical antiseptics, radiographical contrast agents, and amiodarone. Lithium inhibits thyroidal iodide transport and release of thyroid hormones. Long-term lithium treatment results in goiter (up to 50% of patients) and hypothyroidism (20% of patients), usually occurring during the first 2 years of treatment and in patients with preexisting thyroid peroxidase antibodies.

## “PERIPHERAL” HYPOTHYROIDISM

Hypothyroidism due to abnormalities outside the hypothalamus, pituitary, and thyroid is rare. In thyroid hormone resistance, a genomic mutation of the thyroid hormone receptor  $\text{TR}\beta_1$  fails to transmit the proper signal, causing symptoms and signs of thyroid hormone deficiency in certain target tissues. Serum TSH and serum  $\text{T}_4$  and triiodothyronine ( $\text{T}_3$ ) are increased—a remarkable combination of test results. The disease is frequently detected only during adulthood. Three cases of infants with massive hepatic hemangioma and hypothyroidism with markedly elevated serum TSH have been described. The hypothyroidism was caused by high levels of type 3 iodothyronine deiodinase activity in the hemangioma tissue, which catalyzes the degradation of  $\text{T}_4$  into the inactive reverse  $\text{T}_3$  and the degradation of  $\text{T}_3$  into the inactive  $3,3'\text{-T}_2$ .

## See Also the Following Articles

Amiodarone and Thyroid • Craniopharyngiomas • Goitrogens, Environmental • Hypothalamic Hypothyroidism • Hypothyroidism, Congenital • Hypothyroidism, Diagnosis of • Hypothyroidism, Systemic Manifestations of • Hypothyroidism, Treatment of • Iodine

## Further Reading

- Gerstein, H. C. (1990). How common is postpartum thyroiditis? A methodologic overview of the literature. *Arch. Int. Med.* **150**, 1397–1400.
- Huang, S. A., Tü, H. M., Harvey, J. W., *et al.* (2000). Severe hypothyroidism caused by type 3 iodothyronine deiodinase in infantile hemangiomas. *N. Engl. J. Med.* **343**, 185–189.
- Prummel, M. F., and Laurberg, P. (2003). Interferon- $\alpha$  and autoimmune thyroid disease. *Thyroid* **13**, 547–552.
- Regal, M., Paramo, C., Sierra, S. M., and Garcia-Mayor, R. V. (2001). Prevalence and incidence of hypopituitarism in an adult Caucasian population in northwestern Spain. *Clin. Endocrinol.* **55**, 735–740.
- Vance, M. L. (1994). Hypopituitarism. *N. Engl. J. Med.* **330**, 1651–1662.

There remains some controversy regarding the influence of severity of CH on long-term outcome. Most of the previous studies have reported significant differences in the outcome of patients with severe CH reflected by the type of CH, low or undetectable serum thyroxine (T4) concentrations, or delayed skeletal maturation at diagnosis. However, there was a significant overlap of the IQ score ranges of patients with severe and moderate forms of CH. Moreover, the lack of significant difference in studies with an onset of treatment within the first 2 weeks of life indicated that the influence of severity of CH can be compensated by an early onset of treatment, keeping the period without thyroid hormones as short as possible. Earlier studies had already shown that in some patients with CH the lower socioeconomic score is correlated with significantly lower IQ scores. Moreover, the maintenance of euthyroidism, as assessed by normal thyroid-stimulating hormone (TSH) levels reflecting the compliance of the patient or the family, was shown to be correlated with significantly improved development.

## TREATMENT

During the first 10 years of neonatal screening, the recommendations for the initial L-thyroxine substitution dose ranged from 5 to 8  $\mu\text{g}/\text{kg}/\text{day}$ , but soon follow-up studies demonstrated that the use of higher initial doses that were used in many European countries, and that led to a more rapid normalization of T4 and especially TSH levels, were correlated with a better long-term outcome. Subsequently, official treatment recommendations were changed to 10 to 15  $\mu\text{g}/\text{kg}/\text{day}$ . Furthermore, studies comparing higher and lower doses of initial treatment demonstrated that by using the higher doses, no difference in the outcomes of patients with severe and moderate CH is detectable.

From these data, it can be concluded that the majority of patients with CH, in whom thyroid hormone substitution is started within the first 2 weeks of life with a dose of more than 10  $\mu\text{g}/\text{kg}/\text{day}$ , will achieve a normal development and long-term outcome irrespective of the severity of CH at diagnosis, provided that normal TSH levels as well as thyroid hormone levels in the upper normal range are maintained with a good compliance to treatment.

## DEFECTS OF THYROID HORMONE BIOSYNTHESIS

It is assumed that 15 to 25% of patients' primary CH is caused by loss-of-function mutations in genes

encoding proteins that are involved in the synthesis of thyroid hormones in thyroid follicular cells.

TSH is a prerequisite not only for normal thyroid hormone synthesis but also for proliferation. TSH acts via a seven-transmembrane glycoprotein receptor by activating G proteins. Accordingly, the inherited CH in the *hyt/hyt* mouse was shown to be caused by homozygosity for a loss of function mutation of the TSH receptor (P556L), leading to postnatal development of thyroid hypoplasia and severe CH. Partial resistance to TSH was identified in patients with euthyroid hyperthyrotropinemia. Subsequently, homozygosity or compound heterozygosity for several loss-of-function mutations of the TSH receptor has been identified in patients with CH. The phenotype includes a spectrum of severity ranging from mild forms to severe CH with "apparent athyrosis" on thyroid scan. With ultrasound studies, hypoplastic remnants of the thyroid are detectable; therefore, the phenotype is similar to patients with PAX-8 mutations. The mode of inheritance is autosomal recessive, but elevated TSH levels have been described in some of the heterozygous carriers.

Iodide transport at the basolateral membrane of the thyroid cell is the initial step of thyroid hormone biosynthesis. The transport is realized by a 13-transmembrane domain protein. After cloning of the gene encoding this protein, the sodium-iodide symporter (NIS), homozygous and compound heterozygous mutations of the gene have been identified in patients with CH. Different mutations result in reduced binding of  $\text{Na}^+$  or  $\text{I}^-$  to the symporter, disturbed ion translocation, truncated NIS proteins, or a defect of NIS trafficking to the thyroid cell membrane. It seems that thyroid enlargement is not present at birth and that the development of nodular goiter, which has been described in patients with NIS mutations, occurs in later life.

Thyroperoxidase (TPO), located at the follicular side of the apical membrane, is a key enzyme in thyroid hormone biosynthesis catalyzing the iodination of tyrosine residues of the thyroglobulin molecule and the coupling of iodotyrosines to triiodothyronine (T3) and T4 in the presence of hydrogen peroxide ( $\text{H}_2\text{O}_2$ ). The autosomal recessive inheritance of loss of function mutations of the TPO gene has been described in the majority of patients with CH, with a total organification defect demonstrated by perchlorate discharge tests. The systematic study of patients with severe CH and a normally developed or enlarged thyroid also revealed a high proportion (60%) of autosomal recessive inheritance of TPO mutations.

Thyroglobulin (Tg), a homodimeric glycoprotein, is involved in thyroid hormone synthesis and storage.

Autosomal recessive inheritance of Tg defects has been identified in only a few cases of CH, most likely due to the difficulties of performing mutational screening of the large Tg gene (>300 kb, 48 exons, open reading frame of 2748 AA). The human phenotype is characterized by variable degrees of severity of CH and goiter development when thyroid hormone replacement is delayed. Dominant inheritance of Tg mutations has been described as well, not in patients with CH but rather in patients with the development of goiter later in life.

The functional role of pendrin (PDS), an anion transporter, in the thyrocyte is the transport of iodide across the apical membrane into the follicular lumen. Pendrin is also expressed in the inner ear and kidney, where its exact functions are unknown. The phenotype of Pendred's syndrome is characterized by congenital sensorineural hearing loss, goiter, and (in only a minority of patients) overt CH. Autosomal recessive inheritance of mutations of the pendrin gene has been reported in familial cases of Pendred's syndrome in which some newborns presented with congenital deafness, hypothyroidism, and goiter.

The thyroid oxidases 1 (THOX1) and 2 (THOX2) have been identified as components of the H<sub>2</sub>O<sub>2</sub> generating system of the thyroid. The genes encode for two very similar proteins with a fivefold higher expression of THOX 2. The putative structure of the proteins predict seven transmembrane domains, four nicotinamide adenine dinucleotide phosphate (NADPH)-binding sites, and a flavine adenine dinucleotide-binding site. A screening of patients with CH with organification defects revealed one patient with permanent CH and a homozygous mutation resulting in a truncated protein lacking the hydrogen-generating domains. In addition, three patients with transient hypothyroidism during the neonatal period were identified with monoallelic loss-of-function mutations of THOX2.

## MOLECULAR DEFECTS OF THYROID ORGANOGENESIS

Even in patients with putative autosomal recessive defects of thyroid hormone biosynthesis, familial cases have been described only rarely and systematic molecular genetic studies of candidate genes of thyroid hormone biosynthesis in patients with normally developed or enlarged thyroid glands are scarce. Therefore, no data on the epidemiology of the various molecular defects of thyroid hormone biosynthesis or on the possible modes of inheritance are available.

Some studies have described autosomal recessive inheritance and familial dominant occurrence of thyroid dysgenesis. Moreover, an increased frequency of minor abnormalities of the development of the thyroid and pharyngeal derivatives has been described in first-degree relatives of patients with CH due to thyroid dysgenesis. Again, epidemiological data on the prevalence of familial thyroid dysgenesis are scarce, but it has become apparent that thyroid dysgenesis, at least in a subset of patients, is an inherited disorder.

Studies of mouse models with targeted disruption of genes involved in the development of the thyroid gland have provided insight into the molecular mechanisms of organogenesis and, thereby, the basis for molecular genetic studies in human patients affected by thyroid dysgenesis. In mice, normal organogenesis and migration have been shown to be dependent on the normal expression and interplay of at least three different transcription factors: NKX 2.1, PAX-8, and TTF-2. Furthermore, targeted mutagenesis of these transcription factors in mice has demonstrated associated developmental defects of other organs because none of these factors is exclusively expressed in the thyroid.

The PAX-8 gene belongs to a family of genes that is characterized by a highly conserved paired-box DNA-binding domain, which encodes for proteins that play an important role in the entire embryonic development. PAX-8 is expressed in the thyroid primordium, the mid- and hindbrain region, and the developing kidney. Mice homozygous for disruption of the PAX-8 gene are characterized by severe hypothyroidism and small hypoplastic thyroid remnants without follicular structures, whereas in heterozygous mice no abnormalities of thyroid development have been described. Screening for mutations in the PAX-8 gene of patients with CH has led to the identification of several patients with heterozygous mutations that have been inherited in a dominant fashion. Most patients do not present with other developmental defects, but in two unrelated male patients, one hypoplastic kidney and one renal agenesis were observed. The thyroid gland of the affected patients presents with different morphological phenotypes. Thyroid hypoplasia, cystic hypoplastic remnants, and ectopic thyroid were described, and the severity of hypothyroidism was mild to moderate.

FKHL15/FOXE1 (previously named TTF-2) belongs to a family of transcription factors characterized by a forkhead DNA-binding domain. FKHL15 is expressed in the thyroid, Rathke's pouch, pharyngeal structures, and hair follicles. In mice with homozygosity for targeted disruption of the *Ttf2* gene, both

thyroid agenesis and thyroid ectopy were identified, indicating that athyrosis and ectopy may be regarded as different degrees of severity of the same molecular defect in humans as well. Furthermore, these mice have a cleft palate that makes their feeding impossible; therefore, early neonatal death is unavoidable. Although screening of the FKHL15 gene of patients with CH without associated problems failed to demonstrate any mutation, the study of two siblings with so-called Bamforth syndrome, including athyrosis and CH, developmental delay, cleft palate, choanal atresia, bifid epiglottis, and spiky hair, demonstrated homozygosity of a loss-of-function mutation of the FKHL15 gene in both siblings. In another family, patients were described with a less severe and incomplete phenotype, indicating either partial activity of the gene or the presence of other modifiers. FKHL15 mutations seem to be a very rare cause of CH in humans, resulting in a specific syndrome with multiple manifestations in other organs.

The NKX2.1 (TTF-1, TITF-1, or T/ebp) gene encodes for a transcription factor of the homeobox domain containing genes of the NKX2 family. NKX2.1 is expressed in the thyroid, forebrain, basal ganglia, pituitary, and lung. Targeted disruption of both *Titf1* alleles leads to a complex phenotype of newborn mice. These mice die shortly after birth due to respiratory distress resulting from defective lung development with insufficient surfactant production, and they lack any thyroid tissue at birth (athyrosis). Because of the early death of homozygous mice, a study of hypothalamic–pituitary function or neurological testing could not be performed. The search

for mutations in the NKX2.1 gene in patients with CH did not reveal any abnormalities. In the investigation of the NKX2.1 gene in patients with CH, where the long-term outcome, despite an early onset of treatment and adequate doses, was unfavorable due to pulmonary complications, severe muscular hypotonia, and neurological symptoms defined as choreoathetosis, heterozygous mutations of the NKX2.1 gene were identified. The phenotype of thyroid and pulmonary manifestations covers a wide spectrum ranging from hyperthyrotropinemia to severe CH due to thyroid agenesis and severe neonatal RDS requiring ventilation to a slight increase in pulmonary infections, whereas choreoathetosis presents with less phenotypical variation. Familial benign choreoathetosis without accompanying pulmonary or thyroid disorders has been attributed to NKX2.1 mutations as well. In light of the severe phenotype of NKX2.1 knockout mice, homozygosity for NKX2.1 mutations in humans is probably not viable. The mechanism by which heterozygous mutations cause the phenotype is most likely haploinsufficiency. Although heterozygous NKX2.1 mice have been reported to be unaffected previously, a more recent study described abnormalities of thyroid function and neurological development in heterozygous mice. See [Tables I to IV](#).

## CONCLUSION

Neonatal screening has resulted in early diagnosis and treatment of patients with CH. Subsequently, normal mental development and outcome have been documented in more than 90% of the patients who were

**Table I** Molecular Defects of Thyroid Hormone Biosynthesis

	NIS	PDS	Tg	TPO	THOX2
<b>Function</b>	Transport of iodine from the blood into the thyroid cell	Transport of iodide from the cytoplasm to the follicular lumen	Synthesis and storage of iodothyrosines	Iodination of Tg (organification) and coupling of iodothyrosines	Generation of H <sub>2</sub> O <sub>2</sub>
<b>Human thyroid phenotype</b>	Goiter Congenital Hypothyroidism in some patients	Goiter Congenital Hypothyroidism in some patients	Goiter Congenital Hypothyroidism in some patients	Goiter Congenital Hypothyroidism	Permanent congenital hypothyroidism (homozygous) Transient congenital hypothyroidism (heterozygous)
<b>Inheritance</b>	Autosomal recessive	Autosomal recessive	Autosomal recessive and autosomal dominant	Autosomal recessive	Autosomal recessive and autosomal dominant
<b>Chromosome</b>	19p13	7q31	8q24	2p25	15q21
<b>OMIM</b>	601843 600044	274600	188450	274500	607200

Table II Molecular Defects of Thyroid Development

	TSH receptor	PAX-8	TTF-2	NKX2.1
<b>Protein family</b>	G protein-coupled receptor	Paired domain	Forkhead domain	Homeodomain
<b>Expression pattern</b>	Thyroid Pituitary Hypothalamus? Adipose tissue	Transcription factor Thyroid Mid- and hindbrain Kidney	Transcription factor Thyroid Anterior pituitary	Transcription factor Thyroid Forebrain Pituitary Lung
<b>Phenotype in knockout mice</b>	Thyroid hypoplasia CH	Thyroid hypoplasia Early death	Thyroid agenesis or thyroid hypoplasia Cleft palate Early death	Thyroid agenesis Pituitary aplasia  Forebrain defects Disturbed lung development Neonatal death
<b>Human thyroid phenotype</b>	Thyroid hypoplasia Severe to moderate CH	Thyroid hypoplasia Cystic rudiments Ectopy	Thyroid agenesis	Thyroid agenesis Thyroid hypoplasia Normal thyroid Severe to moderate CH
<b>Manifestation in other human organs</b>	Hyperthyrotropinemia —	Severe to mild CH Developmental defect of the kidneys in some patients	Severe to moderate CH Cleft palate Bifid epiglottis Choanal atresia Spiky hair	Hyperthyrotropinemia Choreoathetosis Respiratory distress Pulmonary infections Mental retardation
<b>Inheritance</b>	Autosomal recessive	Autosomal dominant	Autosomal recessive	Autosomal dominant
<b>Chromosome</b>	14q31	2q13-14	9q22	14q12-q21
<b>OMIM</b>	275200	167415	602617	600635

followed up to adolescence. In up to 10% of patients detected by screening programs, neuropsychological development below the normal range was observed. The following factors, which were correlated with a less favorable outcome, have been identified:

- a delay in the onset of therapy,
- an insufficient initial thyroid hormone dose,
- a poor socioeconomic environment,
- poor compliance with therapy,
- other associated defects or complications.

If therapy was initiated during the first 2 weeks of life with an adequate L-thyroxine dose (> 10 µg/kg/day), no difference in the development of patients with severe or milder forms of CH was observed. From recent studies, there is evidence that persistent developmental delay, mental retardation, and

Table III Outcome Studies of CH

Study	Number of patients	Number of controls	IQ of patients (median)	IQ of controls (median)
New England, 1990	72	144	106	109
Glorieux, 1985	36	195	102	106*
Illig, 1986	40	40	104	108
Toublanc, 1990	49	52	116	118
Illicki, 1991	60	68	101	97
Fuggle, 1991	344	443	105	112*
Rovet, 1992	95	108	107	111
Kooistra, 1994	62	133	97	103

Table IV Outcome Studies in Correlation with Severity of CH and Age at Onset of Treatment

Study	Median IQ		Significance	Age at onset of therapy (median + range)
	Severe CH	Moderate CH		
Illicki, 1991	100	101	n.s.	12 (8–22)
Glorieux, 1992	89	104	P < 0.001	37 (10–65)
Kooistra, 1994	95	101	P < 0.001	23 (14–93)
Tillotson, 1994	108	115	P < 0.001	17 (1–173)
Own study, 1997	100	102	n.s.	9 (1–32)

neurological symptoms in some patients with early treatment of CH may be caused by the same molecular defect leading to an impaired development of the thyroid gland rather than by fetal or perinatal hypothyroidism. However, further research is needed to define these molecular mechanisms.

During the era of newborn screening, CH remains an apparent sporadic disease. However, with the normal outcome of patients with early diagnosis and treatment leading to a normal reproduction rate, it is possible that inheritance even of thyroid dysgenesis will become more prevalent. Virtually all industrialized countries now have neonatal screening programs for hypothyroidism in which capillary blood specimens (obtained from heel pricks) collected on filter paper soon after birth are analyzed for TSH or T4.

### See Also the Following Articles

Hypothyroidism, Causes of • Hypothyroidism, Congenital, Long-Term Follow-Up • Hypothyroidism, Congenital, Screening Programs • Hypothyroidism, Diagnosis of • Hypothyroidism, Systemic Manifestations of • Hypothyroidism, Treatment of • Thyroid Disease, Epidemiology of • TSH (Thyroid-Stimulating Hormone; Thyrotropin)

### Further Reading

Biebermann, H., Liesenkotter, K. P., Emeis, M., Obladen, M., and Grüters, A. (2001). Severe congenital hypothyroidism due to

a homozygous mutation of the  $\beta$ -TSH gene. *Pediatr. Res.* **46**, 170–173.

Biebermann, H., Schöneberg, T., Krude, H., *et al.* (1995). Mutations of the human thyrotropin receptor gene causing thyroid hypoplasia and congenital hypothyroidism. *J. Clin. Endocrinol. Metab.* **82**, 3471–3480.

Bikker, H., Vulmsa, T., Baas, F., *et al.* (1995). Identification of five novel mutations in the human TPO gene by denaturing gel electrophoresis. *Hum. Mut.* **19**(6), 9–16.

Clifton-Bligh, R. J., Wentworth, J., Heinz, P., *et al.* (1998). Mutation of the gene encoding human TTF-2 associated with thyroid agenesis, cleft palate, and choanal atresia. *Nat. Genet.* **18**, 399–401.

Coyle, B., Reardon, W., Herbrick, J. A., *et al.* (2001). Molecular analysis of the PDS gene in Pendred syndrome. *Hum. Mol. Genet.* **7**, 1105–1112.

Devos, H., Rodd, C., Gagne, N., *et al.* (1999). A search for possible mechanisms of thyroid dysgenesis: Sex ratios and associated malformations. *J. Clin. Endocrinol. Metab.* **84**, 2501–2506.

Fuijwara, H., Tatsumi, K., Miki, H., Harada, T., Miyai, K., Takai, S., and Amino, N. (1997). Congenital hypothyroidism caused by a mutation in the sodium-iodide symporter. *Nat. Genet.* **16**, 124–125.

Krude, H., Schutz, B., Biebermann, H., *et al.* (2002). Choreotetosis, hypothyroidism, and pulmonary problems due to NKX2.1 haploinsufficiency. *J. Clin. Invest.* **109**, 475–480.

Leger, J., Marinovic, D., Garel, C., Bonaiti-Pellie, C., Polak, M., and Czernichow, P. (2002). Thyroid developmental anomalies in first degree relatives of children with congenital hypothyroidism. *J. Clin. Endocrinol. Metab.* **87**, 575–580.

Macchia, P. E., Lapi, P., Krude, H., *et al.* (1998). PAX8 mutations associated with congenital hypothyroidism caused by thyroid dysgenesis. *Nat. Genet.* **19**, 83–86.

Van de Graaf, S. A., Ris-Stalpers, C., Veenboer, G. J., *et al.* (2001). A premature stop codon in TG mRNA results in familial goiter and moderate hypothyroidism. *J. Clin. Endocrinol. Metab.* **84**, 2537–2542.

colleagues defined a plasma total  $T_4$  at diagnosis of 43 nmol/L as the critical point below which the IQ becomes affected by CH.

Both Tillotson and colleagues and Derksen-Lubsen and Verkerk concluded that treatment factors did not appear to have an impact on developmental outcome. However, several recent more studies of children born during the 1990s and treated at a mean age of 14 days with 10 to 15  $\mu\text{g}/\text{kg}/\text{day}$  of  $T_4$  showed that the developmental gap that existed between children with severe CH and controls has been closed. Both early and high-dose treatment appear to be necessary, but this remains controversial.

## TREATMENT AND FOLLOW-UP

Most centers are now able to start treatment at a mean age of 9 to 14 days. The rationale for increasing the initial dose of  $T_4$  was that, with 5 to 6  $\mu\text{g}/\text{kg}/\text{day}$ , persistent elevations of thyroid-stimulating hormone (TSH) and persistent delay in bone maturation at 3 years were observed. The starting dose at many centers is now 10 to 15  $\mu\text{g}/\text{kg}/\text{day}$ . This regimen promptly normalizes plasma TSH but is associated with plasma-free  $T_4$  levels that can be above the reference range of most laboratories. However, it is important to recognize that the normal ranges of free  $T_4$  and of total triiodothyronine ( $T_3$ ) extend to much higher levels in infants than in older children and adults. In addition, with these starting doses, the mean plasma level of  $T_3$  remains within the normal range, and objective signs of hyperthyroidism have not been documented.

Treatment must be started without waiting for the results of confirmatory tests. (In cases of doubtful diagnosis, treatment can be safely stopped at 3 years of age.) TSH,  $T_3$ , and free  $T_4$  will be checked 2 to 4 weeks later to ensure that thyroid hormones have risen to normal values and that TSH has normalized. Later, clinical and biological follow-up at least at 3, 6, 9, and 12 months will allow adjustment of the treatment during the first year of life. TSH has to be kept in the normal range but does not need to be suppressed, and free  $T_4$  has to be kept near the upper limit of the normal range. Afterward, controls every 6 months will be sufficient so long as TSH and free  $T_4$  remain in the normal range. The dose of  $T_4$  will be titrated upward if the TSH rises and downward if TSH is consistently below the normal range and/or if  $T_3$  is high. Bone mass is not compromised by the large doses of  $T_4$  and high free  $T_4$  levels during the first year of life. There is anecdotal evidence, but no published data based on objective measurements, of

excessive irritability during the first 2 years of life when  $T_4$  levels are high.

Infants with dysmorphogenesis appear to generally need less  $T_4$  than those with dysgenesis. Children with CH fed with soybean formulas may require a higher dose of  $T_4$  because soy products interfere with  $T_4$  absorption.

If an orthotopic thyroid gland was present on the initial nuclear medicine scan, and if there is no secondary rise in TSH requiring an increase in thyroid hormone doses after it has returned to normal, transient hypothyroidism should be suspected. In these cases, treatment should be discontinued at 3 years of age, when thyroid hormone insufficiency no longer has irreversible effects on brain development, and thyroid hormones and TSH should be measured 3 to 4 weeks later. If the TSH has not risen to a level exceeding 10 U/L, it should be tested again in another 6 weeks. If it remains normal, transient hypothyroidism is confirmed.

## CONCLUSION

The most important aspect of the outcome of CH is developmental. Most children with CH have normal neuropsychological development (and normal physical growth) when they are treated early and with high doses of  $T_4$ . Further studies of neurophysiological functions that are more sensitive to under- or overtreatment than is the measurement of IQ will probably lead to greater individualization of initial dose recommendations. In the meantime, starting as early as possible with 10 to 15  $\mu\text{g}/\text{kg}/\text{day}$  appears to be safe and seems to allow all children with CH, including those with a severe form of the disease, to achieve their full intellectual potential. In view of the excellent outcome of children with CH detected and treated shortly after birth, testing of cognitive functioning can be limited to those children who are diagnosed late and/or who have school difficulties.

## See Also the Following Articles

Hypothalamic Disease • Hypothalamic Hypothyroidism • Hypothyroidism, Causes of • Hypothyroidism, Congenital, Screening Programs • Hypothyroidism, Diagnosis of • Hypothyroidism, Subclinical • Hypothyroidism, Systemic Manifestations of • Hypothyroidism, Treatment of • Thyroid Hormone Metabolism • TSH Function and Secretion

## Further Reading

American Academy of Pediatrics, Section on Endocrinology and Committee on Genetics, and American Thyroid Association

- Committee on Public Health. (1993). Newborn screening for congenital hypothyroidism: Recommended guidelines. *Pediatrics* **91**, 1203–1209.
- Bongers-Schokking, J. J., Koot, H. M., Wiersma, D., Verkerk, P. H., and de Muinck Keizer-Schrama, S. M. (2000). Influence of timing and dose of thyroid hormone replacement on development in infants with congenital hypothyroidism. *J. Pediatr.* **136**, 292–297.
- Derksen-Lubsen, G., and Verkerk, P. H. (1996). Neuropsychologic development in early treated congenital hypothyroidism: Analysis of literature data. *Pediatr. Res.* **39**, 561–566.
- Dubuis, J.-M., Glorieux, J., Richer, F., Deal, C. L., Dussault, J. H., and Van Vliet, G. (1996). Outcome of severe congenital hypothyroidism: Closing the developmental gap with early high dose levothyroxine treatment. *J. Clin. Endocrinol. Metab.* **81**, 222–227.
- Hrytsiuk, I., Gilbert, R., Logan, S., Pindoria, S., and Brook, C. G. (2002). Starting dose of levothyroxine for the treatment of congenital hypothyroidism: A systematic review. *Arch. Pediatr. Adolesc. Med.* **156**, 485–491.
- Klein, R. Z., and Mitchell, M. L. (2000). Neonatal screening. In “Werner and Ingbar’s The Thyroid (L. E. Braverman and R. D. Utiger, eds.), pp. 973–977. Lippincott Williams & Wilkins, Philadelphia.
- New England Congenital Hypothyroidism Collaborative. (1981). Effects of neonatal screening for hypothyroidism: Prevention of mental retardation by treatment before clinical manifestations. *Lancet* **2**, 1095–1098.
- Rovet, J. F., and Ehrlich, R. M. (1995). Long-term effects of L-thyroxine therapy for congenital hypothyroidism. *J. Pediatr.* **126**, 380–386.
- Tillotson, S. L., Fuggle, P. W., Smith, I., Ades, A. E., and Grant, D. B. (1994). Relation between biochemical severity and intelligence in early treated congenital hypothyroidism: A threshold effect. *Br. Med. J.* **309**, 440–445.
- Van Vliet, G. (1999). Neonatal hypothyroidism: Treatment and outcome. *Thyroid* **9**, 79–84.
- Van Vliet, G. (2001). Treatment of congenital hypothyroidism. *Lancet* **358**, 86–87.

newborns with sporadic permanent CH. Sensitive TSH assays have been shown to reduce the number of controls. This method is used in most countries of Europe and Japan as well as in developing countries with recent screening programs ongoing.

Some countries have a more ambitious program for screening: recognizing all forms of hypothyroidism by the measurement of T<sub>4</sub>, TSH, and thyroid hormone-binding protein (TBG), but it is expensive and offers fewer advantages of detecting hypothalamo-pituitary hypothyroidism or deficiency of TBG. T<sub>4</sub> level is determined first, and a confirmation by TSH measurement is performed in the lowest T<sub>4</sub> values (10–20%) (in the Netherlands, an assessment of TBG is realized on the 5% lowest T<sub>4</sub> values). This method is also used in the United States. Furthermore, the screening results obtained by these measurements can be used to monitor the iodine supply in the newborn population. This is an important issue because there is still iodine deficiency in many European countries.

## RESULTS

The mean incidence of CH observed in North America, Japan, and Europe was 1 per 3000 to 1 per 4000 births, without seasonal variations. CH had an ethnic background; the highest incidence, 1 per 1600 births, was observed in the Middle East due to the high prevalence of dysmorphogenesis induced by consanguinity. Conversely, some ethnic groups were less affected, for example, people of African ethnic origin (incidence varying from 1 per 10,000 to 1 per 20,000 births in the United States). Etiologies observed in all series were similar, with the type being revealed by ultrasound or radionuclide scans: athyreosis, 30 to 40%; ectopy, 40 to 50%; dysmorphogenesis, 15%; and hypoplasia, 5%. Athyreosis and ectopies had a feminine prevalence of 4 to 1; with dysmorphogenesis being autosomal recessive, the sex ratio was 1 to 1. This genetic background is an argument for a genetic origin of CH, but only 2% were family cases in the French population.

The babies underwent screening on day 3 in combination with other disease screening. They were recalled on days 8 to 10, and L-T<sub>4</sub> treatment was started (10–15 µg/kg/day). The treatment was managed on blood test results; the thyroxine should be in the normal range within 2 weeks, and TSH should be normalized within 4 weeks and should remain in the normal range afterward to obtain the best long-term results. Probably, the first patients detected by screening 20 years ago have not reached an optimal achievement due to late onset of treatment and no

tight biological control; nevertheless, newly detected patients with CH could have a normal outcome.

Two other categories have been raised from the screening results: false-positive and false-negative cases. First, false positives are babies recalled for CH without permanent CH. Causes are multiple (e.g., transient hypothyroidism in preterm infant, iodine overload in mother or baby, autoimmune disease in mother). Isolated elevated TSH persisting over months should be investigated by ultrasonography or scintigraphy to display big ectopies or dysmorphogenesis or to be checked for mutations of TSH receptor (TSHr). Second, false negatives are babies who are genuinely CH but who are not screened and recalled at the proper time; they are the failures of screening. The main causes of screening pitfalls are contaminated samples and human errors in either processing the sample or reporting the result. Furthermore, pituitary TSH deficiency cannot be detected by the measurement of TSH alone. CH could be missed by both methods in dysmorphogenesis with iodine overload or in infants having an exchange transfusion prior to the screening.

## PHYSIOPATHOLOGY

The physiopathology of CH by dysgenesis is still unknown. Recently, four genes—TTF1 and TTF2 (thyroid transcription factors 1 and 2), PAX8, and TSHr—have been shown to be mutated in some very specific instances. Mutations of these genes have been searched extensively in sporadic cases and were not found. Therefore, thyroid dysgenesis might be due either to a cascade of genes controlled by TTF1, TTF2, and PAX8 or to a multifactorial origin related to genetic background. Dysmorphogenesis is related to mutation in genes of the biochemical pathway: NIS (symporter Na/iodine), TPO (thyroid peroxidase), thyroglobulin, desiodases (DI and DII), and Pendred syndrome (Pendrine).

## See Also the Following Articles

Hypothalamic Hypothyroidism • Hypothyroidism, Causes of • Hypothyroidism, Congenital, Long-Term Follow-Up • Hypothyroidism, Diagnosis of • Hypothyroidism, Subclinical • Hypothyroidism, Systemic Manifestations of • Hypothyroidism, Treatment of

## Further Reading

Castanet, M., Polak, M., Bonaiti-Pellie, C., Lyonnet, S., Czernichow, P., and Leger, J. (2001). Nineteen years of national screening for congenital hypothyroidism: Familial cases with

- thyroid dysgenesis suggest the involvement of genetic factors. *J. Clin. Endocrinol. Metab.* **86**, 2009–2014.
- Derksen-Lubsen, G., and Verkerk, P. H. (1996). Neuropsychologic development in early treated congenital hypothyroidism: Analysis of literature data. *Pediatr. Res.* **39**, 561–566.
- Dubuis, J. M., Glorieux, J., Richer, F., Deal, C. L., Dussault, J. H., and Van Vliet, G. (1996). Outcome of severe congenital hypothyroidism: Closing the developmental gap with early high dose levothyroxine treatment. *J. Clin. Endocrinol. Metab.* **81**, 222–227.
- Kopp, P. (2002). Perspective: Genetic defects in the etiology of congenital hypothyroidism. *Endocrinology* **143**, 2019–2024.
- Toublanc, J. E. (1992). Comparison of epidemiological data on congenital hypothyroidism in Europe with those of other parts in the world. *Horm. Res.* **38**, 230–235.
- Working Group on Neonatal Screening of the European Society for Paediatric Endocrinology. (1999). Revised guidelines for neonatal screening programs for primary congenital hypothyroidism. *Horm. Res.* **52**, 49–52.

Table 1 Accuracy of 12 Symptoms and Signs in the Diagnosis of Primary Hypothyroidism (Percentages)

Symptoms and signs	Sensitivity	Specificity	Positive predictive value	Negative predictive value
Symptoms				
Hearing impairment	22	98	90	53
Diminished sweating	54	86	80	65
Constipation	48	85	76	62
Parasthesia	52	83	75	63
Hoarseness	34	88	73	57
Weight increase	54	78	71	63
Dry skin	76	64	68	73
Physical signs				
Slow movements	36	99	97	61
Periorbital puffiness	60	96	94	71
Delayed ankle reflex	77	94	92	80
Coarse skin	60	81	76	67
Cold skin	50	80	71	62

Note. Hypothyroid,  $\geq 6$  points; intermediate, 3–5 points; euthyroid,  $\leq 2$  points.

Source. Zulewski, H., *et al.* (1997). Estimation of tissue hypothyroidism by a new clinical score: Evaluation of patients with various grades of hypothyroidism and controls. *J. Clin. Endocrinol. Metab.* **82**, 771–776. © 1997, The Endocrine Society.

best single assay for detection of hypothyroidism. Test characteristics of the serum TSH assay for the diagnosis of abnormal thyroid function are 98.8% sensitivity, 94.3% specificity, 83.9% positive predictive value, and 99.7% negative predictive value. The high diagnostic accuracy of the TSH assay is caused by the exquisite sensitivity of the pituitary for small changes in serum thyroid hormone concentrations. Because of the negative feedback of thyroid hormone on the TSH release from the pituitary, a fall in serum thyroxine will result in elevated TSH levels in serum, and a rise in serum thyroid hormone concentrations will suppress serum TSH.

Figure 1 depicts a flow diagram for the biochemical diagnosis of hypothyroidism. If TSH is normal, euthyroidism is nearly certain and no further tests are necessary. However, central hypothyroidism (due to TSH deficiency) may be overlooked because serum

TSH in this condition is usually normal or decreased. Fortunately, clinical examination provides sufficient clues to suspect hypothalamic or pituitary disease such as symptoms arising from space-occupying lesions in the sella or from overproduction of pituitary hormones. As a rule, lack of gonadotropins occurs before the onset of TSH deficiency; therefore, the presence of regular menstrual periods in women or normal potency in men renders central hypothyroidism unlikely. The low incidence of central hypothyroidism (2.7 per 100,000 persons per year) does not warrant routine free thyroxine (FT<sub>4</sub>) measurements after a normal TSH test result.

If TSH is elevated, a decreased serum FT<sub>4</sub> value indicates overt hypothyroidism and a normal FT<sub>4</sub> points to subclinical hypothyroidism. In both instances, the thyroid gland itself is at fault (primary hypothyroidism). The very rare combination of elevated TSH and elevated FT<sub>4</sub> allows the diagnosis of thyroid hormone resistance or TSH-producing adenoma.

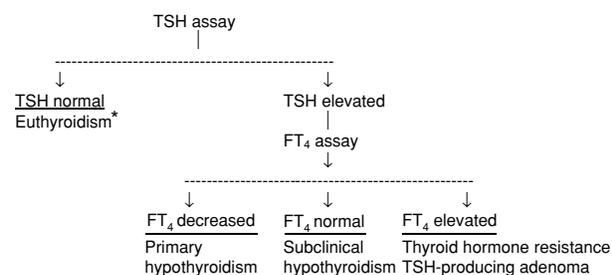


Figure 1 Flow diagram for the biochemical diagnosis of hypothyroidism. \*, see text.

## NOSOLOGIC DIAGNOSIS

The cause of the hypothyroidism may reveal itself in many instances from the history (e.g., recent delivery, exposure to iodine excess, family members with autoimmune thyroid disease, use of antithyroid drugs, thyroid surgery or <sup>131</sup>I therapy) and physical examination (although most patients will have no goiter).

The presence of thyroid peroxidase antibodies in serum indicates chronic autoimmune thyroiditis.

Thyroid scans usually show low and inhomogenous uptake of the radioisotope. Preserved thyroidal radioiodine uptake and homogenous distribution of the tracer increases the likelihood of reversible hypothyroidism (due to iodine excess).

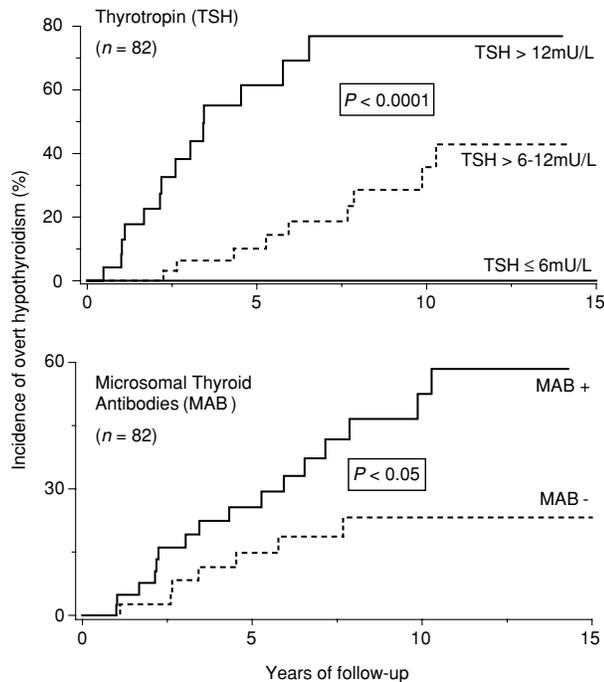
Spontaneous recovery of hypothyroidism occurs in the course of subacute and postpartum thyroiditis and sometimes during the first 6 months after subtotal thyroidectomy or  $^{131}\text{I}$  therapy. It is exceptional in chronic autoimmune thyroiditis.

### See Also the Following Articles

Hypothalamic Hypothyroidism • Hypothyroidism, Causes of • Hypothyroidism, Congenital • Hypothyroidism, Subclinical • Hypothyroidism, Systemic Manifestations of • Hypothyroidism, Treatment of

### Further Reading

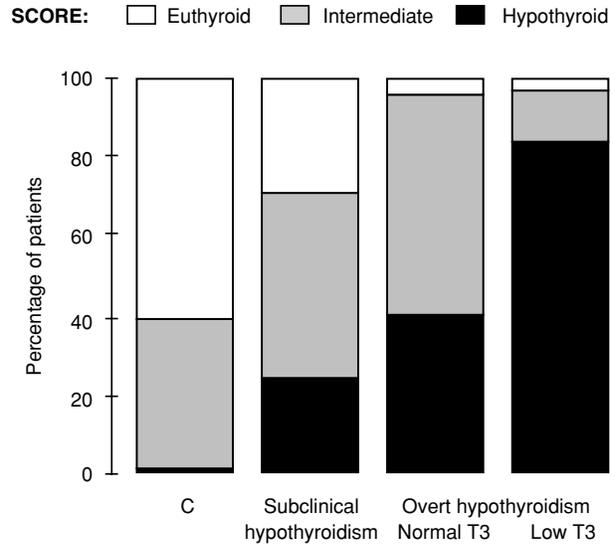
- Billewicz, W. L., Chapman, R. S., Crooks, J., *et al.* (1969). Statistical methods applied to the diagnosis of hypothyroidism. *Q. J. Med.* **38**, 255–266.
- Hollowell, J. G., Staehling, N. W., Flanders, W. D., *et al.* (2002). Serum TSH,  $T_4$ , and thyroid antibodies in the United States population (1988 to 1994): National Health and Nutrition Examination Survey (NHANES III). *J. Clin. Endocrinol. Metab.* **87**, 489–499.
- Seshadri, M. S., Samuel, B. U., Kanagasabapathy, A. S., and Cherian, A. M. (1989). Clinical scoring system in hypothyroidism: Is it useful? *J. Genet. Int. Med.* **4**, 490–492.
- Waise, A., and Belihetz, P. E. (2000). Unsuspected central hypothyroidism. *Br. Med. J.* **321**, 1275–1277.
- Wiersinga, W. M. (1989). The value of sensitive TSH measurements in clinical practice. *J. Endocrinol. Invest.* **9**(Suppl. 4), 67–76.
- Zulewski, H., Müller, B., Exer, P., *et al.* (1997). Estimation of tissue hypothyroidism by a new clinical score: Evaluation of patients with various grades of hypothyroidism and controls. *J. Clin. Endocrinol. Metab.* **82**, 771–776.



**Figure 1** Kaplan-Meier estimates of incidence of overt hypothyroidism according to TSH and microsomal thyroid antibody levels. Graphs show natural course and spontaneous evolution without treatment. Adapted from Huber, G., Staub, J. J., Meier, C., Mitrache, C., Guglielmetti, M., Huber, P., and Braverman, L. E. (2002). Prospective study of the spontaneous course of subclinical hypothyroidism: prognostic value of TSH, thyroid reserve, and thyroid antibodies. *J. Clin. Endocrinol. Metab.* **87**, 3221–3226.

later. In this survey, 97 females with subclinical or overt hypothyroidism have been detected with a prevalence of 9.3% and a calculated incidence of 0.41% per year. The risk factors for the development of hypothyroidism were a raised TSH level and/or positive antithyroid antibodies. For females with a raised serum TSH, a mean annual risk for developing hypothyroidism of 26% over 10 years could be calculated.

In a prospective study of 82 female patients with known subclinical hypothyroidism, we investigated the natural course of this syndrome with regular evaluations at yearly intervals. Over a mean period of 9.2 years, 28% of the patients progressed to overt hypothyroidism, 68% remained in the subclinical state, and 4% reverted to a normal TSH. The incidence of overt hypothyroidism was correlated with the initial serum TSH concentrations. The calculated 10-year rate of overt hypothyroidism was 0% for TSH levels of 4 to 6 mU/L but was 43% for values of 6 to 12 mU/L and 77% for TSH levels greater than 12 mU/L (Fig. 1).



**Figure 2** Clinical assessment of patients and controls with the clinical score by Zulewski and colleagues in subclinical hypothyroidism ( $n = 93$ ), overt hypothyroidism ( $n = 50$ ), and age-matched controls ( $n = 80$ ). Only the patients with the most severe hypothyroidism (with low triiodothyronine [T3]) are clearly hypothyroid. The patients with subclinical hypothyroidism or with less severe hypothyroidism (normal T3) show just a few clinical signs. Adapted from Zulewski, H., Müller, B., Exer, P., Miserez, A. R., and Staub, J. J. (1997). Estimation of tissue hypothyroidism by a new clinical score: Evaluation of patients with various grades of hypothyroidism and controls. *J. Clin. Endocrinol. Metab.* **82**, 771–776.

## CLINICAL MANIFESTATIONS AND BENEFITS OF TREATMENT

### Symptoms

One of the most controversial aspects concerning subclinical hypothyroidism is whether affected patients are symptomatic and so may benefit from thyroid hormone replacement. Based on case control studies, nearly 30% of patients with subclinical hypothyroidism may have symptoms that are suggestive of thyroid hormone deficiency (Fig. 2). Symptoms and signs of hypothyroidism may be very vague and non-specific and so are easily overlooked in an individual patient. Five randomized, placebo-controlled intervention trials over a period of 6 to 12 months were published. These studies support the finding that patients with subclinical hypothyroidism may indeed have, in part, specific clinical signs and symptoms of hypothyroidism (Table I). Three of these studies reported significant improvement in signs and symptoms of hypothyroidism assessed by various clinical scores of hypothyroidism, whereas two other studies found no benefit of thyroxine therapy.

**Table 1** Effect of Thyroxine Replacement Therapy on General Symptoms in Patients with Subclinical Hypothyroidism: Randomized Placebo-Controlled Trials

Treatment duration (months)	n	TSH level (mU/L)		Symptomatic response	Reference
		before L-T <sup>4</sup>	on L-T <sup>4</sup>		
12	17	10.8	2.6	Symptom score (Billewicz; $P < 0.05$ )	Cooper <i>et al. Ann. Int. Med.</i> (1984)
6	17	7.7	1.9	Symptom score (Billewicz; $P < 0.01$ )	Nyström <i>et al. Clin. Endocrinol.</i> (1988)
10	18	12.1	4.6	HRQL questionnaire ( $P = \text{n.s.}$ )	Jaeschke <i>et al. J. Gen. Int. Med.</i> (1996)
12	31	12.8	3.1	Symptom score (Billewicz, Zulewski; $P < 0.05$ )	Meier <i>et al. J. Clin. Endocrinol. Metab.</i> (2001)
6	23	8.0	3.4	HRQL questionnaire ( $P = \text{n.s.}$ )	Kong <i>et al. Am. J. Med.</i> (2002)

*Note.* The symptomatic response to thyroxine treatment was assessed by clinical scores based on Billewicz, W. Z., Chapman, R. S., Crooks, J., *et al.* (1969). Statistical methods applied to the diagnosis of hypothyroidism. *Q. J. Med.* **38**, 255–266, and Zulewski, H., Müller, B., Exer, P., *et al.* (1997). Estimation of tissue hypothyroidism by a new clinical score: Evaluation of patients with various grades of hypothyroidism and controls. *J. Clin. Endocrinol. Metab.* **82**, 771–776, as well as by quality-of-life scales.

### Cognitive Function and Neuropsychiatric Dysfunction

Several studies suggest that subclinical hypothyroidism is associated with neuropsychiatric disease, with higher scores on scales of depression and anxiety in affected patients. Neuropsychiatric parameters (i.e., reaction time and figure identification), as well as memory scores, were significantly improved in four intervention trials assessing the effect of thyroxine treatment in patients with subclinical hypothyroidism (mean TSH levels at baseline: 7 to 12 mU/L).

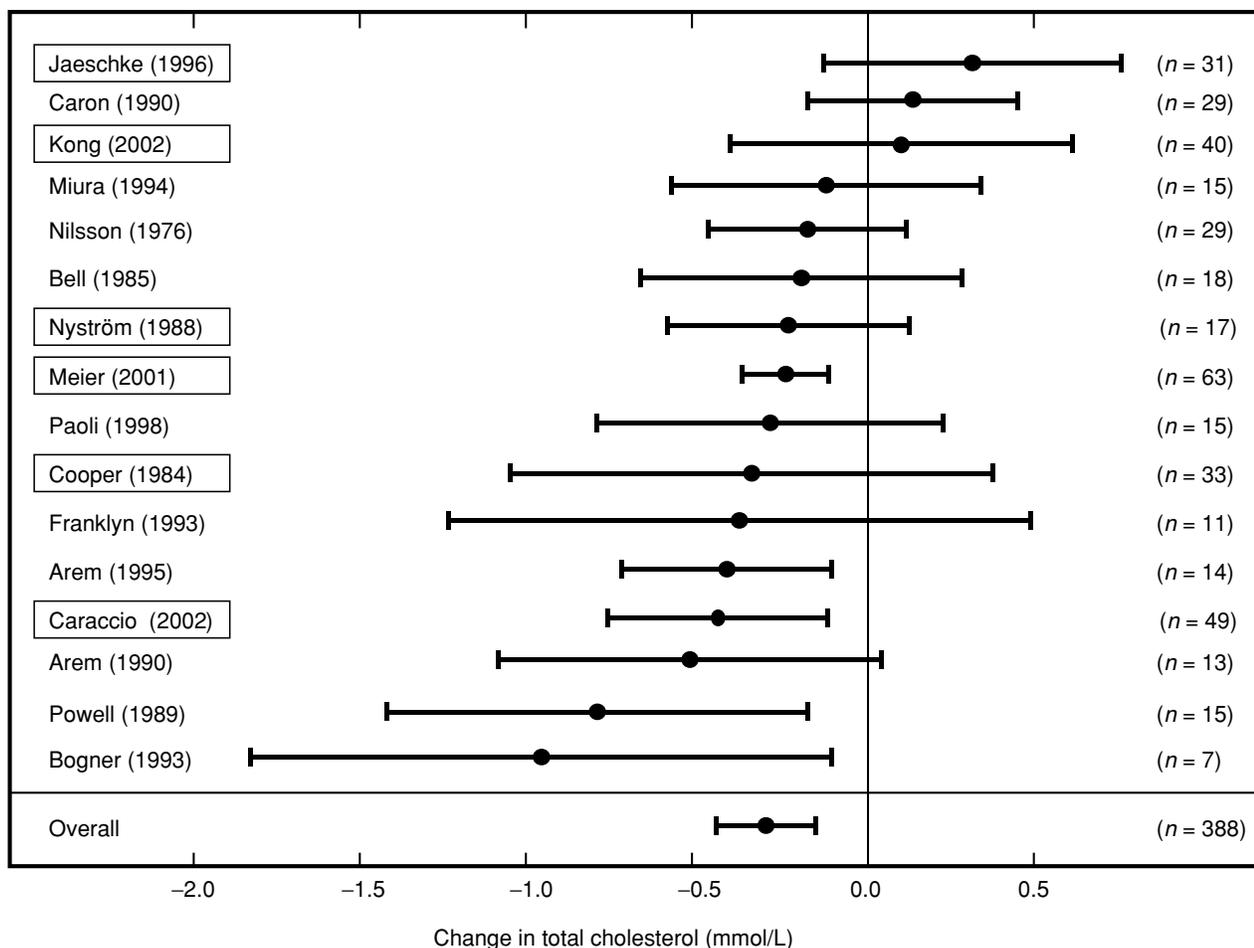
### Atherosclerosis and Cardiovascular Risk Factors

The relationship between subclinical hypothyroidism and atherosclerotic risk factors has been widely investigated, with major interest in serum lipid abnormalities. Several cross-sectional studies found serum lipid concentrations, mainly total cholesterol levels, to be within the normal range, whereas others detected significant elevations of total cholesterol and/or low-density lipoprotein (LDL) cholesterol concentrations, especially in smokers. In addition, reduced high-density lipoprotein (HDL) cholesterol levels were reported in some studies. Two meta-analyses reanalyzed the effect of several intervention trials and could demonstrate a beneficial effect of thyroxine on serum cholesterol levels. In analyzing 13 studies, Danese and co-workers reported favorable reductions of total cholesterol and LDL cholesterol levels, with mean decreases of 0.20 and 0.26 mmol/L, respectively. These findings were confirmed by two randomized, placebo-controlled studies reporting

significant reductions of total and LDL cholesterol concentrations in thyroxine-treated women with subclinical hypothyroidism (Fig. 3). Our “Basel Thyroid Study” is the first randomized trial that demonstrates a significant effect of thyroxine on total cholesterol and atherogenic LDL cholesterol levels. Based on this study, we calculated a relevant risk reduction of cardiovascular mortality of 9 to 31% in relation to the observed improvement in LDL cholesterol levels. Hence, subclinical hypothyroidism must be considered as a risk factor for the development of atherosclerosis and coronary heart disease. Many cross-sectional studies have suggested an association between subclinical hypothyroidism, or autoimmune thyroid disease, and atherosclerosis. A recent population-based survey reported that increased serum TSH is an independent risk factor for the development of aortic atherosclerosis and myocardial infarction. In this study, the risk of having atherosclerotic disease was twice as high in women with subclinical hypothyroidism than in euthyroid controls. The difference persisted after adjustment for age, body mass index, blood pressure, smoking status, and cholesterol levels. Further mechanisms are thought to be involved in the association between mild thyroid failure and cardiovascular disease. These include smoking, a hypercoagulable state, elevated lipoprotein and homocysteine levels, and endothelial effects of thyroid hormones.

### Others

Myocardial function has been shown to be slightly impaired in patients with subclinical hypothyroidism. The identified functional abnormalities include impaired myocardial contractility and diastolic function,



**Figure 3** Effect of thyroxine treatment on total cholesterol levels in subclinical hypothyroidism. Randomized placebo-controlled trials are shown in boxes. Adapted from Danese, M. D., Ladenson P. W., Meinert, C. L., and Powe, N. R. (2000). Clinical review 115: Effect of thyroxine therapy on serum lipoproteins in patients with mild thyroid failure—A quantitative review of the literature. *J. Clin. Endocrinol. Metab.* **85**, 2993–3001.

at rest or with exercise, with beneficial results after thyroid hormone replacement therapy. Furthermore, abnormalities in peripheral nerve function and neuromuscular activity, skeletal muscle abnormalities, intraocular pressure, and ovulatory dysfunction have been described.

## DIAGNOSTIC AND THERAPEUTIC RECOMMENDATIONS

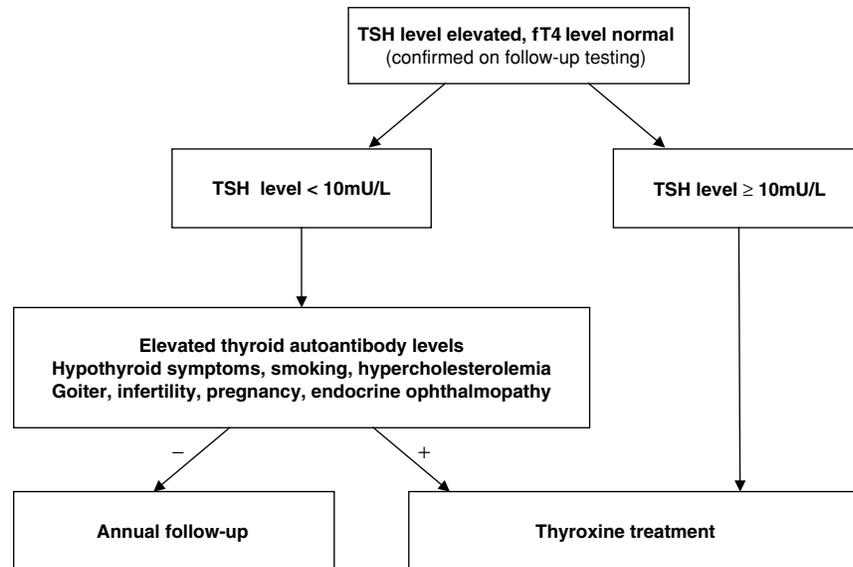
### Screening

Screening for thyroid disease is still a controversial issue. TSH screening in women over 35 years of age has been shown to be cost-effective. Some authors favor TSH screening for thyroid dysfunction in asymptomatic adults, although others do not favor

this policy because the effects of subsequent thyroxine therapy are not beneficial in all patients with subclinical hypothyroidism. Subclinical hypothyroidism is a frequent finding in women over 40 years of age (affecting about 10% of the female population in this age group), and its clinical presentation may be subtle. Therefore, TSH screening should be advocated at least by a case-finding approach, focusing on patients visiting their physicians for unrelated reasons. Because smoking impairs both thyroid hormone secretion and thyroid hormone action, smoking status should be considered in the evaluation of patients in whom hypothyroidism is suspected.

### Replacement Therapy with Thyroxine

The goal of treating patients with mild thyroid failure is to reverse clinical and metabolic alterations by



**Figure 4** Algorithm for the management of patients with subclinical hypothyroidism.

hormone supplementation and to prevent progression of the subclinical form to the overt stage of hypothyroidism, with its considerable morbidity and possible mortality. Thyroxine treatment should be used in patients who have elevated TSH levels greater than 10 mU/L and measurable circulating thyroid autoantibodies. Furthermore, thyroxine therapy is indicated for patients at risk with special clinical conditions such as goiter, thyroidectomy, depression, infertility, and endocrine ophthalmopathy and hypercholesterolemia, particularly in the presence of other cardiovascular risk factors such as smoking and hypertension. In patients with minimal or moderate TSH elevations (<10 mU/L) with no clinical or metabolic changes, treatment can be withheld, but annual follow-ups are required (Fig. 4).

Overtreatment can produce overt or subclinical hyperthyroidism with its effects on morbidity (e.g., atrial fibrillation, osteopenia) and mortality. Therefore, fine-tuning of thyroxine replacement therapy with the goal of restoring serum TSH to physiological euthyroid levels is mandatory.

### See Also the Following Articles

Hypothalamic Hypothyroidism • Hypothyroidism, Causes of • Hypothyroidism, Congenital, Long-Term Follow-Up • Hypothyroidism, Congenital, Screening Programs • Hypothyroidism, Diagnosis of • Hypothyroidism, Systemic

Manifestations of • Hypothyroidism, Treatment of • Thyroid Hormone Action

### Further Reading

- Ayala, A. R., Danese, M. D., and Ladenson, P. W. (2000). When to treat mild hypothyroidism. *Endocrinol. Metab. Clinics North America* **29**, 399–415.
- Chu, J. W., and Crapo, L. M. (2001). The treatment of subclinical hypothyroidism is seldom necessary. *J. Clin. Endocrinol. Metab.* **86**, 4591–4599.
- Cooper, D. S. (2001). Clinical practice: Subclinical hypothyroidism. *N. Engl. J. Med.* **345**, 260–265.
- Helfand, M., and Redfern, C. C. (1998). Clinical guideline, part 2: Screening for thyroid disease—An update. *Ann. Int. Med.* **129**, 144–158.
- Kahaly, G. J. (2000). Cardiovascular and atherogenic aspects of subclinical hypothyroidism. *Thyroid* **10**, 665–679.
- McDermott, M. T., and Ridgway, E. C. (2001). Subclinical hypothyroidism is mild thyroid failure and should be treated. *J. Clin. Endocrinol. Metab.* **86**, 4585–4590.
- Meier, C., Staub, J. J., Roth, C. B., Guglielmetti, M., Kunz, M., Miserez, A. R., Drewe, J., Huber, P., Herzog, R., and Müller, B. (2001). TSH-controlled L-thyroxine therapy reduces cholesterol levels and clinical symptoms in subclinical hypothyroidism (Basel Thyroid Study): A randomized, double-blind, placebo-controlled trial. *J. Clin. Endocrinol. Metab.* **86**, 4860–4866.
- Ross, D. S. (2000). Subclinical hypothyroidism. In “Werner and Ingbar’s The Thyroid” (L. E. Braverman and R. U. Utiger, eds.), pp. 1001–1006. Lippincott Williams & Wilkins, New York.
- Staub, J. J. (1998). Minimal thyroid failure: Effects on lipid metabolism and peripheral target tissues. *Eur. J. Endocrinol.* **138**, 137–138.

patients who develop hypothyroidism rapidly (i.e., when replacement therapy is discontinued in patients with primary hypothyroidism or after the gland is surgically removed) have more symptoms. In such patients, manifestations of overt hypothyroidism are present by 6 weeks. In general, older patients tend to have fewer symptoms and signs of hypothyroidism than do young adults.

In adults, common features of hypothyroidism include easy fatigability, tiredness, coldness, mild weight gain, constipation, menstrual irregularities, and muscle cramps. Drowsiness and slowing of intellectual and motor activity are often reported. Sensitivity to cold is suggested by the use of more blankets on the bed. Women frequently complain of hair loss, brittle nails, and dry skin. Periorbital puffiness may be present. Stiffness and aching of muscles may be attributed to rheumatism. Constipation may occur. Numbness and tingling of the extremities are frequent. Physical findings include cool dry skin, a puffy face and hands, a hoarse husky voice, and slow reflexes.

### Clinical Features of Hypothyroidism in the Elderly

Hypothyroidism in the elderly is often atypical and elusive and often lacks the classic clinical features present in younger patients. This is due to a combination of several factors, including the insidious onset, the ambiguity of several signs and symptoms (e.g., fatigue, weakness, cold intolerance, dry skin, hair loss, constipation, poor appetite, depression and/or mental deterioration, hearing loss, cardiomegaly, congestive heart failure) that may be attributed to normal aging, and the frequent coexistence of several age-associated diseases.

The most relevant clinical findings that lead one to suspect hypothyroidism in the elderly are an unexplained increase in serum cholesterol, severe constipation, congestive heart failure (particularly when it presents as restrictive cardiomyopathy), and macrocytic anemia (as a consequence of folate deficiency or coexistent autoimmune gastritis and pernicious anemia).

## CLINICAL ASPECTS OF HYPOTHYROIDISM DUE TO VARIOUS ETIOLOGIES

### Primary Hypothyroidism

Primary hypothyroidism in adults results mainly from autoimmune thyroiditis, is more common in women than in men, and occurs between 40 and 60 years of

age. In these patients, clinical features of hypothyroidism may be accompanied by the typical goiter of Hashimoto's thyroiditis. When present, the goiter is usually firm in consistency, generally moderate in size, and often lobulated (although well-defined nodules are unusual).

Other organ-specific autoimmune diseases, such as insulin-dependent diabetes mellitus, Addison's disease, premature ovarian failure, celiac disease, hypoparathyroidism, and myasthenia gravis, may also coexist. Patients with primary hypothyroidism may also complain of vitiligo and alopecia. Primary autoimmune hypothyroidism may also occur as a component of either the type I or the type II polyglandular autoimmune syndrome. The specific association of primary hypothyroidism and primary adrenal cortical insufficiency is known as Schmidt's syndrome. The rare type I syndrome consists of at least two of the triad of Addison's disease, hypoparathyroidism, and chronic mucocutaneous candidiasis; other autoimmune disorders, such as alopecia, chronic autoimmune thyroiditis, and malabsorption syndrome, may also be present. Autoimmune thyroid disease is reported in 10 to 12% of these patients. Type I polyglandular autoimmune syndrome presents more often during childhood. The type II syndrome is more common and usually presents during adulthood. Addison's disease, Hashimoto's thyroiditis, and type 1 diabetes are the most common endocrine deficiencies found in these patients, although gonadal failure, pernicious anemia, and vitiligo are observed in a significant percentage.

### Central Hypothyroidism

The clinical picture of central hypothyroidism varies depending on the severity of thyroid failure, the extent of associated hormone deficiencies, the age of the patient, and the nature of the underlying lesion. Central hypothyroidism is due to TSH deficiency caused by either hypothalamic or pituitary disease. The differentiation of central hypothyroidism from primary hypothyroidism is important for the institution of the proper therapy. The clinical features of central hypothyroidism are similar to those of primary hypothyroidism, although the former are generally less pronounced. The skin is pale and cool, but it is not as coarse and dry as in primary hypothyroidism. Periorbital and peripheral edema are uncommon in patients with central hypothyroidism. Loss of axillary, pubic, and facial hair, as well as thinning of the lateral eyebrows, is more pronounced. The tongue is not enlarged, and hoarseness of the voice is not as

prominent as it is in primary hypothyroidism. The heart tends to be small, and blood pressure is low. Atrophic breasts and amenorrhea are found in women. Body weight is more likely to be reduced than to be increased. Defects in growth hormone (GH) and gonadotropin secretion usually precede TSH insufficiency, and in most cases adrenocorticotropin hormone (ACTH) secretion is the last to be affected. Growth failure with delayed skeletal maturation results from GH deficiency in children. Hypoglycemia may occur. Gonadotropin insufficiency results in impotence, loss of libido, and diminished beard growth in men and results in amenorrhea, infertility, and atrophy of the breasts in women. ACTH deficiency leads to weakness, postural hypotension, and depigmentation of the areole and other normally pigmented areas of the skin. Symptoms and signs that arise directly from the hypothalamic or pituitary lesion may precede, accompany, or even obscure manifestations of pituitary failure. The manifestations of a sellar mass include headache and symptoms secondary to compression of adjacent structures with visual field disturbances and ophthalmoplegia.

### DIAGNOSTIC ACCURACY OF THE CLINICAL FEATURES OF HYPOTHYROIDISM

Several attempts have been made to develop a clinical score system that, based on the most frequent symptoms and signs of hypothyroidism, could accurately predict the diagnosis of thyroid failure in individual patients.

During the 1960s, Billewicz and colleagues described a diagnostic index that scored the presence or absence of various signs and symptoms of hypothyroidism. However, at that time, modern laboratory thyroid function tests were not available to validate the diagnostic accuracy of such a score. More recently, Zulewski and colleagues proposed a convenient clinical score that is both easy to determine and sensitive for individual assessment of the severity of thyroid failure. The frequencies of the 14 more common symptoms and signs of overt hypothyroidism are shown in [Table I](#). The most frequent features in hypothyroid patients were prolonged ankle reflex (77%) and complaints about dry skin (76%). A reduced pulse rate and cold intolerance were recorded with a high frequency in euthyroid controls and so were excluded from this score. The sensitivity and specificity of each symptom and sign of hypothyroidism, and the analysis of their positive and negative predictive values, are shown in [Table I](#). [Table II](#) shows the scoring system

**Table I** Sensitivity and Specificity of the 14 Symptoms and Signs of Hypothyroidism and Analysis of Their Positive and Negative Predictive Values (Percentages)

Symptoms and signs	Sensitivity	Specificity	Positive predictive value	Negative predictive value
Ankle reflex	77	93	92	80
Dry skin	76	64	68	73
Cold intolerance	64	65	65	64
Coarse skin	60	81	76	67
Puffiness	60	96	94	71
Pulse rate	58	42	50	50
Sweating	54	86	79	65
Weight increase	54	77	70	63
Paresthesia	52	82	75	63
Cold skin	50	80	71	61
Constipation	48	85	76	62
Slow movements	36	99	96	61
Hoarseness	34	87	73	57
Hearing loss	22	97	90	53

of symptoms and signs of hypothyroidism. Because a correlation analysis revealed a significant correlation of these scores with age, a simple age-correcting factor was defined by adding 1 point to the sum of symptoms and signs in women under 55 years of age. According to this analysis, the following diagnostic ranges for the clinical judgment with the age-corrected score were defined: hypothyroid, more than 5 points; intermediate, 3 to 5 points; euthyroid, 0 to 2 points.

### ORGAN SYSTEM MANIFESTATIONS OF HYPOTHYROIDISM

#### Cutaneous Manifestations and Changes in the Connective Tissues

The cutaneous changes observed in hypothyroidism belong to the most classic and frequent findings of the disease ([Table III](#)). In more than 80% of patients with primary hypothyroidism, the epidermis is dry, rough, cool, and covered with fine superficial scales. This is an expression of decreased cutaneous metabolism, reduced secretion of sweat and sebaceous glands, vasoconstriction, thinning of the epidermis, and hyperkeratosis of the stratum corneum. The face is puffy, pale, and expressionless at rest ([Fig. 1](#)). The palpebral fissure may be narrowed because of blepharoptosis. The tongue is usually large, and some patients will complain of this problem. The tongue is smooth if pernicious anemia coexists. The voice is

Table II Scoring of Symptoms and Signs of Hypothyroidism

		Score	
		Present	Absent
On the basis of			
Symptoms			
Diminished sweating	Sweating in the warm room on a hot summer day	1	0
Hoarseness	Speaking voice, singing voice	1	0
Paresthesia	Subjective sensation	1	0
Dry skin	Dryness of skin, noticed spontaneously, requiring treatment	1	0
Constipation	Bowel habit, use of laxative	1	0
Impairment of hearing	Progressive impairment of hearing	1	0
Weight increase	Recorded weight increase, tightness of clothes	1	0
Physical signs			
Slow movements	Observe patient removing his/her clothes	1	0
Delayed ankle reflex	Observe the relaxation of the reflex	1	0
Coarse skin	Examine hands, forearms, and elbows for roughness and thickening of skin	1	0
Periorbital puffiness	This should obscure the curve of the malar bone	1	0
Cold skin	Compare temperature of hands with that of examiner	1	0

*Note.* For clinical judgment, add 1 point to the sum of symptoms and signs present in women under 55 years of age. Hypothyroid, more than 5 points; intermediate, 3 to 5 points; euthyroid, less than 3 points.

husky, low-pitched, and coarse due to the enlargement of the tongue and the thickening of the pharyngeal and laryngeal mucous membranes. The speech is deliberate and slow.

### Hair Follicles and Nails

The hair is dry, dull, and coarse, and it grows slowly, becoming sparse and falling out readily. Loss of scalp, genital, and beard hair may also occur. Hair may be lost from the temporal aspects of the eyebrows (Queen Anne's sign). The nails, through retardation of growth, become thickened and brittle, striated in both transverse and longitudinal grooves, showing frequent deformities.

Table III Cutaneous Signs and Symptoms of Hypothyroidism (Percentages)

Cutaneous manifestation	Frequency
Cold intolerance	50–95
Nail abnormality	90
Thickening and dryness of hair and skin	80–90
Edema of hands, face, and eyelids	70–85
Change in shape of face	70
Malar flush	50
Nonpitting edema	30
Alopecia	30–40
Pallor	25–60
Decreased sweat secretion	10–70

### Cardiovascular Changes

Lack of thyroid hormones causes multiple alterations in the cardiovascular system (Table IV). Bradycardia, cardiomegaly, and low-voltage complexes on the electrocardiogram (ECG) are well-known features. The decrease in pulse rate approximately parallels the decrease in the body's metabolic rate. Myocardial contractility is reduced. The cardiac output at rest is decreased due to reduction in both stroke volume and heart rate, reflecting loss of the inotropic and chronotropic effects of thyroid hormones. Peripheral vascular resistance at rest is increased, and blood volume is reduced. These hemodynamic alterations

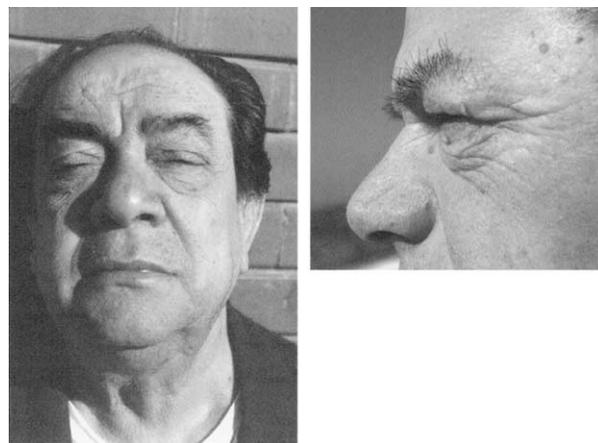


Figure 1 Photographs of a hypothyroid patient.

**Table IV Cardiovascular Signs and Symptoms in Hypothyroidism**

Symptoms
Dyspnea
Decreased exercise tolerance
Angina
Signs
Low pulse rate
Increased systemic vascular resistance
Diastolic hypertension
Cardiomegaly
Pericardial effusion
Peripheral nonpitting edema
Low voltage ECG, nonspecific ST-T changes

cause narrowing of pulse pressure, prolongation of circulation time, and decreased blood flow to the tissues. Arterial blood pressure is often mildly increased. Hypertension is present in 10 to 20% of patients with hypothyroidism. Diastolic hypertension is usually restored to normal after treatment. Few symptoms associated with the cardiovascular system are referred from patients with hypothyroidism. Exertional dyspnea and exercise intolerance are probably due to skeletal muscle dysfunction.

On physical examination, certain findings can suggest hypothyroidism. The heart rate is lowered, the pulse pressure is narrowed, and the carotid upstroke and left ventricular apical impulse are diminished. The heart sounds are diminished in intensity. This finding is due largely to effusion into the pericardial sac of fluid rich in protein and glycosaminoglycans.

### Electrocardiographic Changes

Electrocardiographic changes include sinus bradycardia, prolongation of the PR interval, low amplitude of the P wave and QRS complex, alterations of the ST segment, and flattened or inverted T waves. Although suggestive of myocardial ischemia, these waveform changes often disappear during T<sub>4</sub> substitution therapy. Pericardial effusion is probably responsible for the low amplitude.

### Respiratory Changes

Respiratory troubles are rarely a major complaint in hypothyroid patients. However, hypothyroidism may cause respiratory problems through (1) depression of the respiratory center in the brain, (2) disturbed neural conduction and/or neuromuscular transmission to the

respiratory muscles (due to hypothyroid neuropathy), (3) diseased respiratory muscle function (due to hypothyroid myopathy), and/or (4) changes in the alveolar-capillary membranes and the surfactant lining the alveoli, leading to impaired gas exchange.

Fatigue and dyspnea on exertion are frequent symptoms. Dyspnea is a frequent complaint of myxedematous patients, but it is also a common symptom among well persons. Congestive heart failure of separate origin, pleural effusion, anemia, obesity, and/or pulmonary disease may be responsible for dyspnea.

### Gastrointestinal Changes

The gastrointestinal manifestations of hypothyroidism are listed in Table V. Poor appetite can be a leading symptom in hypothyroid patients. Anorexia can reasonably be interpreted as the reflection of a lowered food requirement. Although two-thirds of hypothyroid patients report weight gain, it is of modest degree and due largely to retention of fluid by the hydrophilic glycoprotein deposits in the tissues. True obesity is not a feature of hypothyroidism per se.

Constipation is frequently present and is the result of lowered food intake and decreased peristaltic activity. Atrophy of the gastric and intestinal mucosa and myxedematous infiltration of the bowel wall may be present at histological examination. Immune gastritis is often observed in hypothyroid patients with autoimmune thyroiditis. As many as 50% of patients with autoimmune hypothyroidism have achlorhydria, 25% have circulating antibodies directed against the gastric parietal cells or intrinsic factor, and 2 to 10% have pernicious anemia caused by impaired absorption of vitamin B<sub>12</sub>.

Symptoms or signs of disturbed liver or exocrine pancreatic function are usually not encountered,

**Table V Gastrointestinal Manifestations of Hypothyroidism**

Symptoms
Anorexia
Gaseous distension
Constipation
Signs
Prolonged gastric emptying
Prolonged intestinal transit time
Slowed intestinal absorption
Ascites
Elevated liver enzymes
Gallbladder hypotonia

but biochemical tests may suggest disease. The association of liver disease and hypothyroidism is suggestive of a multisystem autoimmune disease affecting both the liver (e.g., chronic active hepatitis, primary biliary cirrhosis) and the thyroid. Structural liver damage is unusual in hypothyroidism per se. Serum glutamine–oxaloacetic transaminase (GOT), lactate dehydrogenase (LDH), and creatin–phosphokinase (CPK) levels are elevated in patients with hypothyroidism. These enzymes return to normal over 2 to 4 weeks during treatment. Urinary amylase levels may be increased.

### Cerebral and Neurological Changes

Thyroid hormone is essential for the development of the central nervous system. Deficiency during fetal life or at birth causes hypoplasia of cortical neurons with poor development of cellular processes, retarded myelination, and reduced vascularity. Deficiency of thyroid hormone beginning during adult life causes less severe manifestations that usually reverse after treatment with thyroid hormone.

Table VI lists the numerous symptoms suggesting either neurological or psychiatric disorders in patients with moderate to severe hypothyroidism. In adult and elderly patients, mental changes may go unrecognized for a long time because of their slow development and because they may mimic cerebral atherosclerosis. However, an unusual complacency, fatigue, and pronounced somnolence or even lethargy, together with a prolonged reaction time, should suggest the possibility of hypothyroidism. All intellectual functions, including speech, are slowed. There is loss of

**Table VI Neurological and Psychiatric Manifestations in Hypothyroidism**

Neurological symptoms or signs
Somnolence, lethargy
Slow speech
Impaired cognitive functions
Headache
Paresthesias
Cerebellar ataxia
Deafness
Vertigo
Delayed relaxation of deep tendon reflexes
Psychiatric syndromes
Depression
Bipolar disorders
Affective psychosis

initiative, and slow-wittedness and memory defects are common. Myxedematous dementia in the elderly patient may be mistaken for senile dementia. Memory is undoubtedly impaired, and attention and the desire to think are reduced. The emotional level seems definitely low, and irritability is decreased. Except during the terminal stage, reasoning power is preserved. Sensory phenomena are common. Numbness and tingling of the extremities are frequent. Mono-neuropathies occur in hypothyroidism, as attested to by the high incidence of carpal tunnel syndrome (i.e., compression of the median nerve at the wrist). Nocturnal paresthesia and pain in the median nerve distribution in one hand or both hands is a common manifestation of this condition.

The tendon reflexes are slow, especially during the relaxation phase, producing the characteristic “hung-up reflexes.” This phenomenon is due to a decrease in the rate of muscle contraction and relaxation rather than to a delay in nerve conduction.

Electroencephalographic changes include slow alpha wave activity and general loss of amplitude.

### Muscle–Skeletal Changes

#### **Muscles**

In patients with hypothyroidism, disordered muscle function often is the predominating feature of the clinical syndrome. Generalized muscular hypertrophy, accompanied by easy fatigue and slowness of movements, occurs in some myxedematous children or adults.

Muscle symptoms such as myalgia, muscle weakness, stiffness, cramps, and easy fatigability are very prevalent in hypothyroid patients. The symptoms are aggravated by exposure to cold. They are also prominent during the rapid onset of hypothyroidism after surgery or 131-iodine therapy.

#### **Skeletal System: Calcium and Phosphorus Metabolism**

In the adult skeleton, thyroid hormone deficiency decreases recruitment, maturation, and activity of bone cells, leading to decreased remodeling that is especially reflected in the impaired function of the osteoclasts. Despite this decrease in osteoclastic activity, trabecular bone volume and bone mineral density appear to be comparable to those in age-matched normals, presumably because of the corresponding decrease in osteoblastic activity. The concentrations of calcium and phosphorus in serum are usually normal, but calcium may be slightly elevated. Serum alkaline phosphatase levels are often decreased, as are serum osteocalcine

levels. Because the levels of parathyroid hormone are often slightly increased, some degree of resistance to its action may be present. Serum concentrations of 1,25-dihydroxycholecalciferol are also increased.

### Hematological Changes

Anemia is present in as many as two-thirds of hypothyroid children and adolescents and in approximately one-third of adults with hypothyroidism. Anemia is usually mild. In two reports on a large series of patients with hypothyroidism from various causes, the incidence of anemia ranged from 32% to as high as 84%. Anemia in hypothyroidism may be a normochromic and normocytic anemia due to the diminished oxygen requirements and decreased production of erythropoietin, or it may result from a specific depression of the marrow that lacks thyroid hormone. The anemia may be macrocytic, sometimes from deficiency of vitamin B12. Folate deficiency from malabsorption or dietary inadequacy may also cause macrocytic anemia.

Granulocyte, lymphocyte, and platelet counts are usually normal in hypothyroidism. Hypothyroid patients may have bleeding symptoms such as easy bruising, menorrhagia, and prolonged bleeding after tooth extraction. The most frequent defects in hemostasis are prolonged bleeding time, decreased platelet adhesiveness, and low plasma concentrations of factor VIII and von Willebrand factor.

### Changes in the Reproductive Tract

In both sexes, thyroid hormones influence sexual development and reproductive function. Infantile hypothyroidism leads to sexual immaturity, and juvenile hypothyroidism causes a delay in the onset of puberty followed by anovulatory cycles. Paradoxically, primary hypothyroidism may also cause precocious sexual development and galactorrhea.

In adult men, hypothyroidism may lead to impotence, lack of libido, and/or (rarely) testicular tubular involution. The testicles are histologically immature if hypothyroidism preceded puberty and show tubular involution if the onset of hypothyroidism was after puberty. In adult hypothyroid men, semen analysis is usually normal. In adult women, severe hypothyroidism may be associated with diminished libido and failure of ovulation. In general, hypothyroid women complain of menorrhagia as well as (occasionally) oligo and amenorrhea. Plasma gonadotropins are usually in the normal range in primary hypothyroidism,

and the pulsatile gonadotropin release during the follicular phase is normal, but the ovulatory surge of luteinizing hormone (LH) might not happen. Secretion of progesterone is inadequate, and endometrial proliferation persists, resulting in excessive and irregular breakthrough menstrual bleeding. The anovulation is reflected in the frequent finding of a proliferative endometrium. These changes may be due to a deficient secretion of LH. Mild to moderate hyperprolactinemia is a frequent finding in hypothyroid women with or without galactorrhea. It is attributed to the stimulatory effect of increased thyrotropin-releasing hormone (TRH) on prolactin secretion. Fertility is reduced, and spontaneous abortion may result, although many pregnancies are successful.

The literature contains many reports of pregnancy in untreated hypothyroid women. Euthyroid neonates born to hypothyroid mothers during pregnancy have been reported to achieve lower IQs later in life. When treatment has been started during pregnancy, a normal child is usually produced, but abortion is frequent in myxedematous women. Pregnancy-induced hypertension is two to three times more common in hypothyroid women. Low birthweight may be secondary to premature delivery for gestational hypertension. The incidence of various congenital abnormalities may be increased, but recent studies do not report an increased risk of fetal death or congenital anomalies with proper treatment.

### Other Endocrine Glands

Hypothyroidism decreases GH secretion, and hypothyroid children have a dramatic retardation of growth. Retarded growth caused by hypothyroidism appears to result from deficient secretion of GH as well as from impaired action of GH. Many hypothyroid children have subnormal serum GH response to insulin-induced hypoglycemia. GH secretion is decreased in hypothyroidism related to an increase in hypothalamic-somatostatinergic tone and results in low serum insulin-like growth factor-1 (IGF-1) concentrations. Serum IGF-2, IGF-binding protein-1 (IGFBP-1), and IGFBP-3 also fall, whereas IGFBP-2 rises. These changes are reversible with treatment.

Thyrotroph hyperplasia caused by primary hypothyroidism may result in sellar enlargement, particularly when the condition has remained untreated for a long period of time. Patients with severe hypothyroidism may have an increase in serum prolactin level that

correlates with the level of serum TSH, and some patients develop galactorrhea.

### Metabolic Changes

The decrease in energy metabolism and heat production is reflected in the low basal metabolic rate (BMR), decreased appetite, cold intolerance, and slightly lower basal body temperature. Measurement of the resting energy expenditure is rarely performed nowadays. In patients with complete athyreosis, it falls between 35 and 45% below normal. Both the synthesis and the degradation of proteins are decreased (the latter especially so), resulting in a nitrogen balance that is usually slightly positive.

In hypothyroidism, absorption of glucose from the gastrointestinal tract is slowed and peripheral glucose assimilation is retarded. At the same time, glycerol release from adipose tissue is slowed and the availability of amino acids and glycerol for gluconeogenesis is decreased. The oral glucose tolerance curve is characteristically flat, and the insulin response to glucose is delayed. Degradation of insulin is slow, so the sensitivity to exogenous insulin may be increased. Despite the easily demonstrable abnormalities in carbohydrate metabolism in hypothyroidism, clinical manifestations of these abnormalities are seldom conspicuous. A variety of abnormalities in plasma lipid concentrations occur in hypothyroidism (Table VII). Plasma free fatty acid concentrations are normal, whereas plasma concentrations of triglycerides, phospholipids, and low-density lipoprotein (LDL) cholesterol are well elevated. Biosynthesis of fatty acids and lipolysis is reduced. In general, the changes bear a reciprocal relationship to the level of thyroid activity.

The increased serum cholesterol may represent an alteration in the substrate steady-state level caused by a transient proportionally greater retardation in degradation than in synthesis. The increase of serum cholesterol is largely accounted for by an increase of LDL cholesterol. Interestingly, the LDL particles of

hypothyroid patients are also susceptible to increased oxidizability. The increase of high-density lipoprotein-2 (HDL2) cholesterol but not of HDL3 cholesterol is due to diminished activity of cholesteryl ester transfer protein and hepatic lipase (which is involved in the conversion of HDL2 to HDL3). The occasional modest increase of serum triglycerides has been related to decreased lipoprotein lipase activity in postheparin plasma.

### See Also the Following Articles

Hypothalamic Hypothyroidism • Hypothyroidism, Causes of • Hypothyroidism, Congenital, Long-Term Follow-Up • Hypothyroidism, Congenital, Screening Programs • Hypothyroidism, Diagnosis of • Hypothyroidism, Subclinical • Hypothyroidism, Treatment of

### Further Reading

- Bartalena, L., Placidi, G. F., Martino, E., *et al.* (1990). Nocturnal serum thyrotropin (TSH) surge and the TSH response to TSH-releasing hormone: Dissociated behavior in untreated depressives. *J. Clin. Endocrinol. Metab.* **71**, 650–655.
- Burrow, G. N., Fisher, D. A., and Larsen, P. R. (1994). Maternal and fetal thyroid function. *N. Engl. J. Med.* **331**, 1072–1079.
- Chiovato, L., Lapi, P., Fiore, E., Tonacchera, M., and Pinchera, A. (1993). Thyroid autoimmunity and female gender. *J. Endocrinol. Invest.* **16**, 373–401.
- Duranti, R., Gheri, R. G., Gorini, M., *et al.* (1993). Control of breathing in patients with severe hypothyroidism. *Am. J. Med.* **95**, 29–33.
- Erslev, A. J. (1990). Anemia of endocrine disorders. In "Hematology" (W. J. Williams, E. Beutler, A. J. Erslev, and M. A. Lichtman, eds.), 4th ed., pp. 408–415. McGraw-Hill, New York.
- Gomberg-Maitland, M., and Frishman, W. H. (1998). Thyroid hormone and cardiovascular disease. *Am. Heart J.* **135**, 187–196.
- Mariotti, S., Franceschi, C., Cossarizza, A., and Pinchera, A. (1995). The aging thyroid. *Endocrine Rev.* **16**, 686–715.
- Rosen, T., and Kleman, G. A. (1994). Thyroid and skin. In "Dermatological Signs of Internal Disease" (J. P. Callen, J. L. Jorizzo, K. E. Greer, *et al.*, eds.), 2nd ed., pp. 189–204. W. B. Saunders, Philadelphia.
- Shafer, R. B., Prentiss, R. A., and Bond, J. H. (1994). Gastrointestinal transit in thyroid disease. *Gastroenterology* **86**, 852–856.
- Tachman, M. L., and Guthrie, G. P., Jr. (1984). Hypothyroidism: Diversity of presentation. *Endocrine Rev.* **5**, 456–464.
- Tomasi, P. A., Fanciulli, G., Zini, M., Demontis, M. A., Dettori, A., and Delitala, G. (1997). Pulsatile gonadotropin secretion in hypothyroid women of reproductive age. *Eur. J. Endocrinol.* **136**, 406–409.
- Zulewski, H., Muller, B., Exer, P., Miserez, A. R., and Staub, J. J. (1997). Estimation of tissue hypothyroidism by a new clinical score: Evaluation of patients with various grades of hypothyroidism and controls. *J. Clin. Endocrinol. Metab.* **82**, 771–776.

**Table VII** Changes in Serum Lipids in Hypothyroidism

Total cholesterol	Increase
LDL cholesterol	Increase
HDL2 cholesterol	Modest increase
HDL3 cholesterol	No change
Triglycerides	No change or modest increase

mortality. Many physicians believe that it makes sense to allow hypothyroid patients who are dissatisfied with the outcome of restoring serum T4 and TSH to normal to increase the dose of T4 such that serum TSH is suppressed, in which case serum free T4 is likely to lie between 25 and 30 pmol/L. In this circumstance, it is essential that serum T3 be unequivocally normal.

Serum TSH concentrations are of no value in assessing the correct dose of T4 in patients with central hypothyroidism, and measurements of both serum T3 and T4 combined with a biochemical index of thyroid hormone action, such as soluble interleukin-2 receptors, have been suggested. In practice, however, physicians are likely to be influenced more by the well-being of their patients.

### Starting Dose of Thyroxine

There is no evidence base for determining how T4 therapy should be initiated, but it is customary to prescribe 50 µg daily, increasing to 100 µg daily after 2 to 4 weeks. Measurement of serum T4 and TSH at 2 months after starting will dictate any further adjustment of dose. In the elderly, and particularly in those with symptomatic ischemic heart disease, a starting dose of 25 µg daily is advisable, with increments of 25 µg every 2 to 4 weeks, recognizing that the full replacement dose might not be achieved without coronary artery bypass grafting or angioplasty. In those individuals where hypothyroidism has developed rapidly and been detected early, such as following total thyroidectomy for thyroid carcinoma, a full replacement dose of 100 to 150 µg daily may be given from the outset.

Patients begin to feel better within 10 to 14 days of starting T4, even with doses as little as 25 µg daily. Reduction in body weight (rarely more than 10% and is due largely to fluid loss) and improvements in periorbital puffiness are among the early responses, whereas maximum improvement in hair and skin texture may take up to 3 months and reversal of the rare feature of cerebellar ataxia may take considerably longer.

### Variation in Dose

Once T4 therapy is established, it is good practice to review patients annually and measure serum TSH, not only to ensure compliance but also to determine whether a change in dose is required. There is an increasing list of circumstances in which the dose of T4 might need to be adjusted, usually upward (Table I).

**Table I** Situations where an Adjustment of the Dose of Thyroxine May Be Necessary

Increased dose required	
Use of other medication	
Phenobarbitone	} Increased thyroxine clearance
Phenytoin	
Carbamazepine	
Rifampicin	
Sertraline <sup>a</sup>	
Chloroquine <sup>a</sup>	
Cholestyramine	} Interference with intestinal absorption
Sucralfate	
Aluminum hydroxide	
Ferrous sulfate	
Calcium carbonate	
Dietary fiber supplements	
Pregnancy	Increased concentration of serum thyroxine-binding globulin
Estrogen therapy	
After surgical or iodine-131 ablation of Graves' disease	Reduced thyroidal secretion with time
Malabsorption (e.g., celiac disease)	
Decreased dose required	
Aging	Decreased thyroxine clearance
Graves' disease developing in patient with long-standing primary hypothyroidism	Switch from production of blocking to stimulating TSH receptor antibodies

<sup>a</sup>Mechanism not fully established.

The most recently recognized are use of other medications such as chloroquine, sertraline, calcium carbonate, and estrogens either as the oral contraceptive pill or as hormone replacement therapy. Ingestion of dietary fiber supplements may reduce bioavailability of T4 by its adsorption onto wheat bran.

During pregnancy, most patients require an increase in T4 dose of some 50 µg daily, probably due to the increased concentration of T4-binding globulin. Therefore, it is important to measure free T4 and TSH during each trimester. It is also important to ensure that the appropriate dose is being taken pre-conception if pregnancy is planned because there is evidence that inadequately treated hypothyroidism in the mother may result in a reduction in the IQ of the offspring.

There are numerous manufacturers of T4 preparations, and from time to time there is divergence between the amount of active ingredients specified and the true content of the tablets. This possibility

should be entertained whenever a patient's serum TSH concentrations rise even though no new medication has been prescribed and the dose of T4 has been stable for many years.

### SUBCLINICAL HYPOTHYROIDISM

By definition, affected patients have normal concentrations of serum T3 and T4 but have raised TSH, usually between 5 and 10 mU/L. They are normally asymptomatic but may have consulted because of tiredness and weight gain. Many have a family history of thyroid disease or another organ-specific autoimmune disease such as insulin-dependent diabetes mellitus or pernicious anemia. The consensus is that subclinical hypothyroidism represents the mildest form of thyroid failure and, as such, should usually be treated with T4. Those favoring therapy are most influenced by the knowledge that between 25 and 50% feel better when taking T4 and by the inexorable progression to overt thyroid failure of approximately 5% each year in those who possess antithyroid peroxidase antibodies—a risk that increases with higher serum TSH concentrations. Contrary to expectations, treatment of subclinical hypothyroidism, at least in those with serum TSH concentrations less than 10 mU/L, does not substantially improve the lipid profile.

### TEMPORARY HYPOTHYROIDISM

Most patients with primary hypothyroidism require lifelong treatment with T4. There are, however, well-recognized situations in which the thyroid failure is transient and even short-term treatment with T4 is unnecessary. These include the recovery phase of subacute (de Quervain's), painless, and postpartum thyroiditis; Hashimoto's thyroiditis, particularly if excess iodine or iodine-containing drugs, such as amiodarone, have been implicated in the development of the thyroid failure; the neonatal period in children born to a minority of mothers with autoimmune hypothyroidism due to the transplacental passage of TSH receptor-blocking antibodies; and approximately 5% of patients with chronic autoimmune thyroiditis due to the disappearance of these same antibodies from the serum.

It is often not appreciated that hypothyroidism may be temporary if it occurs within 6 months of subtotal thyroidectomy or iodine-131 for Graves' disease. If T4 therapy is required at an early stage, a suboptimal dose of 75 µg daily should be prescribed. Measurement of serum T4 and TSH at 6 months will

indicate whether continued treatment with a higher dose is required (raised serum TSH) or whether T4 therapy can be stopped (low or normal TSH) and thyroid function can be reassessed 4 to 6 weeks later.

### OTHER TREATMENTS OF HYPOTHYROIDISM

During the first part of the 20th century, the only treatment available for hypothyroidism was animal thyroid extract that contained both T3 and T4. Because of its variable potency, it was largely abandoned in favor of T4 alone. Recently, there has been a renewal of interest in treating hypothyroidism with both T4 and T3. This is the result of a study in 1999 by Bunevicius and colleagues, who showed that T3 and T4, when given in the ratio normally secreted by the thyroid gland, resulted in significant improvement in mood and neuropsychological function as compared with that resulting from a higher dose of T4 alone. Of greatest interest, this improvement occurred without suppression of serum TSH. Further evidence that monotherapy may be inadequate derives from recent observations that excess weight is gained by hyperthyroid patients who are rendered hypothyroid by ablative therapy and given T4 in a dose that simply restores TSH to normal. These results are not altogether surprising given that it is possible to restore euthyroid concentrations of thyroid hormones in most tissues of the hypothyroid rat only if a combination of T4 and T3 is infused. Therefore, it is likely that treatment of hypothyroidism in future years will be with a combination of synthetic T3 and T4 in the ratio of 1:10 and with the T3 in slow-release form to avoid cardiovascular adverse effects.

### See Also the Following Articles

Hypothalamic Hypothyroidism • Hypothyroidism, Causes of • Hypothyroidism, Congenital, Long-Term Follow-Up • Hypothyroidism, Congenital, Screening Programs • Hypothyroidism, Diagnosis of • Hypothyroidism, Subclinical • Hypothyroidism, Systemic Manifestations of

### Further Reading

- Bunevicius, R., Kazanavicius, G., Zalinkevicius, R., and Prange, A. J. (1999). Effects of thyroxine as compared with thyroxine plus triiodothyronine in patients with hypothyroidism. *N. Engl. J. Med.* **340**, 424–429.
- Escobar-Morreale, H. F., Escobar del Rey, F., Obregon, M. J., and Morreale de Escobar, G. (1996). Only the combined treatment with thyroxine and triiodothyronine ensures euthyroidism in

- all tissues of the thyroidectomized rat. *Endocrinology* **137**, 2490–2502.
- Haddow, J. E., Palomaki, G. E., Allan, W. C., *et al.* (1999). Maternal thyroid deficiency during pregnancy and subsequent neuro-psychological development of the child. *N. Engl. J. Med.* **341**, 549–555.
- Sawin, C. T., Geller, A., Wolfe, P. A., *et al.* (1994). Low serum thyrotropin concentrations as a risk factor for atrial fibrillation in older patients. *N. Engl. J. Med.* **331**, 1249–1252.
- Tigas, S., Idiculla, J., Beckett, G., and Toft, A. (2000). Is excessive weight gain after ablative treatment of hyperthyroidism due to inadequate thyroid hormone therapy? *Thyroid* **10**, 1107–1111.
- Toft, A. D. (1999). Thyroid hormone replacement: One hormone or two? *N. Engl. J. Med.* **340**, 469–470.
- Uzzan, B., Campos, J., Cucherat, M., Nony, P., Boissel, J. P., and Perret, G. Y. (1996). Effects on bone mass of long term treatment with thyroid hormones: A meta-analysis. *J. Clin. Endocrinol. Metab.* **81**, 4278–4289.

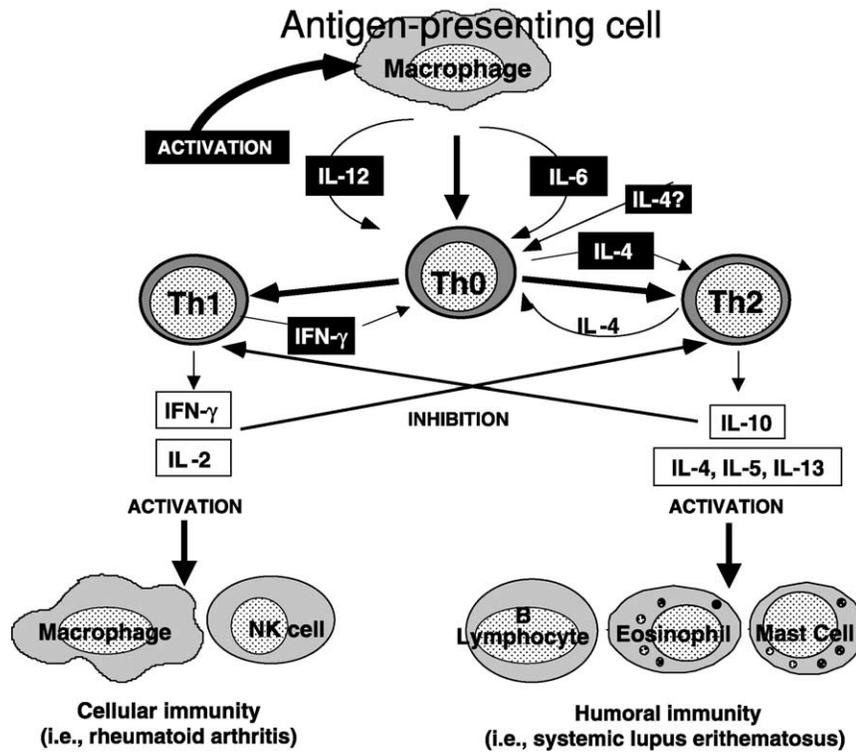


Figure 1 Role of Th1/Th2 cytokines in the regulation of cellular and humoral immunity.

variety of clinical manifestations of the disease as well as resistance to the biological therapies (TNF $\alpha$  blockers).

## ENDOGENOUS STEROIDS AND MODULATION OF THE IMMUNE/INFLAMMATORY RESPONSE

An intricate balance between soluble mediators (cytokines), released by activated cells of the inflammatory/immune system, and products of the neuroendocrine system, such as neuropeptides and steroid hormones, maintains the homeostasis during the immune response. In general, the major inflammatory cytokines (IL-1, IL-6, and TNF $\alpha$ ) stimulate the production of corticotropin-releasing hormone (CRH) in the hypothalamus. CRH release leads to pituitary production of adrenocorticotropic hormone (ACTH), followed by glucocorticoid secretion by the adrenal cortex and indirect perturbations of gonadal function.

Conversely, hormonal products of the hypothalamus–pituitary–adrenal (HPA) axis and hypothalamus–pituitary–gonadal (HPG) axis modulate the cytokine production (Fig. 2). For example, adrenal and gonadal androgens, in particular dehydroepiandrosterone sulfate (DHEAS) and both testosterone and dihydrotestosterone (DHT), have been found to repress

the expression and activity of the human IL-6 gene promoter, thereby supporting the concept of anti-inflammatory/immunosuppressive effects exerted by androgens. Conversely, physiological concentrations of estrogens exert immunoenhancing activities, at least on APCs and generally on the humoral immune response.

Notably, glucocorticoids are the most potent endogenous inhibitors of immune and inflammatory processes, including proinflammatory cytokine production. Peripheral levels of IL-6, and to lesser extent those of TNF $\alpha$  and IL-1 $\beta$ , are tonically inhibited by basal levels of glucocorticoids. In chronic diseases, a condition of subclinical hypoadrenal function is detected. Therefore, in systemic autoimmune/inflammatory diseases, the use of glucocorticoids as low-dose oral administration or high-dose pulse therapy should be regarded as a sort of replacement therapy, with modulating effects on activated T-lymphocytes and macrophages.

## SPECIFIC ROLES FOR VARIOUS HORMONES

### Cortisol (Adrenal Hormone)

The multiple anti-inflammatory effects of cortisol and its synthetic derivatives have been studied extensively

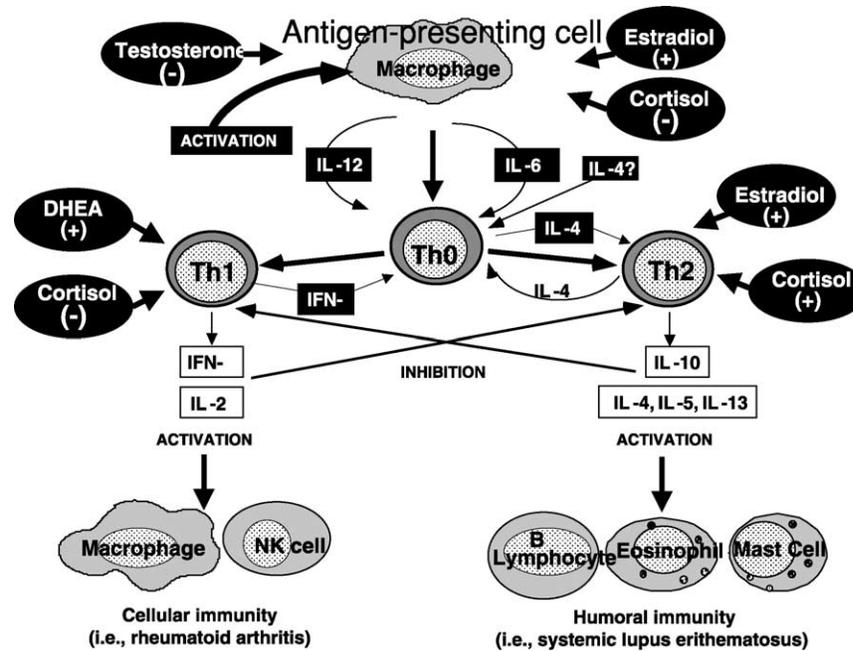


Figure 2 Hormonal effects on Th1/Th2 cytokines in the regulation of cellular and humoral immunity.

over the past 50 years. It is obvious that this hormone has multiple anti-inflammatory activities such as inhibition of cytokines, phagocytosis, the oxidative burst, and inducible cyclo-oxygenase (COX-2). In addition, cortisol *in vivo* induces a more Th2-directed immune response, with an increase of anti-inflammatory IL-10. It has been assessed that an approximately 5- to 10-fold variation in blood glucocorticoids (increase or decrease) is paralleled by oscillation of about the same magnitude in the number of activated B lymphocytes. With respect to the situation *in vivo*, this has been shown many times by glucocorticoid therapy in animals with arthritis and patients with RA. In conclusion, high physiological concentrations of cortisol at  $10^{-6}$  to  $10^{-5}$  M are anti-inflammatory and immunosuppressive (humoral response).

### Dehydroepiandrosterone (Adrenal Hormone)

Two other adrenal hormones, which are secreted on ACTH stimulation, are dehydroepiandrosterone (DHEA) and its sulfated derivative DHEAS. DHEA is the active hormone that can be converted to downstream androgens in peripheral tissue cells such as macrophages. Because androgens such as testosterone have anti-inflammatory properties, DHEA and DHEAS, the androgen precursors, contribute to the hormone pool so as to maintain an anti-inflammatory

situation in the periphery when the gonadal glands undergo gradual involution during aging. Furthermore, DHEA may have its own intracellular receptor to exert direct anti-inflammatory effects in animals and humans.

### Progesterone (Ovarian or Placental Hormone)

Progesterone exerts immunomodulatory activities. An immune response during the luteal phase of the ovarian cycle, as compared with the follicular phase, is shifted toward a Th2 type. Progesterone-mediated promotion of the Th2 response, inducing conversion of Th0 cells into Th2 cells, is indispensable for maintaining pregnancy.

In the presence of progesterone, lymphocytes from pregnant females produce a protein known as progesterone-induced blocking factor (PIBF), which alters the profile of cytokine secretion by stimulation of the cytokine Th2 production and inhibition of the cytokine Th1 production.

### Androgens (Gonadal Hormones)

In general, androgens, in physiological concentrations, tend to suppress the immune responses. Functional androgen receptors have been found on monocytes/macrophages.

Recent studies have shown that both physiological ( $10^{-8}$  M) and pharmacological (or high physiological) ( $10^{-6}$  M) concentrations of testosterone inhibit IL-1 $\beta$  secretion by PBMCs obtained from RA patients. In addition, physiological concentrations of testosterone inhibit IL-1 synthesis in primary cultured human synovial macrophages. In other studies, dihydrotestosterone has been found to repress the expression and activity of the human IL-6 gene promoter in human fibroblasts, thereby supporting the concept of anti-inflammatory/immunosuppressive effects of androgens. Testosterone treatment does not appear to directly affect isolated B or T cells. However, a follow-up study of SLE patients showed that testosterone suppresses both anti-double-stranded DNA IgG and total IgG production of PBMCs. Low serum and synovial fluid androgen levels were observed in RA patients.

### Estrogens (Gonadal Hormones)

In general, estrogens in physiological concentrations serve to enhance immune responses and may act as an important stimulator of human humoral immunity. It was reported that 17 $\beta$ -estradiol enhances IgG and immunoglobulin M (IgM) production by PBMCs in both men and women without altering cell viability or proliferation. A more recent study confirmed that 17 $\beta$ -estradiol has the capacity to increase the production of polyclonal IgG, including anti-double-stranded DNA IgG in PBMCs of SLE patients, by enhancing B-cell activity via IL-10. Concerning the effects of estrogens on proinflammatory cytokine production (IL-1, IL-6, and TNF $\alpha$ ), these effects seem to be bimodal, with decreased synthesis using pharmacological concentrations ( $10^{-6}$  M) and increased synthesis using physiological concentrations ( $\leq 10^{-8}$  M). Recently, it was demonstrated that estrogens are able to enhance secretion of matrix metalloproteinases from human fibroblast-like synoviocytes. Functional androgen receptors have been found on monocytes/macrophages, lymphocytes, and fibroblasts. With respect to serum levels of estrogens in RA patients, they are not changed—in strict contrast to androgen levels. Increased hydroxylation of estrogens is observed in SLE serum and in RA synovial fluids.

### Conversion of Steroid Hormones in Peripheral Target Cells

Because steroid hormones can be converted along defined pathways to downstream hormones in the periphery, multiple secondary effects may arise according to local regulation of steroid conversion. We have

recently demonstrated that conversion of DHEA occurs in macrophages and depends on the state of differentiation and presence of endotoxin. In postmenopausal women and aged men, the availability of adrenal DHEA is important to maintain a significant level of sex hormones in the periphery. It is not known whether, under certain circumstances, macrophages or other peripheral immune cells are able to synthesize steroid hormones in adequate amounts using cholesterol, as has been demonstrated in thymocytes. Furthermore, we do not exactly know the importance of, for example, degradation of cortisol by the 11 $\beta$ -hydroxysteroid dehydrogenase in the synovial tissue, which changes the activity of the steroid hormones in the local microenvironment. Increased concentrations of estrogens are observed at the site of the immune-inflammatory reaction. For example, in synovial fluids of RA patients of both genders, increased estrogens and reduced androgens are found to be related to the increased activity of the metabolizing enzyme aromatase stimulated by local high concentrations of the major inflammatory cytokines (IL-1, IL-6, and TNF $\alpha$ ).

### Prolactin (Pituitary Hormone)

Prolactin modulates lymphocytes by activating T-lymphocyte and B-lymphocyte proliferation, stimulates IgG secretion, and increases cytokine secretion from various cells such as IL-1, IL-6, TNF $\alpha$ , and IFN- $\gamma$ . Because of these stimulatory effects on immune cells, a proinflammatory role of prolactin was suspected in autoimmune diseases. Elevated levels of prolactin were associated with a more severe progressive systemic sclerosis, the active phase of SLE, RA, human autoimmune thyroid disease, and adjuvant arthritis. Hyperprolactinemia was associated with the presence of autoantibodies in otherwise healthy women, indicating the stimulating effect of this hormone on IgG. Furthermore, therapeutic prolactin antagonism with bromocriptine was found to have some beneficial effects on autoimmune disease in humans and animals that do not seem to depend on direct immunomodulating effects of bromocriptine.

In conclusion, the proinflammatory and immunestimulating role of prolactin has clearly been found in animal models of inflammatory diseases; however, this role in human diseases is still open for discussion.

### Melatonin (Pineal Gland Hormone)

The indole hormone melatonin (N-acetyl-5-methoxytryptamine) is the main secretory product of the pineal gland. Melatonin is synthesized in a strictly

nocturnal pattern. Synthesis is directly coupled to secretion; both begin at dusk, continue through the night, and terminate at dawn. This rhythmic pattern of secretion is circadian and depends on the master biological clock that resides in the suprachiasmatic nucleus, the paraventricular hypothalamic nucleus, and its direct or indirect projections to the sympathetic preganglionic neurons in the spinal cord. Direct inhibitory effects of light on pineal activity may contribute to phasing of the onset and termination of melatonin production. Indeed, melatonin may participate in the coordination of circadian responses, reinforcing or complementing other entraining signals, and may coordinate a variety of seasonal photo-periodic responses. Melatonin is also an efficient free radical scavenger and antioxidant, both *in vitro* and *in vivo*.

Moreover, melatonin exerts a variety of effects on the immune system. In several species, pinealectomy or any other experimental procedure that inhibits melatonin synthesis and secretion induces a state of immunodepression that is counteracted by melatonin replacement. In general, melatonin displays an immunoenhancing effect. Melatonin is able to activate T lymphocytes, monocytes, natural killer (NK) cells, and neutrophils. Melatonin even activates Ab-dependent cellular cytotoxicity and enhances Ab responses *in vivo* (8, 12, 21–23). On the other hand, melatonin enhances splenic ornithine decarboxylase activity (an index of lymphatic cell proliferation) in rats. In animal models as well as in human and *in vitro* experiments, melatonin enhances inflammatory cytokine (e.g., IL-1, IL-2, IL-4, IL-5, IL-6, TNF $\alpha$ , GM-CSF, TGF- $\beta$ , IFN- $\gamma$ ) and nitric oxide production.

## IMMUNOMODULATION DURING PREGNANCY

Hormonal and related immune response changes observed during pregnancy and postpartum might contribute to the variations of disease activity in autoimmune diseases such as RA and SLE. During normal pregnancy, both free and total cortisol levels are increased and the circadian secretion persists, but evening levels do not decline as they do in nonpregnant women. Estrogens and progesterone are synthesized in great amounts and decrease metabolic clearance of cortisol by direct action on liver enzymes involved in its inactivation.

Because high concentrations of estrogens seem to exert immunosuppressive activities, this effect should at least partly account for the improvement of arthritic symptoms in pregnant patients. On the other hand, during pregnancy, prolactin reaches its highest levels during the third trimester. Optimal concentrations are required for maximal lymphocyte stimulation; however, lower or higher levels cause decreased or negative responses. Therefore, elevated prolactin concentrations during pregnancy and early lactation might induce immunosuppression.

The maternal production rate of the adrenal androgen DHEAS is increased at least two times during pregnancy, but serum levels seem to be unchanged. On the other hand, serum levels of gonadal androgens increase. Serum total testosterone reaches levels four to five times higher than those in menstruating women, whereas serum-free testosterone increases only during the third trimester.

The hormonal changes characterizing the pregnancy are primarily directed toward a reduction of the immune/inflammatory response and, at least in RA and (more recently) pregnancy and the postpartum period, have been suggested to represent a model of how abrupt changes in steroid hormone levels may regulate the immune response. Recently, the cytokine expression during pregnancy has seemed to shift toward a Th2-type profile; in particular, high levels of anti-inflammatory IL-10 are expressed in placental tissues. Similarly, the administration of corticosteroids seems to increase serum levels of IL-10 and generally seems to polarize toward a Th2-type cytokine pattern. Conversely, RA can flare up during the postpartum period because the levels of IL-10 decrease to normal or even subnormal concentrations following the decrease in gonadal and adrenal steroid serum levels.

In conclusion, it has been suggested that the plasma increase of estrogens, androgens, and cortisol during the third trimester of pregnancy may induce an improvement in RA as consequence of a suppression of the proinflammatory Th1-type cytokine pattern and a potentiation of the anti-inflammatory Th2-type cytokine pattern.

## NEW PERSPECTIVES AND CONCLUSIONS

Recent studies showed that testosterone therapy dramatically suppresses the lymphocyte infiltration in,

and significantly improves the functional activity of lacrimal glands in the MRL/lpr female mouse model of Sjögren syndrome. A subsequent study showed that androgen treatment influences the expression of proto-oncogenes, with decreased bcl-2 and c-myc mRNA levels, as well as apoptotic factors in salivary and lacrimal tissues of the same model of Sjögren syndrome. Conversely, the recent observation of estrogen-induced decrease of apoptosis of PBMCs of SLE patients suggests that the presence of estrogens may allow survival of autoimmune cells. If sex hormones, as well as cytotoxic agents, do modulate synovial macrophage apoptosis in, for example, RA patients, such an approach would promise to be an important pathway of control in the arthritis.

The complex interactions and the molecular basis of common pathways between inflammatory/immunological and neuroendocrine circuits are matters of continuous and intensive research and may offer highly promising strategies for therapeutic manipulations of autoimmune diseases.

### See Also the Following Articles

Aging, Immunology and • Corticotropin-Releasing Hormone (CRH) and Inflammation • Cytokine Actions, Cellular Mechanism of • Cytokine Receptors • Glucocorticoids and Immunity • Thyroid Autoimmunity

### Further Reading

- Bijlsma, J. W. J., Straub, R. H., Masi, A. T., Lahita, R. G., and Cutolo, M. (2002). Neuroendocrine immune mechanisms in rheumatic diseases. *Trends Immunol.* **23**, 59–61.
- Chrousos, G. P. (1995). The hypothalamic–pituitary–adrenal axis and immune-mediated inflammation. *N. Engl. J. Med.* **332**, 1351–1362.
- Cutolo, M., Accardo, S., Villaggio, B., Barone, A., Sulli, A., Coviello, D. A., Carabbio, C., Felli, L., Miceli, D., Farruggio, R., Carruba, G., and Castagnetta, L. (1966). Androgen and estrogen receptors are present in primary cultures of human synovial macrophages. *J. Clin. Endocrinol. Metab.* **81**, 820–827.
- Cutolo, M., Villaggio, B., Candido, F., Valenti, S., Giusti, M., Felli, L., Sulli, A., and Accardo, S. (1999). Melatonin influences IL-12 and nitric oxide production by primary cultures of rheumatoid synovial macrophages and THP-1 cells. *Ann. NY Acad. Sci.* **876**, 246–255.
- Cutolo, M., and Wilder, R. (2000). Different roles for androgens and estrogens in the susceptibility to autoimmune rheumatic diseases. *Rheum. Dis. Clin. North America* **26**, 825–839.
- Elenkov, I. J., Wilder, R. L., Bakalov, V. K., Link, A. A., Minitrov, M. A., Fisher, S., Crane, M., Kanik, K. S., and Crousos, G. P. (2001). IL-12, TNF-alpha, and hormonal changes during late pregnancy and early postpartum: Implications for autoimmune disease activity during these times. *J. Clin. Endocrinol. Metab.* **86**, 4933–4938.
- Masi, A. T., Bijlsma, J. W., Chikanza, I. C., Pitzalis, C., and Cutolo, M. (1999). Neuroendocrine, immunologic, and microvascular systems interactions in rheumatoid arthritis: Physiopathogenetic and therapeutic perspectives. *Semin Arthr. Rheum.* **29**, 65–81.
- Straub, R. H., and Cutolo, M. (2001). Impact of the hypothalamic–pituitary–adrenal–gonadal axes and the peripheral nervous system in rheumatoid arthritis: A systemic pathogenic viewpoint. *Arthr. Rheum.* **44**, 493–507.
- Wilder, R. H., and Elenkov, I. J. (1999). Hormonal regulation of tumor necrosis factor-alpha, interleukin-12, and interleukin-10 production by activated macrophages: A disease-modifying mechanism in rheumatoid arthritis and systemic lupus erythematosus. *Ann. NY Acad. Sci.* **876**, 14–31.
- Worthington, J., Jones, R., Crawford, M., and Forti, A. (1994). Pregnancy and multiple sclerosis: A 3 year prospective study. *J. Neurol.* **341**, 228–233.

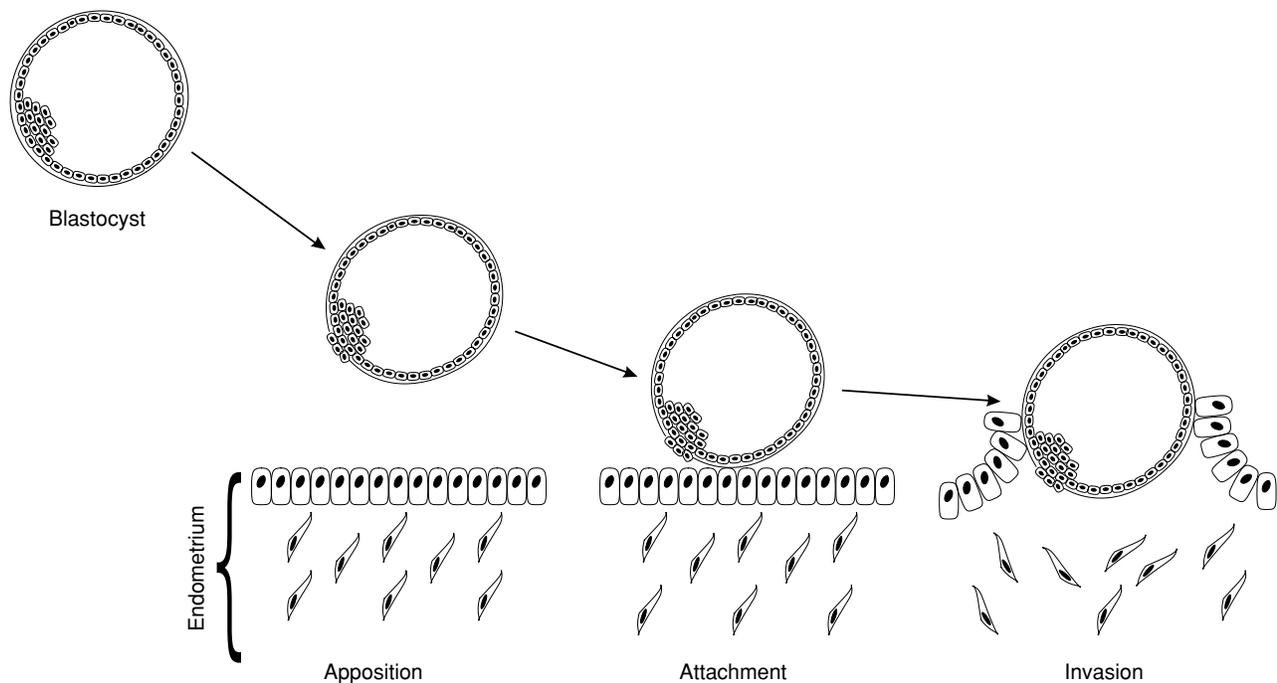


Figure 1 Stage of blastocyst implantation.

the secretory histological type. Progesterone, in cooperation with  $E_2$ , initiates decidual transformation of the endometrium, induces prostaglandin  $E_2$  production by endometrial macrophages, and causes stromal edema and angiogenesis in the endometrium. In addition, progesterone decreases uterine irritability and contractility and, together with hCG, suppresses immunological response that prevents embryo rejection by the maternal immunological system.

During embryo implantation, the uterine endometrium undergoes morphological and physiological changes to the host conceptus. The restricted period of time when the endometrium is receptive for the conceptus is called the "implantation window." The embryo being transferred to the uterus at an inappropriate time is not able to implant. In the normal human menstrual cycle lasting 28 days, the implantation window lasts between 19 and 24 days of the cycle. One of the earliest indexes for transmission of embryonic signals to the endometrium that surrounds the blastocysts is capillary permeability and local stromal edema. This leads to the changes in the endometrium termed "decidualization." During this process, endometrial stromal cells transform into large decidual cells that are rich in glycogen.

Attachment of the embryo to the endometrium is facilitated by the specific cell adhesion structures and molecules located on the apical epithelial plasma membrane.

#### **Pinopodes**

A characteristic feature for the status of endometrial receptivity is formation of the pinopodes. Pinopodes, a large cytoplasmic projection located on the apical surface of endometrial epithelium, are associated with changes in cell membrane morphology during the implantation window opening. Progesterone is responsible for the formation of these structures. Pinopodes are involved in facilitating adhesion of the blastocyst to the endometrial epithelium.

#### **Trophinin/Tastin**

Trophinin, together with its adhesion molecule tastin (trophinin-assisting protein), is additional protein responsible for trophoblast cell adhesion to the endometrial epithelium. Tastin is located in the endometrial cell membrane. The cytoplasmic domain of trophinin is bound to the cytoskeleton. The presence of highly concentrated areas of trophinin "patches" is believed to be adhesion sites for the embryo. The trophinin is expressed exclusively when the implantation window is opened.

#### **Mucin: MUC1**

MUC1 is an integral membrane glycoprotein. MUC1 has high molecular weight, a highly glycosylated extracellular domain, and a relatively short domain associated with the cytoskeleton. The endometrial expression of MUC1 is menstrual cycle dependent.

It is believed that MUC1 prevents adhesion of the embryo to endometrial epithelium and that only healthy blastocysts are able to decrease MUC1 expression leading to the contact of the cells. In addition to the endometrium, MUC1 is expressed in blastocyst during the preimplantation period.

### **Integrins**

Cell-to-cell adhesion seems to be a crucial process in the course of implantation. Integrins are transmembrane glycoproteins responsible for this process. They are members of an immunoglobulin superfamily. Integrins are composed of two units:  $\alpha$  and  $\beta$ . Between several types of integrins, heterodimers  $\alpha_4\beta_3$  and  $\alpha_v\beta_1$  seem to play an important role in the implantation. Integrin  $\alpha_4$  and integrin  $\beta_3$  coexpress in the apical secretive endometrium when the implantation window is opened. For this reason, integrin  $\beta_3$  is proposed to be a marker of uterine receptivity. Integrins are believed to be receptors for extracellular matrix components: fibronectin, collagen, and laminin. Because trophoblasts express oncofetal fibronectin, the presence of integrins on apical endometrial epithelial cells may have a major role in the initial process of embryo adhesion to decidual cells. Contact of the embryo with endometrium trigger the production of proteases by trophoblasts, and this is crucial for the next step of implantation.

## **INVASION**

Proteases degrade the endometrial extracellular matrix, helping the embryo to invade the uterine wall. Matrix metalloproteinases and urokinase-type plasminogen activator are involved in extracellular matrix proteolysis. In addition, first-trimester cytotrophoblasts express specific enzymes called gelatinases, which are responsible for matrix degradation. Enzymes are controlled indirectly by cytokines and by interaction with the extracellular matrix. Proteolytic activity of invading trophoblasts is a result of balance between matrix metalloproteinases secreted by the embryo and specific inhibitors, including tissue inhibitors of metalloproteinases originating from decidua.

## **MARKERS OF IMPLANTATION**

Trophoblasts and decidua produce lots of factors and their receptors affecting events associated with the process of implantation. Changes in the expression

of platelet-derived growth factor (PDGF), fibroblast growth factor (FGF), insulin-like growth factor (IGF), transforming growth factor (TGF), IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-11, IL-12, and colony-stimulating factor (CSF) at the time of implantation were noted; however, their precise role in the implantation processes remains to be elicited. Because there are so many different factors involved in the process of implantation, it is worth determining which ones could be used as predictors of successful implantation. The issue is still open to debate, but leukemia inhibitory factor (LIF) and hemoglobin-binding epidermal growth factor (HB-EGF) are candidates for this function.

LIF expresses in endometrial epithelial cells just before implantation and immunoreactivity of LIF in the endometrium decrease when substances blocking implantation (RU486) are administered. The role of this cytokine in the process of implantation is confirmed by results showing that mouse lacking LIF become infertile. HB-EGF has been found in endometrial epithelial cells close to the place of embryo implantation.

### **See Also the Following Articles**

Assisted Reproductive Technology (ART) • Endometriosis • Fertilization • Infertility, Overview • Menstrual Cycle: An Integrative View • Pregnancy Endocrinology

### **Further Reading**

- Giudice, L. C. (1999). Implantation and endometrial function. *In* "Molecular Biology in Reproductive Medicine" (B. J. C. M. Fauser, ed.), pp. 333–352. Parthenon, New York.
- Lessey, B. A. (2000). The role of endometrium during embryo implantation. *Hum. Reprod.* **15**(Suppl. 6), 39–50.
- Lessey, B. A. (2001). The use of biomarkers for the assessment of uterine receptivity. *In* "Textbook of Assisted Reproductive Techniques: Laboratory and Clinical Perspectives" (D. K. Gardner, A. Weissman, C. M. Howles, and Z. Shoham, eds.), pp. 353–366. Martin Dunitz, London.
- Pablo, J. L., Meseguer, M., Caballero-Campo, P., Pellicer, A., and Simon, C. (2001). Embryonic regulation in the process of implantation. *In* "Textbook of Assisted Reproductive Techniques: Laboratory and Clinical Perspectives" (D. K. Gardner, A. Weissman, C. M. Howles, and Z. Shoham, eds.), pp. 341–352. Martin Dunitz, London.
- Sharkey, A. M., and Smith, S. K. (1999). Local regulators of embryo–endometrial interactions at implantation. *In* "A Textbook of *in Vitro* Fertilization and Assisted Reproduction: The Bourn Hall Guide to Clinical and Laboratory Practice" (P. R. Brindsen, ed.), 2nd ed., pp. 267–273. Parthenon, New York.

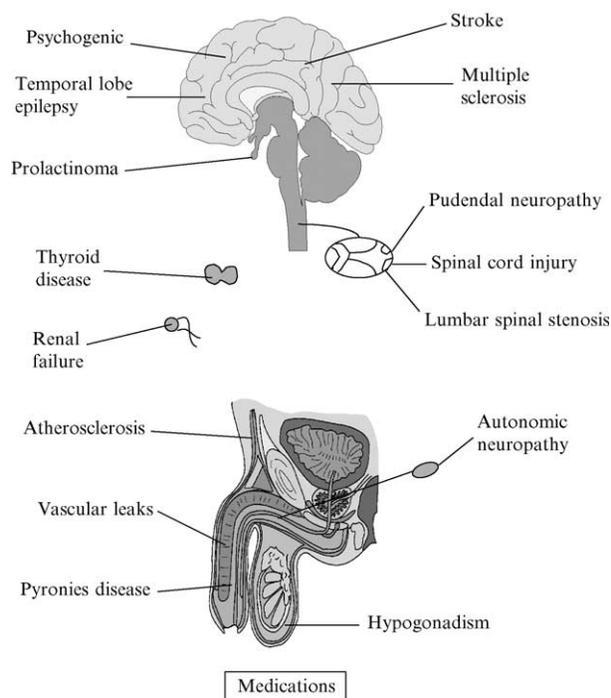


Figure 1 Causes of impotence.

disease are at major risk for developing either a myocardial infarction or a stroke over the subsequent 2 years. Autonomic neuropathy is another important cause. Persons with diabetes mellitus usually have a mixed cause of impotence, including vascular disease and neuropathic causes as well as increased collagen cross-linking.

Medications are a major cause of impotence. Impotence is produced by medications of nearly every category, including antidepressants (with the exception of trazadone), antihypertensives, antipsychotics, anticholinergics, H<sub>2</sub>-receptor antagonists, anticonvulsants, and antiandrogens (digitalis, estrogens, spironolactone, megestrol acetate, ketoconazole, and gonadotropin-releasing hormone [GnRH] antagonists). Because older persons often receive multiple drugs, this cause becomes more common with aging. Tobacco smoking is commonly associated with impotence, both acutely due to the effects of nicotine on erection and chronically secondary to its role in the pathogenesis of atherosclerosis. Alcohol has, since the time of Shakespeare, been known to “provoke desire and inhibit performance.”

Multiple sclerosis, strokes, temporal lobe epilepsy, and spinal cord injury all are associated with impotence. Pituitary tumors (prolactinomas) and hyper- and hypothyroidism all can result in impotence.

Systemic disorders, such as chronic obstructive pulmonary disease (COPD) and renal failure, can lead to impotence. Oxygen therapy can reverse impotence in men with COPD.

Although psychogenic causes of impotence are relatively rare in older persons, they need to be considered in the differential diagnosis. Performance anxiety is commonly seen as an additive cause of impotence in men with early organic disease. Depression is the most common psychogenic cause of impotence. Widower’s syndrome is a unique condition in older men. In this case, the man’s wife dies and a female friend starts providing the widower with companionship as well as help with the housework and cooking. After a while, she suggests that it would be appropriate for him to reward these “services” by having intercourse with her. Although the man does not want the woman’s help to be withdrawn, he does not want to have a sexual relationship with her. To avoid having intercourse, he develops impotence.

### TESTOSTERONE, IMPOTENCE, AND THE ANDROPAUSE

Longitudinal studies with the aging have clearly demonstrated that testosterone levels decline with aging. This is due predominantly to chaotic secretion of GnRH from the hypothalamus, but it also involves altered pituitary secretion of leutenizing hormone and decreased production of testosterone from the testes. In addition, there is an increase in sex hormone-binding globulin and a decrease in the apparent  $K_d$  of testosterone for this globulin. Thus, measurement of total testosterone in older persons gives a spuriously high value in functional terms. For this reason, it is now recommended that bioavailable testosterone be measured to give a true estimation of the tissue-available testosterone in older men.

Based on the measurement of bioavailable testosterone, the prevalence of hypogonadism is estimated to be between 3 and 5% at 40 years of age and between 34 and 70% at 70 years of age. Because testosterone is essential for the production of NO synthase in the penis, very low levels of testosterone are associated with complete impotence. Higher levels of testosterone are associated with the development of a soft erection. In persons with COPD, the erectile problems appear to be secondary to low testosterone and to hypoxia reducing the production of NO.

Low testosterone is also associated with a decline in libido. This is a more important effect of testosterone

than its effect on potency. Other effects of the declining testosterone with aging include decreased muscle mass and strength, increased fat mass, decreased memory, increased leptin, decreased hematocrit, decreased bone mass, and a decline in functional status. Testosterone replacement in older persons has been shown to decrease angina and enhance coronary artery dilatation. Low—not high—testosterone is associated with atherosclerotic disease.

At Saint Louis University, we have developed a screening test for the andropause (Table I). This test has good sensitivity and specificity for detecting persons with low bioavailable testosterone and can be used to follow the outcomes of therapy. The major false positive for this questionnaire is dysphoria, and all men who answer positively on this questionnaire should be screened for depression before being treated for testosterone deficiency. The Massachusetts Male Aging Study has developed a risk factor survey that also has good sensitivity and specificity. The Aging Males' Symptoms Rating Scale was developed in Europe. It is not specific for testosterone deficiency but covers a large range of symptoms experienced by older men.

## MANAGEMENT OF IMPOTENCE

Table II lists the major modalities available for the treatment of impotence in older men. Phosphodiesterase-5 inhibitors represent the major advance in the management of impotence. These agents delay the breakdown of cyclic GMP and,

**Table I The Androgen Deficiency in Aging Male (ADAM) Questionnaire**

(Circle one)		
Yes	No	
		1. Do you have a decrease in libido (sex drive)?
Yes	No	2. Do you have a lack of energy?
Yes	No	3. Do you have a decrease in strength and/or endurance?
Yes	No	4. Have you lost height?
Yes	No	5. Have you noticed a decreased enjoyment of life?
Yes	No	6. Are you sad and/or grumpy?
Yes	No	7. Are your erections less strong?
Yes	No	8. Have you noticed a recent deterioration in your ability to play sports?
Yes	No	9. Are you falling asleep after dinner?
Yes	No	10. Has there been a recent deterioration in your work performance?

*Note.* A positive answer represent yes to question 1 or 7 or any three other questions.

**Table II Management of Impotence in Older Men**

1. Psychological
  - a. Treat depression
  - b. Behavior therapy
  - c. Couples counseling
2. Stop medications that may be causing impotence wherever possible
3. Medications
  - a. Phosphodiesterase inhibitors
    - (1) Pentoxifylline
    - (2) Sildenafil (Viagra)
    - (3) Vardenafil
    - (4) IC351
  - b. Apomorphine slow release (Uprima)
  - c. Zinc (if zinc deficient)
  - d. Yohimbine (poor efficacy)
  - e. Testosterone
  - f. Intracavernosal agents
    - (1) Prostaglandin E<sub>1</sub>
    - (2) Phentolamine
    - (3) Papaverine
    - (4) Vasoactive intestinal peptide
  - g. Intraurethral
    - (1) Prostaglandin E<sub>1</sub>
  - h. Transpenile
    - (1) Nitroglycerine
    - (2) Minoxidil
4. Vacuum tumescent devices
5. Surgery
  - a. Prosthesis
    - (1) Semi-rigid
    - (2) Inflatable
  - b. Arterial surgery
  - c. Venous surgery
  - d. Angioplasty

thus, increase the vasodilatory efficacy of the NO released locally in the penile tissue. We originally demonstrated that pentoxifylline, a nonspecific phosphodiesterase inhibitor, reversed impotence in some men. Sildenafil (Viagra), a specific phosphodiesterase-5 inhibitor, reverses erectile dysfunction in 50 to 70% of men. It is slightly less effective at enhancing erectile dysfunction in older men than in younger men. The major side effects of sildenafil are gastroesophageal reflux, hypotension, alterations in vision, and death when given in conjunction with nitrates.

Direct injections of papaverine, phentolamine, or prostaglandin E<sub>1</sub> into the corpora cavernosa are more efficacious than phosphodiesterase-5 inhibitors at producing erections. Slow-release apomorphin (Uprima)

works through the central nervous system to increase erections. It is less effective than sildenafil.

Vacuum tumescence devices play a role in the management of impotence in older men who are interested in obtaining an erection once or twice a month. Better results with these devices are obtained in men who are in stable relationships.

Penile prostheses remain the gold standard for the treatment of impotence for many older men. In older men, the semi-rigid prosthesis is usually a better choice than the inflatable prosthesis.

## THE PROSTATE AND IMPOTENCE

Radical prostatectomy for prostate carcinoma may result in impotence in approximately half of the patients, an untoward effect of surgery that appears to be operator dependent. In addition, many patients who have a prostatectomy also have either an orchiectomy or receive GnRH analogues to lower their testosterone. Hypogonadism in these patients not only decreases libido but also can cause hot flashes, night sweats, fatigue, muscle loss, osteoporosis, and a decline in hematocrit. In men without metastases following prostatectomy, if their PSA is zero for 3 to 5 years following surgery and they are hypogonadal, testosterone replacement into the low physiological range with careful monitoring of the PSA is becoming an acceptable practice.

## COMPLEMENTARY MEDICINE AND IMPOTENCE

The phosphodiesterase inhibitor canthridin ("Spanish fly") is obtained from blister beetles. It produces

congestion of the corpora but also produces inflammation and hematuria. NO synthase is stimulated by ginsenosides that are derived from *Panax ginseng*. Ambien, an herbal preparation of *Ambra gnsea*, produces a very small stimulation of luteinizing hormone leading to miniscule increases in circulating testosterone. A number of these products are now used by older men.

## See Also the Following Articles

Aging and Longevity of Human Populations • Aging and the Male Reproductive System • Erectile Dysfunction • Fertility in Men with Spermatogenesis Abnormalities • Gonadotropins and Testicular Function in Aging • Neuroendocrine System and Aging • Oxidative Stress and Aging • Prostate Cancer • Sexual Function and Androgens • Stress, Aging, and Central Nervous System Interactions • Thyroid, Aging and

## Further Reading

- Andersson, K. E. (2001). Pharmacology of penile erection. *Pharmacol. Rev.* **53**, 417–450.
- Korenman, S. G. (1998). New insights into erectile dysfunction: A practical approach. *Am. J. Med.* **105**, 135–144.
- Morales, A., Heaton, J. P. W., and Carson, C. C. (2000). Andropause: A misnomer for a true clinical entity. *J. Urol.* **163**, 705–712.
- Morley, J. E. (1986). Impotence. *Am. J. Med.* **80**, 897–905.
- Morley, J. E. (2001). Androgens and aging. *Maturitas* **38**, 61–71.
- Morley, J. E., and Perry, H. M. (2000). Androgen deficiency in aging men: Role of testosterone replacement therapy. *J. Lab. Clin. Med.* **135**, 370–378.
- Slag, M. F., Morley, J. E., Elson, M. K., Trencle, D. L., Nelson, C. J., Nelson, A. E., Kinlaw, W. B., Beyer, H. S., Nuttall, F. Q., and Shafer, R. B. (1983). Impotence in medical clinic outpatients. *JAMA* **249**, 1736–1740.
- von Keitz, A. (2001). The management of erectile dysfunction in the community. *Intl. J. Impotence Res.* **13**(Suppl. 3), S45–S51.

Table I History of in Vitro Fertilization

Year	Investigator(s)	Event
1870–1880	S. Schenk	Observed rabbit sperm stripping away cumulus cells from egg under the microscope
1890	Walter Heap	Successful transfer of fertilized rabbit oocytes flushed from donor rabbit to foster mother rabbit
1930	Gregory Pincus	Observed rabbit sperm penetrating the egg with binocular microscope
1934	Gregory Pincus and E. V. Enzmann	Mixed rabbit sperm and eggs and returned them to the fallopian tube with birth of young; fertilization not observed
1944	John Rock and Miriam Menkin	Harvested human eggs just before ovulation, combined them with human sperm, and observed the results
1954	Charles Thibault	Reported in vitro fertilization of rabbit eggs; first report of mammalian fertilization in vitro
1959	M. C. Chang	Fertilized rabbit eggs in vitro and transferred four-cell embryos with birth of healthy young
1969	Robert Edwards, Barry Bavister, and Patrick Steptoe	Human in vitro fertilization with embryo division
1971	Robert Edwards and Patrick Steptoe	First transfer of in vitro fertilization embryo with intent of establishing a pregnancy
1978	Robert Edwards and Patrick Steptoe	Birth of Louise Brown
1992	Gianpiero Palermo	Development of intracytoplasmic sperm injection

Pincus began to study rabbit eggs and sperm under the binocular microscope. In 1944, Rock and Menkin reported in vitro fertilization of a human egg but did not intend to use the resulting embryo in an attempt to start a pregnancy. In 1967, Edwards and Steptoe began work on a technique for harvesting and fertilizing human eggs *in vitro* with the intention of treating infertile couples. Edwards reasoned that he could “simply pluck the egg from the ovary and fertilize it in the laboratory.” In 1969, Edwards, Bavister, and Steptoe reported that they had achieved human fertilization *in vitro*. The first IVF human pregnancies were achieved in 1975, although they were not successful in reaching term. Then, on July 25, 1978, Louise Joy Brown was born by caesarian section at Oldham General Hospital in England. Her birth was the first to demonstrate that conception in the laboratory could result in the birth of a normal baby. The first IVF baby in the United States was born in 1980. Shortly thereafter, in response to patient demand, numerous clinics offering ART procedures were established all over the world, ushering in a new epoch in helping infertile couples to achieve pregnancy.

## IVF PROCEDURES

At first, most women treated with IVF procedures had damaged fallopian tubes. As confidence with the new technique increased, clinicians began to use IVF to treat other causes of infertility. Today, ART procedures are

used to help women with damaged fallopian tubes, endometriosis, couples who are infertile because of low sperm count or poor sperm function, unexplained infertility, and some genetic disorders.

## Ovulation Induction

Edwards and Steptoe obtained a single egg, which would become Louise Brown, with a retrieval timed just before her mother’s natural ovulation. Today, most women are treated with follicle-stimulating hormone (FSH) to promote the maturation of multiple eggs before their retrieval. This treatment is known as “ovulation induction” or “controlled ovarian hyperstimulation.” FSH treatment for ovulation induction is administered by daily injection for several days, beginning on the third day of menstrual bleeding. Many women are treated with gonadotropin-releasing hormone (GnRH) agonists, GnRH antagonists, or oral contraceptives to suppress their hypothalamic–pituitary–ovarian axis and, consequently, their reproductive cycle. If there is evidence that the hypothalamic–pituitary–ovarian axis is not suppressed, FSH injection is not begun. Women undergo monitoring of their serum estradiol concentration and sonogram imaging to determine the growth of their follicles. When the largest follicle has a diameter of greater than 18 mm, the egg is approaching maturity. Final egg maturation is triggered by injection of human chorionic gonadotropin (hCG) or luteinizing hormone

(LH). The clinician retrieves the eggs 34 to 37 h after this final injection.

### Egg Retrieval

Until the late 1980s, laparoscopy was used as the primary method of egg retrieval. To achieve pregnancy, some women would undergo multiple laparoscopies in a single year. Laparoscopic egg retrieval required about 45 min of general anesthesia and was more difficult when pelvic scars covered the ovaries. Today, physicians use ultrasound imaging to guide a needle through the vaginal wall to the ovary. Introduction of ultrasound-guided egg retrievals during the mid-1980s eliminated the need for laparoscopy and a full operating room to complete egg retrievals, allowing ART services to be provided in outpatient settings. Using ultrasound, each follicle is imaged on the display. The retrieval needle is passed through the needle guide, which is parallel to the vaginal ultrasound probe. The needle passes through the vagina and into the peritoneal cavity. The follicles are punctured, and their fluid is suctioned through the needle to a sterile container. The follicular fluid is brought to the embryology laboratory, where the eggs are identified under low-power magnification. The eggs are removed from the fluid, carefully examined by the embryologist, and evaluated for maturation. An egg must have completed its first meiotic division to be successfully fertilized. After the eggs are inspected, the embryologist places each egg in a small bubble of media along with a suspension of washed sperm.

### Embryology Laboratories

Embryology laboratories have evolved greatly over the years. Most successful embryology laboratories have protocols to ensure quality control and to maintain a sterile, toxin-free environment. Techniques have been developed in the embryology laboratories to assist in fertilization, embryo growth, and embryo selection and safe transfer.

### Fertilization

Expansion of assisted reproductive indications to help couples with poor sperm production or function led to methods of gamete micromanipulation to improve the chance of fertilization in the laboratory. A clear shell called the zona pellucida surrounds human eggs. Zona drilling and subzonal insertion were techniques used to help spermatozoa pass the zona pellucida

barrier and bind to the oocyte cell membrane. Working in Belgium in 1992, Palermo achieved fertilization of a human egg by injecting a sperm cell through the egg cell membrane into the cytoplasm, a procedure that became known as intracytoplasmic sperm injection (ICSI). Until then, many embryologists believed that intracytoplasmic injection of sperm would disrupt the basic mechanisms of fertilization. Palermo showed that an egg could be successfully fertilized by ICSI and that a healthy embryo and successful pregnancy could result.

Incorporation of ICSI techniques into the IVF procedures has allowed successful treatment of couples who in the past could only have been offered adoption or artificial insemination. A few living sperm cells retrieved from testicular biopsy or epididymal aspiration can be used for intracytoplasmic injection of oocytes. Use of laboratory techniques to overcome such natural barriers to fertility provoked concern that children resulting from these procedures might be adversely affected. Studies have shown that men with deletions or mutations of genes needed to produce spermatozoa are likely to pass on those traits when they have children by ART procedures. Some genetic traits associated with male infertility are closely linked to other inherited diseases such as cystic fibrosis. For these reasons, couples with severe male factor infertility are encouraged to seek genetic counseling as part of their preparation for infertility treatment.

### Embryo Culture

Over the first 5 days after fertilization, the embryo grows rapidly from a single-cell zygote to a hollow ball of many cells, called a blastocyst. During this crucial interval, the embryo makes its transition from a metabolism based on maternal genes to its own genetic program. The media used for embryo culture is a simple solution of salts and nutrients. High-quality embryos can grow in this standard media with good implantation and pregnancy rates.

The metabolic needs of the embryo change as the embryo grows. Sequential culture uses media of different types depending on the stage of embryo development. These more complex media are used to respond to the changing metabolic requirements of the embryo. Coculture with human- or animal-derived cells is another way in which to create a dynamic culture system. In coculture, the cultured cells are thought to act as hosts for the developing embryos, conditioning the media in response to signals from the embryo. Although there is evidence from animal and human experiments supporting

the use of these systems, the evidence in humans is empirical.

### **Embryo Transfer**

Over time, embryologists have learned to produce healthier embryos. In 1985, it was not unusual to transfer six to eight embryos in a single procedure. Today, to decrease the chance of multiple gestations, clinics commonly transfer no more than three high-quality embryos. Embryo transfer is typically done after 3 days of culture. In some cases, embryo culture may be extended to allow transfer of a single blastocyst. Some clinics will transfer more embryos if the overall embryo quality is poor. Embryos are transferred in a thin flexible straw that is threaded carefully through the cervix. The optimal transfer places the embryos in the mid-uterine cavity, together with a minimal amount of transfer media. The transfer is performed carefully to avoid uterine bleeding, which can decrease the chance of implantation. In most cases, the transfer procedure is painless. Embryo transfer may fail if the embryos become trapped in the cervical mucus or do not leave the transfer catheter. Each of these problems can be overcome by careful technique. Some clinicians use sonographic imaging to confirm intrauterine placement of the embryos.

A normal uterine cavity and endometrium will encourage implantation. A physician will often evaluate the patient's endometrial cavity with hysterosalpingogram or ultrasound imaging before the ART cycle begins. Endometrial polyps or fibroids inside the uterine cavity are frequently removed in preparation for treatment. Progesterone supplementation is used after embryo transfer to promote normal secretory endometrium. Embryo hatching is another way in which to assist implantation. Hatching is performed by making a slit in the zona pellucida just before the embryo is transferred. Hatching is thought to make it easier for the embryo to emerge from the zona and may improve signaling between the embryo and the endometrium.

### **Cryopreservation**

As mentioned previously, to reduce the risk of a multiple pregnancy, most clinics transfer only two or three embryos. Because an average ART cycle produces more than three embryos, there are often healthy embryos left over after the transfer. It is possible to freeze the leftover embryos to preserve them. This process is called cryopreservation. If a couple does not achieve pregnancy after IVF, the partners can attempt pregnancy again by transfer of thawed embryos remaining from their first treatment cycle.

Pregnancy rates after cryopreservation and thaw of human embryos are less than those following a fresh embryo transfer. The chance of pregnancy will depend on the quality of the embryos and on the maternal age at the time that the embryos were frozen. Live births have occurred after as long as 7 years of cryopreservation, although in most cases couples will use their frozen embryos within a short time. For some couples, cryopreserved embryos have become objects of legal contention. Who maintains custody of the embryos if a couple divorces? Are frozen embryos property? Do frozen embryos themselves have any rights?

### **Treatment Success and Risks**

Even though ART is expensive and invasive when compared with other therapies, ample empirical evidence supports the use of ART procedures for treatment of infertility. Live birth rates per initiated cycle vary internationally from 18 to 25%. Chance of success is greatest with younger maternal age and evidence of normal ovarian function. Women over 40 years of age or with high concentrations of FSH early in the menstrual cycle have little chance of achieving a live born child. Multiple pregnancy rates after ART treatment average 25 to 40% and can be as high as 60% in some series. The high rate of multiple pregnancy after ART treatment contributes to an increase in spontaneous pregnancy loss (25%), ectopic pregnancy (5%), and preterm delivery. Ovarian hyperstimulation syndrome occurs in 5% of cycles and is another significant consequence of ART treatment.

## **FUTURE IMPLICATIONS OF ART TECHNIQUES**

It is likely that we have only just begun to see the direction that these technologies can take us. ART techniques, combined with a growing knowledge of molecular and cell biology, will allow further development of preimplantation diagnosis and possible preimplantation gene therapy. The goal of helping to reverse the effects of aging on reproductive potential may one day be realized through the application of techniques borrowed from cloning and stem cell research. Further improvement in embryo culture technique and embryo selection should increase the number of established pregnancies, reduce the number of multiple pregnancies and pregnancy losses, and further increase the live birth rate per initiated ART cycle.

## ETHICAL CONCERNS

Some have expressed concern that the expense and risks of ART may outweigh the benefit to the few couples who achieve successful live births. Further concern about expense, risk, and commercialization arises if an egg donor or gestational surrogate, who bears procreative risk in return for payment, is involved. There is consensus that surrogates and donors should be paid for their services but not paid so much that the participants are blinded to their potential risks. In many countries, ART procedures are available only to those who can afford them. Does a society have responsibility to provide infertility care to all of its members as it would to any individual suffering with disease? Some governments have dictated the maximum number of eggs to be retrieved and fertilized and the maximum number of embryos to be transferred per case in an effort to decrease the chance of a multiple gestation. Is this intrusion on the right of individuals to have control over their medical care justified by a concern for patients' safety and the safety of their potential fetuses? Do governments have a right to limit the types of reproductive services available to their citizens?

In July 1978, after Louise Brown was born, religious leaders complained that doctors did not have the right to "play god." Steptoe replied, "I don't know what the fuss is all about. I'm just glad we were able to help Mrs. Brown to have such a happy healthy baby."

Today, because of ART, millions of couples and their physicians have had the joyful opportunity of echoing Steptoe's sentiment.

## See Also the Following Articles

Anti-Müllerian Hormone • Assisted Reproductive Technology (ART) • Endometriosis • Fertility in Men with Spermatogenesis Abnormalities • Fertilization • Gonadotropin-Induced Ovulation • Implantation • Infertility, Overview • Ovarian Failure Treatment Strategies: Egg Donation • Pregnancy Endocrinology • Premature Ovarian Failure • Superovulation and Intrauterine Insemination

## Further Reading

- Edwards, R. G. (1990). "Establishing a Successful Human Pregnancy." Raven, New York.
- Edwards, R. G., and Steptoe, P. C. (1980). "A Matter of Life: The Story of a Medical Breakthrough." William Morrow, New York.
- Jones, H. W. (1986). "In Vitro Fertilization—Norfolk." Williams & Wilkins, Baltimore, MD.
- Macklin, R. (1994). "Surrogates and Other Mothers: The Debates Over Assisted Reproduction." Temple University Press, Philadelphia.
- Marsh, M. (1996). "The Empty Cradle: Infertility in America from Colonial Times to the Present." Johns Hopkins University Press, Baltimore, MD.
- Trounson, A., and Gardner, D. K. (1999). "Handbook of in Vitro Fertilization," 2nd ed. CRC, Boca Raton, FL.
- Wisot, A. L., and Meldrum, D. R. (1990). "New Options to Fertility: A Guide to in Vitro Fertilization and Other Assisted Reproductive Methods." Pharos Books, New York.

women due to a higher prevalence of biliary disease) considering that the prevalence of adrenal masses is equal in both sexes in autopsy studies. Autopsies series and CT investigations indicate that both adrenals are affected equally. Bilateral incidentally detected adrenal masses have been reported in 2 to 10% of cases.

## DIFFERENTIAL DIAGNOSIS

The histological characterization of incidentalomas reveals many different pathological entities (see [Table I](#)). Their exact prevalence is not known because only a small portion of the patients require surgical

**Table I** Differential Diagnosis of Incidentally Detected Adrenal Masses According to the Italian Multicentre Study

	n
Adrenal adenoma	198 (52)
Inactive	137 (36)
Cortisol producing	50 (13)
Aldosteronoma	12 (3)
Androgen producing	—
Estrogen producing	—
Adrenocortical carcinoma	47 (12)
Nodular hyperplasia	—
Pheochromocytoma	42 (11)
Ganglioneurinoma	15 (4)
Ganglioneuroblastoma	
Angiomyolipoma	
Lipoma	
Liposarcoma	
Myelolipoma	30 (8)
Lymphangioma	
Hemangioma	
Hematoma/Hemorrhage	
Abscess	
Adrenal cyst	20 (5)
Metastasis	7 (2)
Renal cell carcinoma	
Retroperitoneal sarcoma	
Gastric leiomyoma	
Neurilemmoma	
Accessory spleen	
Other	21 (6)
Total	380

*Note.* Percentages are in parentheses. The table reflects the distribution of tumors in patients undergoing surgery for various reason. Therefore, functional and malignant tumors are overrepresented in this series. A total of 380 of 1004 patients underwent adrenalectomy.

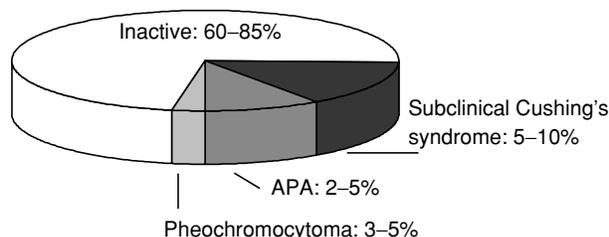
removal of their tumors. The distributions of the pathological origins of adrenal incidentalomas vary with several clinically important factors, including cancer history and mass size. Among patients with a history of cancer, three-fourths of the incidentalomas are metastases. In contrast, in non-cancer patients, more than 75% of adrenal incidentalomas represent benign tumors. In these patients, the most frequent diagnosis by far is that of a benign, hormonally inactive adrenal adenoma. Approximately 20% of all lesions are functional adrenal tumors. The most frequent sub-clinical disease is SCCS (5–17%). Pheochromocytoma is found in approximately 5% of patients, and (normo-kaliemic) Conn's syndrome is found in 2 to 5% of patients. ACC is an extremely rare tumor, with an estimated annual incidence of 0.5 to 2 cases per 1 million population. The prevalence of ACC in incidentaloma patients is related to the size of the mass. In surgical series, ACC accounts for 2% of tumors less than 4.0 cm, 6% of tumors 4.1 to 6.0 cm, and 25% of tumors greater than 6.0 cm. However, the prevalence of ACC is overestimated in surgical series because the suspicion of ACC is an indication for surgery. Moreover, the frequency of ACC is derived from highly selected patient populations and does not reflect the prevalence rate seen in population-based studies. The age and sex of the patient is not helpful in predicting the presence of ACC because it is found equally in both sexes and in all age classes.

## DIAGNOSTIC APPROACH

The patient with an adrenal incidentaloma requires a complete history and physical examination, a biochemical workup, and (possibly) additional radiological studies. Careful evaluation holds the potential of early detection of a harmful underlying disease and can lead to a curative therapy, for example, in secondary hypertension or adrenal carcinoma. The diagnostic approach has two foci: endocrine activity and the risk of malignancy.

## Endocrinological Investigation

Roughly 20% of all incidentalomas are functional (see [Fig. 1](#), showing autonomous endocrine activity in adrenal incidentaloma). Several studies have shown that the risk of subclinical endocrine activity increases with size. Lesions smaller than 1 cm are generally nonfunctional. However, the prevalence of hormone excess increases to 40% in lesions that are 6 cm or larger. The goal of the studies has been to determine whether the patients have mild glucocorticoid excess



**Figure 1** Prevalence of subclinical biochemical disease in patients with adrenal incidentalomas.

(SCCS), primary hyperaldosteronism (Conn's syndrome), or pheochromocytoma (Table II).

Significant autonomous cortisol secretion (SCCS) is found in 5 to 20% of patients. The spectrum of biochemical abnormalities depends on the amount of glucocorticoids secreted and ranges from only a slight attenuated diurnal cortisol rhythm to a complete atrophy of the contralateral gland. SCCS must be excluded in every patient scheduled for surgery to avoid postoperative adrenal crisis. The best means to uncover autonomous cortisol secretion is the short (overnight) dexamethasone suppression test, which rarely fails to detect SCCS. To reduce false positive results, we prefer a higher dexamethasone dose (3 mg instead of 1 mg). A suppressed serum cortisol concentration ( $<5 \mu\text{g/dl}$  or  $110 \text{ nmol/L}$ ) excludes significant cortisol secretion by the tumor. A positive low-dose dexamethasone requires a confirmatory high-dose dexamethasone

**Table II** Stepwise Endocrine Evaluation of a Patient with Adrenal Incidentaloma

Step 1: Screening test: Excluding endocrine activity
Serum cortisol after dexamethasone suppression (3 mg at 11 PM orally)
Serum potassium and plasma aldosterone/plasma renin activity ratio in hypertensive patients
24-h urinary catecholamine excretion
Step 2: Confirming endocrine activity
Confirmation of SCCS by high-dose (8-mg) dexamethasone suppression
Confirmation of Conn's syndrome by determination of plasma aldosterone during saline infusion test
Confirmation of pheochromocytoma by repeat urinary catecholamine excretion
Step 3: Determining outcome/origin of hormone excess
SCCS: CRH stimulation test, urinary-free cortisol
Conn's syndrome: Upright posture test, selective adrenal venous sampling (optional)
Pheochromocytoma: meta-iodine-benzyl-guanidine (MIBG) scintigraphy, genetic testing (optional)

suppression test (8 mg). Further workup should include a corticotropin-releasing hormone (CRH) test to evaluate the degree of hypothalamic-pituitary suppression, the analysis of the diurnal cortisol rhythm, and the measurement of urinary-free cortisol. Patients with low adrenocorticotrophic hormone (ACTH) concentrations who are not responding to CRH are at risk for adrenal insufficiency after surgery and require adequate substitution therapy. Increased values of urinary-free cortisol are a late finding usually associated with emerging clinical signs of Cushing's syndrome. Thus, determination of urinary-free cortisol is not an appropriate screening test for SCCS.

In patients with hypertension, serum potassium and a plasma aldosterone/plasma renin ratio should be determined to evaluate primary aldosteronism. A normal potassium concentration does not exclude primary hyperaldosteronism given that the majority of patients are normokaliemic under random conditions. A plasma aldosterone/plasma renin ratio greater than 30 (ng/dl to ng/ml/h) and a plasma aldosterone concentration (PAC) of greater than 20 ng/dl are highly suggestive of autonomous aldosterone production. The diagnosis of primary hyperaldosteronism has to be confirmed by the failure of fludrocortison to suppress aldosterone (0.1 mg fludrocortison every 6 h for 4 days, PAC on day 5 greater than 5 ng/dl, sampling midmorning after 2 h upright position) or following acute saline suppression (2 L of 0.9% NaCl solution infused i.v. over a period of 4 h; failure to suppress PAC  $<6 \text{ ng/dl}$ ). Patients with an elevated PAC and plasma renin activity (PRA)/renin ratio may require further evaluation in a specialized center. Investigations may even include bilateral adrenal catheterization to prove that the incidentaloma is the source of the mineralocorticoid excess.

Pheochromocytoma among adrenal incidentalomas is found in approximately 5% of patients. Most patients do not show the typical clinical symptoms such as tachycardia, sweating, and headache, and up to 40% are completely asymptomatic. Because of the sequelae of undiagnosed pheochromocytoma, every patient should be evaluated by determination of 24-h urinary catecholamines that have a sensitivity and specificity of more than 90%. Measurement of plasma catecholamines is inferior, and a suppression test (clonidine test) is rarely required. Whether free plasma metanephrines are superior under random conditions has not yet been determined. Patients with elevated catecholamine excretion should undergo  $^{131}\text{I}$ -metaiodobenzylguanidine (MIBG) scintigraphy for preoperative detection of multiple pheochromocytoma or metastasis.

Routine evaluation for androgen and estrogen production in patients with adrenal incidentalomas is not recommended. Especially determination of DHEAS has not been shown to be a very sensitive marker of ACC. In a recently published Italian multiple-center trial, DHEAS values were elevated in 17% of all patients with cortical carcinoma (28% in younger individuals, i.e., under 50 years of age). The sensitivity and specificity were 17% and 93%, respectively; negative and positive predictive values were 95% and 10%, respectively.

### Adrenal Imaging

The size and appearance of an adrenal mass on CT or MRI helps to distinguish between benign and malignant lesions. A homogenous mass with a smooth border and an attenuation value of less than 10 Hounsfield units (HU) on an unenhanced CT study strongly suggests the diagnosis of a benign adrenal adenoma. Unfortunately, this information is often not available. Adrenal carcinomas are generally larger than 6 cm, are nonhomogenous, and show soft tissue density. Irregular margins and central necrosis or hemorrhage increase the probability of malignancy. However, benign pheochromocytoma may also present as a large nonhomogenous tumor with hemorrhage. Whether MRI is superior to CT in diagnosing and differentiating among adrenal masses remains uncertain. The recently developed chemical shift MRI (CSI) identifies benign adrenal adenomas with high lipid content by a typical signal intensity loss on chemical shift imaging relative to the liver. Recently, a sensitivity of 91% and a specificity of 94% were reported using CSI for differentiation between adenoma and metastasis, but its accuracy in ACC is uncertain.

### Fine Needle Aspiration

In patients with adrenal incidentaloma who have no history of malignancy, fine needle aspiration (FNA) has no proven efficacy. FNA is not free of side effects and may lead to pneumothorax, frank retroperitoneal bleeding, or needle track metastasis in the case of ACC. The additional improvement of the new imaging techniques (CT and MRI) in characterization of adrenal masses should restrict FNA to few indications. In patients with a history of malignancy, FNA is a valuable tool (sensitivity and specificity of approximately 90%) and should be performed if the presence of adrenal metastasis may alter the therapy or prognosis. Pheochromocytoma should always be ruled out

prior to FNA because it may cause hypertensive crisis or even death.

## THERAPEUTIC CONSIDERATIONS

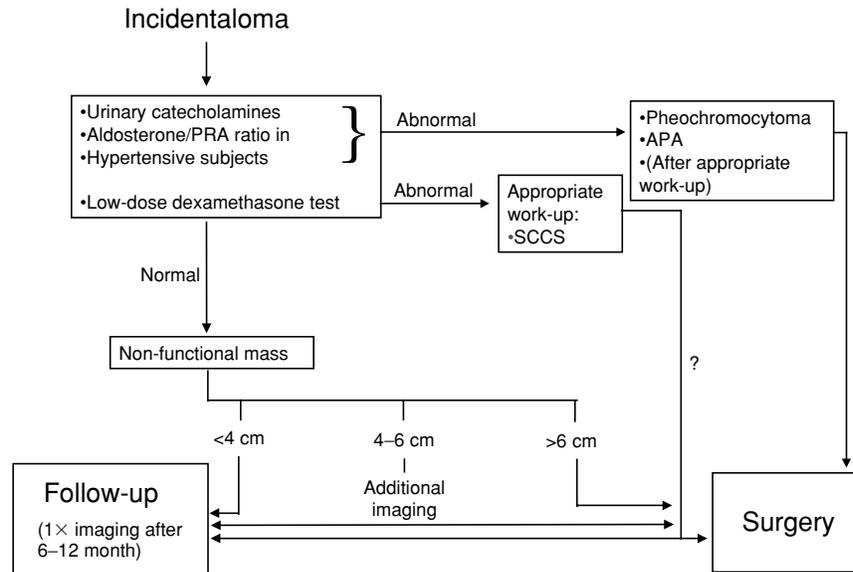
Adrenal surgery has significant morbidity and mortality that must be taken into account during the decision-making process. The newly developed endoscopic adrenalectomy, with its low morbidity (5–10%) and mortality (<1%), may justify earlier surgical intervention. However, it has to be emphasized that in the published series, this operation was performed mostly in specialized centers with experienced laparoscopic surgeons. Surgical outcome may be significantly less favorable in centers without such specialization.

### Treatment of the Functional Incidentally Detected Adrenal Mass

To prevent serious morbidity, hormonally active incidentalomas should be surgically removed. This strategy is undisputed for aldosterone-producing adenoma and pheochromocytoma. However, it remains doubtful whether all patients with SCCS benefit from adrenal surgery. There are data indicating that some patients with SCCS may develop metabolic derangement, including hypertension, weight gain, and insulin resistance, that could be attributable to autonomous cortisol secretion. Rarely, it may progress from subclinical disease to overt Cushing's syndrome. Surgery should be considered in patients with SCCS who have suppressed plasma ACTH and elevated urinary-free cortisol because progression to overt Cushing's syndrome is imminent. Adrenalectomy is also recommended in younger individuals (<50 years of age) and in patients who have metabolic disease of recent onset possibly related to the incidentaloma (e.g., obesity, hypertension, diabetes).

### Treatment of the Nonfunctional Incidentally Detected Adrenal Mass

In patients with nonfunctioning incidentalomas, distinguishing between malignant and benign primary adrenal tumors dictates subsequent management. Variables to consider, according to the National Institutes of Health consensus conference (Fig. 2), are the size of the lesion, its radiographic characteristics, and its growth rate. More than 60% of incidentalomas less than 4 cm are benign adenomas, whereas less than 2% represent ACC. In contrast, the risk of adrenal carcinoma increases to 25% in lesions that are



**Figure 2** Diagnostic and therapeutic considerations for the treatment of patients with adrenal incidentalomas following the recommendations of the National Institutes of Health in its 2002 state-of-the-science statement.

6 cm or more in diameter, whereas benign adenomas account for less than 15%. Therefore, the generally accepted recommendation is to excise lesions that are more than 6 cm. Lesions that are less than 4 cm and appear to be defined as low risk by radiographic criteria are unlikely to have malignant potential and are generally not resected. For lesions between 4 and 6 cm, either close follow-up or adrenalectomy is considered a reasonable approach.

Data from several small series of patients indicate that approximately 85% of lesions remain stable, 5 to 30% of incidentalomas increase in size, and 3 to 4% decrease in size. In these series, the risk of the lesion being an ACC was extremely low. Based on these data, a repeat CT scan between 6 and 12 months after the initial study is reasonable. For data that do not increase in size, continued radiological evaluation is not recommended. Hormone excess may develop in less than 20% of cases when followed for up to 10 years, but this is unlikely in patients with lesions of less than 3 cm. Current evidence suggests repeat dexamethasone suppression and urinary catecholamine excretion testing for up to 3 to 4 years.

### See Also the Following Articles

Adrenal Tumors, Molecular Pathogenesis • Pheochromocytoma

### Further Reading

- Barzon, L., *et al.* (1999). Risk factors and long-term follow-up of adrenal incidentalomas. *J. Clin. Endocrinol. Metab.* **84**, 520–526.
- Boland, G.W., *et al.* (1998). Characterization of adrenal masses using unenhanced CT: An analysis of the CT literature. *Am. J. Roentgenol.* **171**, 201–204.
- Chidiac, R. M., and Aron, D. C. (1997). Incidentalomas: A disease of modern technology. *Endocrinol. Metab. Clin. North America* **26**, 233–253.
- Gordon, R. D. S. A., Stowasser, M., Tunny, T. J., Storie, W. J., and Rutherford, J. C. (1993). How common is primary aldosteronism? Is it the most frequent cause of curable hypertension? *J. Hypertension* **11**(Suppl. 5), S310–S311.
- Kasperlik-Zeluska, A. A., *et al.* (1997). Incidentally discovered adrenal mass (incidentaloma): Investigation and management of 208 patients. *Clin. Endocrinol. (Oxf.)* **46**, 29–37.
- Kloos, R. T., *et al.* (1995). Incidentally discovered adrenal masses. *Endocrinol. Rev.* **16**, 460–484.
- Latronico, A. C., and Chrousos, G. P. (1997). Extensive personal experience: Adrenocortical tumors. *J. Clin. Endocrinol. Metab.* **82**, 1317–1324.
- Mantero, F., *et al.* (2000). A survey on adrenal incidentaloma in Italy: Study Group on Adrenal Tumors of the Italian Society of Endocrinology. *J. Clin. Endocrinol. Metab.* **85**, 637–644.
- National Institutes of Health. (2002). Management of the clinically inapparent adrenal mass (incidentaloma) [state-of-the-science statement]. [http://consensus.nih.gov/ta/021/021\\_statement.htm](http://consensus.nih.gov/ta/021/021_statement.htm)
- Reincke, M. (2000). Subclinical Cushing's syndrome. *Endocrinol. Metab. Clin. North America* **29**, 43–56.
- Ross, N. S., and Aron, D. C. (1990). Hormonal evaluation of the patient with an incidentally discovered adrenal mass. *N. Engl. J. Med.* **323**, 1401–1405.

Cycles strongly suggest that a woman is ovulating. Cycle lengths of roughly 22 to 35 days are usually ovulatory, especially if accompanied by premenstrual symptoms such as bloating and breast tenderness. If a woman has menses more or less often, she probably is not ovulating or is ovulating infrequently. Amenorrhea (no menses) is suggestive of ovulatory dysfunction except in patients with uterine disease such as intrauterine scarring (Asherman's syndrome) or serious congenital uterine abnormality. Ovulation may be confirmed by measuring a serum progesterone level during the luteal phase of the cycle (about 1 week before a woman's next menses is due) or by charting the basal body temperature, by ultrasound of the ovaries, and/or by detection of luteinizing hormone (LH) surge, often using urinary LH detection kits. Absolute proof of ovulation is made by the occurrence of pregnancy.

Polycystic ovarian syndrome (PCOS) is the etiology of anovulation or oligo-ovulation in approximately 70% of cases. Other causes include hypothalamic dysfunction, hyperprolactinemia, age-related ovulation dysfunction, premature ovarian failure, and thyroid disease. Congenital adrenal hyperplasia (CAH) is a relatively rare cause of oligo-ovulation. Extremes of weight, such as that associated with anorexia or with obesity, may also lead to anovulation by producing hypothalamic dysfunction.

It is important to determine the cause of oligo-ovulation before embarking on a treatment plan. A diagnosis can generally be determined by clinical history and by measurement of serum thyroid-stimulating hormone (TSH), prolactin, follicle-stimulating hormone (FSH), and estradiol ( $E_2$ ). Some clinicians find it useful also to measure serum testosterone, fasting blood sugar, and fasting insulin levels to confirm or refute a clinical suspicion of PCOS. However, the value of this additional testing is controversial. Measurement of the follicular phase 17-hydroxyprogesterone level will detect most cases of CAH.

### Tubal Disease

Open and functional fallopian tubes are necessary for conception. Tubal and/or adhesive factors account for about 35% of all infertility cases. Infertility may result from complete blockage of the distal end of the fallopian tube (hydrosalpinx) as a sequelae of sexually transmitted disease, surgical intervention for management of ectopic pregnancy or other intra-abdominal conditions, nongynecological abdominal infection (uncommon), endometriosis

(uncommon), or a congenital anomaly (rare). Proximal obstruction may result from salpingitis isthmica nodosa or other inflammatory conditions, or it may be idiopathic. Peritubal adhesions or damage to the lining of the tube can impair tubal mobility, oocyte pickup, and/or sperm and embryo transport. Iatrogenic infertility occurs secondary to tubal sterilization.

The hysterosalpingogram (HSG) is typically used as the initial screen for the presence of tubal disease. In this fluoroscopic procedure, media placed through the cervix and into the uterine cavity can be seen passing through the fallopian tubes if they are patent. The presence or absence of tubal disease is most definitively confirmed by laparoscopy.

### Endometriosis and Peritoneal Adhesions

Endometriosis is found more commonly in women with infertility than in women with normal fertility. Women with endometriosis may have an earlier onset of diminished ovarian reserve, possibly associated with ovarian endometriosis cysts (endometriomas). Immunological and genetic factors are also likely to be important in the pathogenesis and pathophysiology of endometriosis. Although there has been much debate, it is likely that even minimal endometriosis can reduce a woman's fertility. In cases of extensive endometriosis, tubal function may be compromised by tubal adhesions. Peritoneal adhesions may also interfere with tubal function. Endometriosis and peritoneal adhesions can be definitively diagnosed only by laparoscopy. Diagnostic laparoscopy should be performed only by surgeons who are generally capable of treating the disease they find at the initial diagnostic laparoscopy, a procedure called operative laparoscopy.

### Uterine and Cervical Abnormalities

Abnormalities may be present in the cavity of the uterus where the embryo needs to implant and develop. These abnormalities include intrauterine adhesions, polyps, fibroids, and an abnormally shaped uterine cavity. Problems within the uterus can potentially interfere with implantation of the early embryo or increase the incidence of miscarriage.

The HSG is often used as an initial screen to determine whether the cavity of the uterus is anatomically normal. The saline hysteroogram (SHG) is a pelvic ultrasound done while injecting saline (salt water) through the cervix to outline the uterine cavity. Unlike the HSG, it allows visualization of the wall of the uterus as well as the cavity at the same time, a

difference that may be helpful in assessing the position of fibroids and in determining whether or not they are resectable through the hysteroscope. As a screen for abnormalities in the uterus, the SHG is a more sensitive and more specific screening test than is the HSG, using findings at hysteroscopy as the gold standard.

An inadequate progesterone effect on the uterine lining is called a luteal phase deficiency. Luteal phase defect may be due to either deficient progesterone production or inability of the uterine lining to respond to estrogen during the proliferative phase and/or to progesterone during the luteal phase. It is quite debatable how often a luteal phase defect can explain either infertility or recurrent miscarriage—if luteal phase defect explains these problems at all. Many infertility specialists no longer routinely do an endometrial biopsy because the information has not usually been helpful in improving live birth rates. Problems of implantation (e.g., endometrial receptor defects) may theoretically occur secondary to dysfunctional endometrium. Currently, however, there are no adequate tests to diagnose such theoretical conditions or specific treatments for such conditions.

Conditions within the cervix can contribute to infertility, but they are rarely the sole cause. Risk factors for cervical abnormalities may include cone biopsy and diethylstilbestrol (DES) exposure. Chronic infections may, in some cases, contribute to poor cervical mucus. Abnormalities of the cervical mucus may be diagnosed by the postcoital test (examination of the cervical mucus for the presence of motile sperm after intercourse). Many physicians do not recommend this test because its value in guiding infertility treatment is not clear. Its only role may be in diagnosing inadequate coital technique or frequency that can be based on sexual dysfunction, ignorance, and/or cultural/social/religious issues that can interfere with coitus and impair infertility.

### Abnormalities of Sperm

A male factor is the sole cause for infertility in about 20% of couples and contributes to infertility in approximately 30 to 40% of other couples. Although a cause for decreased sperm quality may be determined in some cases, male subfertility is frequently idiopathic.

Azoospermia (absence of sperm in the ejaculate) may be classified as obstructive (e.g., congenital absence of the vas deferens, ductal obstruction, vasectomy) or nonobstructive. Most cases of nonobstructive azoospermia are due to primary testicular failure. Endocrine abnormalities such as hyperprolactinemia

that secondarily affect spermatogenesis are less common. Genetic disorders associated with azoospermia or severe oligospermia include Klinefelter syndrome (47,XXY) and microdeletions of the Y chromosome. Congenital absence of the vas deferens is associated with cystic fibrosis mutations.

Oligospermia may result from hormonal problems, retrograde ejaculation, varicocele, or a primary testicular problem with sperm production. Genitourinary infection may interfere with sperm function. Environmental factors such as medications, drugs, alcohol, tobacco abuse, and radiation may impair sperm production and function. Deficiency of trace elements such as zinc, folate, and selenium may impair sperm function. Antisperm antibodies may impair fertility in couples with unexplained infertility. In many cases, no treatable cause of poor sperm quality can be found.

A recent review of the effects of male age on semen quality and fertility concluded that increasing age is associated with a decline in semen volume, sperm motility, and sperm morphology but not with sperm concentration. There is most likely some decline in male fertility with age, particularly over 50 years of age, but results are confounded by female partner age. There is no age at which men cannot father a pregnancy. Thus, fertility is much more related to the age of the female partner.

The semen analysis is used as the basic screen for abnormalities of sperm. A semen analysis typically includes assessment of volume, concentration, motility, and morphology of the sperm. Other tests of sperm function may be used in some special circumstances.

### Advancing Female Age

An age-related decline in female fertility begins many years prior to the onset of menopause despite continued regular ovulatory cycles. “Ovarian reserve” is a term often used to describe a woman’s reproductive potential with respect to ovarian follicle number and oocyte quality. The drop in fertility associated with diminished ovarian reserve is due to a depletion of eggs and to a decline in average egg quality. Although there is no strict definition of what may be considered advanced reproductive age, a decline in fertility on average begins for women in their late 20s or early 30s, becoming more pronounced after 35 and very pronounced after 40 years of age. However, the decline in ovarian reserve may occur much earlier for some women. Risk factors for early loss of ovarian reserve include smoking, family history of premature ovarian failure, significant ovarian pathology due to

severe endometriosis and/or pelvic adhesive disease, and previous ovarian surgery.

Once an older woman becomes pregnant, she has a markedly increased risk of spontaneous abortion, with the clinical spontaneous abortion rate being approximately 15% for women under 35 and 40% or higher for women over 40 years of age. The age-associated decline in female fertility and the increased risk of spontaneous abortion are most likely attributable to abnormalities in the oocyte that lead to aneuploid embryos (those with the wrong number of chromosomes). There is good evidence that aneuploid embryos will either arrest very early in development (before a woman even knows she is pregnant) or will miscarry, most commonly during the first trimester.

Because of the marked effect of age on success rates from infertility treatment, it is advisable for older women to begin evaluation sooner than do younger women. It is reasonable to pursue an infertility evaluation after only 6 months of attempting pregnancy because the consequences of undiagnosed infertility factors can be even more detrimental to women who have a limited time frame for achieving a successful pregnancy.

It is advisable to assess ovarian reserve in all women 35 years of age or over and in younger women with risk factors for early onset of diminished ovarian reserve or unexplained infertility. An assessment of ovarian reserve typically includes measurement of FSH and E<sub>2</sub> levels on the third day of a woman's menstrual cycle. A more sensitive marker of ovarian reserve is the clomiphene citrate challenge test (measurement of FSH the day after a 5-day course of clomiphene along with the day 3 measurements) and a count of small antral follicles in the early follicular phase.

### Unexplained Infertility

In approximately 5 to 10% of couples seeking pregnancy, all of the preceding tests are normal, and in a much higher percentage of couples, only minor abnormalities are found. These couples are often said to have unexplained infertility. Couples with unexplained infertility may have problems with egg quality that have not been detected with any of the available tests. In other cases, couples with unexplained infertility may have a problem with the ability of the sperm to fertilize the egg, undiagnosed tubal dysfunction, or implantation failure. Ongoing studies are beginning to try to elucidate a genetic basis for unexplained infertility.

## INFERTILITY TREATMENTS

### Objectives for Managing Infertility

Cost-effective, evidence-based infertility management depends on an organized plan of diagnosis and treatment from the time of initial presentation. This includes the use of simpler, cheaper tests first and the establishment of a correct diagnosis, followed by simultaneous treatment of the man and woman using effective therapy. Whenever possible, treatment choices should be based on the available evidence regarding their efficacy and risks as well as on the wishes of the individual couple.

### Ovulation Induction

The development of a single mature follicle is the goal of anovulatory women. If the cause of infertility is not lack of ovulation, the goal of ovulation induction is the development of multiple follicles (controlled ovarian hyperstimulation). Clomiphene is a selective estrogen receptor modulator (sometimes called an antiestrogen) that is administered orally for 5 days to induce ovulation. In anovulatory women with polycystic ovarian syndrome, clomiphene is expected to induce ovulation in approximately 80% of oligoovulatory or anovulatory women. The monthly pregnancy rate depends on the presence or absence of other infertility factors and may range from less than 5 to 20%. Twins occur in 10% or fewer of pregnancies, and triplets occur with a frequency of less than 1%. In women with PCOS who fail to ovulate on clomiphene alone, metformin plus clomiphene may achieve ovulation, particularly if women are obese or have evidence of insulin resistance.

Gonadotropin therapy involves daily injections of FSH and sometimes LH. This treatment will lead to ovulation in most women with polycystic ovarian syndrome and in patients with hypothalamic amenorrhea. Multiple pregnancy occurs with approximately 25% twins, 5% triplets, and 2% quadruplets. Severe ovarian hyperstimulation syndrome occurs in approximately 1% of patients, with moderate ovarian hyperstimulation occurring in 5 to 7% of patients. Gonadotropin treatment requires intensive ultrasound monitoring, and its cost is much greater than that of clomiphene. Therefore, it is typically used if treatment with clomiphene is unsuccessful.

For women with polycystic ovarian syndrome, ongoing pregnancy rates with laparoscopic ovarian drilling (a procedure that uses an electro-surgical needle or another energy source to destroy

androgen-producing ovarian stroma) after 6 to 12 months are similar to three to six cycles of ovulation induction with gonadotropins. Multiple pregnancy rates are lower with ovarian drilling. However, ovarian drilling requires a surgical procedure and carries a risk of adhesive disease. Furthermore, its impact on long-term ovarian reserve has not been studied adequately.

For women with hyperprolactinemia, the dopamine agonist bromocriptine leads to ovulation in approximately 80% of patients. Cabergoline, a long-acting dopamine agonist, is at least as effective as bromocriptine and appears to be better tolerated. Treatment of thyroid disease may lead to ovulation if abnormalities of thyroid function are detected on initial evaluation. Adrenal gland dysfunction is rare, but treatment may improve ovarian function.

## Surgical Treatments

### Tubal Surgery

Blocked fallopian tubes may be surgically opened and peritubal adhesions may be removed. The most important factor affecting success of tubal surgery may be the degree of damage to the lining of the tubes before surgery and the postoperative reformation of adhesions. Severe underlying intratubal pathology, such as intratubal adhesions, is not surgically correctable. Although some tubal problems are correctable by surgery, women with severely damaged tubes are so unlikely to become pregnant that *in vitro* fertilization (IVF) offers the best hope for a successful pregnancy (Table I). Very badly damaged obstructed tubes may

**Table I** Pregnancy Rates Following Treatment for Tubal Factor Infertility (Percentages)

Salpingolysis/Ovariolysis	25–62
Fimbrioplasty	60–70
Fimbrioplasty (postinfectious)	27
Salpingostomy (reconstruction)	21–39
Tubal anastomosis	52–82
Tubal anastomosis (uterotubal)	50–69
Tubal cannulation	25–35
Salpingostomy (for ectopic)	38–80
Methotrexate (for ectopic)	50–55
Fulguration of endometriosis	40–75
Repeat tuboplasties	6–20
IVF (per cycle)	20–40

Source. Adapted from Silverberg, K. M., and Hill, G. A. (1991). *J. Gynecol. Surg.* 7, 67–82.

**Table II** Pregnancy Rates Following Treatment of Endometriosis-Associated Infertility (Monthly Fecundity Percentages)

Treatment	Stage		
	Minimal/Mild	Moderate	Severe
Expectant	3	3	0
Ovarian suppression <sup>a</sup>	3	4	2
Surgical	5	5	3

Source. Adapted from Adamson, D. (2001). *Int. J. Fertil. Women's Med.* 46, 151–168.

<sup>a</sup>After discontinuation of ovarian suppression medications.

fill with fluid, creating hydrosalpinges and resulting in live birth rates with IVF only half those for women without hydrosalpinges. For this reason, removal of hydrosalpinges should be considered before an IVF cycle. This operation, a salpingectomy, can almost always be performed laparoscopically. Tubal reanastomosis (typically done to reverse a tubal ligation) is generally a surgical procedure associated with a much better outcome than is repair of tubes damaged by infection or severe endometriosis.

### Surgery to Diagnose and Treat Endometriosis

In some cases, the decision may be made to proceed with surgery to diagnose and treat endometriosis. It appears that laparoscopic treatment in skilled hands is as effective as laparotomy. For minimal and mild endometriosis, laparoscopic surgery probably improves live birth rates, although prospective data are limited and not all studies confirm this finding. The short-term increase in monthly fecundity with surgical treatment of minimal or mild endometriosis is probably lower than the monthly fecundity expected with treatments for unexplained infertility such as controlled ovarian hyperstimulation and intrauterine insemination (COH/IUI) and assisted reproductive technology. Therefore, sometimes these treatments are used empirically without laparoscopy, even if it is possible that endometriosis may be present. Although ovarian suppression is effective for management of pain associated with endometriosis, ovarian suppression is not indicated as a treatment for infertility, with the possible exception of ovarian suppression for 2 months in patients with severe endometriosis before IVF. The live birth rate may be increased if moderate or severe endometriosis is diagnosed and treated surgically (Table II). However, sometimes the long-term prognosis for these patients is poor with treatments other than IVF.

### Surgery to Correct Uterine Abnormalities

There are no good trials to actually determine whether or not myomectomy is beneficial with respect to infertility. Myomectomy may improve the success of IVF if the myoma is distorting the uterine cavity. Some studies support, and many clinicians perform, myomectomies for documented uterine cavity distortion whether or not IVF is being considered. A potential complication of myomectomy is postoperative adhesions or subsequent need for cesarean section. Polypectomy may reduce the rate of miscarriage in couples undergoing IVF. Most uterine anomalies do not require surgical treatment of the uterus. A septate uterus should generally be treated hysteroscopically at the time of diagnosis, especially in a couple with recurrent miscarriage. Adenomyosis and DES exposure may affect the live birth rate of various infertility treatments but are not surgically correctable. Severe adenomyosis may contribute to infertility, but diagnosis can be difficult without a pathological specimen, and no medical or surgical treatments have been confirmed to improve pregnancy rates.

### Treatments for Male Factor Infertility

Male factor infertility may be amenable to treatment of a specific cause if one is identified. For example, if a varicocele is palpable, surgery may be considered as a treatment option when semen analyses are abnormal. When a specific cause cannot be identified, male factor infertility is often treated by IUI or IVF. IVF is the treatment of choice for couples with severe male factor infertility. During an IVF cycle, it is possible to perform direct injection of a single sperm into each oocyte, a procedure known as intracytoplasmic sperm injection (ICSI). With this procedure, it has become possible to treat very severe cases of male factor infertility. Donor sperm can be used for women attempting conception who do not have a male partner and in cases of male factor infertility. Donor insemination is used much less frequently for male factor infertility since ICSI has become available.

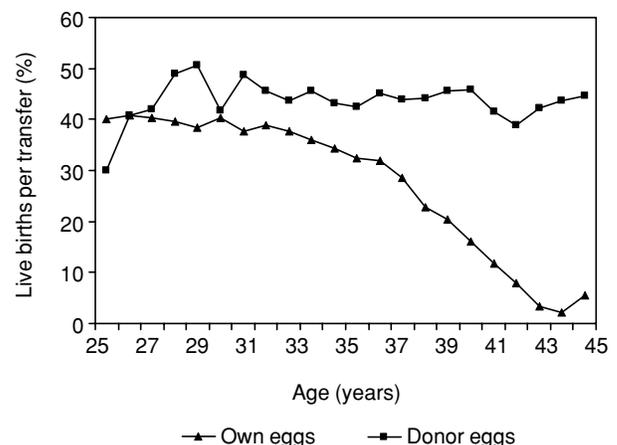
### Treatments for Age-Related Infertility

COH/IUI has limited efficacy for older women, yielding a delivery rate per cycle of 5% or less (range = 1.4–5.2%) for women over 40 years of age with infertility that is unexplained except by age. This compares with live birth rates per cycle of 17 to 22% for women under 35 and of 8 to 10% for women 35 to 40 years of age. Advancing age is associated with

reduced ovarian responsiveness to gonadotropins and poorer egg quality.

The presence of infertility factors such as male factor, tubal disease, endometriosis, and pelvic adhesions would argue for proceeding directly to IVF in women of advanced reproductive age. Pregnancy rates from IVF are higher—usually much higher—per cycle than with other treatments, including COH/IUI, surgery, and treatment of male factor infertility, but do decline significantly with age. According to the 2000 Society for Assisted Reproductive Technology (SART) report, live birth rates per cycle were 33% in women under 35, 27% in women 35 to 37, 18% in women 38 to 40, 10% in women 41 to 42, and 4% in women over 42 years of age. This age-related decline in IVF success is related to a decline in ovarian reserve.

Oocyte donation yields the highest live birth rate of any assisted reproductive technology (ART) treatment and is the treatment of choice for age-related infertility that is not addressed successfully by other treatments. Oocyte donation involves the retrieval of eggs from a woman with good ovarian reserve (typically a woman in her 20s or early 30s), fertilization of these eggs in the laboratory, and subsequent transfer of embryos to the woman who will carry the child. Pregnancy rates with oocyte donation are dependent on the age of the donor, not on the age of the recipient (Fig. 1). Pregnancy rates with donor oocytes are much higher than those with the usual infertile population due to the improved egg quality.



**Figure 1** Live births per transfer for fresh embryos from patients undergoing *in vitro* fertilization with their own eggs or from donor eggs. Data are displayed by age of the recipient of the eggs. From the U.S. Department of Health and Human Services and Centers for Disease Control and Prevention. (2002). 2000 assisted reproductive technology success rates [national summary and fertility clinic reports].

## Treatments for Unexplained Infertility

Proposed treatments for unexplained infertility include IUI, COH/IUI, and ART (Table III). Cycle fecundity with COH/IUI is expected to be lower when the diagnosis is unexplained infertility than when the diagnosis is anovulation. For anovulatory women, the treatment is correcting an identified problem, whereas in couples with unexplained infertility, it is quite possible that oocyte quality or sperm dysfunction may be present. Cycle fecundity is also expected to be lower if significant male factor is present than if the semen analysis is normal. Female age is one of the most important predictors of success.

## Outcomes of Infertility Treatment and ART

Patients undergoing infertility treatment have a medical disease, infertility, with dysfunction of the hypothalamus, pituitary, ovaries, uterus, pelvis, male reproductive system, or other systems. They also are older than the population of women who can conceive without difficulty. Despite these important differences, babies resulting from infertility treatment generally are as healthy, or nearly as healthy, as the normal population. Any significant differences in outcomes result primarily from the increased rate of multiple pregnancy and associated prematurity. Therefore, reducing the multiple birth rate is an important objective of infertility specialists. For most infertile patients, the chances of having a healthy baby when they conceive following infertility treatment are not substantially different from those when they conceive spontaneously.

Table III Pregnancy Rates Following Treatment for Unexplained Infertility (Percentages)

Treatment	Monthly Fecundity
No treatment	3
IUI	4
Clomiphene	6
Clomiphene plus IUI	7
Gonadotropin	8
Gonadotropin plus IUI	18
IVF <sup>a</sup>	23
Gamete intrafallopian transfer <sup>a</sup>	26

Source. Adapted from Guzik *et al.* (1998). *Fertil. Steril.* 70, 207–213.

<sup>a</sup>These monthly fecundities reflect outcomes from prior to 1995. IVF success rates have improved since this time. See text.

Since the first IVF baby was born in 1978, more than 1 million babies have been born via IVF all over the world. Large national and international registries and numerous studies have concluded that IVF results in generally healthy babies. However, recently some studies have found a general increased risk of birth defects, prematurity, lower birthweight, and selected genetic birth defects in IVF babies. All of the studies and data can be faulted for shortcomings in their methodology, but a legitimate concern has been raised that will require better, larger, and longer follow-up studies to confirm that IVF babies have no clinically substantial increased risk of untoward outcomes. Patients should be counseled about these concerns, but data and studies supporting the usually positive outcomes for IVF babies are reassuring for most patients.

## Decision Making in Infertility

Many factors must go into the decision regarding which infertility treatment is best for a given patient. These factors include the value that the patient and her partner place on various outcomes, for example, their own genetic baby, an IVF baby, a donor egg baby, a baby from a gestational carrier, adoption, and the decision to be “child free.” The financial and time cost, as well as the physical and emotional side effects and risks of treatment, must be weighed. Once this is done, the prognosis for each type of treatment and the risks of adverse outcomes need to be balanced. COH/IUI is often successful but carries the major risk of twins in approximately 7 to 10% of pregnancies with clomiphene and 20 to 30% of those with gonadotropins. Triplets occur in less than 1% of pregnancies with clomiphene but occur in up to 5 to 10% of those with gonadotropins. Surgery does not carry an increased multiple pregnancy risk but does carry surgical risk and may result in an increased risk of ectopic pregnancy in 1 to 30% of surgeries, depending on the degree of tubal damage. Furthermore, the monthly fecundity following tubal surgery is generally lower than that with other treatments, so it takes longer on average to conceive. IVF is much more successful than other treatments but is more expensive and carries a risk of approximately 30% for twins and 4% for triplet births for each pregnancy. The risk of multiple pregnancy with IVF can be reduced substantially by limiting the number of embryos that are replaced. Adherence to current guidelines regarding the number of embryos to replace is essential if the number of multiple pregnancies is to be reduced. Additional complications of premature babies and

pregnancy in women of advanced age must be considered, including the increased risk of birth defects and their management. Careful informed discussion with each patient and partner is necessary to review all of the potential options, prioritize them, and develop a treatment plan once the diagnosis is established.

## SUMMARY AND CONCLUSION

Common causes of subfertility include ovulatory disorders, tubal disease, adhesions, endometriosis, uterine abnormalities, sperm dysfunction, and advancing female age. Infertility is unexplained after thorough evaluation in about 5 to 10% of cases, and only minor abnormalities are found in many other couples. The provision of cost-effective infertility treatment that is evidence based is a challenge given the limited number of well-designed trials of infertility treatments, rapid technological advances, and variance in couples' objectives, values, and ethical beliefs. Successful ovulation induction in anovulatory women is possible for nearly all women except in cases of ovarian failure. The effectiveness of tubal surgery is dependent on the degree of damage to the lining of the tube and the extent of pelvic adhesions. Surgical treatment of endometriosis can improve the live birth rate for some infertile couples. Male factor infertility may be amenable to treatment of a specific cause but is often empirical with use of IUI or IVF. Egg donation is currently the most effective treatment available for age-related infertility, ovarian failure, and when other treatments have not been successful. Couples with unexplained infertility may

be effectively treated with ovulation induction plus IUI or IVF.

## See Also the Following Articles

Assisted Reproductive Technology (ART) • Endometriosis • Erectile Dysfunction • Fertility in Men with Spermatogenesis Abnormalities • Fertilization • Gonadotropin-Induced Ovulation • Implantation • In Vitro Fertilization (IVF) • Ovarian Failure Treatment Strategies: Egg Donation • Polycystic Ovary Syndrome (PCOS) • Pregnancy Endocrinology • Premature Ovarian Failure • Superovulation and Intrauterine Insemination

## Further Reading

- Adamson, G. D., and Pasta, D. J. (1994). Surgical treatment of endometriosis-associated infertility: Meta-analysis compared with survival analysis. *Am. J. Obstet. Gynecol.* **171**, 1488–1505.
- American Society for Reproductive Medicine. (2002). Aging and infertility in women: ASRM practice guidelines. [www.asrm.org/media/practice/ageandinfertility.pdf](http://www.asrm.org/media/practice/ageandinfertility.pdf).
- Corfman, R. S., and Badran, O. (2000). Effect of pelvic adhesions on pelvic pain and fertility. *Infertil. Reprod. Med. Clinics North America* **5**, 405–411.
- Grimbizis, G. F., Camus, M., Tarlatzis, B. C., *et al.* (2001). Clinical implications of uterine malformations and hysteroscopic treatment results. *Hum. Reprod. Rev.* **7**, 161–174.
- Guzick, D. S., Sullivan, M. W., Adamson, G. D., *et al.* (1998). Efficacy of treatment for unexplained infertility. *Fertil. Steril.* **70**, 207–213.
- U.S. Department of Health and Human Services and Centers for Disease Control and Prevention. (2002). 2000 assisted reproductive technology success rates. [national summary and fertility clinic reports]. [www.cdc.gov/nccdphp/dhrh/art.htm](http://www.cdc.gov/nccdphp/dhrh/art.htm).
- Witz, C. A., and Schenken, R. S. (1997). Pathogenesis. In "Seminars in Reproductive Endocrinology: Endometriosis" (L. Speroff and G. D. Adamson eds.) Vol. 15, No. 3, pp. 199–208. Thieme, New York.

resistance of gluconeogenesis. Thus, hepatic insulin resistance in type 2 diabetes occurs secondary to chronic hyperinsulinemia.

## Adipose Tissue

White adipose tissue accounts for less than 10% of glucose uptake but plays a major role in whole body insulin sensitivity by releasing into the circulation various modulators of insulin action, including lipid products and hormones. Free fatty acids (FFAs) derived from adipocytes are elevated in obesity and type 2 diabetes and accumulate in muscle and liver, contributing to insulin resistance in these tissues. Mice with a selective deletion of the insulin receptor in brown adipocytes experience an age-dependent loss of intrascapular brown fat but increased expression of uncoupling proteins-1 and -2. In parallel, these mice develop a defect in insulin secretion, resulting in progressive glucose intolerance, without insulin resistance. These data demonstrate that the insulin receptor is required for brown fat formation throughout development and that brown fat regulates insulin secretion and glucose homeostasis.

Histological, immunohistochemical, and ultrastructural analyses of white adipose tissue in newborn fat insulin receptor knockout mice (FIRKO mice) reveal a marked decrease in adipose tissue area. This fat cell depletion resulted from a reduction of adipocyte volume (~90%), with a small decrease in the number of adipocytes. Electron microscopy analysis displays a normal pattern of adipogenesis in FIRKO mice, demonstrating that lack of insulin receptors is not associated with a selective impairment of the adipocyte differentiation process. Thus, the insulin receptor is dispensable for adipocyte differentiation but is necessary for the large extent of lipid storage in these cells. In addition, FIRKO mice display increased longevity, demonstrating an inverse relationship between high rates of adipose metabolism and life span.

Adipocytes also secrete cytokines (adipokines) that modulate peripheral insulin sensitivity. Tumor necrosis factor- $\alpha$  (TNF- $\alpha$ ) is elevated in obese humans and rodents and has been shown to reduce insulin signaling by decreasing insulin receptor and IRS1 tyrosine phosphorylation. Mice harboring deletion of the genes encoding TNF- $\alpha$  or the TNF- $\alpha$  receptor are refractory to insulin resistance associated with diet-induced obesity or when crossed with genetically obese ob/ob mice. Resistin is another adipokine secreted from adipocytes, and the levels of this hormone are increased in obese mice and reduced after administration of the insulin-sensitizing peroxisome proliferator-activated

receptor- $\gamma$  (PPAR- $\gamma$ )-activating thiazolidinediones. Resistin also inhibits adipogenesis and, thus, plays a negative feedback role in fat cell formation.

Adipose tissue also secretes leptin, another hormone that positively influences metabolism and energy expenditure, and has profound effects on appetite. Inactivation of leptin production (ob gene in mice) or defects in the leptin receptor (db gene) results in autophagy, obesity, and severe insulin resistance. Similarly, transgenic ablation of white adipose tissue in mice leads to severe insulin resistance, elevated lipid levels, undetectable leptin concentrations, and diabetes, a phenotype reminiscent of generalized lipotrophic diabetes in humans. In these models, restoration of physiological serum leptin levels, either by leptin infusion, by transgenic overexpression, or by surgical implantation of white adipose tissue, reverses the diabetic and insulin-resistant phenotype. These results demonstrate that insulin resistance and diabetes in lipodystrophic mice can be caused by a deficiency of leptin secondary to a failure of adipocyte differentiation.

Adiponectin is also secreted from adipocytes in a regulated fashion. The expression of this adipokine is decreased in obese humans and rodents, and increased expression positively correlates with insulin sensitivity. Acute treatment of obese mice with adiponectin decreases plasma FFA, muscle, and liver triglyceride content and improves insulin resistance. In a knockout model of lipotrophic diabetes, insulin resistance was completely reversed by a combination of leptin and adiponectin infusion but was only partially reversed by either leptin or adiponectin alone.

## Skeletal Muscle

Surprisingly, specific deletion of the skeletal muscle insulin receptor (MIRKO) yields mice with normal glucose and insulin levels but with dyslipidemia and increased adiposity. Although MIRKO mice display reduced insulin-stimulated muscle glucose uptake *in vitro* and during a euglycemic hyperinsulinemic clamp, glucose tolerance tests show near-normal glucose uptake. This discrepancy is likely due to increased insulin-independent glucose uptake. In this regard, muscle-specific ablation of the insulin receptor, coupled with a loss of insulin-like growth factor-1 (IGF-1) expression, results in severe insulin resistance in muscle. Interestingly, skeletal muscle insulin-responsive glucose transporter (GLUT4) null mice are also insulin resistant, demonstrating that downstream insulin action and glucose uptake, rather than

proximal insulin receptor signaling per se, is critical for muscle insulin resistance in mice.

### Pancreatic Beta Cells

Insulin receptor knockouts in pancreatic beta cells ( $\beta$ -IRKO) display a phenotype similar to that of type 2 diabetes, having a selective loss of glucose-stimulated first-phase insulin secretion and progressive impairment of glucose tolerance. The loss of normal secretory function results from decreased expression of the critical glucose-sensing enzyme glucose kinase. These data directly demonstrate the insulin receptor functions in an autocrine feedback loop necessary for maintaining appropriate beta cell insulin secretion.

### Brain

The insulin receptor is widely distributed throughout the central nervous system (CNS), especially in the hypothalamus and pituitary. Mice with CNS-specific disruption of the insulin receptor gene (NIRKO mice) show increased food intake, resulting in diet-induced obesity and insulin resistance. Thus, insulin action in the CNS appears to provide a negative feedback loop for postprandial inhibition of food intake and plays a central role in the regulation of body weight.

## SIGNALING DOWNSTREAM OF THE INSULIN RECEPTOR

On activation, the insulin receptor can catalyze the tyrosine phosphorylation of numerous cytoplasmic proteins. The best characterized of these is the insulin receptor substrate family, consisting of four closely related members (IRS1 to -4) and three members of the Grb2-associated-binding proteins (Gab1 to -3). The IRS and Gab proteins appear to provide overlapping functions given that each can participate in the actions of various growth factors, cytokines, and developmental signals. The insulin receptor phosphorylates all four IRS family members on numerous tyrosine residues, in the process generating recognition sites for the interaction of effector proteins containing Src homology (SH2) domains. For example, IRS1 contains 21 putative tyrosine phosphorylation sites, several of which are located in amino acid sequence motifs that bind to SH2 domain proteins, including the p85 regulatory subunit of phosphatidylinositol (PI) 3-kinase, Grb2, Nck, Crk, Fyn, Csk, phospholipase C $\gamma$ , and SHP2. IRS1 also contains more than 30 potential serine/threonine phosphorylation sites in motif

recognized by various kinases such as casein kinase II, protein kinase C (PKC), protein kinase B/Akt, Jun amino-terminal kinase (JNK), and mitogen-activated protein (MAP) kinase. The crucial role for these substrate proteins has been documented both in tissue culture model systems and by the use of homologous recombination in mice. In particular, disruption of the IRS1 gene results in growth retardation and mild insulin resistance, whereas IRS2 null mice display peripheral insulin resistance, impaired insulin secretion, and diabetes. Interestingly, the IRS1 null mice do not become diabetic due to pancreatic beta cell compensation. However, the IRS2 mice have a marked reduction in beta cell mass due to a marked inhibition of beta cell development. Although mice carrying an ablated IRS3 gene did not display any detectable phenotype, disruption of IRS4 resulted in modest growth retardation and mild insulin resistance.

Both physiological and pathophysiological regulation of IRS protein function occur at multiple levels. Several studies have shown that serine/threonine phosphorylation of IRS proteins can reduce insulin-stimulated tyrosine phosphorylation, demonstrating that serine/threonine phosphorylation functions in a feedback-inhibitory mechanism. In this regard, members of the MAP kinase superfamily are capable of catalyzing serine phosphorylations of the IRS proteins. Studies indicate that the stress-induced serine kinase JNK is one such kinase. JNK is stimulated during acute or chronic inflammation and in response to inflammatory cytokines such as TNF- $\alpha$ . JNK phosphorylates IRS1 and IRS2, Shc, and Gab1, and both IRS1 and IRS2 contain JNK-binding motifs. This motif mediates the specific association of JNK with IRS1, catalyzing the phosphorylation of a serine residue on the COOH-terminal side of the phosphotyrosine-binding (PTB) domain (Ser<sup>307</sup> in [murine] IRS1, Ser<sup>312</sup> in [human] IRS1). Phosphorylation of this residue inhibits the function of the PTB domain, disrupting the association between the insulin receptor and IRS1 and inhibiting tyrosine phosphorylation. This mechanism might explain, at least in part, the insulin resistance that occurs during trauma and obesity.

IRS serine/threonine phosphorylation may occur downstream of other kinases activated by inflammatory signals. High doses of salicylates are shown to reverse hyperglycemia, hyperinsulinemia, and dislipidemia in obese rodents by sensitizing the insulin-signaling pathway, including IRS protein tyrosine phosphorylation. The effect of salicylates is attributed to inhibition of I $\kappa$ B kinase- $\beta$  (IKK- $\beta$ ), resulting in decreased IRS serine/threonine phosphorylation. Importantly, heterozygous disruption of IKK- $\beta$  protects

against the development of insulin resistance during high-fat feeding and in obese, leptin-deficient (*ob/ob*) mice. Although there is no physical interaction between IRS proteins and IKK- $\beta$ , salicylates increase insulin-stimulated phosphorylation of IRS proteins in the liver, demonstrating that IKK- $\beta$  inhibits insulin receptor function through the activation of an intermediate kinase(s) or through activation of NF $\kappa$ B.

Whereas this type of feedback via serine/threonine phosphorylation accounts for relatively short-term desensitization, the IRS proteins also undergo long-term down-regulation. Prolonged insulin stimulation substantially reduces IRS1 and IRS2 protein levels through a 26S proteasome degradative pathway. Furthermore, insulin stimulates the ubiquitination of IRS2, and reduction of IRS2 by ubiquitin/proteasome-mediated proteolysis in mouse embryo fibroblasts lacking IRS1 dramatically inhibits the activation of Akt and ERK1/2 in response to insulin or IGF-1. The activity of the ubiquitin/proteasome system is also elevated in diabetes, promoting the degradation of the IRS proteins and exacerbating insulin resistance.

## PHOSPHATIDYLINOSITOL 3-KINASE PATHWAY

Tyrosine phosphorylated IRS proteins engage several effectors that contain SH2 domains. In particular, the SH2 domains of type 1A PI 3-kinase regulatory subunits directly interact with tyrosine phosphorylated IRS proteins, resulting in the targeting and activation of the PI 3-kinase catalytic subunit. Once activated and/or appropriately targeted, this kinase generates PI 3,4,5P<sub>3</sub> in the plasma membrane and perhaps a small amount of PI 3,4,5P<sub>3</sub> in other intracellular membrane compartments. Generation of PI 3,4,5P<sub>3</sub> is essential for insulin-stimulated glucose uptake; glycogen, lipid, and protein synthesis; and changes in gene expression. There are several regulatory (p85 $\alpha$ , p85 $\beta$ , p55/AS53, p55<sup>PIK</sup>, and p50) and catalytic (110 $\alpha$  and 110 $\beta$ ) subunits of this lipid kinase. Surprisingly, genetic ablation of the p85 $\alpha$  regulatory subunit results in enhanced insulin sensitivity, probably due to excess levels of the other regulatory subunits compared with the catalytic subunit.

PI 3,4,5P<sub>3</sub> functions as a signaling intermediate that can target and activate the serine/threonine kinase PI-dependent kinase (PDK1). PDK1 contains a pleckstrin homology (PH) domain that has a high affinity and selectivity for PI 3,4,5P<sub>3</sub>. PI 3,4,5P<sub>3</sub> also interacts with the PH domain of Akt. Akt is a substrate for PDK1, and phosphorylation on two carboxyl-terminal serine/threonine residues results in the activation of

the Akt kinase. Similarly, PDK1 can also phosphorylate an activating threonine residue in the atypical PKC isoforms, PKC $\lambda$  and PKC $\zeta$ .

The Akt protein kinases exist as three isoforms, all of which are activated by phosphorylation on T<sup>308</sup> and S<sup>473</sup>. On growth factor stimulation, Akt localizes near the plasma membrane, where it becomes phosphorylated on T<sup>308</sup> by PDK1. Following phosphorylation and activation, Akt can remain associated with the plasma membrane or intracellular compartments, or it can translocate to the nucleus. Expression of constitutively active Akt stimulates glucose uptake in 3T3-L1 adipocytes, whereas Akt inhibition through the use of dominant negative mutants partially inhibits the insulin-stimulated glucose transport. Although Akt1 knockout mice have no significant alteration in whole body glucose homeostasis, Akt2 gene ablation results in moderate insulin resistance and impaired glucose tolerance. These data demonstrate that Akt2 plays an important role in insulin-regulated glucose metabolism but that Akt1 can partially compensate for the loss of Akt2 function.

PKCs are a large family of serine/threonine kinases that have also been implicated in several actions of insulin. There are three subgroups of PKCs: the classical cPKCs, the novel nPKCs, and the atypical aPKCs. Although these different family members appear to play overlapping but distinct roles, the aPKCs have been implicated in the stimulation of glucose uptake and protein synthesis by insulin in tissue culture model systems. Although the classical PKC $\alpha$  isoform has been ablated in mice and displays enhanced insulin sensitivity, atypical PKC knockout mice are not yet available.

## PI 3-KINASE-INDEPENDENT PATHWAY

Although numerous studies have demonstrated that PI 3-kinase activation is necessary, several lines of evidence have suggested that additional signals are required to fully stimulate GLUT4 translocation and glucose uptake. For example, activation of PI 3-kinase by platelet-derived growth factor (PDGF), interleukin-4 (IL-4), or engagement of integrin receptors does not induce GLUT4 translocation in adipocytes despite activation of PI 3-kinase. In addition, two insulin receptor mutants have been identified that are capable of inducing PI 3-kinase activation but not GLUT4 translocation. These data suggest that although the PI 3-kinase pathway is necessary for insulin-stimulated GLUT4 translocation, this pathway is not sufficient and at least one additional signaling pathway is required.

A potential clue toward identifying the PI 3-kinase-independent arm of insulin action emerged from the idea that signal initiation might be segregated into discrete compartments in the plasma membrane. One candidate for such compartments is caveolae, that is, small invaginations of the plasma membrane containing caveolin that are a subset of lipid raft domains. These localized regions are enriched in lipid-modified signaling proteins, glycosylphosphatidylinositol (GPI)-anchored proteins, glycolipids, sphingolipids, and cholesterol. Insulin stimulates the tyrosine phosphorylation of caveolin, the major structural protein in caveolae. Investigation of this pathway reveals the phosphorylation of another insulin receptor substrate, the proto-oncogene c-Cbl. The insulin-stimulated phosphorylation of Cbl occurs only in cell lines that respond to insulin by changes in glucose and lipid metabolism and not in other fibroblast lines, despite the presence of Cbl and an active insulin receptor. The phosphorylation of Cbl requires the presence of the adapter protein APS that recruits Cbl to the insulin receptor. APS is a member of the Ink family of adapter proteins, containing both a PH domain and an SH2 domain. Once phosphorylated, the insulin receptor recruits APS via an interaction between the activation loop of the receptor and the SH2 domain of APS. On binding to the receptor, APS undergoes phosphorylation on a single tyrosine residue. This, in turn, recruits Cbl to the receptor-APS complex via an interaction with the SH2 domain of Cbl and leads to the tyrosine phosphorylation of Cbl on three residues.

The Cbl-associated protein (CAP) is identified in a two-hybrid screen using Cbl as bait and contains three carboxyl-terminal SH3 domains. CAP is expressed predominantly in insulin-sensitive tissues and in differentiated 3T3-L1 adipocytes but not in preadipocytes. Expression of the CAP gene is increased by the insulin-sensitizing thiazolidinedione (TZD) drugs. TZD activation of PPAR- $\gamma$  directly activates the transcription of CAP through a PPAR- $\gamma$  response element in the CAP promoter. Moreover, TZD-stimulated increases in CAP expression lead to a more robust phosphorylation of Cbl in response to insulin, establishing a potential primary link between TZD-induced insulin sensitization and insulin signal transduction.

The CAP protein is recruited with Cbl to the insulin receptor due to an interaction of its third carboxyl-terminal SH3 domain with Cbl and another direct interaction with APS through its amino- and carboxyl-terminal SH3 domains. The localization and stabilization of this signaling complex in lipid rafts appear to result from its association with the hydrophobic protein flotillin. This interaction is localized to

amino-terminal sequences of CAP that contain homology to the gut peptide sorbin, that is, the sorbin homology (SoHo) domain. Dominant interfering SH3 deletion mutants of CAP that bind to flotillin but not Cbl, or SoHo deletion mutants that bind to Cbl but not flotillin, interfere with the localization of Cbl to lipid rafts. Moreover, these mutants specifically block insulin-stimulated GLUT4 translocation and glucose uptake.

The SH2 domain of the small adapter protein CrkII also associates with tyrosine-phosphorylated Cbl and thereby becomes recruited to plasma membrane lipid raft subdomains. In addition to its amino-terminal SH2 domain, CrkII contains two tandem SH3 domains, the first of which constitutively interacts with the proline-rich domain of the guanyl nucleotide exchange factor, C3G. Consistent with this mechanism, insulin also recruits C3G to lipid rafts, a process that is blocked by expression of the dominant-interfering CAP mutant lacking the SH3 domains.

C3G can function as a guanine exchange factor (GEF) for several small guanosine triphosphate (GTP)-binding proteins, including Rap1 and R-Ras/TC21. In screens for small GTP-binding proteins that could potentially regulate insulin-stimulated GLUT4 translocation, the Rho family member TC10 has been identified. Overexpression of this protein has profound effects on insulin-stimulated GLUT4 translocation and glucose uptake. Moreover, C3G activates the exchange of GTP for guanosine diphosphate (GDP) on TC10 both *in vitro* and *in vivo*.

Unlike other members of the Rho family that undergo geranylgeranylation and interact with guanyl nucleotide dissociation inhibitors, TC10 is subjected to both farnesylation and palmitoylation in a manner analogous to that of H-Ras. These posttranslational modifications are responsible for the targeting of TC10 to lipid raft domains. Moreover, mistargeting of TC10 into non-lipid raft regions of the plasma membrane prevents its activation by insulin and alters the ability of TC10 to modulate insulin-stimulated GLUT4 translocation. In addition, disruption of the lipid raft with cholesterol-extracting drugs or by overexpression of inhibitory forms of caveolin completely blocks TC10 activation as well as the stimulation of glucose transport by insulin. Moreover, treatment of cells with the cholesterol-extracting drug  $\beta$ -cyclodextrin also blocks the stimulation of glucose transport.

TC10 also interacts with one of the components of the exocyst complex, Exo70, in a GTP-dependent fashion. Exo70 translocates to the plasma membrane in response to insulin via the activation of TC10,

where it assembles a multiprotein complex that includes Sec6 and Sec8 as well as other proteins that have not yet been identified. Overexpression of an Exo70 mutant blocks insulin-stimulated glucose uptake but not the trafficking of GLUT4 to the plasma membrane. However, this mutant does block the extracellular exposure of the GLUT4 protein. These data demonstrate that the exocyst may play a critical role in the targeting of the GLUT4 vesicle to the plasma membrane, perhaps directing the vesicle to the precise site of fusion.

## CYTOSKELETON AND GLUCOSE TRANSPORT

It is well established that the cytoskeleton plays an important role in the stimulation of glucose uptake by insulin. Although the PI 3-kinase pathway can influence actin polymerization, changes occur in actin function that are PI 3-kinase independent. For example, insulin was reported to enhance the localization of the unconventional myosin (Myo1c) to GLUT4 compartments that are associated with actin filaments independent of PI 3-kinase activity. Overexpression of wild-type Myo1c enhances insulin-stimulated GLUT4 translocation, whereas expression of a dominant negative Myo1c mutant inhibits this action. Thus, Myo1c may function in a PI 3-kinase-independent insulin-signaling pathway controlling GLUT4 trafficking.

Microtubules have also been implicated in the regulation of GLUT4 translocation. Kinesin KIF5B is partially colocalized with perinuclear GLUT4, and expression of dominant negative kinesin mutants blocks outward GLUT4 vesicle movements. These events are not affected by specific inhibition of PI 3-kinase activity. Thus, insulin signaling appears to target and/or initiate the movement of GLUT4-containing membranes on microtubules via conventional kinesin (KIF5B) and subsequently actin through PI 3-kinase-independent mechanisms.

## SNARE-MEDIATED GLUT4 VESICLE DOCKING/FUSION AND TRANSPORT ACTIVATION

It is well established that vesicle-mediated fusion with target membranes requires a high-affinity interaction between specific pairs of target membrane SNAP receptors (t-SNAREs) and those of vesicle membrane SNAP receptors (v-SNAREs). Considerable progress has been made in identifying the v-SNAREs and

t-SNAREs that facilitate GLUT4 vesicle translocation. VAMP2 or syntaptobrevin 2 is the predominant v-SNARE found in GLUT4 vesicles, whereas syntaxin 4 and SNAP23 function as the major t-SNAREs. Many of the accessory components of GLUT4 vesicles have been identified, and intensive efforts are under way to isolate the component of the various subcellular GLUT4 compartments. Adapter molecules that regulate the interaction between VAMP2 and syntaxin 4 have been identified and include Munc18c, Synip, and tomosyn.

As described previously, it is well established that insulin induces a redistribution of the GLUT4 protein from intracellular storage compartments to the plasma membrane. Although it was generally accepted that this was sufficient to account for the increase in glucose uptake, more recent studies have suggested that a second event is required for the acquisition of glucose transport activity, perhaps one dependent on the activation of the p38 MAP kinase.

## INSULIN REGULATES GLUCOSE USE AND STORAGE

In addition to stimulating uptake, insulin directly regulates the use and storage of glucose. Following its transport into the cell, insulin can activate enzymes responsible for the oxidation of glucose and its storage as glycogen or lipid. The hormone also can inhibit enzymes that promote the breakdown of lipid or glycogen. Intracellular glucose undergoes a regulated phosphorylation step and is directed to a storage pathway involving glycogen synthesis or an oxidative pathway culminating in the rate-limiting enzyme pyruvate dehydrogenase in the mitochondrion. The resulting two carbon intermediates serve as substrates for *de novo* FFA synthesis and subsequent esterification with glycerol to form triglyceride. Numerous enzymes that control these pathways are regulated by insulin through processes that involve covalent modifications such as phosphorylation, allosteric regulation, or changes in gene expression.

The storage of glucose as glycogen allows for its rapid mobilization as a metabolic fuel. Glycogen can be metabolized to produce adenosine triphosphate (ATP) in the absence of oxygen and can produce more ATP per oxygen molecule used than does lipid when used under aerobic conditions. Thus, most tissues rely on glucose as their primary energy source under postprandial conditions and during times of stress or the initial stages of starvation. Insulin regulates glycogen metabolism primarily via the phosphorylation states

and activities of the enzymes that catalyze glycogen synthesis or breakdown. Glycogen synthase is inactivated by its phosphorylation at multiple residues. Phosphorylation of glycogen synthase occurs at nine different sites in a hierarchical fashion. These phosphorylations are catalyzed by a number of different kinases, including calmodulin-dependent kinases, casein kinase, cyclic adenosine monophosphate (cAMP)-dependent protein kinase (PKA), AMP-activated kinase, and glycogen synthase kinase.

Glycogen synthase kinase-3 (GSK3) plays a major role in the phosphorylation and inactivation of glycogen synthase. GSK3 is phosphorylated on Ser<sup>9</sup> by the Akt protein kinase, which is activated by insulin via PDK1 as described previously. Phosphorylation at this site reduces the catalytic activity of the kinase. Thus, one mechanism by which insulin can increase the overall activity of glycogen synthase is through the inactivation of GSK3. Phosphorylation of glycogen synthase by GSK3 occurs in a sequential manner, beginning at Ser<sup>652</sup> (site 4), followed by Ser<sup>648</sup> (site 3c), Ser<sup>644</sup> (site 3b), and Ser<sup>640</sup> (site 3a). The phosphorylation of glycogen synthase by GSK3 can occur only after the initial phosphorylation of site 5 (Ser<sup>656</sup>) that is catalyzed by casein kinase II.

Glycogen synthase can also be activated via direct dephosphorylation mediated by protein phosphatase 1 (PP1), the major phosphatase thought to stimulate glycogen synthase activation by insulin. PP1 can also dephosphorylate and inactivate glycogen phosphorylase and phosphorylase kinase. Because PP1 has relatively broad substrate specificity *in vitro*, a number of targeting subunits have been identified that facilitate the recruitment of the phosphatase to specific subcellular locales, including myofibrils, neuronal dendrites, nucleus, and glycogen particles. Four glycogen-targeting subunits of PP1 have been described. Although these share certain conserved structural domains for the binding of PP1, glycogen, and PP1 substrates, they vary in their interactions with the enzymatic machinery of glycogen metabolism, tissue distribution, and roles in glycogen metabolism.

### G<sub>M</sub>-Muscle Glycogen Targeting Subunit

The expression of PPP1R3a (G<sub>M</sub>, R<sub>GL</sub>) is restricted to striated muscle fibers and cardiac muscle. G<sub>M</sub> is unique in that it possesses phosphorylation sites that have been implicated in the regulation of muscle glycogen synthesis by  $\beta$ -adrenergic stimuli and a large C-terminal region that is required for association with the sarcoplasmic reticulum. Phosphorylation of Ser<sup>67</sup>, which lies within the PP1-binding

region, results in the dissociation of G<sub>M</sub> from PP1, and this may contribute to decreases in glycogen synthesis due to increased intramyocellular cAMP. Two mouse models of targeted disruption of the PPP1R3a gene have been reported, and both animal models exhibit greater than 80% decreases in the levels of steady-state muscle glycogen, paralleled by decreased activity of glycogen synthase and increased glycogen phosphorylase. These data demonstrate that G<sub>M</sub> plays a vital role in the regulation of glycogen metabolism in muscle. Paradoxically, one of the reported G<sub>M</sub> knockout models develops severe glucose intolerance and insulin resistance and becomes obese with aging, whereas the other does not. This discrepancy may be due to differences in genetic background, suggesting that dramatic decreases in muscle glycogen alone are not sufficient to produce insulin resistance without the contribution of other genetic or environmental factors.

### G<sub>L</sub>

PPP1R4 (G<sub>L</sub>) is a liver-specific glycogen-targeting subunit. This subunit is unique in that it does not bind glycogen synthase yet still appears to play a role in the regulation of hepatic glycogen synthase phosphatase activity. However, G<sub>L</sub> possesses a unique high-affinity glycogen phosphorylase *a* binding region at its C terminus. The binding of active phosphorylase by G<sub>L</sub> is thought to permit the allosteric inhibition of liver glycogen synthase phosphatase by phosphorylase *a* during periods of net glycogenolysis. Overexpression of G<sub>L</sub> in primary hepatocytes leads to significant increases in steady-state glycogen. The expression of G<sub>L</sub> is down-regulated during prolonged fasting and in streptozotocin diabetes and can be restored by refeeding or insulin treatment.

### PTG

Protein targeting to glycogen (PTG, PPP1R5) is expressed in a variety of tissues that store glycogen. PTG is expressed most highly in adipose, liver, heart, striated muscle, brain, and kidney. PTG is known to bind glycogen synthase, glycogen phosphorylase, and phosphorylase kinase in addition to PP1 and glycogen. The binding of enzymatic substrates to PTG can occur in the absence of glycogen *in vitro* and is mediated by a single binding region in the molecule. Overexpression of PTG in Chinese hamster ovary cells stably expressing the insulin receptor, primary hepatocytes, 3T3-L1 adipocytes, cultured human muscle cells, or intact rat liver results in dramatic

increases in steady-state glycogen levels. However, unlike the other targeting subunits, overexpression of PTG seems to lock cells in a glycogenic mode, where they become resistant to the effects of glycogenolytic stimuli and glycogen feedback inhibition. Appropriate regulation of PTG expression is required to maintain the appropriate balance between glycogen synthesis and glycogenolysis. Targeted disruption of one allele of the gene for PTG results in moderate decreases in steady-state glycogen in adipose, heart, and liver, with some minor effects in white fiber striated muscle. Homozygous deletion of PTG results in embryonic lethality. PTG heterozygous mice develop glucose intolerance and mild insulin resistance, hyperinsulinemia, and hyperleptinemia with aging. This phenotype appears to arise from a repartitioning of fuel stores into lipid due to the decreases in whole body glycogen stores, thereby resulting in the accumulation of triglyceride in muscle and an attenuation of insulin signaling. The expression of PTG is down-regulated in rat liver due to fasting or streptozotocin diabetes and can recover on refeeding or insulin treatment, much like  $G_L$ . However, the changes in PTG expression occur much sooner on the commencement of fasting than do the changes in  $G_L$  expression.

## REGULATION OF LIPID METABOLISM

The biosynthesis and breakdown of lipids are tightly regulated by insulin by modulating the activity of enzymes that catalyze the rate-limiting reactions of these pathways and by controlling expression of the genes encoding these enzymes. The hormone also stimulates the recruitment of fatty acids into tissues by hydrolysis of triacylglycerol derived from circulating very low-density lipoproteins and intestinal chylomicrons, which are by-products of hepatic lipid synthesis or dietary influx.

Under normal conditions, lipogenesis occurs predominantly in adipose depots and in the liver. Two coupled processes are involved: *de novo* fatty acid synthesis from two carbon intermediates and the subsequent esterification of activated fatty acids to glycerol to form triglyceride (the final lipid storage form). The rate-limiting step of lipogenesis is the carboxylation of acetyl-CoA to form malonyl-CoA, catalyzed by the enzyme acetyl-CoA carboxylase (ACC). The activity of this enzyme depends on its state of polymerization, which is favored by binding of the allosteric activator citrate. Aggregation of the monomer into the active polymeric form is also stimulated by the phosphorylation at a specific site in response to

insulin. Conversely, phosphorylation at alternate sites by cAMP-dependent PKA or by the AMP-activated protein kinase results in the depolymerization and deactivation of ACC. Insulin also promotes the dephosphorylation of these sites.

Oxidation of fatty acids occurs primarily in liver, as well as in cardiac and striated muscle, which uses fatty acids as a main fuel in the resting state. Insulin indirectly inhibits fatty acid  $\beta$ -oxidation by increasing malonyl-CoA production (via ACC activation). Malonyl-CoA allosterically inhibits the activity of carnitine palmitoyl transferase I, and this catalyzes the rate-limiting transport of activated fatty acids into the mitochondrion where oxidation takes place. In a similar fashion, feedback inhibition by malonyl-CoA and long-chain fatty acyl CoA esters also affects ACC activity by triggering depolymerization of the active polymer.

The integrated coordination of lipogenesis and  $\beta$ -oxidation by insulin also occurs at the level of gene expression, mediated by two main transcription factors. These *trans*-acting factors include isoforms of the sterol receptor enhancer-binding proteins (SREBPs) and PPAR- $\gamma$ . SREBPs participate with other *trans*-acting factors, such as the upstream stimulatory factors, to mediate the effects of insulin on the expression of various genes involved in hepatic lipid metabolism, such as that encoding fatty acid synthetase. SREBP2 appears to be involved in hepatic cholesterol metabolism because its overexpression results in increased synthesis of hydroxymethylglutaryl (HMG)-CoA reductase and farnesyl pyrophosphate synthase. The role of SREBP1 in fatty acid/triglyceride metabolism seems to be restricted mainly to the liver, although it likely plays an important role in adipocyte cholesterol metabolism. SREBP null mice show no changes in adipose mass, or in the expression of acetyl-CoA carboxylase or fatty acid synthetase, although the up-regulation of hepatic lipogenic genes during fasting/refeeding is prevented.

PPAR- $\gamma$  is a nuclear hormone receptor present mainly in adipose tissue, but it is also expressed at low levels in liver. PPAR ligands include prostaglandins, fatty acids, and the synthetically derived thiazolidinediones. PPAR- $\gamma$  expression is induced by insulin, and its promoter is itself *trans*-activated by SREBP1. PPAR- $\gamma$  expression is critical for adipocyte differentiation and is required for the *trans*-activation of adipocyte gene products involved in lipid metabolism, including acetyl-CoA synthetase, adipocyte fatty acid-binding protein, fatty acid transport protein, the glycerolneogenic enzyme phosphoenolpyruvate carboxykinase (PEPCK), and lipoprotein lipase (LPL).

This nuclear receptor also increases insulin sensitivity in cells and *in vivo*, explaining the antidiabetic effects of the PPAR- $\gamma$ -activating drugs.

The circulating levels of insulin required to inhibit lipolysis are far below those required to stimulate glucose uptake into muscle. This physiological property results in an initial inhibition of fatty acid release into the circulation from adipose depots, and increased uptake of dietary-derived lipids by the peripheral tissues occurs following ingestion of a meal. Lipolysis in the fat cells is catalyzed by the enzyme hormone-sensitive lipase (HSL) in a mechanism that requires the translocation of the enzyme from the cytoplasm to the periphery of the lipid droplet. Upon phosphorylation by PKA, the enzyme undergoes catalytic activation and is translocated to the lipid droplet by a mechanism involving the PKA substrate perilipin. The phosphorylation of perilipin results in a dispersion of the normal perilipin "coat" around the lipid droplet, thereby potentially increasing the accessible surface area of the droplet to HSL activity. Insulin prevents the phosphorylation of HSL primarily by attenuating cAMP signaling via activation of cyclic nucleotide phosphodiesterase 3B, which catalyzes the conversion of cAMP to 5'-AMP. Insulin also increases the dephosphorylation of HSL and inhibits the phosphorylation of perilipin. Regulation of HSL localization also involves the docking protein lipotransin, and HSL is capable of interacting with lipotransin after phosphorylation by PKA, thereby allowing for its docking at the outer surface of the lipid droplet. Lipotransin possesses an intrinsic ATPase activity that allows for a dissociation of the complex after docking, permitting HSL to proceed with the catalysis. Insulin may freeze the HSL-lipotransin complex at the docking site, thereby preventing the dissociation step that is required for HSL to catalyze triglyceride hydrolysis.

The release of dietary- or hepatic-derived fatty acids from circulating lipoprotein complexes is catalyzed

by LPL, which catalyzes the rate-limiting step in the hydrolysis of lipoprotein-associated triglycerides to yield FFAs and 2-monoacylglycerol. Insulin appears to increase the expression and release of LPL from fat cells. Insulin stimulation of LPL activity occurs by both wortmannin-sensitive PI 3-kinase and rapamycin-sensitive p70 S6-kinase-signaling pathways.

### See Also the Following Articles

Insulin and Insulin-like Growth Factors, Evolution of •  
 Insulin Secretion: Functional and Biochemical Aspects •  
 Insulin Secretion, Physiology

### Further Reading

- Boss, A., Guilherme, A., Robida, S. I., Nicoloso, S. M., Zhou, Q. L., Jiang, Z. Y., Pomerleau, D. P., and Czech, M. P. (2002). Glucose transporter recycling in response to insulin is facilitated by myosin Myo1c. *Nature* **420**, 821–824.
- Brady, M. J., and Saltiel, A. R. (2001). The role of protein phosphatase-1 in insulin action. *Recent Prog. Horm. Res.* **56**, 157–173.
- Cantley, L. C. (2002). The phosphoinositide 3-kinase pathway. *Science* **296**, 1655–1657.
- Inoue, M., Chang, L., Hwang, J., Chiang, S. H., and Saltiel, A. R. (2003). The exocyst complex is required for targeting of Glut4 to the plasma membrane by insulin. *Nature* **422**, 629–633.
- Mauvais-Jarvis, F., Kulkarni, R. N., and Kahn, C. R. (2002). Knock-out models are useful tools to dissect the pathophysiology and genetics of insulin resistance. *Clin. Endocrinol.* **57**, 1–9.
- Mora, S., and Pessin, J. E. (2002). An adipocentric view of signaling and intracellular trafficking. *Diab. Metab. Res. Rev.* **18**, 345–356.
- Rosen, E. D., Walkey, C. J., Puigserver, P., and Spiegelman, B. M. (2002). Transcriptional regulation of adipogenesis. *Genes Dev.* **14**, 1293–1307.
- Saltiel, A. R., and Pessin, J. E. (2002). Insulin signaling pathways in time and space. *Trends Cell. Biol.* **12**, 65–71.
- Unger, R. H., and Orci, L. (2001). Diseases of liporegulation: New perspective on obesity and related disorders. *FASEB J.* **15**, 312–321.
- White, M. F. (2002). IRS proteins and the common path to diabetes. *Am. J. Physiol. Endocrinol. Metab.* **283**, E413–E422.

with its receptor, whereas the IGFs act via complex interactions involving the IGF-binding proteins (IGFBPs). These transport the IGFs with high specificity and affinity in serum and extracellular fluid toward their target cells and do not bind insulin. Less than 1% of serum IGF-I is present in unbound form. The IGFBPs show pronounced changes in their expression patterns, from embryonic through fetal, postnatal, and adult life. In mammals, six structurally related binding proteins (IGFBP-1 to IGFBP-6) have been characterized, and recently several groups of proteins with structural and functional similarities to the IGFBPs, designated IGFBP-related proteins, were discovered. IGFBP-1, -3, and -4 bind IGF-I and IGF-II with comparable affinity, whereas IGFBP-2, -5, and -6 have lower affinity for IGF-I than for IGF-II. In addition to their role in transporting, storing, and modulating the bioactivity of the IGFs, the IGFBPs likely also act as modulators of some physiological actions of the IGFs. Several cell types secrete specific proteases that cleave particular IGFBPs, resulting in reduced IGF binding and an increase in IGF availability to the receptor.

Three distinct receptors mediate the actions of insulin and the IGFs at the target cells. The insulin receptor is a tyrosine kinase that specifically binds insulin and shows a considerably low affinity for the IGFs. For the IGFs, two different receptors have been characterized: the so-called type 1 and type 2 IGF receptors. Like the insulin receptor, the type 1 IGF receptor (IGF-1R) is a heterotetrameric glycoprotein with two  $\alpha$ -subunits and two  $\beta$ -subunits. The IGF-1R shows 57% amino acid sequence identity to the insulin receptor. The extracellular  $\alpha$ -subunits have cysteine-rich domains that confer IGF-binding properties. The  $\beta$ -subunits contain a short extracellular domain, a transmembranal domain, and an intracellular tyrosine kinase domain. The IGF-1R has a high affinity for IGF-I, binds IGF-II with about three times lower affinity, and has an affinity for insulin that is 100 times lower. The type 2 IGF receptor is also the cation-independent mannose-6-phosphate receptor (M6P-R), which in mammals exhibits an additional IGF-II binding site. Therefore, it is also called the IGF-2R/M6P-R and is structurally different from the type 1 IGF and the insulin receptor. The IGF-2R/M6P-R shows a high affinity for IGF-II, a minute affinity for IGF-I, and no affinity for insulin. The receptor certainly serves to internalize and degrade IGF-II. Because neither genetic nor biochemical evidence supports a signaling function for the IGF-2R/M6P-R for certain, it is well accepted that its physiological role is primarily to remove IGF-II

from the circulation. Thus, most if not all effects of IGF-II are mediated via the IGF-1R.

In summary, the IGF system, in contrast to that of insulin, shows an extraordinarily high degree of complexity. It consists of the hormones IGF-I and IGF-II (and some variants), at least six different binding proteins, specific IGFBP proteases, and the type 1 and type 2 IGF receptors.

## EVOLUTION OF INSULIN AND THE IGFs

### Insulin

In mammals, the only production sites of insulin are the beta cells in the pancreatic islets that secrete insulin into the systemic circulation. The main stimulus for insulin secretion is meal ingestion. After binding to the insulin receptor, insulin acts as a metabolic hormone and exerts acute anabolic effects. In liver, insulin promotes synthesis and storage of glycogen and inhibits glycogenolysis, gluconeogenesis, and ketogenesis. It further enhances protein and triglyceride synthesis. In addition, insulin promotes protein and glycogen synthesis in muscle and lipid storage in adipose tissue. The most important net result of these anabolic effects is that insulin maintains a normal blood glucose concentration.

Insulin and its receptor occur in representatives of all vertebrate classes. As in mammals, insulin is produced in a subpopulation of the pancreatic islet cells, and many groups have elaborated that the most powerful stimulus for its secretion is food intake. In all vertebrates that exhibit an exocrine pancreas surrounding the islets (e.g., mammals, birds, reptiles, amphibians, osteichthyes, chondrichthyes), the islets contain insulin, glucagon, somatostatin, and PP cells. However, the islets of cyclostomes, which lack an exocrine pancreas, consist only of numerous insulin and infrequent somatostatin cells. At the evolutionary level of protochordates such as *Branchiostoma lanceolatum*, where no endocrine islets have yet evolved, insulin is present in the endocrine cells of the intestinal mucosa.

The tertiary structure of the insulin molecule (A chain, B chain, and cysteine bridges) is fully conserved in vertebrates, but the primary amino acid sequences may differ among the vertebrate classes. Still, the insulin sequences of human and hagfish, a cyclostome, share 61% amino acid identity. Throughout the vertebrates, the essential amino acid residues in the putative receptor-binding region are quite well conserved. The role of insulin in nonmammals is less well defined

than in mammals, but there is increasing evidence that it also acts mainly as a metabolic hormone. This is indicated, for instance, by studies in bony fish that have shown that islectomy results in severe hyperglycemia. However, the blood glucose level-stabilizing effects of insulin develop more slowly in ectotherms than in mammals.

The conservative evolution of the hormone insulin, its receptor, its functional mode, and its presence in all vertebrates and protochordates emphasize the major impact of the chief action of insulin—maintenance of the blood glucose level.

### Insulin-like Growth Factors

The IGFs show pronounced differences in their distribution patterns and physiological roles when compared with insulin. In contrast to insulin, IGF-I and IGF-II are expressed in several organs, do not circulate in free form but rather attached to the binding proteins, and stimulate normal growth and development by selectively promoting mitogenesis and differentiation and by inhibiting apoptosis via the IGF-1R. However, insulin and the IGFs share about 50% amino acid sequence homology, and proinsulin and the IGFs exhibit similar tertiary structures. Furthermore, insulin and IGF-I cross-react with each other's receptor, although with lower potencies. Thus, questions arise as to how and why such a complex hormonal system (hormones, receptors, and binding proteins) like that of the IGFs arose during chordate evolution, whether insulin and the IGFs show specializations in different vertebrate classes, and whether insulin and the IGFs stem from a common evolutionary precursor.

#### **Insulin-like Growth Factor I**

According to the "classical" somatomedin hypothesis already raised in 1972 by Daughaday and others, growth hormone (GH) stimulates skeletal growth via IGF-I produced in the liver under the influence of GH and secreted into the circulation. From there, IGF-I reaches its target tissues or cells, where it interacts with the IGF-1R. Thus, the physiological role of IGF-I is defined as endocrine. However, soon after its discovery, it was demonstrated that IGF-I, in contrast to insulin, is expressed ubiquitously. In most if not all of these tissues, GH also induced IGF-I production. Recently, it was shown that complete and selective inactivation of the IGF-I gene in the livers of mice by using the Cre/loxP recombination system led to reduction of serum IGF-I by 75%. Surprisingly, this pronounced decrease in serum IGF-I did not

significantly affect postnatal body growth. Important conclusions can be drawn from these experiments. First, the results confirm that liver is the principal source of circulating IGF-I, but a significant amount is probably derived from extrahepatic sites. Second, the level of endocrine IGF-I seems to be not essential for postnatal growth and development. Third, autocrine/paracrine effects of local organ-specific IGF-I probably are more important than is generally assumed.

In birds, reptiles, amphibians, and bony fish as well, liver is the major source of endocrine IGF-I but extrahepatic sites also express IGF-I. These have been analyzed comprehensively by our group in the bony fish tilapia, where we identified parenchymal cells, such as islet D cells, gastro-entero-endocrine cells, renal tubular cells, interrenal cells, chondrocytes, gill chloride cells, granulosa cells in the ovary, Sertoli cells in the testis, and brain neurons, as production sites. When combining and comparing the results from studies on the distribution of IGF-I throughout the vertebrate line of evolution, this list seems to be valid for mammals, birds, reptiles, amphibians, bony fish, and cyclostomes as well. Thus, the production of the hormone IGF-I in numerous organs likely is the rule throughout the vertebrate lineage of evolution. Although our knowledge of the distribution of the IGF-1R is more limited because studies on this topic have been performed only in birds, amphibians, and bony fish, evidence is increasing that the receptor may as well show an ubiquitous occurrence. The probable coexpression of the IGF-1R and its ligand may suggest that autocrine/paracrine actions of IGF-I also play a role in nonmammals. This notion is supported by studies in birds that show interactions between LH and intraovarian IGF-I in follicle development. Similarly, in the bony fish ovary, the local production sites of IGF-I depend on the stage of follicle development and the IGF-I binding is seasonally dependent. However, at present, we cannot judge whether paracrine effects of IGF-I in nonmammals are of similar importance as they are in mammals.

As in mammals, GH and the nutritional status regulate the hepatic production of IGF-I in birds, reptiles, amphibians, and bony fish. Limited evidence indicates that the synthesis of IGF-I in extrahepatic sites of nonmammalian vertebrates may also be GH dependent. An increase in IGF-I mRNA after GH application has been reported for cartilage in amphibians and bony fish and for testes and gills in bony fish. As in mammals, target-specific hormones, such as gonadal steroids and glucocorticoids, might

also regulate IGF-I in organ-specific cells of non-mammals, as has been demonstrated in birds and bony fish.

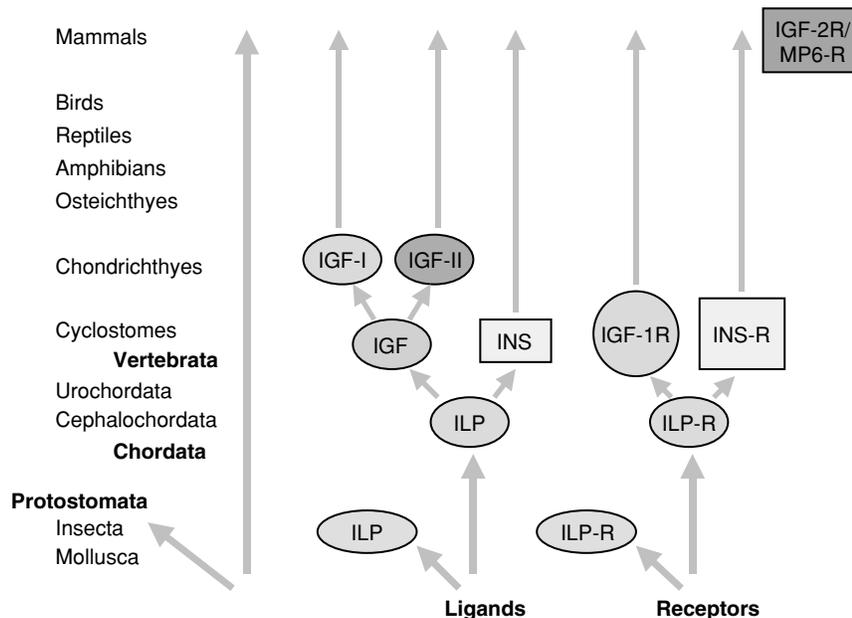
The growth-promoting action of IGF-I and its related role during development seem to be the most prominent functions of IGF-I in nonmammalian vertebrates, as they are in mammals. However, further effects of IGF-I comparable to those in mammals have been established, including the significant role of IGF-I in male and female reproductive processes that has been recognized in birds and bony fish. The IGF-I system also shows specializations in certain vertebrates that are most pronounced in bony fish. Here, IGF-I plays an important role in osmoregulation as well as during parr-smolt transformation of salmonids, that is, the process whereby the metabolism is changed in preparation for the movement from fresh water to sea water. However, our current stage of information is too limited to allow any conclusion on further analogies or potential differences to the situation in mammals or to point to the evolutionary level of appearance of particular types of regulation.

#### **Insulin-like Growth Factor II**

For a long time, it has been thought that in adult humans (and other mammals), the IGF-II gene is expressed solely in liver and choroid plexus, but recent studies indicate further sites of IGF-II production such as ovary, chondrocytes, and endocrine pancreas.

The presence of IGF-II in extrahepatic sites, especially in ovary and endocrine pancreas, has also been demonstrated in birds, reptiles, and amphibians. Most information on IGF-II in nonmammalian vertebrates is available for bony fish. CDNA sequences encoding IGF-II (prepro-)hormones have been characterized from numerous species. Some peculiarities exist regarding IGF-II in osteichthyes. First, the IGF-II prepro-hormones exhibit between 85 and 92% amino acid sequence homology to each other. Thus, among bony fish, the homologies of the IGF-II prohormones are higher than those of the IGF-I prohormones. Second, in osteichthyes, our group has shown that all organs that express IGF-I mRNA also express IGF-II mRNA, which contrasts the situation in mammals. Third, in this context, the findings of the Gutierrez group in Barcelona, Spain, who recently reported specific IGF-II binding to purified receptor preparations from trout embryos, may be of special interest. Whether or not these results in fact indicate a particular physiological impact of IGF-II in bony fish remains to be clarified.

IGF-II appears at the evolutionary level of chondrichthyes (Fig. 1). In 1994, our group identified IGF-II-like peptides by their immunoabsorption properties that were different from mammalian and nonmammalian insulin in a ray, *Raja clavata*, and the Steiner group in Chicago cloned cDNA sequences encoding distinct IGF-I and IGF-II prepro-hormones



**Figure 1** Model of the evolution of the insulin and the IGF system. Adapted from Reinecke, M., and Collet, C. (1998). The phylogeny of the insulin-like growth factors. *Int. Rev. Cytol.* **183**, 70.

from a shark (*Squalus acanthias*). Thus, both IGFs are present in elasmobranchians, and the duplication of the proto-IGF gene probably occurred prior to the divergence of the chondrichthyes (Fig. 1).

The IGF-2R/M6P-R seems to have a unique evolution. Regarding the nonmammalian vertebrates, several studies have identified the cation-independent M6P-R in birds and amphibians, but mammalian or avian IGF-II failed to bind to the receptor. Thus, no binding site for IGF-II seems to reside on the M6P-R at the phylogenetic level of birds and amphibians. Based on these results, Reinecke and Collet proposed that the acquisition of a high-affinity binding site for IGF-II on the M6P-R is a “recent” evolutionary event and appeared with the origin of mammals (Fig. 1). In line with this notion, our group could identify no type 2 IGF receptor, but only the type 1 IGF receptor, in bony fish. In sharp contrast to the other studies is the specific IGF-II binding to purified receptor preparations from trout embryos that was recently described because this would suggest the presence of an IGF type 2 receptor at the evolutionary level of osteichthyes. Whether this result implies that specific and functionally significant IGF-II receptors exist in bony fish must be the topic of further research that also needs to consider adults.

Since its detection, thousands of studies have dealt with the potential physiological role of IGF-II. However, we are far from understanding more than basal parameters. In contrast to IGF-I, the expression of IGF-II in mammals is species dependent. In rodents, the amount of IGF-II mRNA in tissue and the level of serum IGF-II are higher during fetal life than during postnatal life. In agreement, even high serum levels of IGF-II in postnatal rodents exert minimal effects. In contrast, human serum levels of IGF-II are low during ontogeny but increase after birth. Little is also known about the regulation of IGF-II in serum or tissue because GH, in contrast to IGF-I, is a negligible regulator of IGF-II in mammals and studies in nonmammalians are virtually nonexistent. Only in bony fish has some limited evidence been presented that GH may regulate the expression of IGF-II, a possibility that would contrast the situation in mammals. Although after GH treatment the levels of IGF-II mRNA in liver and in several extrahepatic sites of seabream remained unchanged, in rainbow trout liver the amount of IGF-II mRNA was significantly elevated after GH treatment *in vivo* and *in vitro*. Because methodological differences may account for this discrepancy, there is no reason to exclude a possible role for GH in the regulation of IGF-II in bony fish to date.

In general, IGF-II mimics the effects of IGF-I on various target cells. Humbel, whose group was the first to identify the amino acid sequences of both IGFs, wrote in 1990, “Nature is a sphinx presenting us with IGF-II as a riddle.” Later, Gluckman and Ambler stated, “The failure to demonstrate clear roles for IGF-II remains a major dilemma in growth physiology.” Even more recent reviews did not dare to define any physiological role for IGF-II in mammals or to claim laconically, as Le Roith did in 1997, that “the physiological role of IGF-II is unknown.” Nevertheless, when considering the potential physiological role of IGF-II, several ideas come to mind.

1. One might regard IGF-II as a functionless evolutionary relic. In favor of this hypothesis is the fact that throughout the vertebrate line of evolution, both IGF-I and IGF-II bind to the type 1 IGF receptor. After binding, IGF-II exerts the same effects as does IGF-I, although with marked lower potency.

2. We witness the evolution of the peptide hormone IGF-II. Although over some million years, the evolution of the IGF receptors had stagnated, an IGF-II-binding domain appeared on the M6P-R with the origin of mammals (Fig. 1), and we cannot know whether specific physiological functions of IGF-II will evolve during the next million years of evolution.

3. IGF-II at first evolved as a quite meaningless hormone. In other words, it appeared but did not impair any physiological process. However, the differences between the regulation of IGF-II in mammals and that in nonmammalian vertebrates are not trivial. The additional regulation of IGF-II bioavailability by the IGF-2R/M6P-R in mammals may well have its significance. Thus, Reinecke and Collet proposed that “particular aspects about mammalian physiology allow or even demand additional methods of tightly controlling IGF-II levels.” For instance, the IGF-II-binding domain on the M6P-R may have evolved to protect mammals against otherwise harmful effects of IGF-II. This hypothesis is supported by the involvement of IGF-II in tumor genesis and the recent suggestion to consider the IGF-2R/M6P-R as a promising candidate tumor suppressor gene.

### Evolutionary Appearance of (proto-)IGF, Type I Receptor, and IGF-Binding Proteins

The Steiner group has presented strong evidence that all components of the IGF system (ligand, receptor,

and IGF-BPs) appeared concurrently at the evolutionary level of cyclostomes. Several studies of other groups support these findings.

1. Both an insulin gene and an IGF gene have been identified in the Atlantic hagfish *Myxine glutinosa*. Like the mammalian prepro-IGFs, the hagfish prepro-IGF contains a signal peptide; B, C, A, D, and E domains; and all residues necessary to maintain the typical tertiary structure. In its A and B domains, hagfish IGF shows equal (70%) homologies to those of (h)IGF-I or IGF-II and so is considered as proto-IGF. Therefore, the gene duplication that gave rise to the proto-IGF gene from insulin probably occurred subsequently to the divergence of the cephalochordates from the main lineage leading to the vertebrates but prior to the origin of the agnatha some 550 million years ago (Fig. 1).

2. cDNA cloning indicates that two distinct receptor sequences exist in hagfish. One cDNA sequence appears to encode the insulin receptor, and the other appears to encode an IGF type 1 receptor. Thus, the IGF-1R likely also has its origin around the time of the insulin/IGF duplication (Fig. 1). Above the phylogenetic level of cyclostomes, consistently separate insulin receptors and IGF-1R that show a high degree of overlap in the signaling pathways are present.

3. The Australian group of Upton has detected three IGF-BPs in a lamprey, suggesting that the IGF-BPs may have arisen in evolution simultaneously with or soon after the divergence of IGF from insulin.

Thus, two hormones of quite different functions have evolved in the lowest vertebrate class. One is involved in carbohydrate metabolism, and the other is essential for growth, differentiation, and development. The appearance of these separate hormone systems was probably driven by the need to uncouple the short-term regulation of nutrition from the long-term regulation of growth and development.

Furthermore, the concomitant appearance of proto-IGF and the binding proteins suggests that the first IGF-BPs may have fulfilled instant functional roles. Their capability to block the potential of the proto-IGF to bind to the insulin receptor, and so to exert insulin-like activities, certainly is the most fundamental one. However, their potency to provide particular tissue and/or developmental specificities for the newly evolved IGF may have also been of immediate importance.

## Evolutionary Divergence of Insulin, IGF-I, and IGF-II

Two separate duplication events gave rise to insulin and the IGFs as they are present in bony fish, amphibians, reptiles, birds, and mammals. The divergence of IGF-I and IGF-II apparently is the by-product of a large partial chromosomal duplication involving chromosome 11 and its translocation to chromosome 12. As outlined previously, distinct IGF-I and IGF-II genes first appear in chondrichthyes, whereas a single proto-IGF gene with equal sequence identity to and characteristics of mammalian IGF-I and IGF-II has been identified in the agnathan *M. glutinosa*. Thus, the duplication of the proto-IGF gene likely occurred between the radiation of the lineages leading to the extant cyclostomes and cartilaginous fish.

The current idea of the origin of the proto-IGF gene (Fig. 1) is based on studies in protochordates. From the cephalochordate *Branchiostoma californiensis*, the amphioxus, the Steiner group has cloned an insulin-like peptide (ILP) cDNA and shown that it is a single-copy gene. The deduced pro-ILP contained B, C, and A domains as expected for insulin and a C-terminal extension similar to the D and E domains in pro-IGFs. ILP was equally (48%) related to human insulin and IGFs. Amphioxus is estimated as an extant relative of the invertebrate progenitors from which the vertebrates have developed. Thus, it is generally assumed that the ILP gene is ancestral to both vertebrate insulin and IGFs and that the insulin/IGF duplication occurred after the radiation of the cephalochordates but prior to the emergence of the agnatha. The biological role of ILP is indicated by the cDNA cloning of the associated receptor from the amphioxus. The ILP receptor is a membrane-bound receptor kinase, and its amino acid sequence is about 49% and 47% identical to that of the human insulin receptor and IGF-1R, respectively. The ILP receptor contains the structural components necessary for binding and activation by mammalian insulin and IGF-I. Consequently, the Steiner group hypothesized that, in analogy to the ILP gene, the ILP receptor gene is an extant representative of an ancestral receptor gene that had duplicated and diverged into the vertebrate insulin and IGF-1R genes during the transition from protochordates to vertebrates.

The previously described prevailing model of insulin and IGF evolution (Fig. 1) that suggests an origin for the IGFs postdating the cephalochordates appears to be sound. However, in 1997, McRory and

Sherwood reported the characterization of distinct prepro-insulin and a prepro-IGF cDNAs cloned from the urochordate *Chelyosoma productum*. The deduced peptides show characteristics of vertebrate insulin and IGF. However, the pronounced amino acid sequence identity of the urochordate insulin and proto-IGF is higher than that with insulin or IGF sequences in other species. In particular, the sequence identity between insulin and proto-IGF in the evolutionary highly variable C domain is very high. As was first argued by Chan and Steiner in 2000, these results somewhat contradict the assumption that the urochordate genes are true orthologues of the vertebrate insulin and IGF genes. Therefore, further studies in various species of protochordates have to be performed to unravel this discrepancy.

### Evolution of the IGF-Binding Proteins

In mammals, each of the six different IGFbps has its own well-characterized properties in delivering the IGFs to their receptors. Some IGFbps show higher IGF-binding affinity, whereas others promote the IGF-receptor interaction. In general, a particular IGFbp can exert both inhibitory and stimulatory actions. These depend on the particular physiological context and target cell. In contrast, our knowledge of the number and nature of the IGFbps in nonmammals is limited, and the results obtained are not consistent. This may in part reflect differences in IGFbps present among the species investigated, but differences in nutritional status and developmental stage or technical differences may also play a role.

In chicken, four circulating IGFbps with molecular weights of 28, 34, 41, and 60 kDa have been identified. Furthermore, the cDNAs encoding two IGFbps have been cloned. The deduced amino acid sequences show 66 and 83% overall sequence identity to mammalian IGFbp-2 and IGFbp-5, respectively. Like in mammals, the chicken 28-kDa IGFbp is up-regulated by fasting. The binding activity of a chicken 36-kDa IGFbp increased after glucocorticoid treatment, as observed for mammalian IGFbp-1. IGFbps sized between 20 and 60 kDa were identified in representatives of all major reptilian groups. In amphibians, a 28.4-kDa IGFbp that is 97% identical to rat and human IGFbp-5 has been found in the frog *Xenopus laevis*, and a single 50-kDa IGFbp has been found in the toad, *Bufo woodhousei*. Simultaneously with the serum IGF-I level, that of the toad IGFbp changed with the seasons.

Like in mammals, only a minor portion of IGF-I circulates in unbound form in bony fish. As estab-

lished by the Dickhoff group in Seattle, Washington, the amount of free IGF-I is 0.3% in coho salmon. Three different IGFbps have been detected in serum of coho salmon, striped bass, goby, tilapia, barramundi, and trout. Based on their molecular weights (23–28, 30–34, and 40–50 kDa, respectively), it has been postulated that the fish IGFbps are related to mammalian IGFbp-1, -2, and -3, respectively. In striped bass, a fourth IGFbp with a molecular weight of 85 to 90 kDa has been detected. Recently, the Duan group first cloned bony fish IGFbp in zebrafish. The zebrafish IGFbp had a molecular size of 28.4 kDa, showed 52% sequence identity with human IGFbp-2, and significantly inhibited IGF-I-stimulated cell proliferation and DNA synthesis.

Like in mammals, IGFbps are produced not only in bony fish liver but also in extrahepatic sites. In striped bass, a 30-kDa IGFbp was found in brain, gills, bone, heart, skeletal muscle, kidney, gut, spleen, and gonads, whereas liver expressed an additional 24-kDa IGFbp. Increasing evidence indicates that the serum levels of the fish IGFbps are regulated by the metabolic and hormonal status, as they are in mammals. In goby, the serum levels of the 29- and the 31-kDa IGFbps depended on the insulin status, and in coho salmon, the serum level of the 41-kDa IGFbp was GH dependent. Similarly, *in vitro* experiments performed in striped bass by the Bern group in Berkeley, California, showed that the modes of release of the 24- and the 30-kDa IGFbps from liver slices into the medium were differentially regulated by GH and other hormones such as insulin and thyroxin. Because fasting in striped bass and goby also affected the lower molecular-weight IGFbps, further functional similarities between fish IGFbps and their mammalian counterparts exist.

Although no IGFbps have been described to date in chondrichthyes, three IGFbps have been detected in the serum of a lamprey species (*Geotria australis*): a major form at 50 kDa and minor forms at 32 and 28 kDa.

In summary, it can be stated that IGFbps likely occur in all vertebrate classes. At least three IGFbps seem to be present: generally a higher molecular-weight IGFbp of 40 to 50 kDa and two smaller IGFbps in the size range of 24 to 34 kDa. Thus, the molecular weights of the nonmammalian IGFbps are within similar ranges as their mammalian counterparts. Furthermore, nonmammalian IGFbps seem to be expressed in liver and extrahepatic sites, and several similarities in the regulation of the IGFbps in mammals and nonmammals, especially in bony fish, are obvious. Therefore, different classes of IGFbps seem to be

conserved among the vertebrate classes, suggesting that the IGF-BPs play fundamental roles in modulating the bioavailability of the IGFs throughout the vertebrate line of evolution. This is also stressed by the likely simultaneous appearance of proto-IGF and IGF-BPs in evolution. Although the ability of the IGF-BPs to inhibit insulin-like activities of the IGFs probably is their most fundamental role, their potential impact in tissue-specific and developmental regulation as it is present in mammals, as well as the evolutionary level of its appearance, still awaits clarification in nonmammalians.

### INSULIN-LIKE PEPTIDES IN THE PROTOSTOMIAN LINEAGE

ILPs occur not only in the deuterostomian line of evolution but also in protostomian invertebrates. A family of insulin-related peptides has been detected in mollusks and the Bombyxins and other ILPs in insects. The organization of the domains of the various ILPs is similar or identical to that of (pro-)insulin and the IGFs, including the six cysteine residues important for maintenance of the secondary and tertiary structure in vertebrate insulin and IGFs.

#### Molluscan Insulin-Related Peptides

Five different members of the molluscan insulin-related peptide (MIP) family have been characterized from the pulmonate snail *Lymnaea stagnalis*. In contrast to insulin and the IGFs, the MIPs seem to be essentially produced in neuroendocrine cells of the central nervous system, the light green cells (LGCs). Because the LGCs are involved in body growth and metabolism, it is assumed that the MIPs play a role in the regulation of growth and differentiation as their mammalian counterparts are. The MIP genes contain introns at positions matching those in the vertebrate insulin genes. The level of sequence similarity in the B and A domains among the members of the MIP family is lower than that among the members of the deuterostomian insulin/IGF family. Thus, the MIPs likely constitute an ancient family of peptides and, as recently put forward by Smit and colleagues, have experienced multiple gene duplications, mutations, unequal crossing-overs, and (partial) gene conversions. In 2001, Lardans and colleagues characterized an insulin receptor-related receptor in an embryonic cell line from the snail *Biomphalaria glabrata*. The receptor had the typical structure of a tyrosine kinase receptor with extracellular, transmembrane, and cytoplasmic regions

and showed a wide distribution throughout the snail body.

#### Insect Insulin-like Peptides

As the first member of this hormone family, Bombyxin, a small insulin-related peptide of about 5 kDa, was isolated from the brain of the silkworm (*Bombyx mori*). The overall sequence identity of Bombyxin A2 with human insulin is 50% in the A domain and 32% in the B domain. It is produced mainly in medial neurosecretory cells of the central nervous system but also occurs in several extracerebral sites. The Bombyxin genes, which have been cloned from the silkworm, constitute five families of structurally related genes. The members of each family are closely related structurally. The presence of multiple Bombyxin gene copies, coupled with the lack of introns, is in marked contrast to the vertebrate insulin and IGF genes. The widespread expression pattern of Bombyxin(s), which resembles that of IGF-I in vertebrates, led to the assumption that the hormone is involved in insect growth, development, and differentiation. Recently, Masamura and colleagues showed that glucose stimulates the release of Bombyxin and that the hemolymph levels of both glucose and Bombyxin are regulated by starvation and refeeding. These data suggest that Bombyxin probably is a metabolic hormone in insects and that glucose may be a common nutritional stimulus for the release of deuterostomian insulin and insect ILPs.

Thus, the role of insulin/ILPs in regulating the switch from the catabolic to the anabolic state, thereby maintaining the blood glucose level, likely evolved before the divergence of the protostomian and deuterostomian lineages of evolution.

The well-characterized ILPs in the fruitfly *Drosophila* (DILPs) also have a domain structure similar or identical to that of mammalian insulin. Their amino acid sequences and domain structure both are more closely related to mammalian insulin than to IGFs. However, the ubiquitous expression of DILP2, which is most similar to insulin, resembles that of IGF-I in vertebrates. The insulin-signaling pathway has been characterized in *Drosophila* and in a nematode. The whole pathway from the receptor down to the serine/threonine kinases is highly conserved in evolution. The *Drosophila* insulin receptor homologue (DInr) has almost equivalent sequence identity to the vertebrate insulin and IGF type 1 receptors, and its overexpression leads to an increase in growth, cell size, and number. Brogiolo

and colleagues recently showed that the *Drosophila* genome contained seven genes encoding ILPs that were expressed in a tissue- and stage-specific manner. Overexpression of one of the ILP genes enhanced growth in a DInr-dependent manner. Based on these results, it has been proposed that nutritionally regulated ILPs in *Drosophila* coordinate growth at the neurosecretory and local levels.

Thus, the protostomian ILPs seem to function as regulators of growth and body size. Indeed, it is likely that the conformation of the protostomian ILPs is similar to that of the deuterostomian insulin and IGFs, having changed little since the two lineages last shared a common acoelomate ancestry.

### Acknowledgments

The work of the author on the phylogeny of insulin and the IGF system was supported by the Swiss National Foundation (grants 32-33349.92, 32-45351.95, 32-61481.00, and 7UKPJ062172.00).

### See Also the Following Articles

ACTH,  $\alpha$ -MSH, and POMC, Evolution of • Angiotensin, Evolution of • Insulin Secretion: Functional and Biochemical Aspects • Insulin-like Growth Factors • Natriuretic Peptide System, Evolution of • Prolactin, Evolution of • Somatostatin, Evolution of • Steroid Receptors, Evolution of

### Further Reading

- Broggiolo, W., Stocker, H., Ikeya, T., Rintelen, F., Fernandez, R., and Hafen, E. (2001). An evolutionarily conserved function of the *Drosophila* insulin receptor and insulin-like peptides in growth control. *Curr. Biol.* **11**, 213–221.
- Chan, S. J., and Steiner, D. F. (2000). Insulin through the ages: Phylogeny of a growth promoting and metabolic regulatory hormone. *Am. Zool.* **40**, 213–222.
- Daughaday, W. H., Hall, K., Raben, M. S., Salmon, W. D., Jr., van den Brande, J. L., and van Wyk, J. J. (1972). Somatomedin: Proposed designation for sulphation factor. *Nature* **235**, 107.
- Duan, C., Ding, J., Li, Q., Tsai, W., and Pozios, K. (1999). Insulin-like growth factor binding protein 2 is a growth inhibitory protein conserved in zebrafish. *Proc. Natl. Acad. Sci. USA* **96**, 15274–15279.
- Gluckman, P. D., and Ambler, G. R. (1993). What is the function of circulating insulin-like growth factor-2 in postnatal life? *Mol. Cell. Endocrinol.* **92**, C1–C3.
- Humbel, R. E. (1990). Insulin-like growth factors I and II. *Eur. J. Biochem.* **190**, 445–462.
- Hwa, V., Oh, Y., and Rosenfeld, R. G. (1999). The insulin-like growth factor-binding protein (IGFBP) superfamily. *Endocrinol. Rev.* **20**, 761–787.
- Jones, J. I., and Clemmons, D. R. (1995). Insulin-like growth factors and their binding proteins: Biological actions. *Endocrinol. Rev.* **16**, 3–34.
- Kelley, K. M., Haigwood, J. T., Perez, M., and Galima, M. M. (2001). Serum insulin-like growth factor binding proteins (IGFBPs) as markers for anabolic/catabolic condition in fishes. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **129**, 229–236.
- Lardans, V., Coppin, J. F., Vicogne, J., Aroca, E., Delcroix, M., and Dissous, C. (2001). Characterization of an insulin receptor-related receptor in *Biomphalaria glabrata* embryonic cells. *Biochim. Biophys. Acta* **1510**, 321–329.
- Le Roith, D. (1997). Insulin-like growth factors. *Sem. Med. Beth. Israel Deaconess Med. Center* **336**, 633–640.
- Liu, J. L., Yakar, S., and LeRoith, D. (2000). Conditional knockout of mouse insulin-like growth factor-1 gene using the Cre/loxP system. *Proc. Soc. Exp. Biol. Med.* **223**, 344–351.
- Masumura, M., Satake, S., Saegusa, H., and Mizoguchi, A. (2000). Glucose stimulates the release of bombyxin, an insulin-related peptide of the silkworm *Bombyx mori*. *Gen. Comp. Endocrinol.* **118**, 393–399.
- McRory, J. E., and Sherwood, N. M. (1997). Ancient divergence of insulin and insulin-like growth factor. *DNA Cell Biol.* **16**, 939–949.
- Mendez, E., Planas, J. V., Castillo, J., Navarro, I., and Gutierrez, J. (2001). Identification of a type II insulin-like growth factor receptor in fish embryos. *Endocrinology* **142**, 1090–1097.
- Mommsen, T. P., and Plisetskaya, E. M. (1991). Insulin in fishes and agnathans: History, structure, and metabolic regulation. *Rev. Aquat. Sci.* **4**, 225–259.
- Navarro, I., Leibush, B., Moon, T. W., Plisetskaya, E. M., Banos, N., Mendez, E., Planas, J. V., and Gutierrez, J. (1999). Insulin, insulin-like growth factor-I (IGF-I), and glucagon: The evolution of their receptors. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **122**, 137–153.
- Plisetskaya, E. M. (1998). Some of my not so favorite things about insulin and insulin-like growth factors in fish. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **121**, 3–11.
- Reinecke, M., and Collet, C. (1998). The phylogeny of the insulin-like growth factors. *Int. Rev. Cytol.* **183**, 1–94.
- Reinecke, M., Schmid, A., Ermatinger, R., and Loffing-Cueni, D. (1997). Insulin-like growth factor I in the teleost *Oreochromis mossambicus*, the tilapia: Gene sequence, tissue expression, and cellular localization. *Endocrinology* **138**, 3613–3619.
- Smit, A. B., van Kesteren, R. E., Li, K. W., Van Minnen, J., Spijker, S., Van Heerikhuizen, H., and Geraerts, W. P. (1998). Towards understanding the role of insulin in the brain: Lessons from insulin-related signaling systems in the invertebrate brain. *Prog. Neurobiol.* **54**, 35–54.

insulin secretion is observed. A high rate of hormonal release is then eventually achieved.

As emphasized in what follows, the insulinotropic action of D-glucose is causally linked to the capacity of the hexose to act as a nutrient and to increase the rate of adenosine 5'-triphosphate (ATP) generation in insulin-producing cells. This fuel concept also applies to the insulin-releasing action of other nutrients mentioned previously. In the case of amino acids, however, the following more nuanced situation prevails. Certain amino acids (e.g., L-leucine, L-glutamine, L-asparagine) also owe their insulinotropic capacity to their role as nutrients in the islet B cells. However, they may also act as allosteric activator of mitochondrial glutamate dehydrogenase and, as a result, facilitate the catabolism of endogenous amino acids in the islet cells. For instance, such is the case for a transported but nonmetabolized analogue of L-leucine. Moreover, other amino acids, which are potent insulin secretagogues, are poor nutrients but stimulate insulin release as a result of their accumulation in the islet cells. Such is the case for cationic amino acids (e.g., L-arginine, L-ornithine), which, through their accumulation inside the islet cells, provoke a depolarization of the plasma membrane and a resulting gating of voltage-sensitive calcium channels.

### Immediate and Direct Effects of Hormones and Neurotransmitters

Two examples of the immediate and direct effects of hormones on insulin release are mentioned here. First, adrenaline inhibits insulin secretion evoked by D-glucose or other insulin secretagogues. This effect is suppressed by  $\alpha$ -adrenergic blocking agents and coincides with a lowering of the cyclic AMP (cAMP) content in islet cells. It was proposed that, by inhibiting insulin release under conditions of stress and exercise, adrenaline would be better able to mobilize D-glucose from the liver and to mobilize free fatty acids and glycerol from adipose tissue.

Second, certain gastrointestinal hormones, such as glucagon-like peptide 1, potentiates insulin release evoked by D-glucose and other nutrient secretagogues. This is attributable to activation of adenylate cyclase, with a resulting increase in the cAMP content of islet cells. The term "hormonal enteroinsular axis" was proposed to indicate that the enhancement of insulin secretion by hormones released from the intestinal tract may account for the fact that hyperinsulinemia caused by D-glucose is more marked when the sugar is given orally rather than intravenously.

Likewise, vagal stimulation of insulin release could be a significant feature in the process of food intake and digestion. Cholinergic agents indeed augment insulin secretion evoked by D-glucose. This process involves occupancy of islet cells' muscarinic receptors, activation of phospholipase C, hydrolysis of polyphosphoinositides, and liberation of diacylglycerol and inositol 1,4,5-triphosphate with resulting activation of protein kinase C and mobilization of intracellular  $\text{Ca}^{2+}$ .

### Long-Term Regulation of Insulin Release

Insulin release evoked by D-glucose is decreased in islets prepared from starved animals. It occurs at an abnormally high rate in islets from obese animals. Several hormones also affect the secretory responsiveness of the endocrine pancreas in a delayed manner. This may account for alteration of insulin release in situations such as pregnancy and lactation, acromegaly, hypo- or hypercorticism, and hypo- or hyperthyroidism. Finally, the long-term regulation of insulin release also involves ontogenic factors, as documented by the changes in islet secretory behavior found in fetal and neonatal life or aging.

### RECOGNITION OF SECRETAGOGUES

As alluded to previously, the B cell is equipped with a number of receptors involved in the recognition of adrenergic factors, cholinergic neurotransmitters, and gastrointestinal hormones. By analogy, it had been proposed that each nutrient secretagogue may bind to a specific receptor possibly located at the B-cell plasma membrane. For instance, a previously widespread view postulated the presence of a stereospecific glucoreceptor in islet cells. However, a more pedestrian metabolic hypothesis eventually prevailed, postulating that the insulinotropic capacity of D-glucose is causally linked to its role as a nutrient in islet cells. This was then extended to other nutrient secretagogues and defined as the fuel concept for insulin release.

The key argument in support of this metabolic hypothesis for glucose-stimulated insulin secretion resides in the finding that the  $\alpha$ -anomer of D-glucose, which is a more potent insulin secretagogue than  $\beta$ -D-glucose, is also more efficiently metabolized in pancreatic islets. As a matter of fact, the metabolism of D-glucose in isolated pancreatic cells or in purified islet B cells displays a number of uncommon features, all of which optimize the increase in ATP generation

rate resulting from the stimulation of islet cells by this hexose. A few examples of such features follow.

First, the concentration of D-glucose rapidly equilibrates across the plasma membrane of the B cell thanks to the efficiency of the glucose carrier GLUT2 (and GLUT1 in human B cells). Second, the phosphorylation of D-glucose in islet B cells is catalyzed not solely by an ubiquitous low- $K_m$  hexokinase but also by a high- $K_m$  glucokinase. As a result of these two features, the rate of D-glucose phosphorylation increases rapidly and sensibly in response to a rise in the extracellular concentration of the hexose. The regulation of D-glucose phosphorylation in the B cell also involves a number of other mechanisms such as a sequential synarchistic regulation by ATP, induction and repression of glucokinase, feedback inhibition of hexokinase by glucose 6-phosphate, intracellular translocation of glucokinase, intervention of a glucokinase regulatory protein, balance between glucose phosphorylation and dephosphorylation (by glucose-6-phosphatase), and binding of hexokinase isoenzymes to mitochondrial porin, with this process modifying kinetic properties (e.g., sensitivity to glucose 6-phosphate), coinciding with a preferential use of mitochondrial ATP for phosphorylation of D-glucose, and allowing a direct coupling between hexose phosphorylation and mitochondrial respiration.

Second, for the rate of glycolysis to keep pace with the rate of D-glucose phosphorylation, activation of key glycolytic enzymes takes place in glucose-stimulated B cells. For instance, such is the case for the activation of phosphofructokinase by fructose 2,6-bisphosphate. Incidentally, the regulation of phosphofructokinase activity in islet cells also displays the specific feature of an apparent resistance of fructose 6-phosphate,2-kinase to cAMP.

Finally, a preferential stimulation of mitochondrial oxidative events in either isolated islets or purified B cells exposed to increasing concentrations of D-glucose represents an essential feature of the B-cell glucose-sensing device. This phenomenon is attributable mainly to a  $Ca^{2+}$ -induced activation of key mitochondrial dehydrogenases such as the FAD-linked glycerophosphate dehydrogenase and 2-ketoglutarate dehydrogenase complex. Therefore, it should be considered as a secondary phenomenon caused by the accumulation of  $Ca^{2+}$  in the B cells. It optimizes the yield of ATP generated through the catabolism of D-glucose given that the mitochondrial oxidation of pyruvate provides the major fraction of such an ATP generation and is necessarily coupled with the oxidative modality of glycolysis.

## COUPLING MECHANISMS

In the process of nutrient-stimulated insulin release, the coupling between metabolic events and more distal events in the secretory sequence appears to be multifactorial. Emphasis is currently placed on the closing of ATP-sensitive  $K^+$  channels caused by the increase in the cytosolic ATP/adenosine diphosphate (ADP) ratio in nutrient-stimulated B cells. The closing of these channels then leads to depolarization of the plasma membrane and the subsequent opening of voltage-sensitive  $Ca^{2+}$  channels. The facilitated entry of  $Ca^{2+}$  into the B cell and the cytosolic accumulation of this divalent cation then acts as a trigger for the release of secretory granules.

A stimulatory effect of D-glucose on insulin release can still be documented, however, in cells exposed to both diazoxide (to maintain the  $K^+$  channels in their open configuration) and a high extracellular  $K^+$  concentration (to nevertheless cause depolarization of the plasma membrane). This finding points to the participation of other mechanisms in the stimulus-secretion coupling for glucose-induced insulin release. In this perspective, the following considerations should be underlined.

First, an increase in the ATP generation rate might not represent the sole factor coupling the catabolism of nutrient secretagogues to the remodeling of ionic fluxes in the islet cells. For instance, nutrient-induced changes in the redox state of islet cells and their intracellular pH may also participate in nutrient-stimulated insulin release.

Second, the cytosolic accumulation of  $Ca^{2+}$  may lead, via calmodulin, to the activation of adenylate cyclase in pancreatic islets. Pancreatic islets indeed contain calmodulin (about 0.1 pmol per islet). This protein binds to a particulate fraction derived from the islets and stimulates adenylate cyclase activity in this subcellular fraction, with both phenomena being activated by ionized calcium. Thus, a calcium-dependent stimulation of adenylate cyclase by endogenous calmodulin may contribute to the accumulation of cAMP evoked by insulin-releasing agents in the islet cells. cAMP should be considered as a modulator of the insulin secretory response to nutrient secretagogues. Its synthesis and breakdown are catalyzed by adenylate cyclase and phosphodiesterase, respectively. cAMP activates a protein kinase, leading to an apparent increase in the responsiveness of the effector system for insulin release to cytosolic  $Ca^{2+}$ .

Finally, in nutrient-stimulated B cells, as in response to cholinergic agents, breakdown or otherwise accelerated turnover of inositol phospholipids takes

place. The accelerated generation of polyphosphoinositides in nutrient-stimulated islet cells may result from a rise in cytosolic ATP production, whereas the stimulation by nutrients of the hydrolysis of inositol-containing phospholipids is currently ascribed to activation of phospholipase C through an increase in the cytosolic  $\text{Ca}^{2+}$  concentration. Enhanced phospholipid metabolism may then lead to the mobilization of calcium from nonmitochondrial intracellular stores by inositol 1,4,5-triphosphate and activation of protein kinase C by diacylglycerol.

## EFFECTOR SYSTEM

The increase in the cytosolic concentration of ionized calcium caused by D-glucose or other insulin secretagogues triggers the exocytosis of secretory granules by causing the activation of a microtubular-microfilamentous system. The participation of this effector microtubular-microfilamentous system in the process of insulin release has been documented by a number of ultrastructural, biochemical, functional, and cinematographical studies.

Microtubules (21–25 nm in diameter) are considered to represent the cytoskeleton of B cells. They are scattered in the cytoplasm. They can be found between rows of aligned secretory granules. They are thought to provide oriented pathways for back-and-forth saltatory movements of secretory granules. They are also prominent in the ectoplasmic area. At that level, the cell web occupies cytoplasmic areas of variable thickness (50–300 nm) just beneath the plasma membrane and extends into the core of microvillous processes. It consists of a network of actin-like microfilaments (4–7 nm in diameter) generally disposed to form irregularly shaped polygons. This contractile cell web acts as a sphincter, either restricting or favoring the access of secretory granules to their exocytotic site at the plasma membrane.

The tools most commonly used to interfere experimentally with the function of the microtubular-microfilamentous system include mitotic spindle inhibitors (e.g., colchicine, vincristine), microtubule stabilizers (e.g.,  $\text{D}_2\text{O}$ ), and the microfilamentous modifier cytochalasin B.

At the exocytotic site, the fusion between the limiting membrane of insulin secretory granules and plasma membrane and subsequent fission of the fused membrane seem to involve an anion-osmotic process in which the insertion at the plasma membrane of an anion transport system derived from the limiting membrane of the secretory granules allows for the

entry of anions (e.g.,  $\text{Cl}^-$ ,  $\text{OH}^-$ ) into the lumen of these granules, followed by their osmotic fission. This process also accounts for the release of several granules aligned along an oriented microtubular pathway, a phenomenon known as the chain release of secretory granules or compound exocytosis.

## PATHOLOGICAL ASPECTS

Defective insulin release is a typical feature of diabetes mellitus. In insulin-dependent diabetes, it results from an autoimmune destruction of insulin-producing cells. In non-insulin-dependent diabetes, the deficiency of insulin release may be caused by a number of distinct site-specific anomalies. Indeed, virtually each step in the process of insulin secretion may be affected by inherited or acquired factors. For instance, a defect in the conversion of proinsulin to insulin and C-peptide may result in the release of large amounts of proinsulin, which is virtually devoid of hypoglycemic action. On the most distal step of the secretory sequence, a defect of the microtubular apparatus may be responsible for a sluggish secretory response to D-glucose in certain animal models of type 2 diabetes.

In most but not all patients with type 2 diabetes, the endocrine pancreas displays a preferential impairment of its secretory responsiveness to D-glucose, as distinct from other nutrient or nonnutrient insulin secretagogues. Site-specific defects responsible for such a situation include (1) a decrease in the number of B cells suitably equipped with sufficient GLUT2 carriers, (2) a nonsense mutation of the glucokinase gene, (3) excessive activity of glucose-6-phosphatase leading to an ATP-wasting futile cycle in the reaction catalyzed by hexokinase isoenzymes and glucose-6-phosphatase, and (4) a deficient activity of the mitochondrial FAD-linked glycerophosphate dehydrogenase, the key enzyme of the glycerol phosphate shuttle. Moreover, glycogen accumulation takes place in the B cell in situations of chronic hyperglycemia. It is held responsible for the secondary process of so-called B-cell glucotoxicity or incompetence. Two typical phenomenological aspects of this process consist in a paradoxical early and transient decrease in insulin release and the perturbation of its anomeric specificity in response to the rapid intravenous administration of D-glucose to patients with type 2 diabetes.

Excessive secretion of insulin can also occur in certain pathological conditions. Such is the case, for instance, in individuals bearing an insulinoma or in

patients with persistent hyperinsulinemia during childhood.

## PHARMACOLOGICAL ASPECTS

The therapeutic agents currently used to modify insulin secretion belong to two classes. First, to enhance insulin release in non-insulin-dependent diabetic individuals, either hypoglycemic sulfonylureas (e.g., tolbutamide, glibenclamide) or meglitinide analogues (e.g., nateglinide, repaglinide) are given orally. These antidiabetic agents display a common configuration and act by causing a direct closing of ATP-sensitive  $K^+$  channels without affecting the ATP content of insulin-producing cells.

Second, to prevent excessive insulin secretion in certain pathological conditions, diazoxide and related drugs are used. They cause a direct opening of ATP-sensitive  $K^+$  channels and, hence, inhibit insulin release caused by D-glucose and other secretagogues. This effect of diazoxide is abolished in the presence of hypoglycemic sulfonylureas or meglitinide analogues.

The destruction of insulin-producing cells (e.g., in patients bearing an insulinoma) can be achieved by selective  $\beta$ -cytotoxic agents, especially streptozotocin.

## See Also the Following Articles

Diabetes, Type 2 • Hyperosmolar Nonketotic Hyperglycemia • Hypoglycemia • Insulin and Insulin-like Growth Factors, Evolution of • Insulin Secretion, Physiology • Insulin-Resistant States, Role of Free Fatty Acids (FFA)

## Further Reading

Flatt, P., and Lenzen, S. (1994). "Frontiers of Insulin Secretion and Pancreatic B-Cell research." Smith-Gordon, London.  
 Henquin, J. C. (2000). Triggering and amplifying pathways of regulation of insulin secretion by glucose. *Diabetes* **49**, 1751–1760.

Howell, S. L., and Tyhurst, M. (1984). Insulin secretion: The effector system. *Experientia* **40**, 1098–1105.  
 Hutton, J. C. (1984). Secretory granules. *Experientia* **40**, 1091–1098.  
 Lacy, P. E., and Greider, M. H. (1972). Ultrastructural organization of mammalian pancreatic islets. In "Handbook of Physiology" (D. F. Steiner and N. Freinkel, eds.), pp. 77–89. American Physiological Society, Washington, DC.  
 Malaisse, W. J. (1972). Hormonal and environmental modification of islet activity. In "Handbook of Physiology" (D. F. Steiner and N. Freinkel, eds.), pp. 237–260. American Physiological Society, Washington, DC.  
 Malaisse, W. J. (1973). Insulin secretion: Multifactorial regulation for a single process of release. *Diabetologia* **9**, 167–173.  
 Malaisse, W. J. (1986). Stimulus–secretion coupling in the pancreatic B-cell: The cholinergic pathway for insulin release. *Diab./Metab. Rev.* **2**, 243–259.  
 Malaisse, W. J. (1987). Insulin release: The glucoreceptor myth. *Med. Sci. Res.* **15**, 65–67.  
 Malaisse, W. J. (1992). Glucose-sensing by the pancreatic B-cell: The mitochondrial part. *Int. J. Biochem.* **24**, 693–701.  
 Malaisse, W. J. (1994). The beta cell in non-insulin-dependent diabetes: Giving light to the blind. *Diabetologia* **37**(Suppl. 2), S36–S42.  
 Malaisse, W. J. (1996). Metabolic signalling of insulin secretion. *Diab. Rev.* **4**, 145–159.  
 Malaisse, W. J., Herchuelz, A., Devis, G., Somers, G., Boschero, A. C., Hutton, J. C., Kawazu, S., Sener, A., Atwater, I. J., Duncan, G., Ribalet, B., and Rojas, E. (1978). Regulation of calcium fluxes and their regulatory roles in pancreatic islets. *Ann. N.Y. Acad. Sci.* **307**, 562–582.  
 Malaisse, W. J., Malaisse-Lagae, F., and Leclercq-Meyer, V. (1988). Anomeric specificity of hexose metabolism in health and disease. *Netherlands J. Med.* **33**, 247–256.  
 Malaisse, W. J., and Orci, L. (1979). The role of cytoskeleton in pancreatic B-cell function. In "Methods of Achievements in Experimental Pathology" (G. Gabbiani, ed.), Vol. 9, pp. 112–136. S. Karger, Basel, Switzerland.  
 Prentki, M., and Wollheim, C. B. (1984). Cytosolic free  $Ca^{2+}$  in insulin secreting cells and its regulation by isolated organelles. *Experientia* **40**, 1052–1060.  
 Samols, E. (1991). "The Endocrine Pancreas." Raven, New York.  
 Schuit, F. C., Huypens, P., Heimberg, H., and Pipeleers, D. G. (2001). Glucose-sensing in pancreatic  $\beta$ -cells: A model for the study of other glucose-regulated cells in gut, pancreas, and hypothalamus. *Diabetes* **50**, 1–11.

in a linear fashion through the range of physiological glucose concentrations seen in the postprandial state. Maximal insulin output is achievable *in vitro*, but has been difficult to demonstrate in human subjects.

The glucose–insulin relationship is also altered by previous exposure of the beta cell to glucose, which can result in two important consequences, dependent on the duration of exposure. The first is a phenomenon that provides evidence for beta cell memory—if the beta cell is exposed to a prior glucose stimulus, repeat glucose stimulation results in a magnified insulin secretory response. This response is influenced by the magnitude and duration of the prior stimulus, as well as its proximity in time to a subsequent glucose exposure. The second phenomenon occurs after prolonged exposure to elevated glucose concentrations, which results in diminished insulin secretion, so-called “glucotoxicity.” Glucotoxicity has been associated with reduced expression of beta cell genes in experimental *in vitro* models; however, the mechanisms for both of these time- and concentration-dependent effects are not well understood. Nevertheless, the pathways that explain both beta cell memory and glucotoxicity are potential regulatory targets for therapeutic action.

### ***In Vitro* Insulin Secretion**

Distinct patterns of first- and second-phase insulin secretion are clearly observed *in vitro* in models that range from isolated islets of Langerhans to the perfused pancreas. This very characteristic pattern, which has been considered the hallmark of the physiological insulin response to glucose, has been shown to be important in two ways. First, as a marker for normal beta cell function—the loss of first-phase insulin secretion has been used experimentally, and clinically to a lesser extent, to evaluate people at risk for the development of type 1 and type 2 diabetes. Second, first-phase secretion has been shown to be necessary for the normal regulation of hepatic glucose production. Thus, attempts to understand the mechanisms for this typical response to glucose stimulation continue to preoccupy investigators. Two different models have been suggested to explain the first-phase response, a “storage-limited model” and a “signal-limited model.” In the first model, a signal releases a distinct and limited group of secretion granules to form the first phase. In the second model, the signal(s) is (are) biphasic, resulting in a secretory pattern that reflects the cumulative signal(s). These models are not mutually exclusive. Direct observation of insulin secretion granules in beta cells supports the concept

of an immediately releasable pool of granules found close to the plasma membrane. This appears to be a well-defined cluster of granules that are set to fuse or may already be fused with the plasma membrane, the first step in the exocytosis of insulin molecules.

### ***In Vivo* Insulin Secretion**

#### **Complexity of Assessment of Insulin Secretion *in Vivo***

The simplest measure is the circulating insulin concentration, which provides useful information but is limited in its ability to accurately reflect insulin secretion by the pancreas. This is because peripheral circulating insulin, which is the usual measurement obtained in clinical settings and small animal models, is influenced by a number of potential confounders. Three physiological phenomena play a role in influencing peripheral plasma insulin concentrations.

#### **Extraction**

The liver extracts 50–60% of the insulin presented to it from the pancreas in a single pass. Thus, peripheral insulin concentrations are markedly diminished compared with the amount that is secreted into the prehepatic circulation.

#### **Dilution**

Peripheral insulin concentrations are further decreased by the dilution that occurs after the entry of insulin into the portal system from the pancreas and again as the hepatic effluent joins the general circulation. This further alters the quantitative relationship between insulin concentrations in the periphery and their pancreatic source.

#### **Clearance**

Insulin is metabolized by the liver and also the kidneys. As with all circulating molecules, serum concentrations are a function of both the rate of secretion and the rate of clearance. Thus, impaired insulin clearance will increase circulating insulin concentrations. Although decreased clearance of insulin occurs in pathological conditions such as renal failure, it is also a consequence of insulin resistance in otherwise healthy subjects.

#### **Methods to Evaluate Insulin Secretion *in Vivo***

The methods that have been developed to evaluate insulin secretion in whole animals are based on dealing with the three problems outlined above. C-peptide is cosecreted with insulin in equimolar quantities from

the beta cell and is therefore an excellent marker for insulin secretion, especially because this peptide is not extracted by the liver (metabolism of C-peptide occurs predominantly in the kidney). Thus, measurement of C-peptide allows researchers to bypass the effect of the liver in insulin extraction and provides a more direct measurement of beta cell function. Clearance must be addressed as well, all the more so because C-peptide clearance is different from that of circulating insulin. This is factored into the assessment of insulin secretion using a mathematical solution known as deconvolution, first introduced by Eaton and elegantly applied by Polonsky, to distinguish between the secretion and clearance components of circulating C-peptide. The mathematical treatment of insulin secretion assessment also takes into account the dilution phenomenon that takes place after hormone release. This work and similar approaches have shown that the average insulin secretory rate in lean, healthy individuals is close to 30 units per day.

## MODULATORS OF INSULIN SECRETION

### The Local Milieu: Influence of Other Islet Cells

The beta cell is surrounded by other endocrine cells in the islets of Langerhans, which have the distinction of also being involved in the regulation of metabolism and uniquely—as first demonstrated by Samols—with the potential ability to influence the secretion of their neighbors. The alpha and delta cells are the next largest contributors to islet cell mass, secreting glucagon and somatostatin, respectively. Both glucagon and somatostatin modulate insulin secretion: glucagon is an insulin secretagogue and somatostatin is an inhibitor of insulin secretion. It is not clear whether these modulatory effects are achieved locally by paracrine cell-to-cell interactions or by a local endocrine effect when the hormones are released into the islet microcirculation and transported within this vasculature from their secretory origin to target cells. Finally, these hormones might influence each other by the classical endocrine route, by entering the general circulation before returning to the islet and affecting their neighboring cells.

### Endocrine Modulators: Incretins

The complexity and elegance of normal physiology are illustrated further by the incretin phenomenon.

An entero-insular axis was hypothesized when it was found that glucose stimulation of insulin secretion was increased if the same glucose concentrations were achieved by glucose ingestion rather than intravenous infusion. This indicated that enteric factors, humoral (incretins) or nervous, enhance the beta cell response to glucose or meals; indeed, the incretins can be seen as a mechanism to prepare the islets for an influx of nutrients yet to be absorbed from the gastrointestinal tract. Numerous enteric peptides have been considered as candidates for the role as the physiological incretin. These include gastrin, cholecystokinin, glucose-dependent insulinotropic polypeptide (GIP), and glucagon-like peptide-1 (GLP-1). GLP-1, a peptide secreted by gastrointestinal L cells, is the prime candidate and is the basis for a new class of anti-diabetic agents being developed to treat type 2 diabetes. GLP-1-like agents or those that slow the clearance of the endogenous peptide have hypoglycemic activity by enhancing glucose-stimulated insulin secretion and suppressing glucagon secretion.

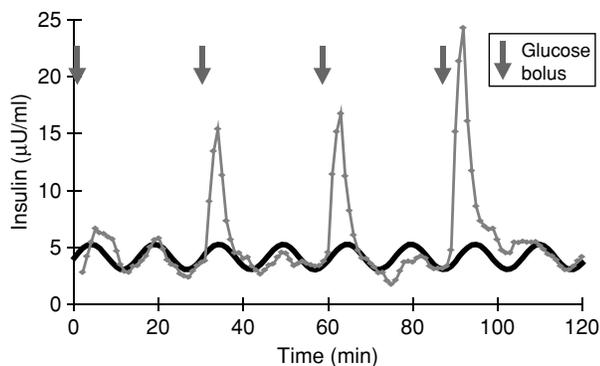
### Nutrient and Neural Modulators

Nonglucose nutrients also modulate insulin secretion in the presence of glucose in the medium or circulation. These include essential amino acids such as arginine and leucine. Free fatty acids (FFA) also have a complex interaction with beta cell function, not dissimilar to that of glucose. Acute elevation of FFA (or ketones) increases insulin secretion, but chronic exposure appears to diminish insulin release. Multiple neural modulators also regulate insulin secretion. The islets of Langerhans are densely innervated. Insulin secretion is enhanced by cholinergic stimulation; adrenergic input inhibits beta cell function. Both  $\alpha$ 1- and  $\alpha$ 2-adrenergic receptors mediate the inhibition of insulin secretion. There is also evidence that many neuropeptides, such as galanin, neuropeptide Y, and vasoactive intestinal polypeptide, may play a role in modulating insulin secretion.

## PATTERNS OF INSULIN SECRETION

The study of the patterns of insulin secretion have contributed to the understanding of the normal physiology of the beta cell and also provide clinicians with tools with which to define dysfunction. An example of this is the characteristic loss of first-phase insulin secretion in both patients with type 2 diabetes and patients in the earliest phases of type 1 diabetes, prior to clinical presentation. Another characteristic feature

of normal beta cell function is the oscillatory pattern of insulin secretion. Unlike many other hormones secreted in pulsatile fashion, insulin secretion occurs in a highly regular pattern under basal conditions. A variety of patterns of insulin secretion have been identified; their characteristics depend on the experimental model and the frequency of sampling (more frequent sampling is necessary to identify high-frequency oscillations). In isolated islets, the period of insulin oscillations ranges from 3 to 15 min and the oscillations are of low amplitude. In human subjects, high-frequency oscillations in insulin concentrations measured in the peripheral circulation also have a period of 6–12 min (Fig. 1). These are highly regular oscillations that are entrainable by glucose both *in vitro* and *in vivo*. *In vivo*, it has also been shown that insulin oscillations are entrainable by nonglucose secretagogues. Low-frequency oscillations (period is 80–150 min) have also been identified in humans. These ultradian insulin oscillations are superimposed on the large postprandial increases in insulin secreted after meals. Both low- and high-frequency insulin oscillations have been shown to lose their regularity in people with diabetes and indeed



**Figure 1** Three-minute moving average of plasma insulin concentrations in a normal subject injected with serial incremental intravenous glucose boluses at 29 min intervals (arrows). The doses of glucose range between 10 and 40 mg/kg. The data are detrended and presented with a fitted sine wave (using the frequency and amplitude of the dominant 15 min period derived from spectral analysis). This figure illustrates a number of points made in this article: (1) High-frequency insulin oscillations are intrinsic to normal human subjects in the fasting state. (2) There is an insulin response to acute elevation in glucose concentrations; even small increases in glucose concentrations result in dramatic increases in insulin secretion in normal subjects that increase in a dose- or concentration-response pattern. (3) There is an entrainment of insulin oscillations by the glucose boluses to the rhythm of the injections. The period for oscillations in this model is usually  $\approx 10$  min. In the example shown here, the oscillations are entrained to a period of 15 min in this normal subject.

both lose their ability to be entrained by glucose in type 2 diabetic subjects.

## RELATIONSHIP BETWEEN INSULIN SECRETION AND INSULIN RESISTANCE

### Role of Insulin Resistance in Regulating Insulin Secretion

Circulating insulin concentrations are greatly influenced by the concomitant degree of insulin sensitivity *in vivo*. It has been known for years that obese subjects who are insulin resistant have elevated serum insulin concentrations. This has led to the concept that insulin resistance must be associated with a signal to the beta cell that enhances insulin secretion. The mechanism for this signal is a subject of debate, though a simple formulation would place glucose at the center of this physiological phenomenon, in the following way: start with a given insulin concentration and effectiveness in an *in vivo* model and increase the level of insulin resistance in this model. Insulin action will be diminished and as a consequence glucose concentrations will rise slightly. Even small changes in glucose concentrations enhance insulin secretion, thus increasing circulating insulin levels and restoring insulin action so that glucose concentrations return toward baseline. A new steady state with hyperinsulinemia is established. This hyperinsulinemia, though a somewhat crude measure of insulin resistance, has been used by epidemiologists as an effective marker of insulin resistance. Despite the logic and simplicity of this hypothesis, data in dogs suggest that glucose may not be the signal for the hyperinsulinemia of insulin resistance. This important question remains to be resolved.

### Assessment of Insulin Secretion *in Vivo* in the Presence of Insulin Resistance

As pointed out above, the close tie-in that exists between the beta cell and the periphery where insulin resistance is played out alters the way in which researchers look at insulin secretion. As an example, consider a prospective study of the physiology of insulin secretion in two subjects, one of whom gained weight after the start of the study. The development of obesity will result in increased insulin resistance and thus also insulin secretion. How could the subjects' insulin secretion rates be compared and how could possible beta cell dysfunction in the obese subject,

who is now at greater risk for diabetes, be tested for? A decrease in insulin secretion as a marker of beta cell dysfunction would not be obvious in the more resistant person if one looked only at the absolute insulin secretion rates or insulin concentrations—these would likely be higher than those of the lean subject even as secretion is beginning to fail. This problem can be dealt with by taking into account the degree of insulin resistance at each time point in the study when insulin secretion is measured. Such an approach was developed by Kahn and Bergman and is known as the disposition index. The disposition index is the product of insulin secretion and insulin sensitivity. The disposition index deals with the problem of evaluating insulin secretion in the presence of different degrees of insulin resistance and allows comparison of insulin secretion in the example described here.

### Acknowledgement

This work was supported in part by MO1-RR-00425 from the NIH.

### See Also the Following Articles

Glucagon and Glucagon-like Peptides • Glucose Physiology, Normal • Glucose Toxicity • Insulin Action, Post-Receptor

Mechanisms • Insulin and Insulin-like Growth Factors, Evolution of • Insulin Secretion: Functional and Biochemical Aspects • Insulin-Resistant States, Role of Free Fatty Acids (FFA)

### Further Reading

- Daniel, S., Noda, M., Straub, S. G., and Sharp, G. W. (1999). Identification of the docked granule pool responsible for the first phase of glucose-stimulated insulin secretion. *Diabetes* **48**, 1686–1690.
- Grodsky, G. (1975). The kinetics of insulin release. In "Handbook of Experimental Pharmacology" (A. Hasselblatt and F. Bruchhausen, (eds.), Vol. 32, pp. 1–19. Springer-Verlag, Berlin, Germany.
- Kahn, S. E., Prigeon, R. L., McCulloch, D. K., Boyko, E. J., Bergman, R. N., Schwartz, M. W., Neifing, J. L., Ward, W. K., Beard, J. C., Palmer, J. P., and Porte, D., Jr. (1993). Quantification of the relationship between insulin sensitivity and  $\beta$ -cell function in human subjects: Evidence for a hyperbolic function. *Diabetes* **42**, 1663–1672.
- Mao, C. S., Berman, N., Roberts, K., and Ipp, E. (1999). High frequency plasma insulin oscillations are entrainable in normal subjects but not in NIDDM patients. *Diabetes* **48**, 714–721.
- Nesher, R., and Cerasi, E. (2002). Modeling phasic insulin release: Immediate and time-dependent effects of glucose. *Diabetes* **51**, S53–S59.
- Polonsky, K. S., Given, B. D., and Cauter, E. (1988). Twenty-four hour profiles and pulsatile patterns of insulin secretion in normal and obese subjects. *J. Clin. Invest.* **81**, 442–448.

reduced, although not to the same extent as is body weight. Whether there are functional deficiencies in these tissues remains to be determined.

To determine the role of serum IGF-I versus the autocrine/paracrine IGF-I in growth and development, liver IGF-I-deficient (LID) mice were generated. Tissue-specific deletion of the IGF-I gene in the liver did not affect postnatal growth and development, either because the residual serum IGF-I was enough to promote growth or because locally produced IGF-I contributes to growth. The ternary complex that carries IGF-I (and IGF-II) in the serum consists of an acid labile subunit (ALS), IGF binding protein-3 (IGFBP-3), and IGF. When LID mice were crossed with ALS nullizygous mice, serum IGF-I levels were reduced by more than 90%, probably because ALS is required to stabilize IGF-I in serum, and the mice exhibited significant growth retardation. This result suggests that there is a minimum threshold of serum IGF-I that must be maintained to support normal growth. Alternatively, or in addition, LID/ALSKO mice may be deficient in delivery of serum IGF-I to tissues. The role of local tissue IGF-I production will be more clearly elucidated as tissue-specific knockouts in extrahepatic tissues are developed.

### Insulin-like Growth Factor-II

Homozygous deletion of the IGF-II alleles in mice results in 40% growth retardation from embryonic day 11 through adulthood. This result is consistent with the concept that IGF-II is primarily a fetal growth factor, an idea initially suggested by the fact that serum IGF-II decreases dramatically during the early postnatal period in rodents. The findings of growth retardation throughout life are also of interest in that IGF-II knockout mice do not show “catch-up” growth, suggesting an enduring effect of embryonic growth retardation. As with IGF-I, IGF-II expression is regulated by tissue-specific tropic agents, including follicle-stimulating hormone (FSH) in ovarian granulosa cells; ACTH in fetal adrenal cells; glucocorticoids, thyroid hormone, and glucose in hepatic cells; and glucose in a pancreatic beta cell line. IGF-II produced in muscle also plays an important role in skeletal muscle myoblast differentiation *in vitro*. In the circulation, IGF-II is normally a 67-amino acid polypeptide. However, patients with nonislet cell tumor hypoglycemia occasionally release a larger precursor form of IGF-II from the tumors that contains a 21-amino acid extension (E peptide). This so-called “big IGF-II” may cause hypoglycemia in these patients by

preventing IGFBPs from neutralizing circulating IGFs. As a result, there is sufficient free big IGF-II to bind to and activate insulin receptors, resulting in fasting hypoglycemia.

## RECEPTORS

### The Insulin-like Growth Factor-I Receptor

The IGF-I receptor (IGF-IR) is encoded by a single gene greater than 100 kilobases long that contains 21 exons. The structure of the IGF-IR closely resembles that of the insulin receptor (IR). In both cases, the gene encodes a single precursor protein in which the amino-terminal  $\alpha$ -subunit is followed by the carboxy-terminal  $\beta$ -subunit. During maturation of the precursor, the  $\alpha$ - and  $\beta$ -subunits are proteolytically separated and then joined together by disulfide bonds. Two such  $\alpha$ - $\beta$  hemireceptors are attached to each other via disulfide bonding of the  $\alpha$ -subunits. Thus, the mature IGF-IR is a heterotetramer, consisting of two extracellular  $\alpha$ -subunits attached to each other, with each  $\alpha$ -subunit attached to a membrane-spanning  $\beta$ -subunit. The  $\alpha$ -subunit binds the ligands extracellularly. The  $\beta$ -subunit consists of the juxtamembrane region, containing sequence motifs that are also found in the IR and that bind to important intracellular substrates, followed by the tyrosine kinase domain. There is approximately 84% similarity between the amino acid sequences of the IGF-IR and IR tyrosine kinase regions. The most divergent region between the two receptors is the cytoplasmic carboxyl-terminus domain of the  $\beta$ -subunit.

### IGF-IR Function

When IGFs bind to the IGF-IR, receptor autophosphorylation of the  $\beta$ -subunit and activation of its tyrosine kinase occur. A family of molecules called insulin receptor substrates (IRS 1–4) bind to the juxtamembrane region of the  $\beta$ -subunit of the ligand-activated receptor and become phosphorylated on tyrosine residues. The phosphorylated IRS molecules then bind to and activate SH2 domain-containing proteins, including Grb2 and PI3K. The adapter protein Grb2 then activates the SOS/Ras/Raf/mitogen-activated protein (MAP) kinase pathway. MAP kinase, a major IGF-I-activated isoform that is termed extracellular signal-regulated protein kinase (ERK), and PI3K activate phosphorylation and dephosphorylation cascades, leading to changes in levels of proteins important in DNA synthesis and cell division. IGF-I increases cyclin D1 levels in a number of cell types by mechanisms such as increased cyclin D1 gene

transcription and increased cyclin D1 mRNA stability, with the latter being effected via a PI3K-dependent mechanism. IGF-I also increases cyclin E expression. In addition, IGF-I decreases the levels of p27 cyclin-dependent kinase inhibitor through PI3K, an effect that could be mediated at several levels, including posttranscriptional control. IGF-I also increases the levels of cdk-2. Through these combined actions, IGF-I speeds the transition from G1 to S in the cell cycle. In addition to IRS molecules, IGF-IR can bind to and tyrosine phosphorylate Shc, which (much like IRS) can activate Grb-2.

IGF-I also inhibits cellular apoptosis by (1) increasing phosphorylation of BAD, (2) increasing phosphorylation of the forkhead transcription factor, and (3) increasing transcription of the anti-apoptotic bcl-2 gene. Mechanisms for these actions of IGF-I include stimulation of ERK, PI3K, and 14-3-3 pathways to promote BAD phosphorylation; stimulation of PI3K/Akt to increase forkhead phosphorylation; and stimulation of both the p38 stress-activated protein kinase (a MAP kinase isoform) and PI3K/Akt pathways, leading to cyclic AMP response element-binding protein (CREB) phosphorylation and increased bcl-2 promoter activity.

Activation of the PI3K pathway has also been shown to mediate certain effects of IGF-I on cellular differentiation. In brown adipose cells, IGF-I induces expression of uncoupling protein and GLUT4 by stimulating PI3K and ERK, respectively. IGF-I promotes osteoblast differentiation, as evidenced by IGF-I activation of the osteocalcin promoter in a protein kinase C (PKC)-dependent mechanism. More recently, IGF-I was found to stimulate myoblast differentiation by promoting calcineurin signaling. Moreover, IGF-I induction of p21, a cdk inhibitor, was shown to be involved in myoblast differentiation.

### Mannose-6-Phosphate/IGF-II Receptor

The IGF-II receptor (IGF-IIR) is characterized by its ability to bind IGFs as well as mannose-6-phosphate (M6P) residues on lysosomal enzymes. Its affinity for IGF-II is much greater than that for IGF-I, and unlike the IGF-IR, the IGF-IIR does not bind insulin at all. This receptor does not appear to transduce IGF-II signals *in vivo*. Rather, it is thought to actually decrease IGF-II signaling by internalizing IGF-II and promoting IGF-II degradation. Supporting this view, M6P/IGF-IIR gene-deleted mice are larger than controls. Consistent with a growth inhibitory role, microsatellite instability in the M6P/IGF-IIR gene

has been described in various cancers, including stomach, colon, and endometrium. Thus, the M6P/IGF-IIR may play a significant role as a tumor suppressor by mediating IGF-II degradation. The IGF-IIR also activates transforming growth factor (TGF)- $\beta$ 1 from a precursor molecule, which could also lead to growth inhibition.

### INSULIN-LIKE GROWTH FACTOR-BINDING PROTEINS

Six distinct IGFBPs that bind IGFs with high affinity have been described. IGFBPs are well conserved, including cysteine residues that may be important for IGF binding. IGFBPs are produced by multiple tissues in highly regulated ways. They are secreted into the circulation, and several lines of evidence from cell culture and *in vivo* studies suggest that IGFBPs also act at or near their site of production. In serum, the overwhelming majority of IGF is found as a 150-kDa complex that includes IGFBP-3 and ALS. Association of IGFs with this complex is thought to protect the IGFs from degradation and tissue clearance and to increase their half-life in serum. IGFBPs other than IGFBP-3 may be part of a 50-kDa complex in serum that, through transcytosis mechanisms not fully characterized, may facilitate transfer of IGFs from the circulation to interstitial space and target cells. Free IGFs may also cross endothelial barriers.

As mentioned previously, IGFBPs also are synthesized to act locally. At the target cell level, IGFBPs modulate the effects of the IGFs and/or act independently of the IGFs and the IGF-IR. Experimentally, both inhibitory and stimulatory effects of IGFBPs on IGF action have been reported. Intensive effort has been devoted to explaining these results. Because IGFBPs in solution have a higher affinity for IGFs than does the IGF-IR, IGFBPs can prevent the binding of IGFs to the IGF-IR and inhibit IGF action. However, modification of IGFBPs lowers their affinity for IGFs. Dephosphorylation of IGFBP-1, attachment of IGFBP-3 to the cell surface, and attachment of IGFBP-5 to the extracellular matrix (ECM) lowers their respective affinities for IGFs, potentially enhancing the delivery of IGFs to the IGF-IR.

Several IGF-independent actions have been reported for IGFBPs. IGFBP-1 contains an Arg-Gly-Asp (RGD) tripeptide that allows it to interact with the integrin- $\alpha$ 5- $\beta$ 1 receptor. Deletion of this RGD sequence from IGFBP-1 prevented it from

stimulating cell migration. Binding of IGFBP-1 to integrin receptors on breast cancer cells activated a signaling cascade in which FAK was dephosphorylated, disrupting attachment of the cells to the ECM and causing apoptosis.

IGFBP-3 was reported to increase the ratio of the pro-apoptotic bax to the anti-apoptotic bcl-2 in p53-deficient breast cancer cells. When these cells were challenged with ionizing radiation, their apoptosis rate was higher than in cells not treated with IGFBP-3. The presence of IGFs in these cultures was apparently excluded, suggesting an IGF-independent effect. In contrast, IGFBP-4 and IGFBP-5 have been reported to promote survival of certain cells in IGF-independent ways, although the mechanisms underlying these effects are not known. Of particular importance will be the definitive identification and characterization of cell surface receptors for IGFBPs. Of great interest also are recent results showing that IGFBP-3 and IGFBP-5 are translocated into the nucleus via the importin- $\beta$  subunit. The consequences of this action are not known, but they suggest novel direct effects of IGFBPs on nuclear events such as gene expression.

## EVOLUTIONARY ASPECTS OF INSULIN AND IGF-I SIGNALING

The nematode *Caenorhabditis elegans* and the fruitfly *Drosophila melanogaster* both contain homologues of the mammalian insulin/IGF signaling system. Recently, interesting phenotypes have been reported to result from mutations in these genes. Loss of function mutations of the IR/IGF-IR orthologue Daf-2 in *C. elegans* gives rise to dauer larvae. The dauer state shows reduced metabolic activity, perhaps an expected result, and also has a prolonged life span. The dauer state is naturally entered during periods of food deprivation and other environmental stresses. These studies suggest an intriguing mechanism in which caloric restriction may increase life span by decreasing insulin signaling. This mechanism has a mammalian counterpart, in which calorically restricted rodents have increased life span in conjunction with reduced levels of serum insulin and IGF-I. Loss of function mutations of age-1, the PI3K homologue, also extend life span in nematodes. In contrast, mutations of Daf-16, a homologue of the mammalian FOXO forkhead transcription factors, and of Daf-19, a homologue of the mammalian PTEN phosphoinositide phosphatase that inhibits PI3K, can negate the effect of the Daf-2

mutation. These results further substantiate the role of insulin/IGF-I signaling in nematodes.

In *D. melanogaster*, the homologue of the IRS family of molecules, known as CHICO, has been shown to play an important role in cell proliferation and overall body growth. Moreover, loss of function mutation of CHICO also extends life span. The precise signaling cascade leading to CHICO action is not fully understood, however, because the *Drosophila* IR homologue (DIR) binds directly to many of its downstream signaling substrates via its carboxyl-terminal extension. This is in contrast to the mammalian IR and IGF-IR, which use the IRS family of docking proteins to activate downstream molecules such as PI3K and Grb-2.

## CONCLUSION

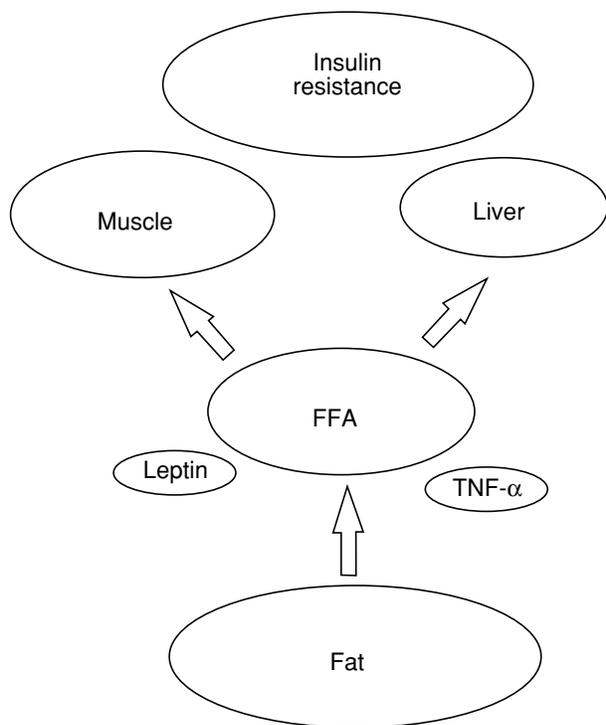
This article has briefly described the essential structural and functional features of components of the IGF system and has described some newer findings on the mechanisms and evolutionary conservation in the insulin and IGF signaling pathways. Reviews listed in the bibliography should be consulted for more in-depth treatment of these and other areas related to IGF physiology.

## See Also the Following Articles

EGF and Related Growth Factors • Fibroblast Growth Factor (FGF) • Growth Factor Receptors • Growth Hormone (GH) • Hepatocyte Growth Factor • Hypoglycemia • Insulin and Insulin-like Growth Factors, Evolution of • Peptide Hormones and Growth Factors: Cellular Signaling Mechanisms

## Further Reading

- Arantes-Oliveira, N., Apfeld, J., Dillin, A., and Kenyon, C. (2002). Regulation of life-span by germ-line stem cells in *Caenorhabditis elegans*. *Science* **295**, 502–505.
- Baxter, R. C. (2000). Insulin-like growth factor (IGF)-binding proteins: Interactions with IGFs and intrinsic bioactivities. *Am. J. Physiol. Endocrinol. Metab.* **278**, E967–E976.
- Clemmons, D. R. (1998). Role of insulin-like growth factor binding proteins in controlling IGF actions. *Mol. Cell. Endocrinol.* **140**(1–2), 19–24.
- Dupont, J., and LeRoith, D. (2001). Insulin and insulin-like growth factor I receptors: Similarities and differences in signal transduction. *Horm. Res.* **55**(Suppl. 2), 22–26.
- Le Roith, D., Bondy, C., Yakar, S., Liu, J. L., and Butler, A. (2001). The somatomedin hypothesis: 2001. *Endocrine Rev.* **22**, 53–74.
- Le Roith, D., Werner, H., Beitner-Johnson, D., and Roberts, C. T., Jr. (1995). Molecular and cellular aspects of the insulin-like growth factor I receptor. *Endocrine Rev.* **16**, 143–163.



**Figure 1** Proposed mechanism of obesity-induced insulin resistance. Excessive fat accumulation (obesity) is associated with increased plasma free fatty acid (FFA) levels that cause insulin resistance in skeletal muscle and in the liver. Other less likely mediators of obesity-induced insulin resistance are leptin and tumor necrosis factor- $\alpha$  (TNF- $\alpha$ ).

elevated plasma FFA levels were followed (after a delay of several hours) by an increase in intramyocellular triglyceride concentration (IMCL-TG) suggested another explanation, namely that FFAs needed to enter muscle cells, be reesterified there, and accumulate as IMCL-TG before causing insulin resistance. There has been some progress in understanding how accumulation of IMCL-TG can interfere with insulin action. For instance, the FFA-mediated increase in IMCL-TG and the development of insulin resistance are accompanied by a several-fold increase in diacylglycerol (DAG) concentration and protein kinase C (PKC) activity in muscle (PKC  $\beta$  and  $\delta$  in human muscle and  $\zeta$  in rat muscle). DAG is likely to accumulate during triglyceride formation or breakdown and is a potent allosteric activator of PKC. PKC, on the other hand, is a serine/threonine kinase and has been shown to interfere with insulin signaling by serine phosphorylating the insulin receptor substrate 1 (which in turn will decrease tyrosine phosphorylation of IRS-1, causing insulin resistance). PKC, however, is probably not the only serine kinase activated by FFAs. A role for I $\kappa$ B- $\beta$

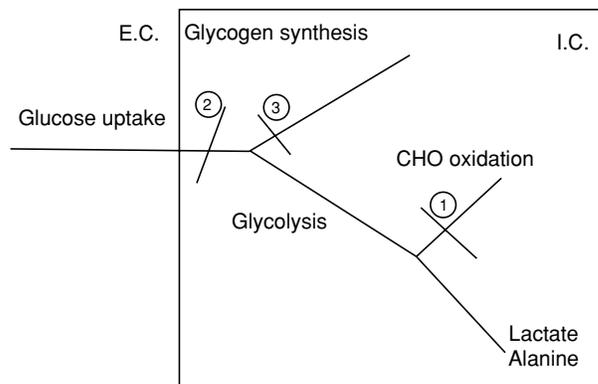
kinase (IKK- $\beta$ ), another serine kinase, was suggested by the recent demonstration that FFA-induced insulin resistance could be prevented in IKK- $\beta$  knockout mice with high doses of acetyl-salicylic acid (an inhibitor of IKK).

### FFAS: PUTATIVE PROMOTERS OF ATHEROSCLEROSIS

Acute increases of plasma FFAs also result in dramatic decreases of I $\kappa$ B in human muscle. I $\kappa$ B is the inhibitor of NF $\kappa$ B. On phosphorylation (by IKK or perhaps by PKC), I $\kappa$ B is released from NF $\kappa$ B, ubiquitinated, and degraded. Detachment of its inhibitor frees NF $\kappa$ B to translocate from the cytosol to the nucleus and to initiate transcription of several genes that have been linked to the onset of atherosclerosis. Thus, it appears that FFAs, and perhaps other substrates as well, may be able to cause both insulin resistance and atherosclerosis. This could provide part of the explanation for the well-established increase in atherosclerotic vascular diseases in insulin-resistant states.

### FFAS AND INSULIN SECRETION

Acute elevations of plasma FFA concentrations have long been known to stimulate insulin secretion. Long-term effects, on the other hand, have remained somewhat controversial. Some studies with animal islets and isolated perfused pancreas have shown a biphasic effect of FFAs: initial stimulation of insulin secretion

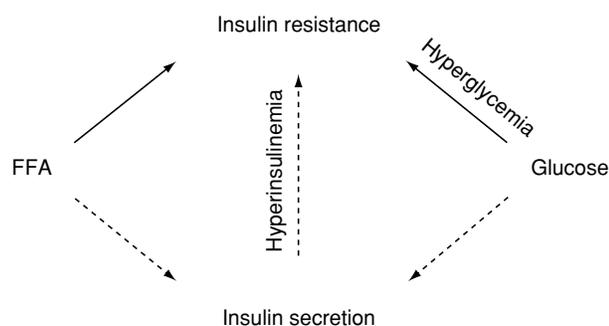


**Figure 2** Defects of glucose use produced by FFAs. The inhibition of carbohydrate (CHO) oxidation (defect 1) is the earliest demonstrable defect. It develops during the initial hour of lipid infusion but does not affect insulin-stimulated glucose uptake. The inhibition of glucose transport and/or phosphorylation (defect 2) develops after 2 to 4 h, whereas the inhibition of glycogen synthesis (defect 3) develops after 4 to 6 h of high plasma FFAs. E.C., extracellular; I.C., intracellular.

followed by inhibition after 24 to 48 h. Other *in vitro* studies, however, have shown the opposite, namely, potentiation of glucose-stimulated insulin secretion for as long as 72 h. In normal human subjects, FFAs potentiate glucose-stimulated insulin secretion for as long as 48 h. Moreover, acute suppression of basal FFAs decreases insulin secretion in nondiabetic and diabetic individuals, indicating that plasma FFAs support between 30 and 50% of basal insulin secretion. It has been suggested that high plasma FFA levels are lipotoxic. However, it is unlikely that physiological elevations of plasma FFA levels (<1.5 mmol/L) are lipotoxic to beta cells. If that were the case, one would expect that individuals with elevated plasma FFA levels (i.e., nearly all obese individuals) would develop type 2 diabetes when in fact only approximately 15 to 20% ever become diabetic. This argument is based on the following consideration. Given that increased plasma FFAs cause insulin resistance, either insulin or glucose concentrations, or both, need to increase to compensate for this resistance. Because compensation by hyperglycemia (i.e., diabetes) occurs in only 15 to 20% of individuals, it follows that FFA-mediated insulin resistance is compensated by FFA-mediated insulin secretion in the remaining majority.

### FFA AND TYPE 2 DIABETES

In healthy individuals, FFA-mediated insulin resistance is matched quantitatively by FFA-induced stimulation of insulin secretion (Fig. 3). (For example, a 10%



**Figure 3** Proposed role of free fatty acids (FFA) in the development of type 2 diabetes. In obese individuals with a genetic predisposition for development of type 2 diabetes (e.g., first-degree relatives of patients with type 2 diabetes), high plasma levels of FFAs cause insulin resistance (indicated by heavy black lines) that cannot be completely compensated by increased insulin secretion (indicated by broken lines). Plasma glucose levels will increase as a result of the partially uncompensated insulin resistance. The resulting hyperglycemia will increase glucose uptake by mass effect and by enhancing insulin secretion.

increase in insulin resistance is associated with a 10 to 15% increase in insulin secretion. The slightly greater degree of insulin secretion than insulin resistance is needed because of the hyperbolic relationship between insulin resistance and insulin secretion.) Because of the perfect match between FFA-mediated insulin resistance and insulin secretion, there is no need for blood glucose to rise. Pre-diabetic individuals (e.g., first-degree relatives of patients with type 2 diabetes mellitus) might not be able to compensate adequately because of impaired insulin secretory responses to FFA. (In patients with overt type 2 diabetes mellitus, deficient insulin secretory responses to FFA have been demonstrated.) Thus, in obese pre-diabetic individuals, increased plasma FFA levels may add another burden of insulin resistance to an already weakened beta cell system, accelerating its breakdown and the development of hyperglycemia.

### PHYSIOLOGICAL AND PATHOPHYSIOLOGICAL CONSIDERATIONS

The FFA-associated peripheral and hepatic insulin resistance can have important physiological functions. For instance, during starvation, rising plasma FFAs produce peripheral insulin resistance that preserves precious carbohydrate resources for oxidation by the central nervous system. The rise in FFAs, caused by an increased maternal fat deposition during the early part of pregnancy and by increased plasma levels of lipolytic gestational hormones, results in insulin resistance of the mother and preserves carbohydrates for the growing fetus. In obesity, on the other hand, the FFA-induced insulin resistance becomes counterproductive because there is no need to spare glucose. Moreover, the chronic insulin resistance of obesity is associated with several cardiovascular risk factors, including hypertension, dyslipidemia, abnormal blood coagulation, and fibrinolysis. FFA-mediated activation of the  $\text{I}\kappa\text{B}/\text{NF}\kappa\text{B}$  pathway may be one connection between FFA-induced insulin resistance and these cardiovascular risk factors.

### See Also the Following Articles

Atherosclerosis • Diabetes, Type 2 • Glucocorticoid Resistance Syndromes and States • Insulin Secretion: Functional and Biochemical Aspects • Obesity and Diabetes, Regulation of Food Intake • Obesity, Treatment of • Resistance to Thyroid Hormone (RTH)

## Further Reading

- Barnes, P. J., and Karin, M. (1997). Nuclear factor-kappaB: A pivotal transcription factor in chronic inflammatory diseases. *N. Engl. J. Med.* **336**, 1066–1071.
- Boden, G. (1997). Role of fatty acids in the pathogenesis of insulin resistance and NIDDM. *Diabetes* **46**, 3–10.
- Boden, G. (2001). Obesity, free fatty acids, and insulin resistance. *Curr. Opin. Endocrinol. Diab.* **8**, 235–239.
- Collins, T., and Cybulsky, M. I. (2001). NF- $\kappa$ B: Pivotal mediator or innocent bystander in atherogenesis? *J. Clin. Invest.* **107**, 255–264.
- Mokdad, A. H., Bowman, B. A., Ford, E. S., Vinicor, F., Marks, J. S., and Koplan, J. P. (2001). The continuing epidemics of obesity and diabetes in the United States. *JAMA* **286**, 1195–1200.
- Randle, P. J., Garland, P. B., Hales, C. N., and Newsholme, E. A. (1963). The glucose-fatty acid cycle. Its role in insulin sensitivity and the metabolic disturbances of diabetes mellitus. *Lancet* **1**, 785–789.
- Yuan, M., Konstantopoulos, N., Lee, J., Hansen, L., Li, Z. W., Karin, M., and Shoelson, S. E. (2001). Reversal of obesity- and diet-induced insulin resistance with salicylates or targeted disruption of IKKbeta. *Science* **293**, 1673–1677.

pulsatile release of PGF into uterine venous blood. A countercurrent transport mechanism between the uterine vein and the ovarian artery ensures efficient movement of PGF from the uterine vein to the ovarian artery and, hence, to the CL. PGF then triggers the release of additional amounts of oxytocin from the CL, which in turn heightens the amplitude of PGF pulses. The sequel to this positive feedback loop is prolonged high-amplitude pulses of PGF and functional and structural loss of the integrity of the CL.

### SIMILARITIES AND DIFFERENCES BETWEEN THE IFN- $\tau$ AND RELATED IFNS

IFNs were discovered because of their abilities to protect cells from viral infection. There are two kinds. IFN- $\gamma$  is a dimeric protein with two identical subunits of approximately 146 amino acids and is unrelated to the type I IFN to which the IFN- $\tau$  belongs. IFN- $\gamma$  is an inflammatory cytokine produced mainly by immune cells but also, curiously, by the pig trophoblast, where its function remains unknown. The type I IFNs are an extensive grouping of related genes found throughout vertebrates and include the IFN- $\alpha$ , - $\beta$ , - $\omega$ , - $\kappa$ , and - $\delta$  as well as the IFN- $\tau$ . Many subtypes, including the IFN- $\tau$ , are represented by multiple genes. The three-dimensional structure of their proteins is based on a bundle of five  $\alpha$ -helices. The hallmark of the entire grouping is their potent antiviral activity, which is usually evident in the picomolar range of concentrations, but they usually also possess antiproliferative and complex immunomodulatory properties. Subtle variations in properties, which have been widely documented, are likely based on the manner in which each IFN interacts with the two subunits of the common receptor and activates the intracellular signal transduction system.

Table I lists some of the properties that the IFN- $\tau$  has in common with other type I IFNs where comparisons have been made. Importantly, there is as yet no evidence that the IFN- $\tau$  is superior in this regard. Where the IFN- $\tau$  differs from the IFN- $\alpha$ , - $\beta$ , and - $\omega$ , in particular, is in the lack of inducibility of its genes in response to virus, its localized expression in one epithelial cell layer (trophectoderm), and its very high rate of production over a few days of conceptus development. It is likely that what sets the IFN- $\tau$  apart and has provided it with the status of a hormone of pregnancy is its presence in the appropriate place, in the required amounts and at the correct time, to prevent CL regression. As a general rule, *cis*-regulatory

Table I IFN- $\tau$  Properties of IFN- $\tau$  That Are Typical of Other Type I IFNs

• Antiviral activity
• Antiproliferative activity
• Immunomodulatory activity
• Binds to the same (IFNAR1/IFNAR2) receptor as do other IFNs
• Activates STAT1/STAT2 signal transduction pathway
• Up-regulates a similar set of genes as do other IFNs

elements of genes govern expression patterns in space and over time. What makes the IFN- $\tau$  uniquely different from other IFNs, therefore, is likely to be embedded in the control elements of its genes.

### EVOLUTION OF THE IFN- $\tau$ AND ITS PROMOTER ELEMENTS

Based on base substitution rates, the first IFN- $\tau$  gene, *IFNT*, originated by a single duplication from an IFN- $\omega$ , *IFNW*, some 36 million years ago in the lineage leading to the current cattle, deer, giraffes, and their relatives. As a consequence, these genes do not occur, nor are they necessary components of the pregnancy recognition systems that are encountered, outside of the ruminant, even-toed ungulates. The duplication event that provided the progenitor *IFNT* was either accompanied by or closely followed by changes in the promoter regions of the gene that led to expression in the trophectoderm. The initial alteration in the promoter was presumably followed by selection of additional mutations that provided the current genes with the proper combination of *cis*-regulatory elements to drive high rates of transcription within the trophectoderm. Certainly, the *IFNT* genes do not possess well-organized viral response elements comparable to those found in IFN- $\alpha$  and - $\beta$  genes, and their promoter regions diverge markedly in overall sequence from those of their close relatives, the *IFNW* genes.

Two regions of the promoter unique to *IFNT* have been inferred to be important for expression in the trophectoderm. One is a distal A/T-rich element situated approximately -358 to -322 bp from the transcription start site, which has not been studied in detail, and the other is a complex Ets-2/AP1-binding site in the proximal region (-91 to -69) of the promoter. Neither element is encountered in the *IFNW*. The transcription factor Ets-2 strongly transactivates *IFNT* promoters through the proximal

site, whereas mutation of the Ets-2 *cis*-element blocks transcription.

*IFNT* expression is detectable but low at the blastocyst stage of bovine embryo development, but it increases markedly *in vivo* a few days later, coincident with the elongation of the conceptus and steeply rising progesterone concentrations in the mother. Expression then begins to fall as the trophoblast cells attach to the uterine wall. Little is known about the underlying causes of these changes, although the initial up-regulation may be driven by maternal factors present in uterine secretions.

The author has argued previously that the emergence of the *IFNT* in the pecoran ruminants and of contrasting systems in other groups of mammals for maintaining the life span of the CL is the result of intense genetic conflict at the trophoblast–maternal interface. There is insufficient space to discuss the arguments here. Nevertheless, it seems likely that the progenitor *IFNT* gene was, at its birth, a minor player in the mix of mechanisms used to combat transmission of pathogens to the fetus. The elevation of the IFN- $\tau$  to an essential placental hormone must have coincided with the gradual acquisition of powerful trophoblast-specific enhancer elements in its genes. This unique mechanism for trophoblast signaling appears to have evolved hand in hand with the unique synepitheliochorial placentation of the ruminants.

## See Also the Following Articles

Cytokines, Evolutionary Aspects and Functions • Implantation • Interferons

## Further Reading

- Demmers, K. J., Derecka, K., and Flint, A. (2001). Trophoblast interferon and pregnancy. *Reproduction* **121**, 41–49.
- Ezashi, T., Ealy, A. D., Ostrowski, M. C., and Roberts, R. M. (1998). Control of interferon- $\tau$  gene expression by Ets-2. *Proc. Natl. Acad. Sci. USA* **95**, 7882–7887.
- Fleming, J. A., Choi, Y., Johnson, G. A., Spencer, T. E., and Bazer, F. W. (2001). Cloning of the ovine estrogen receptor- $\alpha$  promoter and functional regulation by ovine interferon-tau. *Endocrinology* **142**, 2879–2887.
- Hughes, A. L., and Roberts, R. M. (2000). Independent origin of IFN- $\alpha$  and IFN- $\beta$  in birds and mammals. *J. Interferon Cytokine Res.* **20**, 737–740.
- Leaman, D. W., Cross, J. C., and Roberts, R. M. (1994). Multiple regulatory elements are required to direct trophoblast interferon gene expression in choriocarcinoma cells and trophoblast. *Mol. Endocrinol.* **8**, 456–468.
- Mann, G. E., Lamming, G. E., Robinson, R. S., and Wathes, D. C. (1999). The regulation of interferon-tau production and uterine hormone receptors during early pregnancy. *J. Reprod. Fertil. Suppl.* **54**, 317–328.
- Roberts, R. M., Cross, J. C., and Douglas, W. L. (1992). Interferons as hormones of pregnancy. *Endocrine Rev.* **13**, 432–452.
- Roberts, R. M., Ealy, A. D., Alexenko, A. P., Han, C-S., and Ezashi, T. (1999). Trophoblast interferons. *Placenta* **20**, 259–264.
- Roberts, R. M., Xie, S., and Mathialagan, N. (1996). Maternal recognition of pregnancy. *Biol. Reprod.* **54**, 294–302.

## INTERFERON- $\alpha$ AND THE ENDOCRINE SYSTEM

Previous studies indicated that the action of IFN- $\alpha$  requires its binding to specific receptors on the cell surface. This binding may share sites common to those of peptide hormones such as adrenocorticotropin hormone (ACTH) and thyroid-stimulating hormone (TSH). The elevation and detection of ACTH and endorphin-like substances from lymphocytes infected with the Newcastle disease virus (NDV) were the first demonstration that the immune system is capable of producing peptides that can signal the neuroendocrine system. Cells of the immune system communicate with each other through direct contact and through secreted cytokines. However, production of cytokines is not restricted to immune cells. Cytokines, including IFN- $\alpha$ , are also produced in the brain. They exert direct effects on the brain and endocrine organs, and they cause the production and secretion of ACTH by inducing the corticotropin-releasing factor (CRF). Intracerebroventricular (i.c.v.) injection of IFN- $\alpha$  decreases the plasma level of glucocorticoid, which can be prevented by naloxone pretreatment. Moreover, i.c.v. injection of CRF and CRF antagonist has been reported to mimic and block, respectively, the IFN- $\alpha$ -induced suppression of splenic NK activity. These results suggest that IFN- $\alpha$  in the brain activates the CRF system through central opioid receptors and thereby suppresses NK cytotoxicity predominantly through splenic sympathetic innervation.

Both clinical observations and experimental studies in the laboratory have demonstrated a stimulative effect of IFN- $\alpha$  on adrenocortical secretion. In addition, sequential similarities among human leukocyte IFN- $\alpha$ , ACTH, and melanotropic-stimulating hormone (MSH) have been observed. These structural similarities may explain the presence of common functional characteristics found between immunological activity and MSH-ACTH-like activity of a pepsin-derived peptide of IFN- $\alpha$ . The proposed homology between MSH-ACTH and human leukocyte IFN- $\alpha$  is consistent with many research findings and would explain the observed cross-reactivity between MSH-ACTH antisera and human leukocyte IFN- $\alpha$ .

It has been shown that systematic (i.e., intraperitoneal [i.p.]), central (i.c.v.), and local administration of IFN- $\alpha$  within the hypothalamic paraventricular nucleus (PVN) elicited both acute and chronic suppression of hypothalamus-pituitary-adrenocortical secretion in the rat. In view of the inhibitory effects of glucocorticoid hormones on immune activity, this

suggests that IFN- $\alpha$  may exert positive feedback effects on its own immunopotentiating activity by decreasing adrenocortical secretion. In another study, the preceding observations were confirmed by demonstrating that adrenocortical activity was inhibited following hybrid recombinant human IFN- $\alpha$  preparation. It was also shown that IFN- $\alpha$  decreased the electrical activity of electrophysiologically identified neurosecretory PVN neurons. These PVN neurons project to the median eminence that regulates the adrenocortical secretion; that is, IFN- $\alpha$  participates in the regulation of adrenocortical secretion.

The endocrine system, the immune system, and the CNS are affected by psychological factors, including stress and mental illnesses, that can lead to altered immunocompetence and incidence of disease. Two mechanisms have been proposed to explain this phenomenon: (1) an indirect neuroendocrine communication mediated by the secretion of glucocorticoid and/or other hormones that regulate the production and secretion of IFN- $\alpha$  and (2) a direct neural influence on the endocrine and immune systems mediated by the autonomic and central nervous systems. In turn, activation of the endocrine system, the immune system, and IFN- $\alpha$  generates signals that alter CNS neural activity and modulate neuroendocrine secretory events, thereby providing feedback regulation of immune activity.

## EFFECT OF INTERFERON- $\alpha$ ON OTHER SYSTEMS

IFN- $\alpha$  was given to healthy human volunteers, resulting in the slowing down of CNS activity, as assessed by increased reaction time to stimuli presented at random intervals. In rodents, IFN- $\alpha$  injections depressed open-field activity and caused weight loss. The brain areas participating in feeding behavior are thought to be in the hypothalamus. IFN- $\alpha$  may reach high concentrations in the pons and hypothalamus, where the BBB is more permeable. IFN therapy elicits sensory and motor disorders, fever, anorexia, confusion, depression, and sleep disorder. From these reports, it is apparent that IFN- $\alpha$  has remarkable effects on the endocrine system, the immune system, the CNS in general, and the hypothalamus in particular as well as on sites of the somatosensory, motor system, and limbic system. It has been suggested that IFN- $\alpha$  is one of the messengers responsible for the bidirectional communication that links both the immune and endocrine systems with the CNS.

## Thermoregulation

Fever is initiated as a result of the activation of thermosensitive neurons located mainly in the preoptic/anterior hypothalamus (PO/AH) area by unknown mechanisms. When IFN- $\alpha$  is intravenously (i.v.) or i.c.v. injected in rabbits, cats, and mice, it produces fever that does not involve production of the endogenous pyrogenic substance interleukin-1 (IL-1). Therefore, IFN- $\alpha$  has also been identified as an endogenous pyrogen. The PO/AH area is suggested as the most probable site of pyrogenic actions. IFN- $\alpha$  is believed to modulate the thermosensitive neuronal activity in the PO/AH neurons by decreasing the neuronal activity of warm sensitive neurons while increasing the activity of cold sensitive neurons. The same IFN- $\alpha$  doses had no effect on thermally insensitive neurons. The effects of IFN- $\alpha$  on the thermosensitive neurons were blocked by the opiate antagonist naloxone but were not blocked by a dose of sodium salicylate (an antipyrogen agent), which had effectively blocked the neuronal responses to endotoxin and leukocytic pyrogen. The fever induced by IFN- $\alpha$  injection may be explained in part by the effects of IFN- $\alpha$  on PO/AH thermosensitive neurons that involved the opiate receptor mechanism. It appears that IFN- $\alpha$  in the brain may affect the activity of PO/AH neurons and produce fever by a two-step mechanism. The first step involves the immediate action on the opioid receptors. The second step consists of the release of prostaglandins. Another study that investigated the effects of IFN- $\alpha$  on the phenomenon of morphine-induced hypothermia found that IFN- $\alpha$  prevented the development of tolerance to the hypothermic effects caused by morphine.

## Appetite Regulation

Appetite regulation is a profoundly complex process that involves the endocrine system and the CNS via an endogenous metabolite that regulates food intake. The major CNS sites participating in the control of food intake are thought to be the lateral hypothalamus (LH), the PVN, and the ventromedial hypothalamus (VMH). It has been suggested that the VMH is a "satiety center" responsible for producing the sensation of fullness, whereas the LH is a "feeding center" involved in initiating food intake. The hypothalamus also acts as a neuroendocrine transducer site in relation to the control of food intake, involving a balance between a number of neuropeptides and neurotransmitters. Together, these sites are also known to contain neurons that sense endogenous metabolic parameters, such as glucose, and participate

in the control of food intake as well as in energy balance.

In general, IFN- $\alpha$  therapy was shown to elicit anorexia that resulted in the loss of more than 15% of the patients' body weight. A similar decrease in food intake following IFN- $\alpha$  treatment was found in animals. Specifically, IFN- $\alpha$ -induced feeding suppression may involve excitation of glucose-sensitive neurons in the VMH and decreased neuronal activity in glucose-sensitive LH neurons (considered a "hunger center"). A study investigated whether the loss of appetite, one of the major side effects of IFN- $\alpha$  therapy, is mediated by the operation of IFN- $\alpha$  on LH neurons. The investigators used central (i.c.v.) and systemic (i.p.) IFN- $\alpha$  application and simultaneously recorded LH neuronal activity and food intake in conscious rats treated for 3 weeks with IFN- $\alpha$ . In therapeutic doses, IFN- $\alpha$  elicited a reversible dose-related decrease of both food intake and body weight. The decrease in food intake following IFN- $\alpha$  injections was correlated with a depression of LH neuronal electrical activity. Because direct brain application (i.c.v.) and systemic (i.p.) IFN- $\alpha$  treatment caused identical responses, it is postulated that IFN- $\alpha$  suppresses food intake by a direct action on LH neurons. These results suggest that IFN- $\alpha$ , which is endogenously produced by the brain, is involved in the regulation of feeding.

Local application of IFN- $\alpha$  to identify glucose-responsive cells in the VMH was altered in appropriate ways to explain the cytokine-induced anorexia. Others have reported that microiontophoretic application of IFN- $\alpha$  produced both suppression and excitation in VMH neurons, depending on the dose, and suppression in the LH neurons. Similarly, opposing observations between LH and VMH, induced by morphine, have been reported. The reciprocal interaction between LH and VMH in mechanisms of hunger and satiety has been documented extensively; therefore, it is possible that IFN- $\alpha$  affects both VMH and LH in a reciprocal (push-pull) manner in regulating food intake. This agrees with the suggestion by some researchers that cytokine-induced suppression of food intake involves the excitation of glucose-sensitive excitatory neurons in the VMH and the inhibition of glucose-sensitive neurons in the LH. IFN- $\alpha$  elicited excitatory effects on some glucose-sensitive neurons that were stimulated by morphine and by endogenous opioid peptides. Furthermore, naloxone prevented the excitatory neuronal responses to microiontophoretic applications of IFN- $\alpha$  in the VMH recorded from electrophysiologically identified glucose-sensitive neurons.

## Opioid System

The effects of opiate peptides on the secretion of IFNs and other cytokines have generally been described as stimulatory and bimodal.  $\beta$ -endorphin and Met-enkephalin increase the production of IFN- $\gamma$ , whereas morphine suppresses release of IFN- $\gamma$ . Similarities have been demonstrated among the structures of pro-opiomelanocortin,  $\beta$ -endorphin, and IFN- $\alpha$  that link IFN- $\alpha$  with opiates. Other evidence linking IFN- $\alpha$  with opiates is provided by experiments showing that lymphocytes stimulate IFN- $\alpha$  inducers to produce ACTH and endorphin-like substances. Opiates and ACTH exhibit similar antigenic reactions, suggesting that these peptides have some common structural properties as well as an affinity for  $^3\text{H}$ -morphine-binding sites in mouse membranes obtained from the brain. IFN- $\alpha$ , but not IFN- $\beta$  or IFN- $\gamma$ , binds to opiate receptors and shares some pharmacological properties of opiates such as production of analgesia, reduction in motor activity, and catatonia. These effects are reversed or prevented by the opiate antagonist naloxone.

Interaction between IFN- $\alpha$  and morphine has been demonstrated with single injections of morphine in mice that reduced the level of serum IFN- $\alpha$  induced by poly I:C or endotoxin. Moreover, the opiate antagonist naloxone, when given prior to IFN- $\alpha$  treatment, prevents the analgesia and catatonic behavior produced by IFN- $\alpha$  given alone. This evidence suggests that the effects of IFN- $\alpha$  on CNS activity are mediated via a number of different mechanisms, some of which may be antagonized by naloxone. To test the possibility of whether IFN- $\alpha$  exerts its effects via opiate receptors, electrophysiological recordings were made from two preparations: (1) the guinea pig ileum and (2) several neurophysiological procedures (evoked field potentials and single-cell recording) from different brain areas. In the neurophysiological experiment using single-unit recording, IFN- $\alpha$  elicited excitation in cortical neurons and in many VMH neurons. However, in some of VMH cells, IFN- $\alpha$  elicited the opposite effect; that is, it decreased the firing rates of these cells. Naloxone failed to antagonize the IFN- $\alpha$ -induced excitation but was able to antagonize the IFN- $\alpha$ -induced inhibition. This observation suggests that IFN- $\alpha$  action on these two preparations might be mediated not via opiate receptors but rather via IFN- $\alpha$  receptors. However, not all possibilities were tested in these experiments. In brain slice preparation, IFN- $\alpha$  elicited both an increase and a decrease in the firing rate recorded from the anterior hypothalamus (AH) and VMH, respectively. Both responses were blocked

by naloxone. These studies suggest that there are at least three different functional and/or receptor sites for IFN- $\alpha$  within the CNS: (1) an inhibitory site where IFN- $\alpha$  produces reduction in the neuronal activity and naloxone reverses the IFN- $\alpha$  effects that may be mediated by the  $\mu$ -receptor type, (2) an excitatory site where IFN- $\alpha$  produces excitation in the neuronal activity that the opiate antagonist naloxone reverses (i.e., a naloxone-dependent site) and may represent the kappa or delta receptor sites, and (3) an excitatory site where IFN- $\alpha$  produces excitation in the neuronal activity that is not antagonized by naloxone.

## CLINICAL UTILITIES OF INTERFERONS

### Interferon- $\alpha$

IFN- $\alpha$  administered subcutaneously (s.c.) has demonstrated a surprisingly wide range of efficacy in hematological malignancies, including tumors of presumed B-cell, T-cell, and myeloid lineages. In some diseases, such as hairy cell leukemia and chronic myelogenous leukemia, IFN- $\alpha$  is broadly effective.

#### **Kaposi's Sarcoma**

High-dose (27–36 MU) human recombinant IFN- $\alpha$ -2a every day (q.d.) achieved a major response in 12 of the 26 evaluable Kaposi's sarcoma (KS) patients, with 5 of these showing histologically confirmed complete responses with dose and blood levels important factors in responses. The addition of zidovudine in AIDS-associated KS is sometimes efficacious.

#### **Lymphoma**

IFN- $\alpha$ -2b plus chlorambucil is superior to chlorambucil alone in advanced follicular lymphoma and low-grade lymphoma. IFN maintenance therapy in diffuse large cell lymphoma improves duration of complete remission and survival and prolongs remission after fludarabine monophosphate therapy in low-grade non-Hodgkin's lymphoma. However, IFN- $\alpha$  as maintenance therapy is not useful in aggressive malignant lymphoma when more intensive chemotherapy (CHOP-BLEO regimens) has been employed during induction.

#### **Polycythemia Vera**

IFN- $\alpha$  is superior to phlebotomy in controlling the pathological expansion of erythroid elements and all of the clinical aspects of polycythemia vera (PCV), and it corrects peripheral thrombocytosis in essential thrombocythemia.

### **Chronic Myelogenous Leukemia**

IFN- $\alpha$  has significant activity in chronic myelogenous leukemia (CML), with best results at dosages of 5 M IU/m<sup>2</sup>/day. Early-stage, Philadelphia + CML, hematological response rates are 70 to 80%, and cytogenetic response rates are 50% (approximately 20% of which are complete). IFN- $\alpha$ -2a daily continuously and intermittent low-dose cytosine arabinoside (Ara-C) are superior to the results using IFN alone in CML. IFN also effectively treats CML relapse during the chronic phase after bone marrow transplantation. Cytoreductive/myeloablative treatment followed by unmanipulated peripheral blood stem cell transplantation and low-dose IFN- $\alpha$  can result in long-term survival in newly diagnosed CML patients.

### **Hairy Cell Leukemia**

IFN- $\alpha$ , by its direct action on hairy cells, has been shown to be extraordinarily effective with doses as low as 150,000 to 500,000 IU/m<sup>2</sup>/day, sufficient to maximize biomarker neopterin levels.

### **Virus Infections**

IFN- $\alpha$  acts too slowly to arrest acute viral infections, but IFN- $\alpha$  has proved to be useful in some chronic viral infections. Some patients form neutralizing antibodies that block the effects of the IFN; these appear to be relatively more common after recombinant IFN- $\alpha$ -2 than after IFN derived from human cells.

### **Hepatitis C**

In randomized controlled trials using s.c. IFN, 3 MU IFN- $\alpha$ -2b three times per week (t.i.w.) normalized ALT levels and eliminated hepatitis C virus (HCV) RNA in 53% of the treated and none of the untreated patients. Comparing recombinant IFN- $\alpha$ -2a with no treatment, 20 treated cases (66.7%) normalized serum aminotransferase levels within the first 4 months of treatment, compared with 1 case in the untreated group. However, reactivation or breakthrough frequently occurred afterward (20% in both cases). Reinforced regimens (6 MU q.d.  $\times$  12 days, 6 MU t.i.w.  $\times$  22 weeks, 3 MU t.i.w.  $\times$  24 weeks) as opposed to standard regimens (3 MU t.i.w.  $\times$  24 weeks) produced sustained response (ALT normalization) in a limited number (18%) of previously untreated patients and reduced the risk of cirrhosis (1%) after 18 months. In patients with compensated HCV-related cirrhosis, IFN- $\alpha$ -2b 3 M IU t.i.w. showed biochemical response in 6 of 47 treated patients and 0 of 39 control patients. However, a 48-week course of IFN therapy usually fails to achieve sustained response and did not significantly improve the 3-year outcome. Serum

HCV RNA is better than ALT at predicting long-term cure after IFN- $\alpha$  therapy in chronic HCV. Pegylated IFN- $\alpha$ -2b provides a higher virological response in patients infected with genotype 1. In patients with chronic HCV, once-weekly peginterferon  $\alpha$ -2a plus ribavirin was tolerated, as was IFN- $\alpha$ -2b plus ribavirin, and produced significant improvements in the rate of sustained virological response, as compared with IFN- $\alpha$ -2b plus ribavirin or peginterferon  $\alpha$ -2a alone. Certain features of HCV infections predict success or failure of IFN therapy. Clinical and serum biochemical response to IFN- $\alpha$  in chronic HCV is associated with a loss of detectable HCV genome from serum. Short-term and sustained responses were predicted independently by lobular structure on pretreatment liver biopsy and by short disease duration. IFN does not seem to affect overall or event-free survival of patients with HCV-related cirrhosis, whereas it seems to prevent the development of hepatocellular carcinoma.

### **Hepatitis B**

In a randomized controlled trial, IFN- $\alpha$ -2a (10 M IU/m<sup>2</sup> t.i.w.  $\times$  6 months) was significantly better than no treatment in producing a sustained loss of HBeAg in hepatitis B virus (HBV) chronic carriers. A total of 10 of 31 (32%) 4-month-treated patients and only 1 of 14 (7%) control patients became negative for serum HBV DNA/DNAp, which can induce remission in approximately one-third of chronic HBV patients. In addition, 5 M IU recombinant IFN- $\alpha$ -2b for 12 weeks decreased viral replication and tripled the spontaneous seroconversion rate observed in chronic HBV/active viral replication patients. However, IFN- $\alpha$  therapy induces HBeAg seroconversion in only approximately one-third of patients, leaving the majority of patients with persistent disease. However, data from three previously published multicenter, randomized, controlled trials were insufficient to establish a presumed beneficial effect of IFN- $\alpha$ -2a on disease progression and survival.

### **Multiple Sclerosis**

Recently, high-dose recombinant IFN- $\alpha$ -2a (9 M IU intramuscular [i.m.] every other day [q.o.d.]) reduced the relapse rate by 83% and active magnetic resonance imaging (MRI) scans over 6 months. However, attack rates rapidly returned to pretreatment values when therapy was stopped. The apparent difference between the results of this small but well-conducted Italian study and prior studies with IFN- $\alpha$  in multiple sclerosis (MS) relate to IFN dose used in earlier trials (1 M IU nIFN- $\alpha$  s.c. q.d.; 3 M IU nIFN- $\alpha$  s.c. twice

weekly [t.i.d.] for 2 months and then weekly; 2 M IU recombinant IFN- $\alpha$ -2 s.c. t.i.w., differences in trial design, or the selection of more severely affected patients with SP MS).

### Interferon- $\beta$

The use of IFN- $\beta$  has been largely confined to MS. A phase III study in RRMS showed that 8 M IU of recombinant IFN- $\beta$ -1b (Betaseron) s.c. q.o.d. decreases attacks by 30% and decreases brain inflammation as assessed by serial quantitative MRI. Another phase III clinical trial using recombinant IFN- $\beta$ -1a (6M IU weekly i.m. [30  $\mu$ g Avonex]) showed a 38% reduction in the proportion showing progression, an 18% reduction in relapse rate, and a 33% reduction in average number of active (gadolinium [Gd]-enhancing) lesions on MRI. Patients receiving IFN- $\beta$ -1a (Rebif) (22 or 44  $\mu$ g t.i.w. s.c.) in a phase III study had reduced relapse rates for 4 years with reduced new T2 lesion number and lesion burden. However, 6 to 40% of IFN- $\beta$ -treated patients generate neutralizing antibodies in those patients who appear to lose both clinical benefits and MRI-defined responses.

### Interferon- $\gamma$

Immune IFN- $\gamma$  acts in the defense of the organism against foreign pathogens via cellular immunity mediated by macrophages. It is a product of T-helper 1 (Th1) cells that can modulate the function of Th2 lymphocytes. In general, the Th1 cytokines stimulate the body's cellular immunity in viral infections. INF- $\gamma$  is potentially antiviral, antibacterial, antiproliferative, antitumor, and antiallergic. However, the clinical use of s.c. IFN- $\gamma$  has been relatively limited. INF- $\gamma$  can decrease healing time by 50 to 60% in human papillomaviruses (*Condylomata acuminata*). It can clear the damage caused by refractory nontuberculous mycobacterium infection. There can be a beneficial effect of IFN- $\gamma$  in chronic granulomatous disease. INF- $\gamma$  is contraindicated in MS. In general, MS lymphocytes secrete excess amounts of IFN- $\gamma$ .

### CONCLUSION

INF- $\alpha$  produced in the brain exerts direct effects on the endocrine system by activating the neurosecretory hypothalamic neurons and regulates the hypothalamic-pituitary-adrenocortical axis by activating the CRF system via a positive feedback effect. INF- $\alpha$  injection produces fever and regulates thermosensitive

single PO/AH neurons. The effects of IFN- $\alpha$  on the thermosensitive neurons are prevented by naloxone. This suggests that IFN- $\alpha$  mediates the activity of PO/AH neurons and produces fever via opiate receptors and subsequently by prostaglandins release. INF- $\alpha$  injection activates the sympathetic nerves innervating components of the immune system, resulting in suppression of splenic NK cells' activity. Furthermore, using electroencephalogram (EEG), evoked potential and single-cell recording procedures demonstrate that IFN- $\alpha$  has a direct effect on neuronal populations as well as on hypothalamic single-cell activity. INF- $\alpha$ -induced suppression of feeding (anorexia) is correlated with IFN- $\alpha$ -induced excitation of VMH glucose-sensitive neurons and decreased activity in LH glucose-sensitive neurons. Therefore, IFN- $\alpha$  may serve as a regulatory mediator among the CNS, the endocrine system, and the immune system. Immunological therapy uses cytokines such as IFNs to treat various hematological malignancies and infectious ailments as well as autoimmune diseases. Overall, the IFNs, primarily IFN- $\alpha$  and IFN- $\beta$ , have surprisingly wide applications in biology because of their pleotropic activities. New parenteral delivery systems that allow sustained serum levels, or novel delivery systems that avoid systemic side effects, may well increase the applicability of these interesting biological compounds.

### See Also the Following Articles

ACTH (Adrenocorticotrophic Hormone) • Immune System, Hormonal Effects on • Interferon- $\tau$  • Janus Kinases and Cytokine Receptors • TSH Function and Secretion

### Further Reading

- Aguet, M., and Mogensen, K. E. (1983). Interferon receptors. In "Interferon" (I. Gresser, ed.), Vol. 5, pp. 1-22. Academic Press, London.
- Baron, S., Tyring, S. K., Fleischmann, W. R., Jr., Coppenhaver, D. H., Niesel, D. W., Klimpel, G. R., Stanton, G. J., and Hughes, T. K. (1991). The interferons: Mechanisms of action and clinical applications. *JAMA* **266**, 1375-1383.
- Blalock, J. E., and Smith, E. M. (1981). Structure and function of interferon (IFN) and neuroendocrine hormones. In "The Biology of the Interferon System" (E. De Maeyer, G. Galasso, and H. Schellekens, eds.), pp. 93-99. Elsevier, Amsterdam.
- Bocci, V. (1992). Physicochemical and biologic properties of interferons and their potential uses in drug delivery systems. *Crit. Rev. Ther. Drug Carrier Sys.* **9**, 91-133.
- Brassard, D. L., Grace, M. J., and Bordens, R. W. (2002). Interferon- $\alpha$ . *J. Leukocyte Biol.* **71**, 565-581.
- Dafny, N. (1985). Interferon as an endocoid candidate preventing and attenuating opiate addiction. In "Endocoids" (H. Lal, F. Labella, and J. Lane, eds.), pp. 269-276. Alan R. Liss, New York.

- Dafny, N., Prieto-Gomez, B., and Reyes-Vazquez, C. (1985). Does the immune system communicate with the central nervous system? Interferon modifies central nervous system activity. *J. Neuroimmunol.* **9**, 1–12.
- Dougherty, P. M., and Dafny, N. (1991). Interaction of immune cytokines and CNS opioids: A possible interface for stress-induced immune suppression. In "Stress and Immunity" (J. Wybran, R. Faith, H. W. McCain, and N. P. Plotnikoff, eds.), pp. 373–385. Plenum, New York.
- Finter, N. B., Chapman, S., Dowd, P., Johnston, J. M., Manna, V., Sarantis, N., Sheron, N., Scott, G., Phua, S., and Tatum, P. B. (1991). The use of interferon-alpha in virus infections. *Drugs* **42**, 749–765.
- Gramenzi, A., Andreone, P., Fiorino, S., Camma, C., Giunta, M., Magalotti, D., Cursaro, C., Calabrese, C., Arienti, V., Rossi, C., Di Febo, G., Zoli, M., Craxi, A., Gasbarrini, G., and Bernardi, M. (2001). Impact of interferon therapy on the natural history of hepatitis C virus related cirrhosis. *Gut* **48**, 843–848.
- Plata-Salamán, C. R. (1992). Interferons and central regulation of feeding. *Am. J. Physiol.* **263**, R1222–R1227.
- Reyes-Vázquez, C., Prieto-Gomez, B., and Dafny, N. (1994). Alpha-interferon suppresses food intake and neuronal activity of the lateral hypothalamus. *Neuropharmacology* **33**, 1545–1552.
- Saphier, D., Roerig, S. C., Ito, C., Vlasak, W. R., Farrar, G. E., Broyles, J. E., and Welch, J. E. (1994). Inhibition of neural and neuroendocrine activity by  $\alpha$ -interferon: Neuroendocrine, electrophysiological, and biochemical studies in the rat. *Brain Behav. Immunol.* **8**, 37–56.
- Stuart-Harris, R. C., Lauchlan, R., and Day, R. (1992). The clinical application of the interferons: A review—NSW Therapeutic Assessment Group. *Med. J. Austral.* **156**, 869–872.

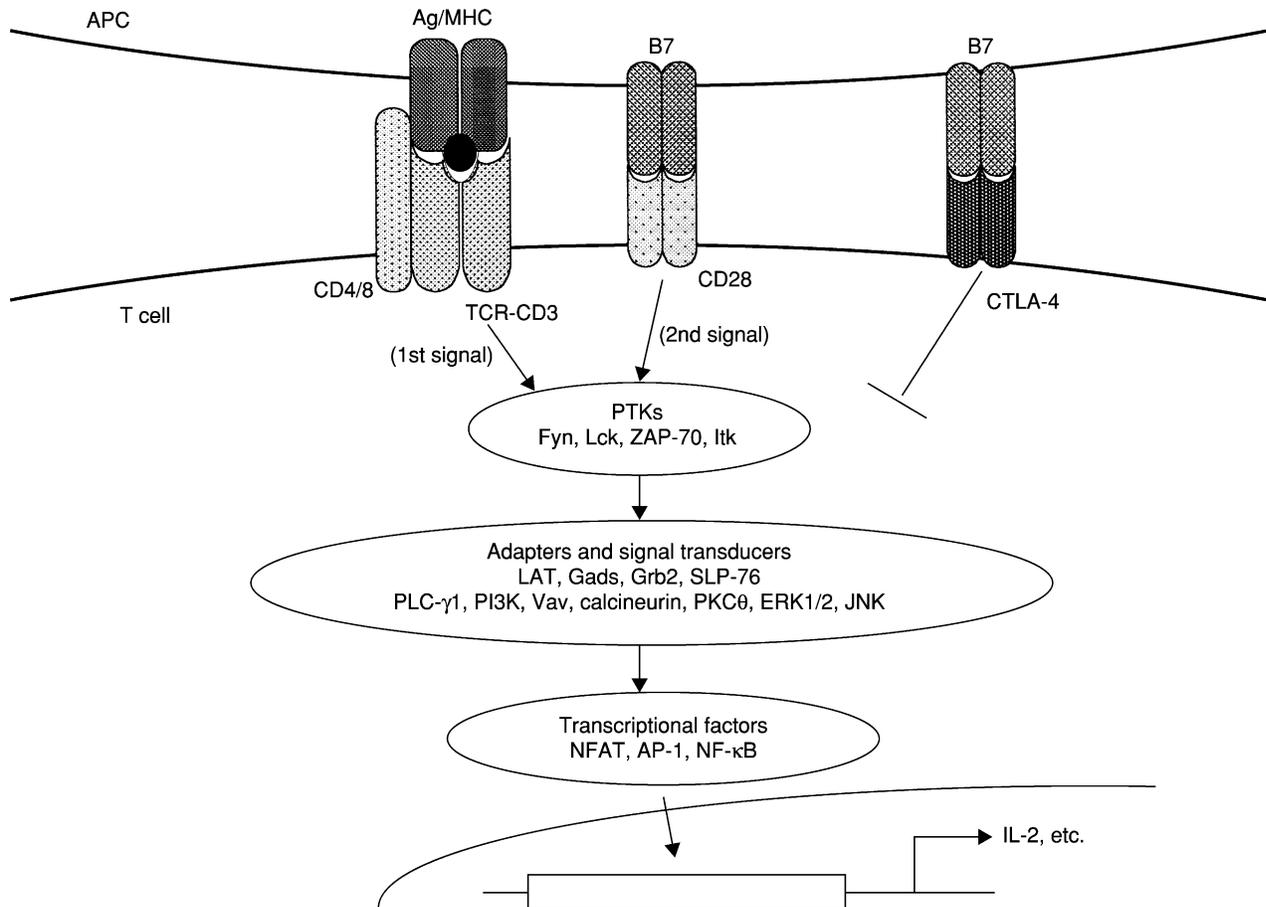


Figure 1 Regulation of IL-2 production through TCR and CD28/CTLA-4 signaling.

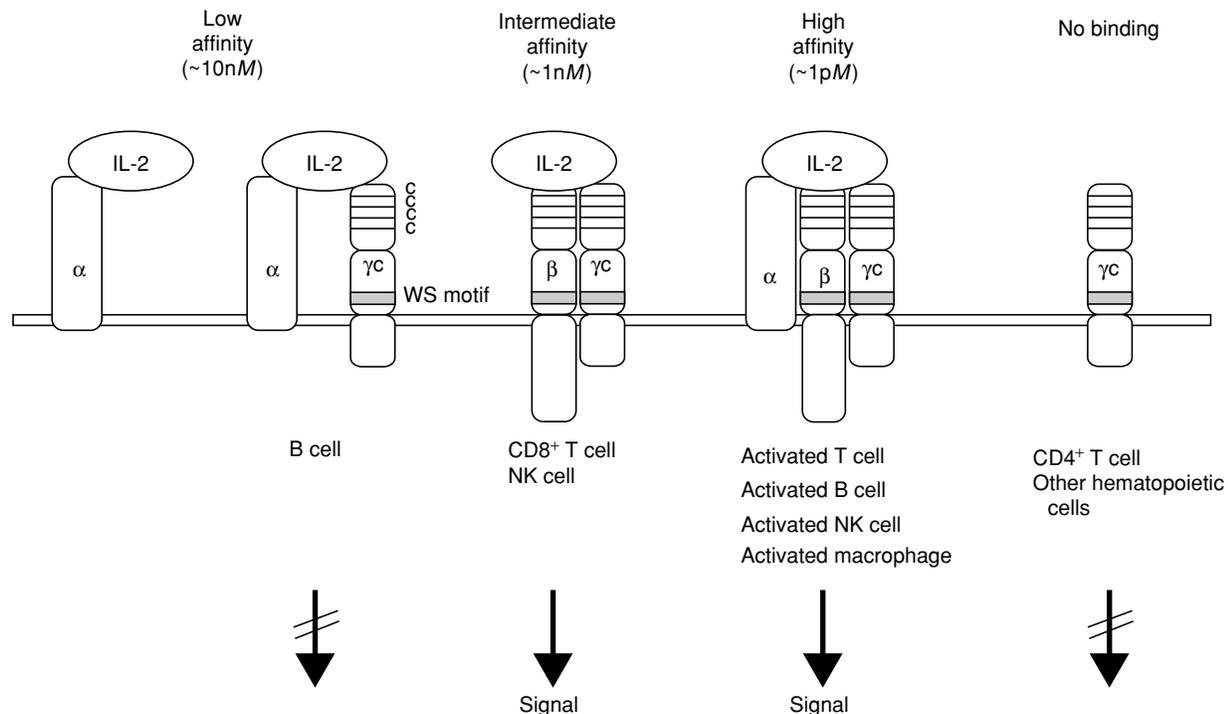
Gads, Grb2, and SLP-76 to catalytic enzymes such as phospholipase  $C\gamma 1$ , phosphatidylinositol 3-kinase, and Vav. These enzymes activate calcineurin, protein kinase  $C\theta$ , ERK1/2, JNK, and other molecules, leading to further activation of a set of transcriptional factors, including nuclear factor of activated T cells, activator protein 1, and nuclear factor  $\kappa B$ , which are involved in IL-2 promoter activation. In addition to these transcriptional regulators, stabilization of IL-2 mRNA is another important role of the second signal through CD28. This leads to greatly enhanced IL-2 production in the presence of the second signal. These activation signals also induce the expression of the CTLA-4 molecule, which is another B7 receptor with much higher affinity for the B7 molecule than CD28, on the activated T cell surface. Unlike CD28, CTLA-4 induces the T cell inactivation signal and down-regulates IL-2 production to limit the excess immune response. Clinically, the immunosuppressive drugs cyclosporin A and FK506 inhibit IL-2 production

by disrupting the TCR-induced signal cascade on calcineurin.

Interestingly, TCR signals in the absence of the second signal do not induce IL-2 production but instead lead to inactivation of the T cells, resulting in an anergic state.

## INTERLEUKIN-2 RECEPTORS

Interleukin-2 receptor  $\alpha$  (IL-2R $\alpha$ ) chain (CD25), which was initially identified as IL-2R, is a type I membrane protein whose molecular weight is 55 kDa. The expression of the IL-2R $\alpha$  chain is tightly restricted to peripheral activated T and B cells, triple-negative thymocytes, and bone marrow pre-B cells. Whereas the IL-2R $\alpha$  chain itself has no signal-transducing activity, functional high-affinity IL-2R is composed of the IL-2R $\alpha$  chain and other two indispensable subunits, the IL-2R $\beta$  chain and the IL-2R $\gamma$



**Figure 2** Schematic structure of the IL-2 receptor complex on human peripheral blood cells. Two pairs of cysteine residues and WS motifs that are conserved in the cytokine receptor family are indicated as horizontal lines and shaded boxes, respectively.

chain (Fig. 2). IL-2R $\beta$  chain (CD122) and IL-2R $\gamma$  chain (CD132) are also type I membrane proteins of 75 and 64 kDa, respectively, and they belong to the cytokine receptor family, which is characterized by the presence of a cytokine receptor domain of approximately 210 amino acids in the extracellular region. This domain is separated by two homologous type III fibronectin subdomains. The first subdomain is characterized by the presence of two pairs of cysteine residues at the N-terminal part and the second subdomain has a consensus tryptophan-serine-X-tryptophan-serine motif (WS motif). Several patterns of IL-2 receptor complex and human peripheral blood cells expressing these receptors are shown in Fig. 2. Cytokine receptor family members are also characterized by the absence of intrinsic tyrosine kinase activity. Non-receptor-type tyrosine kinases Janus kinase 1 (JAK1) and JAK3 are associated with IL-2R $\beta$  chain and IL-2R $\gamma$  chain, respectively, and these PTKs trigger the IL-2-induced signaling cascade. Since the IL-2R $\gamma$  chain was found to be shared with other cytokines, including IL-2, IL-4, IL-7, IL-9, IL-15, and IL-21, it is called the common  $\gamma$  chain ( $\gamma$ c). Furthermore, mutations of the  $\gamma$ c gene cause human X-linked severe combined immunodeficiency, which is characterized by a complete or profound T and NK cell defect.

## BIOLOGICAL ACTIVITIES OF INTERLEUKIN-2

IL-2 was identified as a growth factor for T cells activated by antigens and its biological functions have been extensively investigated. IL-2 is a survival factor of activated T cells and acts by inducing the expression of anti-apoptotic proteins, such as Bcl-2 and Bcl-X<sub>L</sub>. In addition to its growth promotion and survival factor functions, IL-2 induces CTL activation and differentiation. In addition to its effects on T cells, IL-2 can promote the growth and differentiation of activated B cells. During B cell differentiation, IL-2 is involved in the regulation of immunoglobulin J-chain expression. IL-2 also augments the cytolytic activity of NK cells.

Previous reports have shown a different dimension to IL-2. Negative actions of IL-2 on T cells, such as an induction of activation-induced cell death (AICD), have been demonstrated. Indeed, mice with defective IL-2 or a functional IL-2R complex do not show signs of immunodeficiency. However, an accumulation of hyperactivated T cells was observed and various autoimmune diseases developed in these mice. Thus, the major role of IL-2 *in vivo* is thought to be the termination of T cell responses and maintenance of tolerance

by inducing AICD, leading to the homeostasis of T lymphocytes.

## THERAPEUTIC USE OF INTERLEUKIN-2

Human PBL treated with IL-2 *in vitro* show cytotoxic activity against some tumor cells. These MHC-independent cytotoxic cells are called LAK cells. Furthermore, IL-2-activated tumor-infiltrating lymphocytes (TIL) from human tumors seem to have more a potent killing activity against autologous tumors *in vitro* than LAK cells. In some cases of advanced metastatic melanoma and renal cell carcinoma, high-dose IL-2 therapy, with or without adoptive immunotherapy using LAK, TIL, or chemotherapy, is effective at regressing tumor growth. However, further clinical trials are required to improve survival rates and the toxicities accompanying high-dose IL-2 treatment.

In addition to clinical research on IL-2 treatment for malignant tumors, clinical trials using IL-2 as adjunctive therapy with highly active antiretroviral therapy (HAART) against human immunodeficiency virus (HIV)-infected patients is under way.

Immune-based therapy with IL-2 combined with HAART increased CD4<sup>+</sup> T cell counts, leading to improved immune responses, and it is hoped that it may improve the clinical symptoms and survival of HIV-infected patients.

## See Also the Following Articles

Cytokine Actions, Cellular Mechanism of • Cytokine Receptors • Cytokines, Constitutive Secretion • Interleukin-6 • Nuclear Factor- $\kappa$ B and Glucocorticoid Receptors

## Further Reading

- Atkins, M. B. (2002). Interleukin-2: Clinical applications. *Semin. Oncol.* **29**, 12–17.
- Lin, J. X., and Leonard, W. J. (1997). Signaling from the IL-2 receptor to the nucleus. *Cytokine Growth Factor Rev.* **8**, 313–332.
- Rosenberg, S. A. (2001). Progress in the development of immunotherapy for the treatment of patients with cancer. *J. Int. Med.* **250**, 462–475.
- Schimpl, A., Berberich, I., Kneitz, B., Kramer, S., Santner-Nanan, B., Wagner, S., Wolf, M., and Hunig, T. (2002). IL-2 and autoimmune disease. *Cytokine Growth Factor Rev.* **13**, 369–378.
- Sugamura, K., Asao, H., Kondo, M., Tanaka, N., Ishii, N., Nakamura, M., and Takeshita, T. (1995). The common  $\gamma$ -chain for multiple cytokine receptors. *Adv. Immunol.* **59**, 225–277.

Table I Summary of Molecular Properties of IL-6 and Its Receptors

Molecule	Species	Amino acid length		Homology to human (percentage)	Molecular mass (kDa)		Chromosome	Database accession
		Precursor	Mature		Predicted	Observed		
IL-6	Human	212	185	—	20.8	21–28	7p21	X04602
	Mouse	211	187	42	—	—	5	J03783
	Rat	211	187	41	—	—	4	M26744
IL-6R $\alpha$	Human	468	449	—	49.9	80	1q21	X12830
	Mouse	460	441	54	—	—	3	X51975
	Rat	462	443	53	—	—	2	M58587
gp130	Human	918	896	—	101.0	130–150	5q11	M57230
	Mouse	917	895	77	—	—	13	M83336
	Rat	918	896	78	—	—	2q14–q16	M92340

human AF372214, mouse M20572) (see also Table I). Several potential transcriptional control elements, glucocorticoid-responsive element (GRE), an AP-1 binding site, multiple response element (MRE), a *c-fos* serum-responsive element (*c-fos* SRE) homology, a *c-fos* retinoblastoma control element (RCE) homology, a cyclic AMP (camp)-responsive element (CRE), a nuclear factor for IL-6 (NF-IL6, also known as CCAAT/enhancer-binding protein  $\beta$  [C/EBP $\beta$ ] binding site), and an NF- $\kappa$ B-binding site, were identified within the IL-6 promoter region. Several polymorphisms within IL-6 gene and its promoter region have been reported. One of these, -174C/G promoter region polymorphism, is suggested to associate with serum level IL-6 concentration and pathogenesis of some diseases such as rheumatoid arthritis and human immunodeficiency virus-dependent Kaposi's sarcoma.

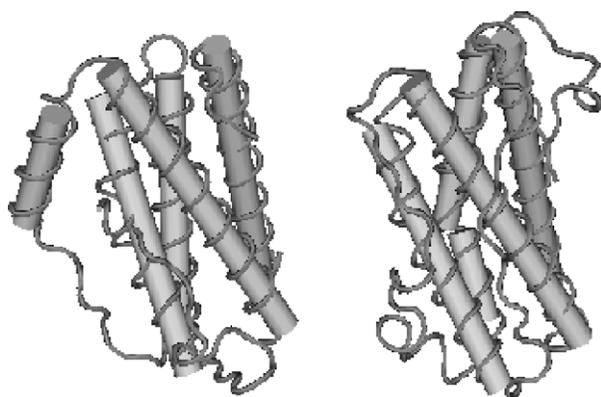


Figure 1 Structure of IL-6 (left) and LIF (right). Protein backbones were drawn in worm style.  $\alpha$ -helices are depicted as tubes. Modified figure based on National Center for Biotechnology Information structure database.

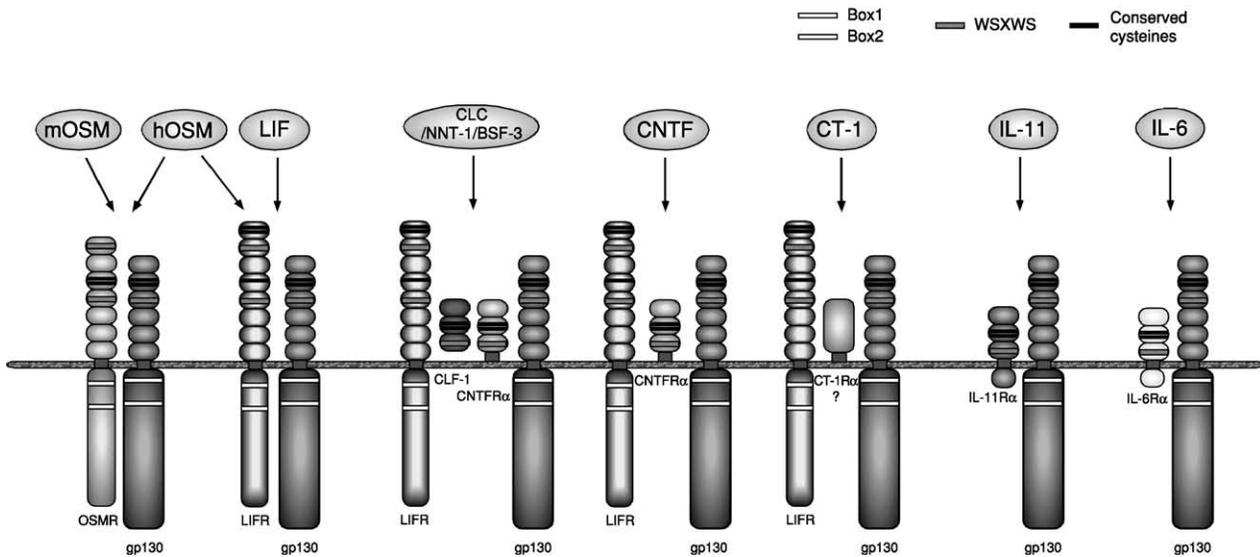
## RECEPTOR FOR IL-6 AND SIGNALS

### Receptors for IL-6 Family Cytokines

The IL-6 receptor consists of an  $\alpha$  chain (IL-6R $\alpha$ ) (CD126) and gp130 (CD130) (Fig. 2). Extensive mutagenesis studies revealed that IL-6 has three distinct receptor-binding sites referred to as sites 1, 2, and 3. Site 1 is important for the binding of IL-6R $\alpha$ . Sites 2 and 3 are used for recruitment of gp130. Both IL-6R $\alpha$  and gp130 are required for the high-affinity binding site for IL-6. IL-6 family cytokines, including IL-11, LIF, CNTF, cardiotropin-1 (CT-1), OSM, and cardiotropin-like cytokine/neurotrophin-1/B-cell stimulatory factor-3 (CLC/NNT-1/BSF-3), share gp130 as a signal-transducing subunit for the receptor complex (Fig. 2). Functional redundancy is one of the characteristic features of cytokines. For example, IL-6, LIF, and IL-11 can induce acute phase protein production in hepatic cells or the differentiation of the mouse leukemic cell line M1. Sharing the common signal transducer gp130 is one of the mechanisms through which the functional redundancy of the IL-6 family cytokines is mediated. The binding of IL-6 family cytokines leads to the homodimerization of gp130 or the heterodimerization of gp130 with other gp130-related receptors (i.e., the LIF receptor and the OSM receptor). IL-6R $\alpha$ , IL-11R $\alpha$ , CNTFR $\alpha$ , and cytokine-like factor-1 (CLF) are not thought to transmit signals given that they have very few or no amino acid residues in their cytoplasmic domains.

### Receptor Conversion Model

A soluble form of cytokine receptors often functions as a competitive inhibitor for the ligand. However, sIL-6R $\alpha$  (soluble form of IL-6R $\alpha$ , which can be released



**Figure 2** Illustration showing that gp130 is a common signal transducer for IL-6 family cytokines. Receptor complexes of IL-6 family cytokines consist of specific  $\alpha$  chain (IL-6R $\alpha$ , IL-11R $\alpha$ , CNTFR $\alpha$ , and CT-1R $\alpha$ ) and  $\beta$  chain (gp130, LIFR, and OSMR), which mediates signaling. Individual domains of extracellular regions of receptors are presented as globular drawings. The binding of IL-6 to the receptor induces the formation of hexameric ligand–receptor complex composed of two each of IL-6, IL-6R $\alpha$ , and gp130. Precise stoichiometry for other cytokines are not well studied. CT-1R $\alpha$  has not yet been identified. m, mouse; h, human.

by receptor shedding or secretion after translation of an alternatively spliced mRNA), when complexed with IL-6, can activate signals in the cells expressing only gp130 receptor subunit on which IL-6 alone cannot act. CLF was found as a secreted protein; therefore, it was considered as a ligand rather than as a receptor. However, CLF can form a complex with CLC/NNT-1/BSF-3 and acts as a ligand to signals through CNTFR $\alpha$ , LIFR, and gp130. This system shows a striking similarity to that of IL-12 (Fig. 3). IL-12 consists of a disulfide heterodimer of a 35-kDa (p35) subunit, which is a cytokine, and a 40-kDa (p40) subunit, which is a soluble form of the cytokine receptor. Therefore, IL-12, generally accepted as a kind of cytokine, is actually a complex of a cytokine and a soluble form of its receptor. Thus, in certain cases of cytokine and receptors, the complex of soluble form of cytokine receptor and its ligand acts as a novel cytokine.

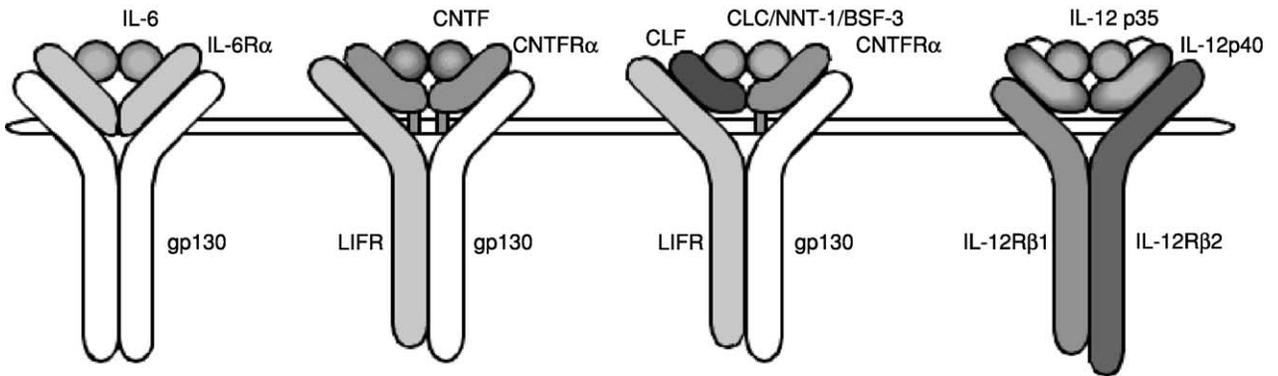
### gp130 and Signaling Molecules

gp130 is a single transmembrane glycoprotein with a molecular mass of 130 to 150 kDa. Based on the presence of conserved domains, gp130 is defined as belonging to the type I cytokine receptor superfamily. The other receptors for the IL-6 family cytokines, including the IL-6R $\alpha$  chain, all belong to the type I

cytokine receptor superfamily, with the exception of CNTFR $\alpha$ , which is a GPI-anchored receptor. In its intracellular region, gp130 contains regions known as box1 (I<sup>651</sup>WPNVDP of human sequence) and box2 (V<sup>691</sup>SVVEIEANDKKP), which are conserved among the members of the type I cytokine receptor superfamily (Fig. 2). The homodimerization or heterodimerization of gp130 results in the activation of the gp130-associated JAKs (JAK1, JAK2, and TYK2) through these two regions. Subsequently, gp130 is phosphorylated on tyrosines and the phosphorylated gp130 recruits signal-transducing molecules, such as SHP2 and STAT3, and activates them to transmit signals downstream (Fig. 4).

Any one of four tyrosines (Y767, Y814, Y905, or Y915) of gp130, each of which has a glutamine at position +3 of the tyrosine motif (YXXQ), is required for the tyrosine phosphorylation of STAT3. Y905 and Y915 in the YXPQ motif are required for the tyrosine phosphorylation of STAT1. After tyrosine phosphorylation, STAT3 forms a homo- or heterodimer with STAT1, enters the nucleus, and regulates the expression of a set of genes.

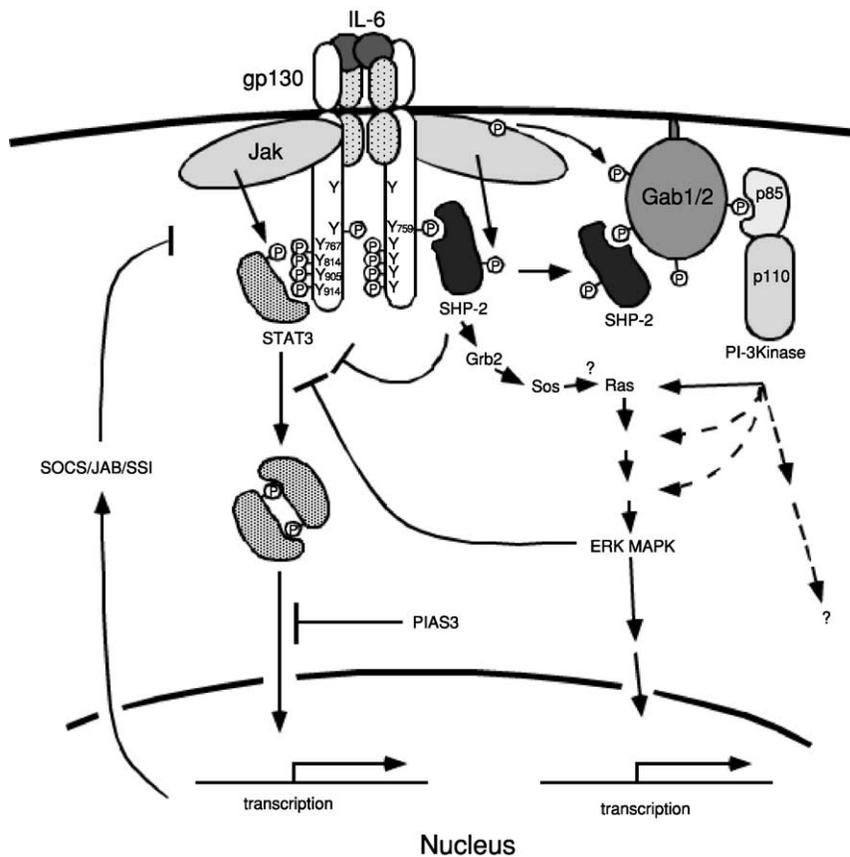
On stimulation of gp130, SHP2 is tyrosine phosphorylated and interacts with Grb2, which is constitutively associated with Sos, a GDP–GTP exchanger for Ras. Gab family proteins are tyrosine phosphorylated and interact with SHP2 and PI-3 kinase in



**Figure 3** Illustration of ligand–receptor system of IL-12 resembling that of IL-6, CNTF, and CLC/NNT-1/BSF-3. IL-12 is a heterodimer of 70 kDa, consisting of disulfid linked subunits of p35 and p40. p35 is significantly homologous to IL-6 and is predicted to have bundled four  $\alpha$ -helices structure. p40 is homologous to CNTFR $\alpha$  and IL-6R $\alpha$ . IL-12R $\beta$ 1 and IL-12R $\beta$ 2, which contain conserved cysteine sequences, WSXWS motifs, and box1 and box2 motifs, are closely related to gp130 and LIFR.

response to gp130 stimulation. A mutation of tyrosine759 of gp130 reduces the interactions between SHP2 and Grb2 and between SHP2 and Gab1 and

diminishes the activation of ERK MAP kinase, suggesting that SHP2 mediates signals to the ERK MAP kinases through Grb2 and the Gab proteins.



**Figure 4** Schematic representation of IL-6 signaling. Distinct tyrosine residues of gp130 are involved in different signal transduction pathways. Y759 elicits SHP2-mediated pathways, whereas Y767, Y814, Y905, and Y914 each can elicit STAT3-mediated pathways. Y759 was also shown to interact with SOCS3. SHP2, MAPK, SOCS, and PIAS proteins were shown to be involved in negative regulation of signals.

STAT3 is required for the induction of SOCS1 and SOCS3 expression in gp130 signaling. Members of the suppressor of cytokine signaling (SOCS) family, also referred to as JAK binding (JAB) and as STAT-induced STAT inhibitor (SSI), are characterized by a conserved SOCS box and an SH2 domain. SOCS3 may act as a negative regulator by interacting with gp130 directly, whereas SOCS1 inhibits the signaling by interacting with JAKs. Mutating tyrosine 759 of gp130 resulted in enhancement of gp130-mediated signaling both *in vivo* and *in vitro*, suggesting that SOCS3, in addition to SHP2, is involved in the tyrosine 759-dependent negative regulation of gp130 signaling.

An Src family tyrosine kinase, Hck, associates with an acidic region of gp130. Protein inhibitor of activated STAT3 (PIAS3) interacts directly with phosphorylated STAT3 and reduces its DNA-binding activity, thereby inhibiting the transcription of its target genes.

## BIOLOGICAL FUNCTIONS OF IL-6

### Biological Functions Observed *in Vitro*

IL-6 was identified as one of the factors acting on B cells to induce immunoglobulin production. Besides its role in inducing B-cell differentiation, IL-6 induces T-cell growth and differentiation, differentiation of the myeloid leukemic cell line M1 into macrophages, megakaryocyte maturation, neural differentiation of PC12 cells, development of osteoclasts, and acute phase protein synthesis in hepatocytes. IL-6 acts as a growth factor for myeloma/plasmacytoma, keratinocytes, mesangial cells, renal cell carcinoma, and Kaposi's sarcoma. STAT3 signaling plays a central role in several biological functions of the IL-6 family cytokine.

### Acute Phase Response *in Vivo*

A large body of work using transgenic and knockout mice has addressed the *in vivo* functions of IL-6 family cytokines (Table II). These experiments confirmed some of the biological functions of IL-6 observed *in vitro*. For example, IL-6-deficient mice showed impaired acute phase response. This result clearly demonstrated the importance of IL-6 in acute phase response. Furthermore, blocking the individual signals elicited from gp130 by the knock-in technique showed that STAT3 plays the central role in acute phase response and that SHP2 negatively regulates it.

### Immune Responses *in Vivo*

The significance of IL-6 in B-cell development was implicated in experiments *in vitro*. IL-6 transgenic mice of C57BL/6 origin develop massive plasmacytosis. Abnormal B-cell expansion was also observed in the IL-6 and sIL-6R $\alpha$  double transgenic mice. All of these data confirmed the importance of IL-6 in B-cell development. Because gp130 is involved in both cell growth and the differentiation of B cells into plasma cells through STAT3 activation, chronic B-cell activation by deregulated IL-6 expression may be one of the major pathogenesises of these diseases. In support of these observations, IL-6-deficient mice showed impaired antigen-specific antibody production and mucosal immunoglobulin A (IgA) production. IL-6-deficient mice also display dysfunction in diverse systems (e.g., type 1 helper T-cell [Th1] development) and protection against *Listeria monocytogenes* and *Mycobacterium tuberculosis*.

### Hematopoiesis

IL-6 and some colony-stimulating factors, such as IL-3, M-CSF, and GM-CSF, synergistically promote the growth of hematopoietic progenitor cells *in vitro*. Consistently, IL-6-deficient mice showed decreased number of hematopoietic progenitor cells by testing *in vivo* spleen colony-forming assays (Table II). IL-6 induces the *in vitro* megakaryopoiesis synergistically with IL-3. The observations that IL-6-deficient mice had reduced megakaryocyte progenitor cells and that IL-6 transgenic mice had increased mature megakaryocytes support *in vitro* results.

### Bone Metabolism

Osteoporosis is a major cause of morbidity in older people. A decrease in sex hormones with aging is suggested to contribute the age-related increase of IL-6 and to stimulate osteoclastic bone resorption. In the presence of dexamethasone, IL-6 could induce osteoclast formation *in vitro*, suggesting that the increase of bone resorption caused by estrogen withdrawal may be explained by an effect of IL-6. The finding that IL-6-deficient mice are resistant to bone loss induced by ovariectomy supports this view.

### Effects on Nervous System

Central nervous system (CNS) cells express both IL-6 and IL-6R $\alpha$ . To demonstrate the role of IL-6 in CNS, transgenic mice expressing IL-6 in CNS

Table II Phenotypes of Knockout Transgenic, and Knock-in Mice

	Phenotypes
IL-6 KO	Reduced TD antibody production and mucosal IgA response BFU-E ↑, CFU-S ↓ Th1 response ↓ Acute phase response ↓ Bone loss by estrogen depletion Impaired hepatocyte regeneration
IL-6 TG	Plasmacytosis, plasmacytoma
sIL-6R $\alpha$ TG	Acute phase response ↑
IL-6/sIL-6R $\alpha$ double TG	Plasmacytoma Extramedullary hematopoiesis Hepatocellular hyperplasia
gp130 (dominant negative) TG	Reduced TD antibody production Normal hematopoiesis
gp130 KO	d12.5-perinatal lethality BFU-E, CFU-GM ↓ Heart abnormality
gp130 (postnatally) KO	Not severe defect on antibody response Thrombopoiesis ↓ Acute phase response ↓ Defect in lung, heart, liver, and nerves
gp130 (heart specific) KO	Normal heart development
gp130 (SHP2 signal deficient) KI	Augmented TD antibody production Normal hematopoiesis Th1-type cytokine production Acute phase response ↑ Arthritis dependent on age
gp130 (STAT3 signal deficient) KI *point mutation	Perinatal lethality Reduced TD antibody production Normal hematopoiesis Th2-type cytokine production
gp130 (STAT3 signal deficient) KI *truncation	Impaired TD antibody production and mucosal IgA response Acute phase response ↓ Gastrointestinal ulceration Severe joint disease
gp130 (all signal deficient) KI	Perinatal lethality Reduced TD antibody production Normal hematopoiesis

*Note.* KO, knockout; TG, transgenic; KI, knock-in; d, days postcoincidence; TD, thymus dependent; BFU-E, erythroid burst-forming unit; CFU-GM, granulocyte-macrophage colony-forming unit; CFU-S, spleen colony-forming unit.

were generated. These mice showed reactive gliosis throughout the CNS. Furthermore, significant neurodegeneration, including the loss and damage of neurons in the hippocampus formation and cerebellum, was observed.

The IL-6 family cytokines may play an important role in nerve regeneration after trauma *in vivo*. In double transgenic mice expressing both IL-6 and IL-6R $\alpha$ ,

regeneration of the autoximized nerve was strikingly accelerated compared with nontransgenic controls.

## DISEASES

### IL-6-Related Diseases

A possible involvement of IL-6 in disease was first seen in patients with cardiac myxoma. These patients

have a variety of autoimmune symptoms, including hypergammaglobulinemia, the presence of autoantibodies, and an increase in acute phase proteins, all of which disappear after the resection of the tumor cells. The finding that myxoma cells produce high levels of IL-6 led to speculation that overproduction of IL-6 might play a critical role in the development of autoimmune symptoms. Rheumatoid arthritis patients have high levels of IL-6, LIF, IL-11, and OSM in their synovial fluid, and increased amounts of sIL-6R $\alpha$  are implicated in the pathogenesis of juvenile chronic arthritis. The possible involvement of IL-6 in autoimmune diseases is suggested; collagen-induced arthritis (CIA) is mild in the IL-6-deficient mice. Neutralizing anti-IL-6R $\alpha$  antibody ameliorates the joint disease in murine CIA. Experimental autoimmune encephalomyelitis is suppressed in the IL-6-deficient mice. Pristane-induced autoantibody production is dependent on IL-6. Patients with multiple myeloma, a malignant plasma cell tumor, have higher than normal levels of IL-6, sIL-6R $\alpha$ , and OSM, suggesting the involvement of IL-6 in plasmacytogenesis. In fact, IL-6 transgenic mice developed plasmacytoma in the Balb/c genetic background. Furthermore, pristane-induced plasmacytogenesis is dependent on IL-6. The overexpression of IL-6 has also been shown in connection with other diseases, including osteoporosis, Castleman's disease, and several autoimmune diseases. Lamina propria cells from patients with Crohn's disease or ulcerative colitis show increased production of IL-6. Blocking IL-6 signaling suppressed the experimental colitis in various animal models of Crohn's disease. Interestingly, human herpes virus 8 (HHV8 or Kaposi's sarcoma-associated herpes virus), which has been found in Kaposi's sarcoma lesions, carries the viral counterpart of IL-6, vIL-6, and vIL-6 is implicated in promoting the course of Kaposi's sarcoma.

## Therapies

Several approaches have been taken to interfere with the formation of the receptor complex to prevent the pathogenic activity of IL-6 family cytokines. These approaches include the design of gp130 antagonist based on the mutational studies and the development of monoclonal antibody targeting IL-6 and IL-6R $\alpha$ . A cocktail of three different anti-IL-6 antibodies or a chimeric anti-IL-6 antibody consisting of the antigen-binding variable region of the murine

anti-IL-6 antibody and the constant region of a human IgG1  $\kappa$  immunoglobulin have been given to patients with rheumatoid arthritis, multiple myeloma, and Castleman's disease. These strategies enabled the rapid clearance of serum IL-6. The endogenous IL-6 production never reached its pretreatment value during the treatment period, and C-reactive protein (CRP) levels decreased to below the detection level in nearly every patient. No human antibodies to the chimeric antibody were induced. Thus, this therapy greatly improved the patients' disease status with low toxicity, low immunogenicity, and a long half-life. A similar approach was taken with a humanized anti-human IL-6R $\alpha$  antibody. This therapy improved the conditions of patients with Castleman's disease and certain autoimmune diseases. Therefore, these therapies using chimeric antibodies could be useful tools in treating diseases that are caused by the deregulation of IL-6.

## See Also the Following Articles

Cytokine Actions, Cellular Mechanism of • Cytokine Receptors • Cytokines, Evolutionary Aspects and Functions • Interleukin-2

## Further Reading

- Grotzinger, J., Kernebeck, T., Kallen, K. J., and Rose-John, S. (1999). IL-6 type cytokine receptor complexes: Hexamer, tetramer, or both? *Biol. Chem.* **380**, 803–813.
- Heinrich, P. C., Behrmann, I., Muller-Newen, G., Schaper, F., and Graeve, L. (1998). Interleukin-6-type cytokine signaling through gp130/Jak/STAT pathway. *Biochem. J.* **334**, 297–314.
- Hirano, T. (1998). Interleukin 6 and its receptor: Ten years later. *Int. Rev. Immunol.* **16**, 249–284.
- Hirano, T. (1999). Molecular basis underlying functional pleiotropy of cytokines and growth factors. *Biochem. Biophys. Res. Comm.* **260**, 303–308.
- Hirano, T., Ishihara, K., and Hibi, M. (2000). Roles of STAT3 in mediating the cell growth, differentiation, and survival signals relayed through the IL-6 family of cytokine receptors. *Oncogene* **19**, 2548–2556.
- Hirano, T., Nakajima, K., and Hibi, M. (1997). Signaling mechanisms through gp130: A model of the cytokine system. *Cytokine Growth Factor Rev.* **8**, 241–252.
- Ihle, J. N. (2001). The Stat family in cytokine signaling. *Curr. Opin. Cell. Biol.* **13**, 211–217.
- Kopf, M., Ramsay, A., Brombacher, F., Baumann, H., Freer, G., Galanos, C., Gutierrez-Ramos, J. C., and Kohler, G. (1995). Pleiotropic defects of IL-6-deficient mice including early hematopoiesis, T and B cell function, and acute phase responses. *Ann. N.Y. Acad. Sci.* **762**, 308–318.
- Leonard, W. J., and O'Shea, J. J. (1998). Jaks and STATs: Biological implications. *Annu. Rev. Immunol.* **16**, 293–322.

**Table 1 Etiopathogenic Factors Associated with IUGR**

<b>Maternal conditions</b>
Maternal starvation and chronic malnutrition
Maternal disease
Hypertension
Severe chronic infections, systemic disease (mainly with hypoxemia), diabetes mellitus
Anatomical abnormalities (uterus)
Environmental factors
Smoking, alcohol, substance abuse
High altitude, low socioeconomic status
Other determinants
Maternal height, prepregnancy weight, pregnancy weight gain
Ethnicity
Low maternal age
Prior low birth weight, multiple gestation, short interpregnancy interval
Reproductive technologies
<b>Placental factors</b>
Vascular abnormalities, abnormal placentation
Chromosomal mosaicism
<b>Fetal conditions</b>
Genetic
Chromosomal abnormalities
Inborn errors of metabolism
Paternal determinants
Polymorphism genetic variations
Malformations
Infections

## Postnatal Growth and Development

### **Growth in Infancy**

At birth, premature neonates (who account for 30% of neonates showing IUGR) tend to show more symmetric IUGR than asymmetric IUGR. Neonates from alcohol-consuming mothers demonstrate a more severe degree of symmetric growth retardation. Nevertheless, there is no convincing evidence that either the auxological data or the body proportionality at birth among infants with a given degree of growth retardation varies as a function of the cause of that growth retardation.

Longitudinal studies of postnatal growth in infants born SGA have shown that catch-up growth (recognizable by an increase in height velocity) is identified immediately after birth, with a variability of magnitude among individuals. Although the majority of catch-up growth occurs during the first 6 months, catch-up growth for height is often more gradual and can take up to 2 years, followed by a normal pattern of development. However, approximately 10% of the children

born SGA will remain short ( $<-2$  SD) at 2 years of age, and most will have a higher risk of short stature later in life.

Mechanisms regulating catch-up growth are still poorly understood. Improvements in perinatal care and nutrition are largely responsible for the decrease in the incidence of growth failure, which fell from 25 to 10% during the past 25 years or so.

Birth weight, birth length, gestational age, and parental height have been identified as factors influencing catch-up growth. Although the serum levels of IGF-1, insulin-like growth factor-binding protein-3 (IGFBP-3), and leptin are significantly reduced and growth hormone (GH) is significantly increased at birth compared with those of normal birth-weight neonates for all gestational ages studied, none of these hormonal parameters is predictive of catch-up growth. These hormonal parameters are within normal limits at 1 month of age and thereafter. In addition, no relationship has been found between hormones levels at birth and any given etiology of IUGR. However, lowered serum IGF-1, IGFBP-3, and leptin are found in IUGR showing a low PI, suggesting that intrauterine nutrition strongly influences the levels of these parameters. During early neonatal life, serum IGF-1 levels also appear to be regulated by nutritional factors given that an increase in serum IGF-1 levels during the first 3 months of life has been found to be correlated with weight gain during that time and, therefore, with the increased nutrient intake after intrauterine nutrient deprivation.

### **Growth in Childhood until Adulthood**

After the end of puberty, the mean adult height, weight, and head circumference are significantly lower in the SGA population than in the normal birth-weight population. When adjusted for target height, a significant deficit ( $-4$  cm) in final height is found for both male and female individuals, with 14% of them presenting with short final stature ( $<-2$  SDS) compared with individuals of normal birth weight. Parental height and birth length SDS are the strongest determinants of individual final height and height gain in individuals born SGA. No other variables, such as sex, gestational age (37–42 weeks), birth-weight SDS, PI at birth, and risk factors during pregnancy associated with IUGR (e.g., pregnancy-induced hypertension, smoking, a history of SGA in offspring), are seen to be predictive of adult height. Thus, the major influence of parental height (with a greater influence of maternal height) on adult height is evident.

Data concerning the onset, duration, and pace of puberty are not available in large series of individuals.

It seems that puberty in girls occurs at a normal age given that their mean age of menarche is comparable to that of the normal birth-weight population. It has also been shown that the differences between the final height SDS and the prepubertal height SDS are similar for SGA and normal birth-weight groups, suggesting that final height in the SGA population is not influenced by puberty and that the pubertal growth spurt is normal in this population.

### **Growth Hormone Treatment of Short Stature Children Born SGA**

Therapy with recombinant human GH in a supraphysiological dose (0.06 mg/kg/day) has been shown to increase significantly growth velocity in non-GH-deficient prepubertal children with marked short stature and born SGA. In most studies, the first-year treatment induces a mean result of a nearly doubling of the pretreatment growth rate. During the subsequent years of GH treatment, although height velocity is increased significantly compared with baseline values, therapeutic efficacy tends to diminish. Nevertheless, the growth rate after 3 years of treatment remains higher than the baseline growth rate prior to treatment. In most studies, mean height is increased by nearly 2.0 SDS over the first 3-year treatment period in prepubertal children. It has been demonstrated that after 5 to 6 years of GH treatment, nearly every child achieves a height within the normal range and in conformity with the target height SD score without excessive bone maturation advancement. The amplitude of the initial growth response to treatment depends on the GH dose (positive dose-response correlation) and the age of the child (with a younger child being more responsive). However, the impact of early treatment on adult height remains to be determined.

Current clinical experience has shown that on cessation of treatment during the prepubertal period, some children maintain a normal growth rate, whereas others show a slowing in growth rate resulting in a lower final height prognosis compared to those at the end of the treatment period. This suggests that a second GH treatment period may be necessary.

Treatment of short children born SGA does not have a current optimal treatment protocol. A 0.5-SD increment in adult height, even when treatment is given later in patients close to pubertal onset, has been demonstrated. Continuous versus intermittent regimens, as well as higher versus lower dose GH regimens, are under investigation to evaluate the safety and efficacy of the growth response so as to optimize final height of short children born SGA.

The effects of GH are not limited to skeletal growth. Indeed, GH has a profound effect on carbohydrate, lipid, and protein metabolism. In these children, GH treatment is well tolerated, particularly with regard to glucose as evaluated by the levels of glycosylated hemoglobin and glycemia during an oral glucose tolerance test. A potential positive metabolic effect, with an increase in muscle mass observed in these children following 3 years of GH treatment, has also been shown. However, although a reversible decrease in insulin sensitivity after the interruption of treatment has been shown, there is concern regarding the possibility of persistent effects of long-term GH treatment on insulin sensitivity in these children. Further studies are required to define a treatment regimen that will optimize safely the long-term outcome on final height and on metabolic aspects.

### **Neurodevelopmental Outcome**

Subtle neurological and developmental disabilities have been described in individuals born SGA, and a small head circumference at birth has been found to be associated with poorer neurodevelopmental outcome during childhood. Moreover, a small adverse effect of IUGR on cognitive performance has been shown during childhood and early adulthood, even after adjustment for head circumference and for parental socioeconomic and educational status.

The exact causes of these developmental abnormalities are not known with certainty. Experimental data show that undernutrition during vulnerable periods of brain development may have long-term effects on brain architecture and differentiation that can affect learning and memory. Whether prolonged IGF-1 deficiency induced by fetal undernutrition *in utero* might alone cause deleterious effects on normal brain and cognitive development remains to be explored. Furthermore, the role of recurrent episodes of asymptomatic hypoglycemia that might be observed during the neonatal period should also be investigated, as should measures to prevent hypoglycemia during this critical period so as to minimize the risk of subsequent neurodevelopmental deficits.

### **Long-Term Metabolic Consequences**

Over the past decade or so, evidence that demonstrates a significant association between reduced fetal growth and the later development of metabolic and cardiovascular diseases has grown. However, despite several hypotheses that have been proposed, the mechanisms

underlying this unexpected association remain to be clarified.

**Cardiovascular and Metabolic Complications Associated with Reduced Fetal Growth: Knowledge from Epidemiological Studies**

The unexpected and long-term deleterious effect of reduced fetal growth was initially observed by Barker and colleagues. This group first reported that a low birth weight was significantly associated with an increased risk for the later development of cardiovascular disease or type 2 diabetes. Other epidemiological studies performed in Pima Indians and Caucasian individuals later confirmed this observation, suggesting that the association holds true regardless of whether the study populations are genetically predisposed to type 2 diabetes.

The independent effect of thinness at birth, as assessed by the PI, and the lack of association with gestational age on these complications observed in several studies strengthen the deleterious effect of reduced fetal growth, rather than prematurity, on this association.

Numerous epidemiological studies have demonstrated that being born with a low birth weight is also associated with hypertension or dyslipidemia. In all epidemiological studies, reduced fetal growth has been shown to be an independent risk factor for these complications; more specifically, it has been found to be independent of the usual confounding factors such as obesity and family history of metabolic diseases. Nevertheless, it should be pointed out that this association is sharply amplified by obesity and/or familial history of type 2 diabetes. Thus, reduced fetal growth should be considered as a contributing factor, whereas the other known risk factors remain unequivocal.

**What Are the Mechanisms Underlying This Association?**

***Does Insulin Resistance Play a Role in the Metabolic Complications Associated with Reduced Fetal Growth?***

The hypothesis of a critical role of insulin resistance in the metabolic complications associated with *in utero* undernutrition has been well documented during the past few years. This hypothesis was first proposed by the San Antonio group, which pointed out the importance of insulin resistance in this association. These authors suggested that insulin resistance could be the central pathological process underlying the development of the whole syndrome, similarly to what is observed in syndrome X. The significant association between a low birth weight and hyperinsulinemia during adulthood observed in several epidemiological

studies strongly suggests the development of insulin resistance. Moreover, using the euglycemic hyperinsulinemic clamp method, it has been demonstrated that type 2 diabetes associated with a low birth weight requires insulin resistance as described in the classical form of type 2 diabetes.

In a case-control study, our group showed that young adults (20 years of age) born SGA demonstrated hyperinsulinemia under oral glucose tolerance tests (OGTTs) in comparison with age-matched individuals born with normal birth size. Using the euglycemic hyperinsulinemic clamp method, we later confirmed that hyperinsulinemia observed in these young adults reflects insulin resistance. Individuals born SGA demonstrated a peripheral glucose uptake that was significantly decreased in comparison with controls. Using the minimal model method, Flanagan and colleagues also confirmed that insulin resistance is associated with intrauterine growth retardation as early as 20 years of age. Other epidemiological studies have shown that hyperinsulinemia associated with low birth weight was observed in Caucasian and Indian children, suggesting that it could develop from the prepubertal period. These observations point to the early development of insulin resistance in this situation.

***The Role of the Adipose Tissue***

These observations suggest that the insulin resistance is the primary defect responsible for the metabolic abnormalities associated with reduced fetal growth. Hence, the original mechanism should be searched among those directly responsible for the development of insulin resistance itself. In keeping with this assumption, considering a strong contribution of adipose tissue represents an original and reasonable hypothesis.

It is now well established that adipose tissue plays a key role in the development and worsening of insulin resistance and its complications. The strong interaction between these two components is clearly demonstrated by the close relationship that exists between obesity and the development of syndrome X.

Reduced fetal growth severely alters the perinatal development of adipose tissue. Newborns born SGA show dramatically reduced body fat mass. This reduction reflects mostly decreased fat accumulation in the adipocytes, as suggested by the low triglyceride content. However, most of these newborns will show catch-up growth during infancy. This catch-up growth also involves the adipose tissue, as demonstrated by the noticeably increased body mass index (BMI) during the first year of life. There is now accumulating

evidence that the increased growth velocity of the adipose tissue could persist beyond catch-up. This would suggest that reduced fetal growth and the subsequent catch-up would alter body composition. In our study population, we could demonstrate that the relative increase in BMI persists until adulthood. This persistent development of adipose tissue leads to a significantly increased percentage of body fat mass at 21 and 24 years of age in comparison with controls.

In addition, we demonstrated that this abnormal development of adiposity is associated with hormonal and metabolic abnormalities. The insulin resistance observed in young adults born SGA also involves the adipose tissue, and leptin production was dysregulated either during the catch-up growth period or during adulthood. These observations are of interest because leptin, known to regulate body fat mass, has been shown to be involved in the control of insulin sensitivity.

#### *The Role of Catch-Up Growth*

Catch-up growth is a physiological process observed in the majority of children born SGA during the first years of life. However, it has been speculated that catch-up might increase the risk of insulin resistance later in life. Indeed, it has been shown that children born SGA who displayed efficient catch-up growth during infancy showed increased body fat mass with a more central fat distribution in comparison with children born with normal birth size. This increased growth velocity of adiposity persists until adulthood, as evidenced by the relative increase in BMI (expressed in SDS) from childhood to adulthood, leading to a significantly increased percentage of body fat mass at 25 years of age in individuals born SGA in comparison with controls.

In persons born SGA, insulin resistance is sharply amplified by obesity. Several studies have clearly demonstrated that individuals who were thin at birth but subsequently developed obesity during childhood or adulthood are at the highest risk for developing insulin resistance and cardiovascular disease. On the contrary, a study performed in rural Gambian adults did not show any evidence of metabolic or cardiovascular disease in persons born who had a low birth weight and who remained lean during adulthood and were fed with a low-fat diet.

Although the relationship between catch-up in weight and the later development of insulin resistance has been clearly demonstrated, catch-up in height does not appear to influence the long-term metabolic outcome. Insulin resistance parameters are not related to linear growth in children born SGA. It has been

shown in prepubertal children that insulin resistance is found only in those children born SGA in whom catch-up growth in height is associated with a high BMI. This points to the deleterious effect of preferential catch-up of weight over height, rather than the catch-up phenomenon by itself, on the long-term metabolic consequences.

Although the catch-up growth process needs to be carefully followed in children born SGA with respect to the usual clinical practices tending to promote catch-up during early childhood, there is no evidence that these practices might be modified.

#### ***Thrifty Phenotype, Thrifty Genotype, Genotype-Phenotype Interactions: How Can This Association Be Explained?***

Two major hypotheses came into view from the initial epidemiological studies to explain such an unexpected association. Both of them come from the original theory of the “thrifty genotype” proposed by Neel in 1962. To explain the high prevalence of type 2 diabetes in Western populations, Neel hypothesized that genes that would favor survival during a time of famine would become detrimental when food supply is abundant. This theory encompasses the two major components of the current association: undernutrition and the later development of metabolic complications.

The Barker group first proposed an alternative to the thrifty genotype hypothesis: the “thrifty phenotype” hypothesis, in which environmental factors play a critical role. These authors introduced the concept of “fetal origins.” According to the “programming”-process, they proposed that “alteration in fetal nutrition and endocrine status results in developmental adaptations that permanently change structure, physiology, and metabolism, thereby predisposing individuals to cardiovascular, metabolic, and endocrine diseases in adult life.”

From data collected on Pima Indians, McCance and colleagues proposed another alternative to the thrifty genotype and thrifty phenotype: the “surviving small baby genotype.” They hypothesized that, given the high mortality associated with reduced fetal growth, selective survival of children genetically predisposed to insulin resistance and type 2 diabetes provides an explanation for this association. This hypothesis clearly favors the genetic contribution. More recently, the same group reported data that support this hypothesis: in Pima Indians, a strong relationship was observed between a low birth weight in children and an increased prevalence of type 2 diabetes in their fathers, suggesting a paternal inheritance of genes controlling either fetal growth or glucose

homeostasis. In keeping with the idea of the contribution of genetic factors, Hattersley and Tooke proposed the “fetal insulin hypothesis,” suggesting that genetically determined insulin resistance could result in low insulin-mediated fetal growth as well as in insulin resistance during childhood and adulthood. However, the same authors acknowledged the importance of both genetic and environmental factors in the control of fetal growth.

## CONCLUSION

Investigations performed during the past decade or so have identified the independent association between reduced fetal growth and later development of metabolic and cardiovascular complications. By means of a prospective research program based on case-control studies, our group has shown the early development of insulin resistance during adulthood that seems to be the key component underlying this unexpected association. Although the mechanisms responsible for the development of insulin resistance in this situation remain unclear, we collected some evidence in favor of an active contribution of adipose tissue. From a broader point of view, the hypothesis suggesting that this association could be the consequence of interactions between a detrimental fetal environment and genetic susceptibility remains most an attractive one but needs to be confirmed.

## See Also the Following Articles

Diabetes, Type 2 • Growth and Chronic Disease • Growth Hormone (GH) • Growth, Normal Patterns and Consti-

tutional Delay • Postnatal Normal Growth and Its Endocrine Regulation • Puberty: Physical Activity and Growth • Short Stature and Chromosomal Abnormalities • Skeletal Development During Childhood and Adolescence

## Further Reading

- Barker, D. J., Osmond, C., Golding, J., Kuh, D., and Wadsworth, M. E. (1989). Growth *in utero*, blood pressure in childhood and adult life, and mortality from cardiovascular disease. *Br. Med. J.* **298**, 564–567.
- Botero, D., and Lifshitz, F. (1999). Intrauterine growth retardation and long-term effects on growth. *Curr. Opin. Pediatr.* **11**, 340–347.
- Czernichow, P. (2001). Treatment with growth hormone in short children born with intrauterine growth retardation. *Endocrine Rev.* **15**, 39–42.
- Hales, C. N., Barker, D. J., Clark, P. M., *et al.* (1991). Fetal and infant growth and impaired glucose tolerance at age 64. *Br. Med. J.* **303**, 1019–1022.
- Hattersley, A. T., and Tooke, J. E. (1999). The fetal insulin hypothesis: An alternative explanation of the association of low birthweight with diabetes and vascular disease. *Lancet* **353**, 1789–1792.
- Jaquet, D., Gaboriau, A., Czernichow, P., and Lévy-Marchal, C. (2000). Insulin resistance early in adulthood in subjects born with intra-uterine growth retardation. *J. Clin. Endocrinol. Metab.* **85**, 1401–1406.
- Kramer, M. S. (1987). Determinants of low birth weight: Methodological assessment and meta-analysis. *Bull. WHO* **65**, 663–734.
- Léger, J., Lévy-Marchal, C., Bloch, J., *et al.* (1997). Reduced final height and indications for insulin resistance in 20 year olds born small for gestational age: Regional cohort study. *Br. Med. J.* **315**, 341–347.
- Neel, J. V. (1962). Diabetes mellitus: A “thrifty” genotype rendered detrimental by “progress”. *Am. J. Hum. Genet.* **14**, 353–362.
- Pollack, R. N. (1992). Intrauterine growth retardation: Definition, classification, and etiology. *Clin. Obstetr. Gynecol.* **35**, 99–107.

antiarrhythmic amiodarone and the radiographic contrast agent ipanoic acid.

Because iodine is a scarce but essential element, iodine deficiency occurs in many geographic areas, creating serious defects in thyroid function. It has been estimated that one-third of the world's population may be at risk. Therefore, widespread programs for supplementing iodine intake have been instituted. Daily intake of 100 to 200  $\mu\text{g}$  I is considered optimal for humans. It is common in many societies to take supplemental iodine in the form of iodized salt (0.001–0.008% I as KI or  $\text{KIO}_3$ ) to ensure adequate iodine intake.

## THYROID IODINE

### Hormonogenesis

#### **Iodide Transport**

Iodine is a major component (65.3%) of thyroxine ( $\text{T}_4$ ), the primary hormone secreted by the thyroid gland. Thus, it is indispensable for thyroid hormonogenesis. Given the very low natural abundance of iodine, it is not surprising that the thyroid possesses an efficient mechanism for concentrating iodide from the circulation. The concept that iodide itself is concentrated by the thyroid independently of its subsequent metabolic fate developed as a result of early studies employing radioiodide together with drugs that block organic iodine formation. The iodide transport system is present in the thyroids of all vertebrates and normally maintains a concentration of iodide in the thyroid 10- to 50-fold greater than in the circulation; this value can reach several hundred in highly stimulated glands. Inorganic iodide, however, normally comprises less than 1% of the total iodine in the gland. The thyroid contains large stores of hormone, and more than 95% of thyroid iodine is organically bound as hormone or hormone precursors. Nevertheless, iodide transport is an essential first step in the biosynthesis of hormone. Pharmacological blockade of iodide transport by the competitive inhibitor perchlorate can completely inhibit hormonogenesis. Iodide entry into the thyroid is probably rate limiting for organic iodine formation under normal iodide loads.

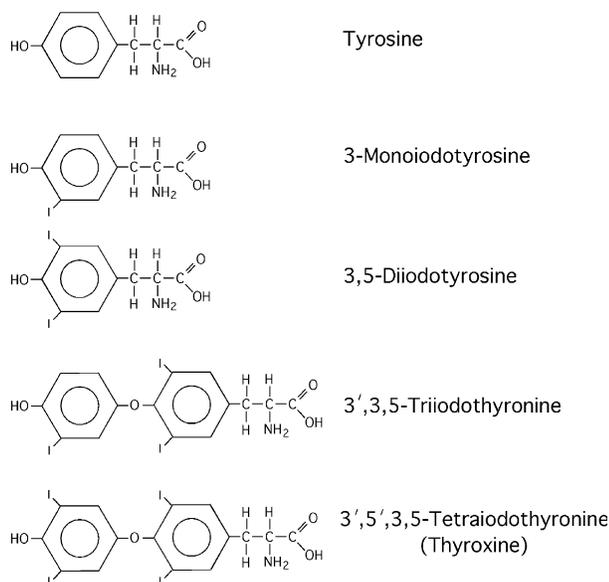
Evidence for a sodium-dependent iodide transporter had become available by the 1970s, but it was not until 1995 that Carrasco and co-workers, after several years of intensive investigation, succeeded in cloning the rat thyroid NIS. They showed it to be an intrinsic membrane protein of 618 amino acids with 13 presumptive transmembrane domains. Subsequently,

human NIS (hNIS) was found to have 84% sequence identity with rat NIS. hNIS has been mapped to chromosome 19p. Immunofluorescence shows NIS to be localized in the basolateral membrane of thyroid follicular cells. NIS is present in other iodide-concentrating tissues such as salivary, gastric mucosa, and mammary. Nontransporting cells transfected with NIS show iodide uptake with many of the properties previously shown for thyroid cells, for example, inhibition by perchlorate and a strict requirement for  $\text{Na}^+$ . Iodide transport is an energy-requiring process. The driving force, as in other  $\text{Na}^+$ -coupled cotransporters, is the energetically downhill movement of  $\text{Na}^+$  across the membrane. The  $\text{Na}^+$  gradient is maintained by  $\text{Na}^+/\text{K}^+$  ATPase in the membrane. Excess iodide, which inhibits iodide transport and other thyroid functions, decreases the expression of NIS mRNA. NIS alone cannot explain all of the iodide-accumulating properties of the thyroid. The thyroid has a second iodide compartment, the follicular lumen, not directly regulated by NIS. Pendrin is the product of the gene defective in Pendred syndrome, a genetic disorder associated with goiter and deafness. Pendrin has been characterized as a sodium-independent ion transporter that localizes to the apical membrane of the thyroid but that is not expressed in extrathyroidal tissues that concentrate iodide. It has been suggested that pendrin is the apical iodide transporter.

#### **Iodination and Coupling**

The major function of the thyroid gland is to synthesize and secrete the iodine-containing hormones  $\text{T}_4$  and triiodothyronine ( $\text{T}_3$ ). These are formed from diiodotyrosyl (DIT) and monoiodotyrosyl (MIT) residues within the backbone of thyroglobulin (Tg), a large (660-kDa), dimeric, thyroid-specific glycoprotein. The structural formulas of the iodinated amino acids found in Tg are shown in Fig. 1.

The iodotyrosines are the most abundant iodinated amino acid components of Tg, but they have no biological activity. They act as precursors to the biologically active hormones. Iodotyrosines are not normally secreted but rather are deiodinated within the gland to release iodide, which then becomes available for reuse in the iodination of tyrosyl residues in Tg. Diiodotyrosine, isolated from coral in 1905, was the first naturally occurring organic iodine compound to be identified. Normal human Tg varies widely in iodine content, depending on iodine intake; values as low as 0.1% and as high as 1.1% have been reported. In one study, the numbers of iodo-amino acid residues per molecule of human Tg containing 0.52% iodine were as follows: MIT, 6.5; DIT, 4.8;  $\text{T}_4$ , 2.3;  $\text{T}_3$ , 0.3. It



**Figure 1** Structures of the major iodinated amino acids in thyroglobulin. Published with permission from Taurog, A. (2000). Hormone synthesis: Thyroid iodine metabolism. In "Werner and Ingbar's The Thyroid" (L. E. Braverman and R. D. Utiger, eds.), 8th ed., pp. 61–90. Lippincott Williams & Wilkins, Philadelphia.

is evident that formation of T<sub>4</sub> is greatly favored over that of T<sub>3</sub>. The hormone content of Tg varies with its iodine content, but within wide limits normal regulatory processes ensure that the hormones are secreted in appropriate amounts. This regulation breaks down when iodine intake becomes rate limiting.

Iodide, the form in which iodine enters the thyrocyte, must first be oxidized to a higher oxidation state before it can iodinate tyrosine. Both the oxidation and iodination are catalyzed by thyroid peroxidase (TPO), a ~110-kDa, transmembrane, glycosylated hemoprotein. Autoradiographic evidence indicates that iodination occurs in the follicle lumen, probably at the apical surface of the follicle cells. Iodination of Tg requires that four separate components (TPO, H<sub>2</sub>O<sub>2</sub>, iodide, and Tg) be brought into close proximity. Evidence from various sources indicates that the apical membrane is the most likely site for this to occur. The transmembrane region of TPO is located near the carboxyl end of the molecule, with most of the molecule oriented toward the follicular lumen. hTPO contains 933 amino acids and approximately 10% carbohydrate. The hTPO gene resides on chromosome 2 and consists of 17 exons and 16 introns, covering at least 150 kb. TPO is a member of a family of animal peroxidases that includes myeloperoxidase and lactoperoxidase. The crystal structure of myeloperoxidase has been determined and serves as a useful guide for the structure of TPO.

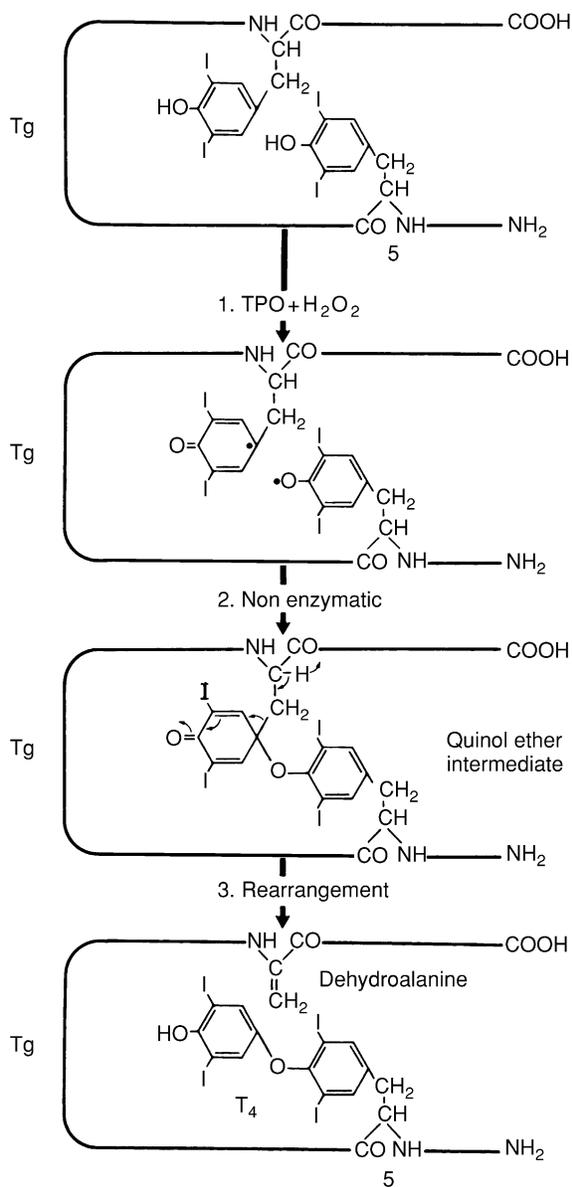
Like all hemoprotein peroxidases, TPO is completely dependent on a source of hydrogen peroxide for its activity. The H<sub>2</sub>O<sub>2</sub> generating system in the thyroid involves a membrane-bound, flavoprotein enzyme, NADPH oxidase, with some similarity to one of the components (gp91<sup>Phox</sup>) of the NADPH oxidase in neutrophils, a system that has been characterized in great detail. Further details of the thyroid H<sub>2</sub>O<sub>2</sub>-generating system remain to be established.

The heme iron in native TPO and other peroxidases is in the ferric form (FeIII). The reaction product of the native enzyme with H<sub>2</sub>O<sub>2</sub>, known as compound I, is two oxidation equivalents above the native state. Compound I reacts with iodide in a two-electron oxidation to form an iodinating species. The best evidence indicates that the active iodinating species is enzyme-bound hypoiodite (EOI) and that iodination of tyrosine occurs by electrophilic substitution.

In addition to iodination, TPO catalyzes the coupling reaction. In the thyroid gland, coupling refers to the reaction between two DIT residues or between one MIT residue and one DIT residue within Tg to form T<sub>4</sub> or T<sub>3</sub>, respectively. The suggestion that DIT is the biological precursor to T<sub>4</sub> was made in 1927 by Charles Harington, the English chemist who first elucidated the structure of T<sub>4</sub> and synthesized it. T<sub>4</sub> had been isolated and crystallized by E. C. Kendall in 1914 at the Mayo Clinic. Several years later, Kendall erroneously reported that T<sub>4</sub> was an indole derivative. He named it thyroxyindole, later shortened to thyroxin and then modified to thyroxine. Thus, the name "thyroxine" is based on a misconception of its structure. It remained for Harington to demonstrate that T<sub>4</sub> was an amino acid with a unique diphenyl ether structure (Fig. 1).

A mechanism for the coupling of two molecules of DIT to form T<sub>4</sub> was first proposed by Johnson and Tewksbury in 1942 and was extended soon after by Harington. It involved oxidation of DIT to a free radical and interaction of two DIT radicals through a quinol ether intermediate to form T<sub>4</sub>. This formulation of the coupling reaction involved free DIT. Later studies indicated that peptide-linked DIT is the precursor to T<sub>4</sub> and that coupling of two molecules of DIT occurs within the matrix of the Tg molecule. The native structure of Tg plays a major role in the coupling reaction. It is likely that the unique structure of Tg has been selected for maximum T<sub>4</sub> formation at low levels of iodination. A hypothetical coupling scheme for the formation of T<sub>4</sub> is shown in Fig. 2.

The scheme proposes that two DIT radicals, generated within the protein matrix through the oxidative



**Figure 2** Proposed coupling scheme for intramolecular formation of T<sub>4</sub> within the thyroglobulin (Tg) molecule. The major hormonogenic site at tyrosyl residue 5 is indicated. Published with permission from Taurog, A. (2000). Hormone synthesis: Thyroid iodine metabolism. In "Werner and Ingbar's The Thyroid" (L. E. Braverman and R. D. Utiger, eds.), 8th ed., pp. 61–90. Lippincott Williams & Wilkins, Philadelphia.

action of TPO, couple to form a quinol ether intermediate. This unstable intermediate then rearranges to form T<sub>4</sub> at the site contributing the inner ring (known as the acceptor or hormonogenic site) and dehydroalanine at the site donating the phenolic ring (the donor site). Favored acceptor and donor sites are most likely dictated by the structure of Tg. Tyr 5 is well established as a favored acceptor site in several species. Other

proposed acceptor sites in human Tg are Tyr 2554 and Tyr 2747. Donor sites are more difficult to identify. Evidence supports Tyr 130 as one likely possibility. It is noteworthy that the hormonogenic residues are located near the extreme ends of the molecule. This may facilitate proteolytic cleavage and release of T<sub>4</sub> and T<sub>3</sub>.

### Why Iodine?

The only established function of iodine in vertebrates is its role in the biosynthesis of thyroid hormones. To earlier investigators, this unique association implied that iodine played some essential role in the action of thyroid hormones. In 1957, Szent-Gyorgi proposed that the hormonal effect of T<sub>4</sub> might be related to the heavy atom perturbation effect of its constituent iodine atoms, an effect that involves the specific electronic properties of the higher halogens, iodine and bromine. However, during the next decade, it was demonstrated that entirely halogen-free analogues of T<sub>4</sub> and T<sub>3</sub>, in which iodine is replaced with methyl or isopropyl groups, showed biological activity. The most active of these, 3,5-dimethyl-3'-isopropyl-L-thyronine, showed approximately 20% of the activity of T<sub>4</sub>. These observations suggested that it is the size of the iodine atoms, rather than their electronic properties, that is most important for hormonal activity. With the development of the receptor hypothesis of hormone action, attention was focused more on the overall three-dimensional shape of the thyroid hormones than on the specific chemical properties of the iodine atoms. The binding of the hormone, inducing a specific and functional conformational change in the receptor, was considered to be the initiating event in hormone activity.

However, the strong dependence of an essential hormone such as T<sub>4</sub> on a scarce resource such as iodine raises the question: Why iodine? One can only speculate, but it seems likely that biosynthetic mechanisms for producing a molecule with the particular conformation of T<sub>4</sub> are much better adapted for iodine-substituted aromatic rings than for alkyl- or bromine-substituted rings. The diphenyl ether structure so characteristic of T<sub>4</sub> occurs only rarely among compounds produced by animals. Biosynthesis of such a molecule is greatly facilitated by the ortho diiodo-phenolic grouping in DIT, the precursor to T<sub>4</sub>.

### Autoregulatory Effects of Iodide

The major regulator of thyroid function is thyroid-stimulating hormone (TSH), acting through its receptor in the basolateral membrane. Its effects are

generally stimulatory. An opposing type of regulation, unique to the thyroid gland, is the ability of the thyroid to protect itself against a sudden excess of iodide through iodide inhibition of processes involved in hormonogenesis. These include TSH-induced cyclic AMP (cAMP) accumulation, iodide organification, iodide transport, H<sub>2</sub>O<sub>2</sub> generation, hormone secretion, thyroid blood flow, thyroid growth *in vivo* and cell proliferation *in vitro*, glucose and amino acid transport, and protein and RNA biosynthesis.

### **Wolff–Chaikoff Effect**

Administration of small to moderate acute doses of iodide to rats or humans does not influence the uptake of simultaneously administered radioiodide by the thyroid. As the iodide doses become progressively larger, however, inhibition of iodination of tyrosyl residues begins to occur. The decreasing yield of organic iodine from increasing doses of inorganic iodide is known as the Wolff–Chaikoff effect. The inhibitory effect depends on the maintenance of a critical intracellular level of iodide in the thyroid. Normally, the inhibition is relieved within hours after the administration of a single large dose of iodide because the thyroid iodide level decreases rapidly as plasma iodide is lost in the urine. Because the inhibition is transient, the phenomenon is frequently referred to as the acute Wolff–Chaikoff effect.

### **Chronic Iodide Excess**

However, when rats or humans are exposed chronically to high doses of iodide, thyroid iodination does not remain inhibited. Escape from Wolff–Chaikoff inhibition occurs after about 48 h through adaptation of the iodide transport system. The adaptation occurs through an iodide-induced decrease in iodide transport activity such that intrathyroidal iodide concentrations become insufficient to sustain a Wolff–Chaikoff effect even at high plasma iodide concentrations. The decrease in iodide transport activity is attributable to inhibition of expression of the NIS symporter gene by excess iodide.

### **Mechanisms of Iodide Inhibition**

Most, if not all, of the inhibitory effects of iodide are reversed by thionamides such as MMI and PTU; therefore, it has been proposed that these effects are mediated by organic iodocompounds. Such compounds were sought for many years, and the major candidates are now considered to be iodolipids. Addition of iodine to the double bonds of unsaturated fatty acids, particularly arachidonic acid and 22:6 ( $\omega$ 3) fatty acid, can produce a variety of products, mainly

lactones, that by themselves can mimic the inhibitory effects of excess iodide *in vitro*. Prominent among these is the arachidonic acid product, 6-iodo-8,11,14-eicosatrienoic acid- $\delta$ -lactone ( $\delta$ -iodolactone), which has been proposed to mediate the iodide-inhibitory effects on thyroid cell growth.

Another type of inhibitory lipid is  $\alpha$ -iodohexadecanal ( $\alpha$ -IHDA), formed by the addition of iodine to the vinyl ether group of plasmalogen ethanolamine.  $\alpha$ -IHDA inhibits H<sub>2</sub>O<sub>2</sub> generation in thyroid cell and membrane preparations, and it has been proposed as the mediator of the Wolff–Chaikoff effect. It would be of interest to know whether it also mediates adaptation to chronic iodide excess (escape from the Wolff–Chaikoff effect). Although the biosynthesis of iodolipids in the thyroid requires iodide oxidation, it is uncertain whether thyroid peroxidase is involved. Many of the inhibitory effects of iodide occur at the basolateral membrane, whereas TPO-catalyzed oxidation and iodination typically occur at the apical membrane. Unsaturated fatty acids probably require a higher iodide level for iodination than do tyrosyl residues of proteins.

## **Iodine-Induced Thyroid Disorders**

Although the thyroid normally has intrinsic autoregulatory mechanisms that allow it to adapt to excessive iodine intake, these mechanisms appear to fail in some individuals. Two human disorders that appear to involve a malfunction of the thyroid autoregulatory mechanism are briefly discussed in what follows.

### **Iodine-Induced Hyperthyroidism**

Iodine-induced hyperthyroidism is seen most commonly when iodine supplementation is introduced in iodine-deficient areas and also in individuals receiving iodine-containing drugs. In both cases, the vast majority of individuals do not develop hyperthyroidism. Those with multinodular goiters seem to be most vulnerable, but the condition also arises in individuals with diffusely enlarged, or even seemingly normal, glands. It has been proposed that some defect in the autoregulatory mechanism of the thyroid must be involved. Presumably, the Wolff–Chaikoff effect is not operative in affected individuals because there is a failure of the autoregulatory signal that normally turns off iodination and hormonogenesis in the presence of excess iodide. The failure of this turn-off signal might be due to a diminished sensitivity to iodide. The signal that normally down-regulates iodide transport may also be faulty so that excess iodide continues to be transported into the gland.

### Iodine-Induced Hypothyroidism

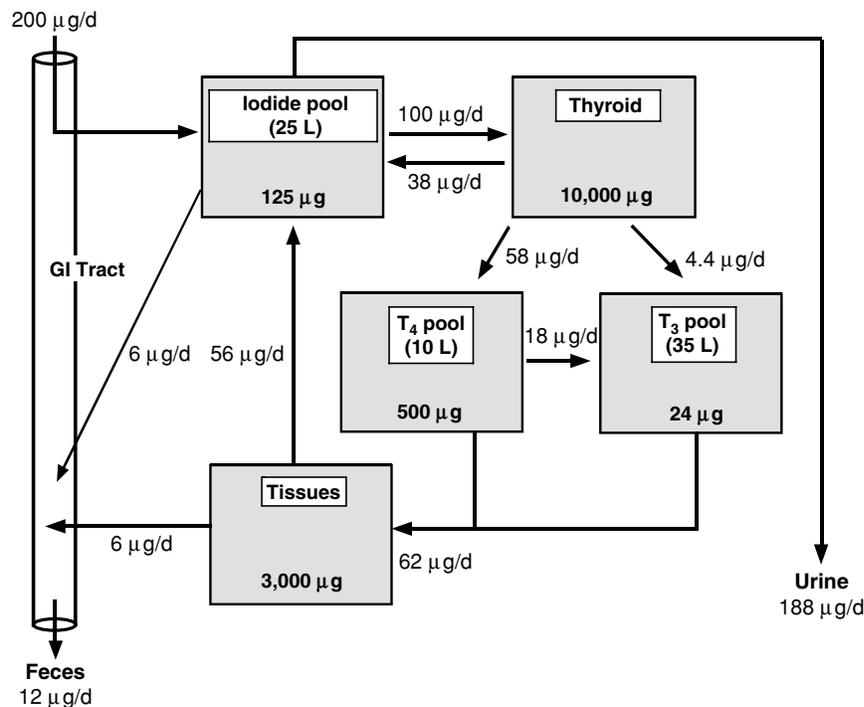
Most humans exposed to chronic high doses of iodine remain euthyroid. However, some euthyroid individuals become hypothyroid in response to such treatment, very likely as a result of some underlying thyroid disease. Included in this group are euthyroid patients with Hashimoto's thyroiditis, a history of postpartum or subacute thyroiditis, or a history of treatment for Graves' disease. Iodine-induced hypothyroidism may be regarded as a failure of escape from Wolff-Chaikoff inhibition in individuals with high concentrations of circulating iodide. Presumably in such cases, the iodide transport mechanism is not down-regulated by excess iodide as would normally be expected, and the thyroidal intracellular iodide concentration continues to exceed the level required for inhibition of iodination.

## EXTRATHYROIDAL IODINE

### Deiodination of $T_4$ and $T_3$

$T_4$  is the major thyroid hormone secreted by the thyroid, which is the sole source of  $T_4$ . The metabolically active form of the hormone,  $T_3$ , is secreted at

only about 10% the rate of  $T_4$ . However,  $T_3$  secretion from the thyroid comprises only about 20% of the total  $T_3$  production in humans. The remaining 80% is formed extrathyroidally by 5'-monodeiodination of  $T_4$  (outer ring deiodination). This reaction is catalyzed by two different deiodinase enzymes: type I (D1) and type 2 (D2). D1, located mainly in liver, kidney, and thyroid, also catalyzes removal of an inner ring iodine from  $T_4$  and  $T_3$  to form the inactive metabolites 3',5',3-triiodothyronine (reverse  $T_3$ ) and 3-3'-diiodothyronine ( $T_2$ ), respectively. D1 serves to provide  $T_3$  to plasma and to inactivate  $T_4$  and  $T_3$ . D2 is located primarily in the central nervous system (CNS), pituitary, and brown fat and acts to release  $T_3$  intracellularly in these tissues as well as to provide  $T_3$  to plasma. Both D1 and D2 contain a critical selenocysteine residue at the active center. A third selenocysteine-containing deiodinase, D3, catalyzes only inner ring deiodination of  $T_3$  and  $T_4$ . It is located in the CNS, placenta, and skin and acts only to inactivate thyroid hormones. Through the action of the various deiodinases on  $T_4$ ,  $T_3$ , and their metabolites, most of the hormonal iodine secreted by the thyroid is returned to the body iodide pool (see Fig. 3).



**Figure 3** Iodine metabolism in 70-kg human in a state of iodine balance. d, day. Adapted from Larsen, P. R., Davies, T. F., and Hay, I. D. (1998). The thyroid gland. In "Williams Textbook of Endocrinology" (J. D. Wilson, D. W. Foster, H. M. Kronenberg, and P. R. Larsen, eds.), 9th ed., pp. 389–515. Saunders, Philadelphia.

## Iodine Kinetics

Figure 3 shows the distribution of iodine among various compartments of the body as well as the daily flux of iodine from one compartment to another. The numerical values are approximations of normal means for individuals on an iodine intake of 200  $\mu\text{g}/\text{day}$ . Such values can vary widely, even within a given study, and the numbers in Fig. 3 should be viewed with discretion. The values for the secretion rate of  $T_4$  (58  $\mu\text{g I}/\text{day}$ ) and  $T_3$  (4.4  $\mu\text{g I}/\text{day}$ ) may be the least variable because homeostatic mechanisms in the pituitary–thyroid axis and within the thyroid normally prevent wide fluctuations in hormone secretion.

Most of the iodine in the body is stored in the thyroid in the form of iodotyrosines and iodothyronines within Tg. Iodine used for thyroid hormone synthesis is drawn from the iodide pool, which is replenished primarily by iodide absorbed from the gastrointestinal tract but also by iodide coming from the thyroid (iodide leak) and from deiodination of iodothyronines in various tissues. The iodide taken up by the thyroid (estimated at 100  $\mu\text{g}/\text{day}$ ) is used largely for the synthesis of  $T_4$  and  $T_3$ , with the rate of  $T_4$  secretion exceeding that of  $T_3$  by about 10-fold. The distribution volume of  $T_4$  is approximately that of the extracellular fluid (about 10L), whereas  $T_3$  enters cells much more readily and has a distribution volume of about 35L. In the circulation, both hormones are tightly bound, largely to specific proteins, such that only about 0.02% of the total  $T_4$  and 0.3% of the total  $T_3$  are present as free hormones. Reflecting this tight protein binding, the half-life of  $T_4$  in the circulation of humans is about 7 days, whereas that of  $T_3$  is about 1 day.

Direct measurement of the iodine content of extrathyroidal tissues is very tedious, and reliable values based on confirmed methodology are unavailable. Mathematical modeling suggests that extrathyroidal iodine may comprise about 30% of total body iodine. It is not clear in what tissues or in what form most of this iodine exists.

The main avenue for excretion of iodine from the body is the kidney. Fecal excretion normally accounts for less than 10% of total iodine excretion. Urinary iodine is almost completely in the form of iodide. Fecal iodine in humans includes both iodide and organic iodine. Because urinary iodine accounts for more than 90% of total iodine excretion, it closely reflects iodine intake; therefore, measurement of

urinary iodine is a very convenient means of assessing iodine intake in human populations.

Renal clearance of iodide in adult humans is about 30 to 40 ml/min and is independent of the plasma iodide concentration. Thus, on low iodine intakes, the kidney does not adapt to the decreased plasma iodide level. All of the adaptation occurs in the thyroid and in the thyroid–pituitary axis. Thyroid clearance of iodide, normally approximately 20 ml/min in humans, increases markedly in iodine deficiency, allowing more efficient use of the available iodine.

## Acknowledgments

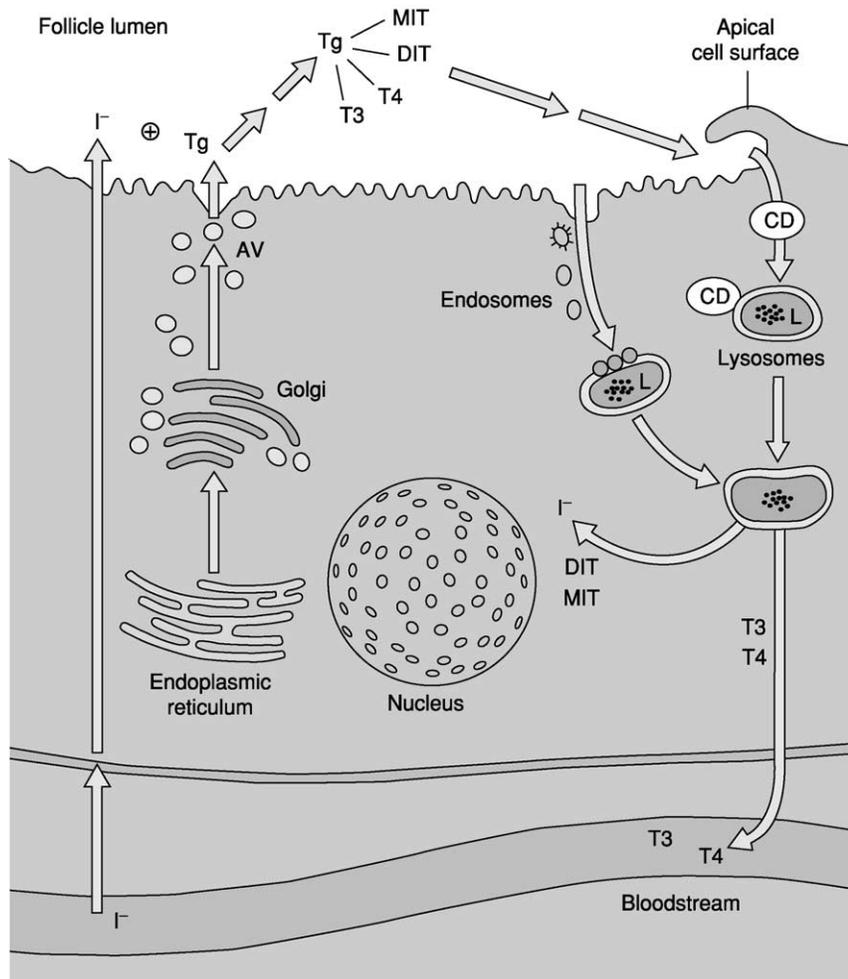
I am indebted to Dr. Jan Wolff for many helpful suggestions and to Dr. Marguerite Hays for assistance in the preparation of Fig. 3.

## See Also the Following Articles

Hypothyroidism, Causes of • Iodine Deficiency • Iodine, Radioactive • Sodium Iodide Symporter • Thyroid Hormone Action • Thyroid Hormone Metabolism

## Further Reading

- Braverman, L. E. (1994). Iodine and the thyroid: 33 years of study. *Thyroid* **4**, 351–356.
- Carrasco, N. (2000). Thyroid iodide transport: The  $\text{Na}^+/\text{I}^-$  symporter (NIS). In “Werner and Ingbar’s The Thyroid” (L. E. Braverman and R. D. Utiger, eds.), 8th Ed., pp. 52–61. Lippincott Williams & Wilkins, Philadelphia.
- Dugrillon, A. (1996). Iodolactones and iodoaldehydes: Mediators of iodine in thyroid autoregulation. *Exp. Clin. Endocrinol. Diabetes* **104**(Suppl 4), 41–45.
- Dunn, J. T., and Delange, F. (2001). Damaged reproduction: The most important consequence of iodine deficiency. *J. Clin. Endocrinol. Metab.* **86**, 2360–2363.
- Fradkin, J. E., and Wolff, J. (1983). Iodide-induced thyrotoxicosis. *Medicine* **62**, 1–20.
- Hays, M. T. (2001). Estimation of total body iodine in normal young men. *Thyroid* **11**, 671–675.
- Jorgensen, E. C. (1976). Structure–activity relationships of thyroxine analogs. *Pharmacol. Therap. B* **2**, 661–682.
- Larsen, P. R., Davies, T. F., and Hay, I. D. (1998). The thyroid gland. In “Williams Textbook of Endocrinology” (J. D. Wilson, D. W. Foster, H. M. Kronenberg, and P. R. Larsen, eds.), 9th Ed., pp. 389–515. Saunders, Philadelphia.
- Taurog, A. (2000). Hormone synthesis: Thyroid iodine metabolism. In “Werner and Ingbar’s The Thyroid” (L. E. Braverman and R. D. Utiger, eds.), 8th Ed., pp. 61–90. Lippincott Williams & Wilkins, Philadelphia.
- Wolff, J. (1989). Excess iodine inhibits the thyroid by multiple mechanisms. In “Control of the Thyroid Gland” (R. Ekholm, L. Kohn, and S. H. Wollman, eds.), pp. 211–244. Plenum, New York.



**Figure 1** Diagram of a thyroid cell to show thyroid hormone synthesis and release. AV, apical villi; CD, colloid droplets; DIT, diiodotyrosine; L, lysosomal proteases; MIT, iodotyrosine; T<sub>3</sub>, 3,5,3'-triiodothyronine; T<sub>4</sub>, thyroxine. Reprinted from Dunn, J. T., and Dunn, A. D. Thyroid physiology. *In* Comprehensive Clinical Endocrinology (C. M. Besser and M. O. Thorner, eds.), 3rd ed., p. 132, Copyright 2002, with permission from Elsevier.

On release into the circulation, the thyroid hormones attach to several binding proteins made by the liver and are then transported to their major targets.

The only clearly established role for iodine is as a component of the thyroid hormone molecules. However, iodine may also influence the breasts, since experimental iodine deficiency reportedly may be associated with mammary dysplasia.

### Amounts of Iodine Required

Experimental and clinical observations have produced recommendations for daily iodine intake. These are based on calculations of daily thyroid hormone production and metabolism in balance studies, the

amount of thyroxine required for adequate replacement in subjects who are athyreotic, and metabolic disposal rates. [Table I](#) shows recommendations by the Food and Nutrition Board of the U.S. National Academy of Sciences for daily iodine intake. The World Health Organization, the International Council for the Control of Iodine Deficiency Disorders (ICCIDD), and UNICEF have recommended almost identical values. Setting an upper limit for iodine ingestion has been more difficult. The Food and Nutrition Board recommended 1100  $\mu\text{g}/\text{day}$  as a safe dose for adults. Most people can tolerate much higher doses without apparent ill effects. Large amounts of iodine inhibit the thyroid's ability to make thyroid hormone, producing goiter and hypothyroidism, but many people can tolerate several milligrams per

**Table I** Recommendations on Dietary Intakes of Iodine

	RDA ( $\mu\text{g/day}$ )	AI ( $\mu\text{g/day}$ )	UL ( $\mu\text{g/day}$ )
Adult men	150		1100
Adult women	150		1100
Pregnancy	220		1100
Lactation	290		1100
Children			
1–3 years	90		200
4–8 years	90		300
9–13 years	120		600
14–18 years			900
Infants			
0–6 months		110	
7–12 months		130	

*Source:* Food and Nutrition Board, Institute of Medicine, U.S. National Academy of Sciences (2001).

*Note.* RDA = recommended daily allowance. AI, adequate intake (estimated when insufficient data for RDA; values set high to avoid possible insufficiency). UL, tolerable upper limit (infants not included; insufficient data).

day. People with autoimmune thyroid diseases, such as Graves' disease and Hashimoto's disease, and many of their relatives are more sensitive to iodine excess.

## Assessment of Iodine Nutrition

### Urinary Iodine Concentration

Over 90% of ingested iodine eventually appears in the urine. Thus, in the steady state, the urinary iodine excretion is a good marker of iodine nutrition. For epidemiological purposes, the concentration itself, in micrograms per liter, is usually satisfactory for gauging the nutritional status of a community. Changes in hydration affect the urinary iodine concentration in individuals but these tend to smooth out in the population median. Occasionally, urinary iodine is expressed as micrograms of iodine per gram of creatinine, but other nutritional abnormalities, such as low protein intake, influence creatinine excretion and the concentration alone is satisfactory and simpler.

The urinary iodine concentration is the single most useful test in assessing populations. Table II relates median urinary iodine concentration to different degrees of iodine nutrition; for deficiency it also presents an index of its severity and, therefore, the priority for corrective action. Urinary iodine concentration also detects excess as well as deficiency.

### Thyroid Size

Iodine deficiency diminishes thyroid hormone production. The pituitary responds with increased TSH

secretion, causing the thyroid to increase hormone production, and the gland enlarges as part of this adaptive response. The degree of enlargement, called "goiter," is proportional to the severity of the iodine deficiency and can be assessed by palpation, classifying it as "no enlargement," "palpable goiter," or "visible goiter," the latter two constituting the "total goiter rate." Thyroid enlargement is one of the earliest and most apparent features of iodine deficiency and palpation is a simple technique that requires some skill but no special instruments. Rapid surveys can be carried out in schools. When the iodine deficiency is severe, such surveys provide a quick and satisfactory basis for recognizing the presence of iodine deficiency and the need for its correction. This was the most common form of assessment for much of the 20th century. However, as iodine nutrition has improved, the degree of thyroid enlargement in such surveys has become more subtle and the results have become correspondingly less accurate.

Ultrasonography provides a quantitative and reproducible assessment of thyroid size. The technique is simple. Portable instruments can easily be used in the field and require only several minutes for each measurement. Normative data exist for iodine-sufficient individuals, related both to age and to body surface area, and can be used for comparison.

### TSH

Serum TSH rises with iodine deficiency and the threat of hypothyroidism. In populations the median serum TSH increases, but frequently remains within the normal range of the TSH assay, unless the iodine deficiency is quite severe. For this reason, the TSH is not a very sensitive indicator of iodine deficiency in the general population, but its use in newborns is an exception. Most developed countries have neonatal screening for early detection and treatment of congenital hypothyroidism, which occurs in approximately 1 in 4000 births in iodine-sufficient

**Table II** The Median Urinary Iodine Concentration as an Index of Community Iodine Nutrition

Median urinary iodine concentration ( $\mu\text{g/liter}$ )	Corresponding approximate iodine intake ( $\mu\text{g/day}$ )	Iodine nutrition
<20	<30	Severe deficiency
20–49	30–74	Moderate deficiency
50–99	75–149	Mild deficiency
100–199	150–299	Optimal
200–299	300–449	More than adequate
>299	>449	Possible excess

populations. Iodine deficiency causes transient neonatal hypothyroidism and a resulting rise in TSH, so the incidence of transient hypothyroidism, as detected on neonatal screening, increases and correlates with the degree of deficiency in the community. Introducing neonatal screening only to detect iodine deficiency is not cost-effective, but if this program is already in place, valuable information about iodine nutrition can be obtained.

### **Serum Thyroglobulin**

The normal thyroid secretes small amounts of thyroglobulin into the bloodstream, detectable by serum assays. Increased thyroid activity from any cause (including iodine deficiency) increases serum thyroglobulin and this correlates well with the degree of iodine deficiency. Urinary iodine is a more practical measure for iodine nutrition, but the serum thyroglobulin is worth considering if blood is being obtained for other purposes.

### **Serum Thyroid Hormone Levels**

Iodine deficiency causes the median serum T4 to decrease and the serum T3 to increase. However, these changes may be subtle and the altered values may still be within the normal range. In practice, these have not been cost-effective measures to assess population iodine deficiency.

### **Radioactive Iodine Uptake**

Iodine deficiency increases the thyroid's uptake of iodine and this can be demonstrated by administering a tracer dose of radioactive iodine ( $^{131}\text{I}$  or  $^{123}\text{I}$ ). The test is cumbersome, particularly in the field, and is not recommended for assessing the iodine nutrition of a population.

### **Survey Technique**

School-age children are a convenient group to study, because they are easy to assemble and reflect recent rather than remote iodine nutrition in a community. The study sample must be representative, take into account the degree of homogeneity within the population being sampled, and address school attendance, rural location, and socioeconomic status. Failure to consider these factors may lead to an unrepresentative survey that does not sufficiently assess the poorer, remote rural areas that are likely to harbor the worst iodine deficiency. The most important sample is the urinary iodine. The most spartan surveys can simply collect urine samples and transfer them to a central laboratory. More extensive data may include an estimate of thyroid size, particularly if obtainable by

ultrasonography. Many countries select a few sentinel sites and monitor them periodically to detect trends.

## **CONSEQUENCES OF IODINE DEFICIENCY**

Thyroid hormone, and therefore iodine, is essential for health. The thyroid hormones modulate many chemical reactions in the body and are necessary for development of the brain and proper growth. Failure to have adequate iodine causes goiter, hypothyroidism, cretinism, reproductive failure, decreased child survival, iodine-induced hyperthyroidism, and socioeconomic deprivation. Some details follow.

### **Goiter**

Thyroid enlargement is an early and easily recognized feature of iodine deficiency. It results from increased TSH stimulation. It can be regarded as an adaptation, and if the iodine deficiency is mild, further damage does not necessarily follow. Goiter is not the worst result of iodine deficiency, but it is not inconsequential. The enlarged thyroid may compress adjacent structures, requiring eventual surgery or other treatment. Prolonged thyroid hyperplasia causes thyroid nodules (lumps), which may occasionally be malignant or become autonomous and overproduce thyroid hormone.

### **Damaged Reproduction**

This is the most severe consequence of iodine deficiency. Its spectrum begins with infertility and includes complications of pregnancy and neonatal death. The need for iodine increases in pregnancy to supply the fetus and also because maternal renal losses are greater. The iodine-deficient pregnant woman may experience gestational hypertension, abnormal fetal presentation, stillbirths, and goiter.

Her offspring faces additional severe risks, with death being the worst. Several epidemiological studies have shown dramatic improvement in child survival with correction of iodine deficiency. For example, child mortality decreased by 50% after iodization of irrigation water in China. Other studies found that child survival increased when iodine was administered during pregnancy and during the first 6 months of life. An extensive literature shows similar results in animals, where stillbirths and abnormal progeny are a common consequence of iodine deficiency.

The developing brain needs thyroid hormone for proper maturation. It is particularly vulnerable

during the second and third trimesters of pregnancy and the first 2 years of life, a period when myelination of the central nervous system is most active. Failure to have adequate thyroid hormone during this critical period can result in permanent mental retardation. In fact, iodine deficiency is the most common cause of preventable mental impairment.

The extent of the brain damage varies over a wide spectrum. Occasionally, it is subtle, requiring specific testing to show impairment. At the other end is cretinism, characterized by severe mental retardation, various other neuromuscular abnormalities, short stature, and deaf-mutism. The incidence of cretinism in severely iodine-deficient areas is uncertain, because many die early or are secluded by defensive family members. Some severely deficient villages have reported incidences of >10% and in such communities the probability is great that most of the population has some degree of impairment.

### Iodine-Induced Hyperthyroidism

This complication occurs when thyroids that were previously iodine deficient produce excess hormone on exposure to increased iodine intake. It happens most frequently in older individuals whose long-standing goiters have developed autonomous nodules. These nodules try to use every atom of iodine to make thyroid hormone and are independent of pituitary control. When exposed to an increased iodine supply, they make thyroid hormone at an accelerated rate and the individual becomes hyperthyroid. Countries or areas that correct previous iodine deficiency can expect some increase in hyperthyroidism in the following years. Fortunately, this trend is usually short-lived and the individual can be adequately treated by the usual techniques available for hyperthyroidism from other causes. Although this complication is regrettable, its damage to the community is mild compared to that from continuing iodine deficiency and iodine-induced hyperthyroidism should never be used as an argument for preventing the correction of iodine deficiency. This complication occurs because of preceding iodine deficiency, so it is reasonable to regard it as another of the iodine deficiency disorders.

### Socioeconomic Damage

Iodine deficiency not only hurts the individual, but also the community in which he or she lives. Hypothyroid people are less efficient, less productive, and more prone to illness. Their domestic animals risk

similar complications from iodine deficiency and they have more stillbirths and produce fewer eggs, less meat, and less wool. Correction of iodine deficiency in a community can result in dramatic improvement in work performance, per capita income, and school performance. For example, one meta-analysis concluded that an average of 13.5 IQ points were lost from moderately severe iodine-deficient communities when compared with iodine-sufficient peers.

### MEANS FOR CORRECTING DEFICIENCY

Vehicles available for delivering iodine to deficient communities include salt, vegetable oil, water, tablets, drops, and occasionally others. The following features are important in selecting a vehicle: (1) provision of physiological amounts of iodine at regular, preferably daily, intervals; (2) easy distribution; (3) affordable cost; (4) acceptability to target population; and (5) safety. In principle, the problem is straightforward: the population is deficient, so iodine must be provided; the critical issue is to see that the prescribed iodine reaches the target correctly.

#### Iodized Salt

Salt is the preferred vehicle for fortification with iodine in most circumstances. Everyone needs salt. It is limited in location and one of the few items that many otherwise self-sustaining communities cannot provide for themselves. The technology for iodization is simple and its implementation is safe.

Potassium iodide (KI) and potassium iodate ( $KIO_3$ ) are the usual fortificants, containing, respectively, 77 and 59% of their weight as iodine. Iodate is generally preferred, especially in developing countries, because it is more stable under conditions of storage and distribution, especially in poorer quality crude salt. Most countries with warm climates or low-quality salt specify iodate; iodide is used mostly by a few countries in North America and Europe.

A variety of techniques for fortifying salt with iodine exist. The most basic is to mix dry potassium iodide or potassium iodate with salt by hand or simple machinery. More often, a solution of potassium iodate or iodide is sprayed or dripped onto salt as it moves on a conveyor belt.

The most common levels for fortification are 20–50 mg iodine/kg salt (20–50 ppm). Daily salt consumption varies widely among populations, but in most instances is between 5 and 10 g/day. Thus, 5 g

iodized at 30 ppm provides the 150  $\mu\text{g}$  recommended for iodine sufficiency. Several other factors affect these calculations. In many countries, especially developed countries, most ingested salt comes from processed foods and that added at the table is relatively minor. For example, estimates for the United States are that only approximately 15% of dietary salt is added at the table and the salt for most processed food is not iodized. Another factor is the loss of iodine between production and consumption of salt. Losses are heavily influenced by the quality of the salt and conditions and duration of its storage before consumption. Under most conditions, such losses are <30%, but occasionally virtually all the iodine will have disappeared by the time it is consumed. Different cooking practices may also cause losses from iodized salt.

Success in implementing the use of iodized salt varies tremendously. Countries that import all their salt can achieve effective iodization by enforcing strict quality control at the borders or at designated processing plants. This situation has made it possible for a number of countries, e.g., Nigeria and Zimbabwe, to achieve universal use of iodized salt fairly quickly. Other countries may have numerous salt deposits scattered throughout their territories and introducing iodization into the local salt trade is a much more daunting task.

### Iodized Oil

Iodinated vegetable oils have been used as X-ray contrast materials for over 70 years. The most common of these, Lipiodol, has 480 mg of iodine per milliliter. Its first use for prophylaxis in iodine deficiency, in the 1950s, proved dramatically effective. For the next 20 years, its administration was principally by intramuscular injection, with a single dose providing adequate coverage for several years. Subsequently, oral use has been more popular. Capsules containing 200 mg iodine, one or two capsules per administration, provide satisfactory coverage for approximately 1 year. Iodized oil has the advantage of bringing iodine supplementation instantly to the subject on administration and avoids the complexities of altering commercial patterns of salt. Its greatest use has been for women and young children with moderate to severe deficiency, to buy time while awaiting the implementation of effectively iodized salt. The disadvantages are that iodized oil requires direct contact with each person, is more expensive than iodized salt, and provides uneven amounts of iodine over the duration of its effects, with a large amount being lost in the first few days. Well over 50 million people have received

iodized oil and it remains a useful alternative under specified conditions.

### Iodized Water

Water, like salt, is necessary for all humans, regardless of geographic or socioeconomic situation. As in salt, the addition of iodine to water can provide a constant physiologic dose.

Several approaches have been used. One of the simplest is to add a concentrated iodine solution to vessels containing drinking water. A program using this approach has been used in schools and homes in northern Thailand for several decades. Addition of iodine to running drinking water has been used in several countries (e.g., Italy, Malaysia, Thailand) by diverting some water through a bed containing iodine as crystals or on a solid support and then reintroducing it to the main stream. Another approach uses diffusers, commercially available polypropylene baskets containing iodide that is slowly released into a water source, such as wells. Finally, an innovative system in western China periodically drips potassium iodate into irrigation water. Follow-up studies have shown impressive benefits to people and domestic animals, at a cost of approximately US \$0.02 per person per year.

Molecular iodine ( $\text{I}_2$ ) has powerful bactericidal properties and is frequently used for water purification. Many communities needing water purification are also iodine deficient, so a system delivering molecular iodine can solve two problems at the same time.

### Iodine Tablets and Solutions

Tablets containing potassium iodide are occasionally used for iodine supplementation. They can be taken daily at 100 or 200  $\mu\text{g}$ , weekly at 1 mg iodine, or at less frequent intervals up to several months, in correspondingly larger doses. The aim is to provide approximately 150  $\mu\text{g}$  iodine per day. The smaller doses may be used as supplements in special circumstances, such as pregnancy, and are also frequently incorporated into antenatal vitamin and mineral preparations. Some countries, such as the United States, stockpile tablets containing 65 or 130 mg potassium iodide to block thyroidal uptake of radioactive iodine in case of a nuclear event, but these amounts are far too excessive for routine prophylaxis against iodine deficiency.

Even remote health posts in developing countries frequently have iodine solutions for topical antiseptics. These can be appropriately diluted and used for iodine supplementation, again aiming at approximately

150 µg iodine per day. Such programs can be quite effective, but require responsible oversight to see that the correct dose is actually taken.

### Choice of Iodine Vehicle

Iodized salt is by far the favored vehicle for most situations. It does not require individual contact, is technologically simple, is low in cost, and provides a constant daily supplement. However, implementation of an effective iodized salt program is not always easy. In some countries, it requires a massive reorganization of the salt trade, so that an iodization step can be reasonably interpolated. For countries that import salt or have only a few large producers, introducing iodization is usually straightforward. Other countries may have many small salt farmers scattered over a large area who harvest the product simply and sell it locally. Some progress has been made by iodizing their product through cooperatives and other methods, but with only partial success.

When iodized salt cannot be effectively introduced within a reasonable time, other methods need to be recruited. Oral iodized oil can be distributed rapidly through community health systems and provides a stopgap while awaiting salt iodization. Iodized water should be considered, depending on water sources and appropriate supervision. Iodized tablets and solutions are effective for individuals, but require careful administration.

Many other vehicles have been occasionally used and may have a niche in certain locales. Examples are brick tea, sugar, candy, fried bananas, and fish sauce. For each, one must ask whether its iodization will reach the right part of the population, especially the poor, women, and children, and in the right amounts. A possible danger is that several different forms of iodine supplementation may converge in an individual and produce iodine excess.

## IODINE DEFICIENCY DISORDERS CONTROL PROGRAMS

### Organization

Correction of iodine deficiency occasionally occurs passively by so-called “silent prophylaxis.” Better transportation and better availability of iodine-containing foodstuffs can bring this about, but it is unrecorded, uncontrolled, and subject to many changes and commercial trends that will not be recognized. A much more effective approach is to have an organization, usually the Ministry of Health in the government,

take charge of the program and its components, especially implementation, education, and monitoring.

Both the problem of iodine deficiency and the means for correcting it are issues of public health and are usually the responsibility of the national government. Their place within the Ministry of Health may vary depending on its structure, but usually fall within a unit for nutrition or a unit for maternal and child health. The most successful programs have had a defined structure within the Ministry of Health, with a specific program for control of iodine nutrition, a designated officer responsible, and a specified budget and work plan. Many countries have decentralized health activities and budgets to provinces and regions and large states may have their own separate units, but these will depend on the central government for direction and support.

Although the Ministry of Health is usually the leader in the government programs, many other sectors, both governmental and private, are involved in issues relating to iodine nutrition. Within the government, they include other ministries, especially education, agriculture, and commerce, and, from the private sector, health professionals, the salt industry, public advocacy groups, and the general population itself, particularly in those areas affected by iodine deficiency. Many countries have developed national committees or coalitions for the elimination of iodine deficiency. These can take different forms, according to the unique political and cultural characteristics of a country. Often in developing countries, the Ministry of Health chairs a committee of members representing other governmental ministries and appropriate parts of the private sector. Such committees are usually advisory, but provide political support for the national effort and promote advocacy. In other countries, the nongovernmental members of the committee may assume a more important role. At the far end of the spectrum, some countries have entirely private groups that have taken responsibility for certain aspects of the program, especially monitoring. In this capacity, they inform the government of the situation and urge appropriate action when necessary. International organizations strongly advise the formation of such national coalitions, as a means for keeping the government and public informed and for championing corrective action as necessary.

### Communication and Education

These are key activities in a program of iodine deficiency disorder (IDD) elimination. Education should take place at all levels, especially for political authorities,

health professionals, the salt industry, and the iodine-deficient population itself. Messages should be appropriate to the audience. For the political authorities, the economic consequences of iodine deficiency often have more impact than does human suffering. For others, such as mothers, awareness of the damage to their children provides a strong incentive for using iodized salt. Health professionals should be taught the grave consequences of iodine deficiency and the relatively easy means for its correction. The general public must be made aware of the damage that iodine deficiency can bring, to create a demand for iodized salt. This, together with enforcement of laws, can make universal salt iodization a reality.

The content of the communication message needs attention. Goiter is the most obvious result of iodine deficiency and is a good marker for it, but is by no means the worst consequence: the damage from neonatal mortality and mental retardation greatly exceeds that from goiter. Therefore, the message must turn from spotlighting goiter and emphasize damage to the developing human instead. The communication campaign must be quick to refute myths and rumors, such as iodized salt causing sterility. Usually iodization will increase the cost of salt and the community must be persuaded that the benefits are worth this investment.

The most effective means of communication vary with the local situation and culture and should be pitched accordingly. Frequent activities are an annual IDD Day with parades, speeches, and demonstrations, radio and television spots, promotional leaflets and literature, and appearances at political and professional meetings.

## Monitoring

The availability of iodine can change frequently and unpredictably with alterations in the physical, political, commercial, and cultural landscape. Natural disasters and civil unrest can disrupt salt plants and distribution patterns, a new government may overturn established patterns of health care and regulatory enforcement, changes in the management and structure of salt companies may profoundly affect the market for iodized salt, and understanding and concern about iodine nutrition and the importance of iodized salt may wax and wane with policy makers and consumers. Because at least some of these factors are likely to occur in any country, careful monitoring of iodine nutrition is essential to maintain effective optimal iodine intake in a population.

## Nutrition

The techniques for monitoring nutrition are the same as those for assessment, described above. The most

useful is the urinary iodine concentration in representative members of the population. Surveys should be carried out periodically, every 2–5 years. The results should be made public. Deviations from acceptable levels should be recognized and the causes sought.

National coalitions can contribute greatly to monitoring, especially in countries that have weak or non-existent central programs. The national coalition can urge the government to carry out mandated monitoring, review the results, and insist on corrective action when necessary. In some countries, professional groups in the private sector have taken the responsibility upon themselves to carry out monitoring and alert the country to the results through the media. For example, in Argentina, a committee of the Federated Endocrine Societies oversees and reports on cyclical monitoring in the different provinces of the country.

Lack of monitoring has been a major cause of failure in programs that were initially successful. Almost all countries in Latin America had developed programs for correction of iodine deficiency through iodized salt over a generation ago, but virtually all of them failed, in large part because of inadequate monitoring. Fortunately, the past 15 years have seen a great improvement and most of those countries appear to have become iodine sufficient, but they still need data and regular monitoring to confirm these impressions.

## Salt

The other important target of monitoring is the iodine content of salt. This is best measured at several levels, first at production, with standard quality-assurance procedures, and more importantly, at the household level. Sanitary inspectors can visit market places and collect salt samples as part of their routine work. These can be measured in a laboratory, the results can be shared with the producer, and any deviations can be corrected. Government inspectors at all levels—production, market, importation—should reject batches of salt without the prescribed iodine content.

A popular monitoring approach is to select representative schoolchildren, obtain urine samples, palpate for goiter (although this is less important), and have them bring salt samples from their homes for testing of iodine content. Such surveys can be carried out quite rapidly and assess both the process (iodized salt) and its impact (iodine nutrition). In large diverse countries, sentinel sites can be used to reflect changes in specific regions.

## GLOBAL STATUS

With only a few possible exceptions, every country in the world has had iodine deficiency at some time in its

history. Seawater contains less than 0.05 ppm iodine, but because of its extremely large volume, it is the world's largest deposit. On land, distribution is quite uneven. The largest concentrations are found in occasional natural deposits, especially in northern Chile, and in association with oil and gas deposits (e.g., in Japan). Severe iodine deficiency occurs in new mountains (e.g., the Andes, Alps, Himalayas), in zones with frequent flooding (e.g., India, Bangladesh), and in far inland areas (e.g., Central Asia, Central Africa). However, other places are not immune and even inhabitants of small islands (e.g., the Azores, Cape Verde) may have significant deficiency. Thus, inadequate iodine is the baseline condition for most parts of the world and poses a constant threat, unless corrected either by natural means, i.e., availability of foods from iodine-sufficient regions, or by intention—i.e., prophylactic programs, usually with iodized salt.

The global extent of iodine deficiency has been estimated periodically over the years. Initially, most figures were based on the presence of goiter by palpation, a fairly crude marker. The World Health Organization in 1999 offered a summary of 191 countries, based largely on goiter; it concluded that iodine deficiency was a public health problem in 130 of those countries and that another 41 could not be

categorized because of insufficient data, leaving only 20 countries iodine sufficient. The people at risk, i.e., living in countries with iodine deficiency, numbered 2.2 billion, approximately 38% of the world's population. Thirteen percent, or 740 million, had goiter. Seventy-five percent of the countries had legislation for iodized salt and 68% of households used iodized salt.

The ICCIDD maintains a database ([www.iccidd.org](http://www.iccidd.org)) that attempts to classify the iodine nutrition status of each country, based principally on representative urinary iodine concentration data. Many countries have incomplete information, so their designation is arbitrary and needs constant adjustment. Table III offers a rough overview of global iodine nutrition, classified by region. This compilation summarizes more detailed data that distinguish different levels of deficiency and also lumps the category of "likely deficient" with "deficient," and "likely sufficient" with "sufficient." Assignment to a category is based on available monitoring data, with the greatest emphasis on urinary iodine, but also taking goiter prevalence into account. Here each country is considered as a unit. If a significant part of the country had iodine deficiency, the country was labeled deficient, even though not all inhabitants have iodine deficiency. For that matter, not all of the inhabitants of countries listed

Table III Global Iodine Nutrition by Region

	Afr (SS)	Amer	As/Pac	E. Eur/Ca	China/FE	Mid E/N Afr	SE Asia	W/C Eur	Total	Percentage of total
Population (millions)	633	835	662	287	1309	514	1269	580	6089	
Number of countries	44	25	14	15	3	19	7	32	159	
Iodine nutrition										
By population (millions)										
Deficient	262	49	467	284	25	304	1267	376	3034	
Sufficient	311	757	68	3	1284	210	2	204	2839	
Excess	54	29	127	0	0	0	0	0	210	
Unknown	6	0	0	0	0	0	0	0	6	
By population, distribution (%)										
Deficient	41	6	71	99	2	59	99	65	50	
Sufficient	49	91	10	1	98	41	1	35	47	
Excess	9	3	19	0	0	0	0	0	3	
Unknown	1	0	0	0	0	0	0	0	0	
By number of countries										
Deficient	25	6	9	14	2	9	6	13	84	53
Sufficient	18	18	4	1	1	10	1	19	72	45
Excess	1	1	1	0	0	0	0	0	3	2

Source. International Council for the Control of Iodine Deficiency Disorders (2002). Available at [www.iccidd.org](http://www.iccidd.org).

Note. Afr (SS), sub-Saharan Africa; Amer, Americas; As/Pac, Asia/Pacific; E. Eur/Ca, Eastern Europe and Central Asia; China/FE, China and Far East; Mid E/N Afr, Middle East and North Africa; SE Asia, Southeast Asia; W/C Eur, Western and Central Europe.

as sufficient are necessarily free of deficiency. China is a good example: the great majority of the population appear to be iodine sufficient, but Tibet remains deficient.

With these limitations, [Table III](#) gives a broad picture of where iodine deficiency persists. Overall, approximately 50% of the world's population lives in a country with significant iodine deficiency and should be considered at possible risk for its consequences. In all regions except the Americas and the China/Far East Region, most of the world's population live in iodine-deficient countries. Large countries, such as India (some deficiency) and China (largely sufficient), weigh heavily in these calculations. Of the 159 countries considered in this survey, 84 are deficient, 72 sufficient, and 3 have excess iodine.

Household use of iodized salt has remained at approximately 70% worldwide. Data collected by UNICEF in 2000 give figures by region (the designation of regions is slightly different than that in the ICCIDD database and the UNICEF compilation does not have entries for developed countries). The regional breakdown is as follows: Latin America and the Caribbean, 81%; East Asia and the Pacific, 80%; Middle East and North Africa, 70%; sub-Saharan Africa, 68%; South Asia, 55%; and Central and Eastern Europe, 28%.

## OBSTACLES TO THE GLOBAL ELIMINATION OF IODINE DEFICIENCY

The past decade has seen remarkable progress in working toward the elimination of iodine deficiency. At the same time, approximately half the world's population still live in countries that harbor some deficiency. In 1990, the World Summit for Children, sponsored by the United Nations and attracting heads of state from most countries in the world, pledged the virtual elimination of iodine deficiency by 2000. That goal failed. A second conference, in May 2002, repeated the pledge, changing the target date to 2005. Reaching it is theoretically possible, but depends on accelerated action from countries and international agencies.

The correction of iodine deficiency is conceptually straightforward. People lack iodine, so they need to obtain it. A choice of good methods for providing iodine exists. Of these, iodized salt is preferred, but the alternatives are also acceptable. Iodized salt has been in use for 80 years. Some countries have adopted it with dramatic success. Why, then, does iodine deficiency continue in so much of the world?

Countries and their inhabitants are unique. Even though a country may share the same geography and ethnic mix as its neighbors, it may differ in many ways that influence the success of an iodization program. Some of the more common reasons that programs are not successful are as follows:

### Difficulties in Implementing Iodized Salt

As already pointed out, some countries have numerous remote salt deposits that are easily mined and sold locally. Other countries have only a few large salt producers, but those do not comply with the law. Still others import their salt, but its iodization is not enforced, and much also enters as contraband.

The compliance and cooperation of the salt industry are crucial to successful iodization. Producers need education on its importance, assistance with technical issues, and constructive actions to make it economically attractive. They must be assured of a market for iodized salt, especially if it costs much more than the non-iodized products. The consumer must learn to demand iodized salt because of its health benefits, even if the price is greater.

### Lack of Awareness

Consumers, salt producers, and government officials must appreciate the damage from iodine deficiency and the importance of its correction. All levels need vigorous communication efforts.

### Insufficient Data

Information about iodine nutrition is fragmentary, not only in the developing world but also in many socio-economically advanced countries. Baseline surveys are necessary in order to plan action. Countries such as Australia and New Zealand have recognized their current iodine deficiency only recently. Other countries, such as Canada, though unlikely to be deficient, do not have recent national data. If a government is unwilling or unable to obtain such information, then other groups, such as endocrinologists, universities, or independent national coalitions, should.

### Inadequate Government Support

The government is ultimately responsible for the public health of its citizens, including their iodine nutrition. Policy makers must be convinced of the importance of iodine nutrition and of the means for its correction and this effort may require extensive

lobbying. The government should have specific laws and enabling regulations to support universal salt iodization and should designate responsibility to a specific unit with sufficient budget to carry out the program. Often government support is strong during periods of great publicity and activity by the unit, but this enthusiasm may taper with new personnel or budget constraints.

### Poor Monitoring

Because iodine deficiency is a chronic condition, it needs regular long-term monitoring. The government must provide for follow-up surveys and respond promptly to any reports that the control program is not working. National coalitions can help to keep iodine nutrition in constant view and to prod the government when lapses in control occur.

### SUMMARY

People need iodine because it is an essential component of the thyroid hormones, which influence a variety of vital biochemical reactions in the body. Insufficient iodine, and therefore insufficient thyroid hormone production, causes pregnancy complications, neonatal and infant death, mental retardation, goiter, hypothyroidism, deaf-mutism, and low economic productivity. The recommended daily intake for an adult is 150  $\mu\text{g}$ , with larger amounts recommended during pregnancy and lactation. The best measure of iodine nutrition is the urinary concentration of iodine.

Salt is the most effective vehicle for delivering iodine to a deficient population and can be iodized by simple inexpensive techniques. Other delivery systems are iodized vegetable oil, iodized water, and iodine tablets. Success in achieving iodine sufficiency depends on an effective program, usually the responsibility of the Ministry of Health and supported by many other sectors, both public and private, especially health providers and the salt industry. Key components in the program are communication, education, and monitoring of both iodine nutrition and salt iodine content.

Approximately half of the world's population lives in countries with iodine deficiency and risks its consequences. Most of these individuals are in the developing world, but many in Western Europe also continue to be deficient. Approximately 70% of households in the developing world use adequately iodized salt. The United Nations and heads of states have pledged virtual elimination of iodine deficiency by the year 2005, but achieving this goal requires an intensified effort.

### See Also the Following Articles

Goitrogens, Environmental • Hyperthyroidism, Childhood and Adolescence • Iodine • Iodine, Radioactive • Nontoxic Goiter • Thyroglobulin • Thyroid Hormone Action • TSH Function and Secretion

### Further Reading

- Bleichrodt, N., and Bonn, M. P. (1994). A meta-analysis of research on iodine and its relationship to cognitive development. *In* "The Damaged Brain of Iodine Deficiency" (J. B. Stanbury, ed.), pp. 195–200. Cognizant Communication, New York.
- Delange, F. (2000). Iodine deficiency. *In* "The Thyroid: A Fundamental and Clinical Text" (L. E. Braverman and R. D. Utiger, eds.), Lippincott, Philadelphia, PA.
- Delange, F., Burgi, H., Chen, Z. P., and Dunn, J. T. (2002). World status of monitoring of iodine deficiency disorders control programs. *Thyroid* **12**, 915–924.
- DeLong, G. R., Leslie, P. W., Wang, S. H., Jiang, X. M., Zhang, M. L., Rakeman, M., Jiang, J. Y., Ma, T., and Cao, X. Y. (1997). Effect on infant mortality of iodination of irrigation water in a severely iodine-deficient area of China. *Lancet* **350**, 771–773.
- Dunn, J. T. (2003). Iodine deficiency. *In* "Diseases of the Thyroid" (L. E. Braverman, ed.), 2nd Ed., Humana Press, Clifton, NJ.
- Dunn, J. T., and Dunn, A. D. (2002). Thyroid physiology. *In* "Comprehensive Clinical Endocrinology" (C. M. Besser and M. O. Thorner, eds.), 3rd Ed. Mosby, St. Louis, MO.
- Glinoe, D., and Delange, F. (2000). The potential repercussions of maternal, fetal and neonatal hypothyroxinemia on the progeny. *Thyroid* **10**, 871–887.
- International Council for the Control of Iodine Deficiency Disorders web site. (August 2003). Available at [www.iccid.org](http://www.iccid.org).
- Panel on Micronutrients, Food and Nutrition Board, Institute of Medicine (2001). "Dietary References Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc." National Academy Press, Washington, DC.

$^{123}\text{I}$  is excellent for thyroid imaging because it delivers far less radiation to the thyroid than does  $^{131}\text{I}$ , has a short half-life (13 h), and does not emit destructive beta radiation. Furthermore, its gamma radiation emission energy (159 keV) is ideally suited to thyroid scanning. The use of  $^{123}\text{I}$  has been superseded for most routine thyroid imaging by Technetium-99m, an isotope that is actively transported by the sodium iodide symporter, with radiation characteristics similar to those of  $^{123}\text{I}$  but with the advantages of minimal cost and ready availability due to its widespread use in nuclear medicine.

### Biological Effects of $^{131}\text{I}$

When  $^{131}\text{I}$  is incorporated into thyroid tissue, the emission of beta particles results in ionization of the thyroid cell, leading directly to the damage of DNA and essential proteins, such as enzymes, and indirectly to the production of free radicals. The early effects of  $^{131}\text{I}$  include necrosis of follicular cells and vascular occlusion that fully develop over a period of weeks to months. Long-term effects include shorter cell survival, impaired replication of surviving thyroid cells with atrophy and fibrosis, and a chronic inflammatory response resembling Hashimoto's thyroiditis. These later effects account for the development of hypothyroidism even years after treatment. The extent of loss of thyroid function, and therefore of the disease state, can be controlled to some extent by the amount of  $^{131}\text{I}$  administered.

### RADIOACTIVE IODINE UPTAKE IN THYROID FUNCTION TESTING AND IMAGING

In the past, measurement of radioactive iodine uptake by the thyroid was commonly used to assess thyroid function. The development of sensitive *in vitro* thyroid function tests and the decrease in normal values for thyroidal radioiodine uptake consequent to an increase in dietary iodine intake have largely rendered this technique obsolete.

Radioiodine uptake measurement, however, remains an important technique in evaluating the hyperthyroid patient. Its main roles are to distinguish subacute or silent thyroiditis from Graves' disease, to provide information regarding the feasibility of radioiodine therapy, and to aid in dose calculation at centers using calculated doses of  $^{131}\text{I}$ .

Thyroid scanning (scintigraphy) allows visualization of the thyroid gland and functioning tissue elsewhere in the body. Therefore, it is valuable in

many clinical situations, including identifying the cause of hyperthyroidism, evaluating thyroid enlargement, the functionality of thyroid nodules, postoperative evaluation of differentiated thyroid cancer, and localization of ectopic thyroid tissue.

### HYPERTHYROIDISM

$^{131}\text{I}$  is increasingly used as first-line therapy for Graves' hyperthyroidism and is the treatment of choice for patients with relapsed hyperthyroidism after antithyroid drug treatment and toxic nodular goiter. Because adequate thyroidal iodine (and hence  $^{131}\text{I}$ ) uptake is a prerequisite for  $^{131}\text{I}$  therapy, it is not an appropriate treatment modality for thyroid-stimulating hormone (TSH)-dependent hyperthyroidism and for other causes of hyperthyroidism with low iodine uptakes such as thyroiditis. The objective of  $^{131}\text{I}$  therapy is to destroy sufficient tissue to cure hyperthyroidism. This may be achieved by rendering the patient either euthyroid or hypothyroid.  $^{131}\text{I}$  typically takes 6 to 8 weeks to produce its effects, and euthyroidism is expected 3 to 6 months after  $^{131}\text{I}$  administration. If this is not achieved, a second dose of  $^{131}\text{I}$  is usually administered. The degree of lasting remission is close to 100% after one or more doses, and there is much more predictable shrinkage of goiter, often with complete disappearance, than can be achieved with antithyroid drugs.

Much attention has focused on achieving euthyroidism by adjusting the dose of  $^{131}\text{I}$ , but there is little consensus regarding the most appropriate dose regimen. The regimens include repeated low doses (80 MBq), fixed doses of 200 or 400 MBq, and doses calculated on the basis of the size of the thyroid, the uptake of radioiodine, or the turnover of radioiodine. There is no evidence that giving a calculated dose of  $^{131}\text{I}$  will achieve euthyroidism but not hypothyroidism, and calculated doses have the disadvantages of inconvenience (> 1 hospital visit required) and higher cost, so many centers use a single fixed dose. The lifetime chance of becoming hypothyroid following  $^{131}\text{I}$  is at least 50% with a lifelong requirement for thyroxine replacement.

### THYROID CANCER

The affinity of differentiated thyroid cancer for iodine is the basis for the use of  $^{131}\text{I}$  for both the detection and the treatment of recurrent thyroid cancer in patients following initial surgical treatment (thyroidectomy). The main indications for  $^{131}\text{I}$  therapy are the

ablation of residual thyroid tissue after thyroidectomy, the treatment of locoregional recurrence, and distant metastases that involve the lung and bones almost exclusively. The precise protocols for  $^{131}\text{I}$  scanning and treatment are complex and controversial and have been reviewed extensively elsewhere.

## NONTOXIC GOITER

Although thyroidectomy remains the standard treatment for large nontoxic goiters that compress the trachea, it may be contraindicated in the elderly or in goiter recurrence after previous thyroid surgery. In such cases,  $^{131}\text{I}$  represents an effective alternative to surgery. Studies have shown that  $^{131}\text{I}$ , administered in doses similar to those used in hyperthyroidism, is effective in reducing goiter size by approximately 40% in patients with nontoxic goiter. Initial concerns that  $^{131}\text{I}$  treatment may lead to thyroid enlargement and worsening of airways obstruction in the short term have proved to be groundless. However, hypothyroidism may occur, as may the development of Graves' disease, which has been described in 5% of cases.

## SAFETY AND SIDE EFFECTS

$^{131}\text{I}$  treatment has been shown to be safe. Concerns about increased cancer risk are largely unfounded, with population studies showing no overall increase in the incidence of cancer or cancer mortality except for a small increase in the absolute risk of small bowel and thyroid cancer. Furthermore, there is no evidence to suggest any reduction in fertility or any problems in the offspring of women treated before pregnancy. The only absolute contraindications are pregnancy and breast-feeding. Pregnancy should be excluded before  $^{131}\text{I}$  administration and should be delayed for 4 months after therapy. Relative contraindications include the use of  $^{131}\text{I}$  in patients under 20 years of age, in whom it should be used with caution, and the presence of thyroid eye disease, which can potentially be worsened by  $^{131}\text{I}$  but easily prevented by the use of corticosteroids.

Permanent hypothyroidism is the only common complication of  $^{131}\text{I}$  treatment, occurring in at least 50% of patients given high doses ( $\geq 400$  MBq) by 1 year and in at least 50% of those given lower doses ( $\leq 200$  MBq) by 25 years. It is dose dependent, and its

incidence remains at 2 to 3% per year many years after therapy. Rarer complications include radiation thyroiditis, thyrotoxic crisis, and early paradoxical increase in goiter size.

## See Also the Following Articles

Graves' Disease, Hyperthyroidism in • Iodine • Iodine Deficiency • Irradiation, Thyroid and • Nontoxic Goiter • Thyroid Carcinoma

## Further Reading

- Bartalena, L., Marcocci, C., Bogazzi, F., Manetti, L., Tanda, M. L., Dell'Unto, E., Bruno-Bossio, G., Nardi, M., Bartolomei, M. P., Lepri, A., Rossi, G., Martino, E., and Pinchera, A. (1998). Relation between therapy for hyperthyroidism and the course of Graves' ophthalmopathy. *N. Engl. J. Med.* **338**, 73–78.
- Bonnema, S. J., Bertelsen, H., Mortensen, J., Andersen, P. B., Knudsen, D. U., Bastholt, L., and Hegedus, L. (1999). The feasibility of high dose iodine 131 treatment as an alternative to surgery in patients with a very large goiter: Effect on thyroid function and size and pulmonary function. *J. Clin. Endocrinol. Metab.* **84**, 3636–3641.
- Cavalieri, R. R. (1996). Nuclear imaging in the management of thyroid carcinoma. *Thyroid* **6**, 485–492.
- Franklyn, J. A., Maisonneuve, P., Sheppard, M., Betteridge, J., and Boyle, P. (1999). Cancer incidence and mortality after radioiodine treatment for hyperthyroidism: A population-based cohort study. *Lancet* **353**, 2111–2115.
- Glinoe, D. (1994). Radioiodine therapy of non-toxic multinodular goitre. *Clin. Endocrin. (Oxf.)* **41**, 715–718.
- Kaplan, M. M., Meier, D. A., and Dworkin, H. J. (1998). Treatment of hyperthyroidism with radioactive iodine. *Endocrinol. Metab. Clinics North America* **27**, 205–223.
- Lazarus, J. H. (1995). Guidelines for the use of radioiodine in the treatment of hyperthyroidism: A summary. *J. Royal College Physicians London* **29**, 464–469.
- Maurer, A. H., and Charkes, N. D. (1999). Radioiodine treatment for nontoxic multinodular goiter. *J. Nuclear Med.* **40**, 1313–1316.
- Mazzaferrri, E. L., and Kloos, R. T. (2001). Current approaches to the primary therapy for papillary and follicular thyroid cancer. *J. Clin. Endocrinol. Metab.* **86**, 1447–1463.
- Meier, D. A., and Kaplan, M. M. (2001). Radioiodine uptake and thyroid scintiscanning. *Endocrinol. Metab. Clinics North America* **30**, 291–313.
- Nygaard, B., Knudsen, J. H., Hegedus, L., Scient, A. V. C., and Hansen, J. E. M. (1997). Thyrotropin receptor antibodies and Graves' disease: A side-effect of  $^{131}\text{I}$  treatment in patients with non-toxic goiter. *J. Clin. Endocrin. Metab.* **82**, 2926–2930.
- Spitzweg, C., Heufelder, A. E., and Morris, J. C. (2000). Thyroid iodine transport. *Thyroid* **10**, 321–330.
- Wartofsky, L., Sherman, S. I., Gopal, J., Schlumberger, M., and Hay, I. D. (1998). The use of radioactive iodine in patients with papillary and follicular thyroid cancer. *J. Clin. Endocrinol. Metab.* **83**, 4195–4203.

membrane is stretched. Other gating mechanisms include interaction with substrates ( $\text{Ca}^{2+}$ , ATP) and other proteins (G proteins, calmodulin), chemical modification (phosphorylation), and physical stimuli (temperature, pH).

Generally speaking, voltage-gated channels have strict ion selectivity. The  $\text{Na}^+$  channels distinguish  $\text{Na}^+$  from  $\text{K}^+$  and the  $\text{K}^+$  channels discriminate  $\text{K}^+$  from  $\text{Na}^+$ . In contrast, ligand-gated channels show much weaker selectivity among ions ( $\text{Li}^+$ ,  $\text{Na}^+$ ,  $\text{K}^+$ , and  $\text{Cs}^+$  in the case of cation channels) and some of the cation channels also conduct  $\text{Ca}^{2+}$ . This may be related to the fact that changes in membrane potential influence all of the voltage-gated channels, whereas functional sites of the ligand-gated channels are localized mainly to small postsynaptic regions.

## MOLECULAR STRUCTURE

Ion channels are transmembrane proteins. Over the past 20 years, molecular cloning has revealed the primary structures of a number of ion channels. Many of them are composed of subunits. The voltage-gated  $\text{Na}^+$  channel is composed of the main  $\alpha$ -subunit and the smaller  $\beta$ -subunits. The voltage-gated  $\text{K}^+$  channel and nicotinic AChR are formed by four and five homologous subunits, respectively. The ATP-sensitive  $\text{K}^+$  channel is a hetero-octamer comprising the pore-forming subunits and the regulatory subunits (sulfonylurea receptors). Structures composed of several subunits may have been advantageous during evolution to allow subtle tuning of the channel functions.

The three-dimensional (3D) structure of ion channels had been only speculative, based on analyses of the predicted amino acid sequences (hydrophobicity,  $\alpha$ -helicity, etc.), effects of site-specific mutations on channel functions, and interactions with toxins or proteins. Studies using X-ray crystallography, however, elucidated the 3D structure of ion channels, including the  $\text{Cl}^-$  channel, the voltage-gated  $\text{K}^+$  channel, and the glutamate receptor channel (partial structure), at atomic resolution. Furthermore, analyses using electron microscopy, together with a computational method, revealed the 3D structure of the nicotinic AChR.

## Ion Permeation Pathway

The ion conduction pathway is not a simple cylindrical pore. At the mouth of the pore, there are charged amino acid residues attracting permeating ions (for example, negatively charged glutamic acid residues for  $\text{Ca}^{2+}$ ). The pore wall is somewhat hydrophilic, so

most of the water molecules surrounding the ion (hydration shell) are taken away. Then, the ion passes through the narrowest part of the channel pore. The amino acid residues in the narrowest part are critical determinants of ion selectivity. The diameter of the narrowest portion of the ion-conducting pore is usually 3–7 Å, which can be estimated by measuring the permeability of organic cations of various sizes.

## Gating Mechanism

Primary amino acid sequences of many voltage-gated channels show a characteristic arrangement of several positively charged amino acid residues in the fourth hydrophobic segment (S4). To sense the changes in membrane potential, there must be a charged structure in the electrical field, i.e., within the membrane. Therefore, the structure for voltage sensing was assumed to be a transmembrane segment. An X-ray study of the voltage-gated  $\text{K}^+$  channel identified the structure of the “voltage-sensor paddle,” which contains S4. The voltage-sensor paddle operates like a hydrophobic cation attached to a lever, enabling the change in membrane potential to open and close the pore.

As for the ligand-gated channels, the structural studies of AChR suggested that binding of a ligand to the receptor site (located on the intersubunit interface) causes a twist in the subunits (rotation around the central axis), which opens the gate.

## FUNCTIONS OF ION CHANNELS

The major functional roles of the ion channels can be categorized as follows: (1) The opening of ion channels leads to rapid breakdown of the electrochemical gradient prebuilt by transporters, which results in rapid changes in cellular potential. (2) Ions are taken into cells through the ion channels down the electrochemical gradient and pumped out to the other side of the cells. This results in a net flux across the epithelial cells (transepithelial transport). (3) Many ion channels mediate  $\text{Ca}^{2+}$  influx, which evokes various  $\text{Ca}^{2+}$ -dependent cellular responses. (4) Some channels function as molecular sensors (for mechanical stress, temperature, etc.), which are essential for cellular homeostasis.

As the history of ion channel research shows, the best-known functions of ion channels are the generation and propagation of action potentials in excitable cells and the synaptic transmission in the nervous system. The classical example is the neuromuscular junction, which was a favorite preparation for

electrophysiologists for many years. When acetylcholine is released from presynaptic nerve terminals of the motor nerve, acetylcholine molecules diffuse across the synaptic cleft and bind to the nicotinic AChR on the muscle cell membrane. Opening of the AChR then causes  $\text{Na}^+$  influx, which results in depolarization of the potential of muscle cells. The neighboring  $\text{Na}^+$  channels sense the depolarization and are activated to allow more  $\text{Na}^+$  influx, spreading out the depolarization and finally generating the action potential. The voltage-gated  $\text{Ca}^{2+}$  channels also contribute to the action potentials. Although the principle of synaptic transmission in the central nervous system (CNS) is essentially the same as in the neuromuscular junctions, some synapses are inhibitory in the CNS. When the postsynaptic ligand-gated channels are  $\text{Cl}^-$  channels (GABA<sub>A</sub> receptors and glycine receptors), the cellular potential response is usually hyperpolarizing (becoming more negative), thus inhibiting the generation of action potentials.

Ion channels are also essential for the release of neurotransmitter or the excretion of hormones. When the action potentials reach the nerve terminal, the voltage-gated  $\text{Ca}^{2+}$  channels open to mediate  $\text{Ca}^{2+}$  flux, which through a complicated cascade induces the fusion of synaptic vesicles to the plasma membrane and induces neurotransmitter release. Hormones are released in a similar way.  $\text{Ca}^{2+}$  is also essential for the contraction of muscles. In the case of the heart,  $\text{Ca}^{2+}$  that enters cardiac muscles through the voltage-gated  $\text{Ca}^{2+}$  channels activates the ryanodine receptors (also called the  $\text{Ca}^{2+}$  release channel), which in turn release  $\text{Ca}^{2+}$  from the sarcoplasmic reticulum ( $\text{Ca}^{2+}$ -induced  $\text{Ca}^{2+}$  release). Note that in the case of skeletal muscles, the  $\text{Ca}^{2+}$  channels are directly linked with the ryanodine receptor; the conformational changes of the former induced by cell excitation (action potential generation) trigger openings of the latter, releasing  $\text{Ca}^{2+}$ , which causes muscle contraction (excitation–contraction coupling).

The *N*-methyl-D-aspartic acid (NMDA) receptors (one type of GluR) are usually blocked by  $\text{Mg}^{2+}$ . But when the membrane potential becomes depolarized, the  $\text{Mg}^{2+}$  block is relieved. Because the NMDA receptors pass  $\text{Ca}^{2+}$  well, repetitive activation of the NMDA receptor leads to substantial  $\text{Ca}^{2+}$  influx into neurons. This activity-dependent  $\text{Ca}^{2+}$  influx is essential for some types of neuronal plasticity, including long-term potentiation in the hippocampal CA1 region.

Ion channels are also essential for transepithelial transport in many tissues. When  $\text{Na}^+$  is reabsorbed in the kidney, for example,  $\text{Na}^+$  enters from the apical (luminal) side into the epithelial cells down

the electrochemical gradient through the amiloride-sensitive epithelial  $\text{Na}^+$  channel (which is different from the voltage-gated  $\text{Na}^+$  channel) and then  $\text{Na}^+$  is pumped out to the basal side.

When cells experience environmental changes, such as osmotic and mechanical pressures, metabolic stress, temperature, and various chemical stimuli, they must handle the changes. When cells are exposed to a low osmotic pressure, they become swollen, but their cellular volume returns to normal. This volume regulation is mediated by a volume-sensitive  $\text{Cl}^-$  channel, which is activated by stretches of the cellular membrane.

The ATP-sensitive  $\text{K}^+$  (KATP) channels are inhibited by intracellular ATP and activated by MgADP. Increased glucose metabolism elevates the cytoplasmic concentration of ATP, which closes the KATP channels. In the case of the pancreatic beta cells, the closure of the KATP channels leads to exocytosis of insulin-containing granules through elevated  $\text{Ca}^{2+}$  mediated by the voltage-gated  $\text{Ca}^{2+}$  channels.

There are also ion channels that are activated by high and low temperatures, high  $\text{Na}^+$  concentration, high  $\text{H}^+$  concentration, and reduction–oxidation states.

## REGULATIONS

Similar to other proteins, ion channels are subject to regulation by G proteins, protein kinases, calmodulin, and other proteins. Studies of the ion channels tagged with green fluorescent protein indicate that the activity of ion channels is regulated by changing the number of functional channels by inserting them into or removing them from the plasma membrane.

## RELATED MOLECULES

Ion channels are proteins imbedded in the lipid bilayer and they can move anywhere unless they are anchored by other proteins. Such scaffold proteins are important not only for anchoring but also for putting functional components together so that signal transduction will be rapid and efficient. One good example is PSD95, which is located in the postsynaptic densities. There are many kinds of functional molecules, including cytoskeletal molecules attached to PSD95, that make a functional complex (referred to as the “signalplex”).

## RELATED DISORDERS

Because ion channels are closely related to the ion transport mechanism and electrical activity, it is natural to

assume that dysfunction of ion channels leads to various kinds of disorders. Ion channels are targets of autoimmune disorders, including myasthenia gravis (against nicotinic AChR at neuromuscular junctions; causes muscle weakness), Lambert-Eaton syndrome (against  $\text{Ca}^{2+}$  channels at nerve terminals; causes muscle weakness), and Isaacs' syndrome (against  $\text{K}^+$  channels of motor nerves; causes spontaneous muscular twitching).

Genetic studies of inherited diseases have identified an increasing number of gene mutations associated with a wide variety of diseases. The first major discovery was the identification of the cystic fibrosis transmembrane regulator (CFTR). CFTR has a complicated molecular structure similar to that of transporters, but is an anion channel conducting both  $\text{Cl}^-$  and  $\text{HCO}_3^-$ . CFTR is expressed in the apical membrane of epithelial cells of the intestine, airways, sweat glands, pancreas, etc. Because CFTR is crucial for various transepithelial transport processes, mutations of CFTR cause impairments in epithelial salt and fluid secretion as well as reabsorption.

Mutations of voltage-gated channels are also associated with several kinds of diseases. Mutations of the skeletal muscle  $\text{Na}^+$  channels cause myotonia and periodic paralysis. Mutations of the cardiac  $\text{Na}^+$  channel cause the long QT syndrome and Brugada syndrome. Other types of long QT syndrome result from mutations of  $\text{K}^+$  channels (HERG, minK, and KCNQ1). Mutations of the brain  $\text{Na}^+$  channels are associated with seizure disorders (generalized epilepsy with febrile seizure plus, severe myoclonic epilepsy of infancy). There are multiple channel phenotypes that arise from the  $\text{Na}^+$  channel mutations, but a relatively common channel phenotype is impaired inactivation, which allows  $\text{Na}^+$  currents to flow in continuously. The persistent  $\text{Na}^+$  influx results in depolarization of muscle cells or neurons, leading to excessive muscle contraction, arrhythmia, and seizures.

Mutations of the gene encoding the voltage-gated  $\text{Ca}^{2+}$  channels are associated with neurological disorders. They include familial hemiplegic migraine, episodic ataxia 2, and spinocerebellar ataxia 6. There are mouse counterparts of the ataxic disorder and some mice also exhibit seizure disorders. Analyses of neuronal network functions of the mouse models suggest that mutations of the  $\text{Ca}^{2+}$  channels impair transmission at some synapses, but leave others relatively spared. This uneven involvement of the mutational effects may disrupt the well-balanced interactions between excitatory and inhibitory neuronal networks,

thus causing the paroxysmal disorders, in addition to the chronic progressive disorders.

## THERAPEUTIC VIEW

Ion channels are usually on the cell surface and relatively easily accessible and thus are good pharmacological targets. Actually, many commonly used drugs act on ion channels.

The  $\text{Na}^+$  channels have been a good target for therapeutic purposes. Blocking of nerve conduction in peripheral nerves prevents, for example, pain signals from entering the CNS. Lidocaine and other local anesthetics can block the  $\text{Na}^+$  channels when they are open and temporarily prevent the channels from re-opening. A similar mechanism is hypothesized for the use of lidocaine to prevent ventricular arrhythmia. Once the cardiac  $\text{Na}^+$  channels open, lidocaine inhibits them from opening successively. As seen from these examples, drugs acting on the  $\text{Na}^+$  channels should be state-dependent, that is, acting on channels only in the open or related states. If they act on  $\text{Na}^+$  channels in both an open and a closed state, they would cause muscle weakness, drowsiness, and other side effects. The blockers of the voltage-gated  $\text{Ca}^{2+}$  channels (" $\text{Ca}^{2+}$  blockers") are widely used for the treatment of hypertension. The  $\text{K}^+$  channel openers are used for the treatment of hypertension and ischemic heart disease.

Ligand-gated channels are also targets for commonly used drugs, which include muscle relaxants (succinylcholine, pancuronium) acting on nicotinic AChR and sedative-hypnotic drugs (diazepam) acting on GABA receptors.

## PERSPECTIVES

The study of ion channels has a relatively long history. In the past, electrophysiology (function) used to lead biochemistry/molecular biology (structure/molecule). However, this order has become reversed. Genome analyses have revealed many candidates for ion channels whose functions are yet unknown. Surprisingly, some of these "orphan" ion channels have turned out to be molecules playing essential roles, for example, in sensing temperature. Further analyses of novel channels, together with the development of specific drugs, will contribute to a better understanding of the basic mechanism of the human body and also to better medical treatments.

### See Also the Following Articles

Bartter's Syndrome • G Proteins and Effectors • Potassium Homeostasis, Regulation of

### Further Reading

- Ashcroft, F. M. (2000). "Ion Channels and Disease: Channelopathies." Academic Press, San Diego, CA.
- Hille, B. (2001). "Ion Channels of Excitable Membranes," 3rd Ed. Sinauer, Sunderland, MA.
- Hodgkin, A. L., and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* **117**, 500–544.
- Jiang, Y., Lee, A., Chen, J., Ruta, V., Cadene, M., Chait, B. T., and MacKinnon, R. (2003). X-ray structure of a voltage-dependent K<sup>+</sup> channel. *Nature* **423**, 33–41.
- Miyazawa, A., Fujiyoshi, Y., and Unwin, N. (2003). Structure and gating mechanism of the acetylcholine receptor pore. *Nature* **423**, 949–955.
- Numa, S. (1989). A molecular view of neurotransmitter receptors and ionic channels. *Harvey Lect.* **83**, 121–165.
- Sakmann, B., and Neher, E. (eds.) (1995). "Single-Channel Recording," 2nd Ed. Plenum, New York.

after a mean dose as low as 100 mGy to the thyroid. There was no evidence for a threshold dose below which the effect disappeared, although at doses higher than 1500 cGy the risk per gray decreased, probably because of cell killing, but the overall risk remained elevated.

In a study of 2634 patients at a Chicago hospital whose thyroids received a mean dose of 590 mGy for benign diseases during childhood, approximately 60% developed thyroid nodules and 15% developed thyroid carcinoma during 40 years of follow-up.

In most studies, the latency period between the time of radiation exposure and the appearance of the thyroid nodule ranges between 5 and 15 years. In the pooled analysis of seven studies mentioned previously, only two cases were seen within 5 years of exposure; the risk clearly increased 5 to 9 years after exposure, peaked at 15 to 19 years after radiation exposure, and then declined, although an excess risk was still apparent at 40 years follow-up.

Taken together, these studies demonstrate that the risk of thyroid cancer after exposure to external radiation is indeed very high and that the thyroid gland is very sensitive to radiation, especially during childhood.

In case of external radiation to the head and neck for malignant disease, the dose delivered to the thyroid is usually greater than that delivered for the treatment of benign conditions. In this situation, the risk of thyroid tumor may still be increased, but the risk per gray decreases. This finding has been attributed to cell killing, which decreases the number of cells that may become neoplastic and explains the high frequency of hypothyroidism observed after large dose irradiation.

High-dose radiation therapy to the neck (>20 Gy), such as the one used in Hodgkin's disease, results in high rate of hypothyroidism but also in an increased risk of thyroid cancer. The final outcome of the thyroid damage is probably related to the distance between the thyroid gland and the radiation field. If this is far from the thyroid, as in the case of thoracic or abdominal radiation fields in children, the thyroid gland may receive radiation doses of some 100 mGy, not enough to produce hypothyroidism but sufficient to trigger thyroid cancer.

Radiation-induced thyroid carcinoma from external sources has also been noted after accidental contamination. Indeed, thyroid carcinoma was the first solid malignant tumor found to be increased among Japanese atomic bomb survivors. After the bombing in Hiroshima and Nagasaki in 1945, the body dose was due mainly to external irradiation (X rays and

neutrons). The health consequences were studied in a cohort of 94,000 survivors and of 26,000 individuals who resided in Hiroshima and Nagasaki shortly after the bombing. A total of 225 thyroid cancers were diagnosed between 1958 and 1987 among the 79,972 survivors who were alive and free of cancer as of January 1958 and who had radiation dose estimates. From the Lifespan Study of atomic bomb survivors in Hiroshima and Nagasaki, it is known that little risk is carried for exposures after 20 years of age and that almost none is carried for exposures after 40 years of age.

## Factors Affecting Sensitivity to Develop Radiation-Induced Thyroid Cancer

### **Age and Sex**

A major risk factor is a young age at the time of irradiation. The risk of thyroid cancer after external irradiation in children under 5 years of age is twofold higher than that in children treated between 5 and 9 years of age and is fivefold higher than that in children treated between 10 and 14 years of age. In the Japanese atomic bomb survivors, the excess risks (ERs) of thyroid cancer were 9.5, 3.0, 0.3, and 0.2 in the age categories of 0–9, 10–19, 20–39, and over 40 years of age, respectively, at the time of the bombing. The ERs were not significant for survivors exposed over 15 to 20 years of age.

Gender does not seem to influence the risk of developing radiation-induced thyroid cancer. Although females are two to three times more likely to develop both benign and malignant thyroid nodules after irradiation, this finding reflects the higher natural incidence of thyroid nodules and cancer in the female general population.

### **Genetic Predisposition**

Several clinical observations suggest that genetic predisposition, such as defects in the DNA repair mechanisms, may affect the risk of developing radiation-induced thyroid cancer. Patients who experience one radiation-related cancer are more likely to develop a second radiation-related cancer. Sibling pairs exposed to radiation develop thyroid tumors more often than would be expected by chance. The risk of thyroid cancer in individuals treated with radiotherapy during childhood for a cancer (other than neuroblastoma) is 3 to 10 times higher than that in children treated for benign conditions. Those treated for neuroblastoma have a fivefold risk of thyroid cancer compared with those treated for other

cancers, suggesting a common predisposition for neuroblastoma and thyroid cancer.

A search for genes predisposing to radiation-induced thyroid cancer is in progress in pedigrees with the recurrence of thyroid cancer. No linkage has been found with genes already known to be involved in thyroid tumorigenesis such as *ras*, *p53*, and *RET/PTC*.

## THYROID DISORDERS AFTER EXPOSURE TO RADIOACTIVE IODINE

The role of radioactive iodine administered for medical uses in the development of thyroid cancer has been addressed in a few studies that found no significant risk and led to the conclusion that <sup>131</sup>I is safe both as a diagnostic and a therapeutic tool. However, most patients of these studies were treated during adulthood, whereas the post-Chernobyl epidemic of thyroid cancer occurred mainly in children and adolescents, suggesting that the young thyroid is particularly sensitive to the effect of radiation. This event has renewed concern about the carcinogenic risk of medical use of <sup>131</sup>I, at least in young patients.

Radioactive iodine (<sup>131</sup>I) has been used in clinical practice for many years, both for the diagnostic evaluation of thyroid disorders and for treatment of benign and malignant conditions (hyperthyroidism and metastatic thyroid cancer, respectively). The radiation dose delivered by <sup>131</sup>I to the thyroid is 1000- to 10,000-fold higher than that delivered to other tissues. Thus, even a relatively low amount of <sup>131</sup>I may deliver a significant radiation dose—potentially carcinogenic—to the thyroid gland. Increasing the radiation dose beyond a few hundred MBq increases the likelihood of obtaining cell killing and decreases the possibility of tumoral changes.

The dose of radioiodine administered to the patient is very large (on the order of >100 mCi) in the treatment of thyroid cancer when the dose used is intended to kill all thyroid cancer cells. The dose is smaller (<20–25 mCi) in the treatment of thyrotoxicosis when the intent is to produce hypothyroidism. In these conditions, the radiation doses are sufficiently high to kill the cells; thus, no unwanted secondary thyroid disease occurs.

Low doses of iodine isotopes are used as a tracer for diagnostic evaluation of the thyroid gland. In this situation, no cell killing is observed and thyroid cell damage is theoretically possible. However, no convincing evidence of subsequent thyroid disorders has been provided.

The most informative analysis in this setting was performed in Sweden with 34,104 people exposed to diagnostic doses of <sup>131</sup>I during the period from 1950 to 1969, for a mean thyroid dose estimate of 110 cGy. A small increase in the number of observed thyroid cancers ( $n=67$ ) was found with respect to the expected number ( $n=50$ ). However, the increase was confined to individuals undergoing thyroid scans for suspicion of thyroid cancer. When the analysis was limited to individuals tested for reasons other than thyroid cancer, no increase was observed.

In the same Sweden cohort, a subset of 1005 women was compared with 248 matched controls regarding the incidence of thyroid nodules. The average length of follow-up was 26 years, and the average age at exposure was 26 years. No difference was found in the two groups; the incidences of nodules were 10.6 and 11.7, respectively.

The conclusion drawn from the study with the Sweden cohort is that diagnostic use of <sup>131</sup>I has no health effect on the thyroid. However, it should be noted that only a minority of the individuals exposed were children.

Also after treatment of hyperthyroidism with <sup>131</sup>I, no evidence of an increased risk of thyroid cancer has been reported. In another study conducted in Sweden with 10,552 adult patients (mean age 57 years) followed for a mean period of 15 years, the relative risk (RR) of thyroid cancer was not increased significantly (RR = 1.29, 95% confidence interval [CI]: 0.76–2.03). The average estimated radiation dose to the thyroid was 100 Gy.

In a study carried out in the United States, it appears that after a mean follow-up of 21 years, <sup>131</sup>I treatment was not linked to thyroid cancer deaths (standard mortality ratio [SMR] = 1.02) or to the development of any specific cancer with the exception of thyroid cancer (SMR = 3.94, 95% CI: 2.52–5.86). The SMRs were 2.08 in patients with Graves' disease and 6.53 in those with toxic nodular goiter. The excess number of deaths was small (observed/expected ratio of 27/10), and the underlying disease, rather than radiation, seemed to play the major role.

This conclusion is not surprisingly. The high dose delivered in the case of hyperthyroidism is frequently sufficient to produce hypothyroidism through cell killing. Indeed, the risk of hypothyroidism at 2 years increases linearly with the thyroid dose for radioactive concentrations ranging from 0.9 to 8.3 MBq/g.

As in the case of diagnostic use of <sup>131</sup>I, also in the case of hyperthyroidism, the patients treated were mainly adults. In a few hundred children treated with <sup>131</sup>I, no significant increase in the incidence of

thyroid cancer has been observed. However, in view of the high sensitivity of the young thyroid gland to radiation, it is probably advisable not to use  $^{131}\text{I}$  in treating hyperthyroidism in young children and adolescents.

So far as treatment of differentiated thyroid cancer with  $^{131}\text{I}$  is concerned, there is a low but significant risk of secondary effects on other organs (mainly testis, ovary, and bone marrow) when cumulating high radiation doses during several treatment courses.

## THYROID EFFECTS OF THE CHERNOBYL NUCLEAR DISASTER

The explosion of one of the reactors at the nuclear power plant in Chernobyl, Ukraine, occurred in April 1986 and released large amounts of radioactive particles into the atmosphere, including  $^{131}\text{I}$  (32–46 Mci),  $^{132}\text{I}$  (27 Mci, resulting from the decay of technetium-132), and  $^{133}\text{I}$  (68 Mci). Most likely, radioiodines were released intermittently over a period of 10 days after the accident. The most contaminated territories were southern Belarus, northern Ukraine, and the Bryansk and Kaluga regions of southern Russia.

Volatile radioactive isotopes could be first inhaled and, after being deposited on the ground, ingested. The time at which ingestion occurred varied considerably, but the milk chain, particularly in children, was the major route of ingestion; at this time, short-lived isotopes of iodine were no longer present. Several factors contributed to the high radiation exposure of the population. Immediate protective countermeasures, such as advising and evacuating the people at risk and distributing iodine prophylaxes, were not undertaken. Furthermore, the most contaminated regions were in a state of moderate iodine deficiency, and this is responsible for increased iodine uptake. All of these factors combine to give enough of an explanation as to why the most serious health consequence of the disaster was thyroid cancer and why children were affected the most.

In the case of radioactive contamination, the thyroid gland is a critical organ at risk. Its contamination depends on the magnitude of contamination, the amount of radioactive iodine taken up by the gland, and the thyroid mass itself. Furthermore, whatever the level of contamination, the thyroid dose is always higher in children than in adults. The thyroid dose is dependent on the final concentration, namely the ratio between radioiodine uptake and thyroid mass. The uptake in children is similar to that in adults except that the thyroid mass is smaller and the thyroid

dose per gram of tissue is greater—and is extremely higher in newborns and very young children.

## Post-Chernobyl Thyroid Cancer

As a result of the accident, a tremendous increase in the number of cases of childhood papillary thyroid cancer has occurred in the subsequent years, starting in 1992. The magnitude of this increase and the geographical and temporal distribution of the cases strongly suggest that thyroid cancer was due to the reactor explosion and particularly to the huge amount of iodine radioisotopes released. The increase in the number of cases of thyroid carcinoma in children and adolescents has been observed starting from 1990, only 4 years after the Chernobyl accident, in southern Belarus and northern Ukraine, and from 1994 in southern Russia. More than 1000 cases of thyroid cancer have been reported among approximately 2 million children younger than 15 years of age who were exposed to radioactive fallout. In the Gomel region, the most contaminated area of Belarus, the incidence between 1986 and 1996 was 13 per 100,000 children per year compared with a baseline incidence of less than 1 per 100,000 children per year.

Although dosimetric data are imprecise, the mean thyroid dose has been estimated to be nearly 700 mSv in Belarus. In Ukraine, 79.0% of the children received a thyroid dose less than or equal to 300 mSv, 10.5% received between 300 mSv and 1 Sv, and 10.5% received more than 1 Sv. As a means of comparison, in children exposed to external irradiation, the risk of thyroid cancer was significant even for thyroid doses as low as 100 mSv.

In most of the children who developed thyroid cancer, the estimated thyroid dose has been less than or equal to 300 mGy. However, an excess thyroid cancer incidence has been observed even in areas where the mean thyroid dose in children was estimated to be 50 to 100 mGy.

## Genetics of Post-Chernobyl Thyroid Cancer

Molecular biology investigation of post-Chernobyl tumors shows some peculiarities. Ras and p53 genes are not involved in their pathogenesis. Rearrangements of the RET proto-oncogene are found in nearly 70% of the cases, a percentage higher than that observed in naturally occurring papillary carcinomas. Also, the subtype of RET/PTC rearrangement has a peculiar pattern. Several studies have reported that

RET/PTC 3 (and more rare variants of RET/PTC 3) is the form more frequently expressed in radiation-induced tumors, rather than RET/PTC 1, which is predominant in naturally occurring cases. A correlation has also been established between the solid variant of papillary tumors and the activation of RET/PTC 3.

These data indicate that RET/PTC activation may be the direct result of radiation-induced DNA damage. Because RET/PTC is also frequently found in pediatric papillary thyroid cancer without known exposure to radiation, it is also possible that age per se may play an important role. Alternatively, one can speculate that virtually all pediatric papillary thyroid cancers are radiation-induced cancers that develop in individuals with an increased susceptibility to spontaneous background radiation.

## OTHER THYROID CONSEQUENCES OF RADIATION EXPOSURE

Thyroid cancer and benign thyroid nodules after radiation exposure occur as stochastic effects. Depending on the radiation dose, deterministic effects, hypothyroidism, and acute thyroiditis are also possible. Another documented consequence of radiation is the possibility of developing autoimmune thyroid disorders.

Hypothyroidism is caused by radiation doses on the order of more than several gray to the thyroid. Such levels are observed mainly after radioiodine therapy for benign thyroid diseases, such as Graves' disease and toxic nodular goiter, and in this case hypothyroidism is the objective of treatment.

Spontaneous hypothyroidism (or subclinical hypothyroidism) has been reported in atomic bomb survivors in Nagasaki. Among 2587 individuals, spontaneous hypothyroidism was diagnosed in 43 (27 of whom were thyroid antibody positive and 16 of whom were thyroid antibody negative) without sex differences. Because an association was observed between thyroid dose and the prevalence of antibody positive (but not antibody negative) spontaneous hypothyroidism, the last conceivably stemmed from an underlying autoimmune thyroid disorder. This concept is supported by the finding that the frequency of radiation-associated hypothyroidism in patients treated for malignant lymphoma decreased significantly when radiation therapy was combined with chemotherapy and immunosuppressive agents.

After external irradiation to the head and neck, the occurrence of thyroid autoimmunity has been

reported in several studies. An increased incidence of thyroid antibodies was found in individuals who received radiation for benign disorders during childhood. Variable degrees of thyroid lymphocytic infiltration have been reported in more than two-thirds of individuals who received radiation several years before thyroidectomy for nodular thyroid lesions. In patients who received radiation of the neck for Hodgkin's disease, 3% or more developed Graves' disease (a 7- to 20-fold excess risk) and 1% developed thyroiditis.

Hypothyroidism has also been reported after exposure to internal radiation (radioactive iodine). In the people exposed to the fallout from the Marshall Islands accident, hypothyroidism was noted within 10 years after the accident. In this instance, most of the cases were not associated with autoimmune thyroid reaction.

On the contrary, an increased prevalence of antithyroid antibodies (19.5%) without hypothyroidism has been reported in children living in a Belarus village heavily contaminated by the post-Chernobyl radioactive fallout, compared with children living in a noncontaminated village (3.8% prevalence). The susceptibility to developing thyroid autoimmunity increased with age at the time of exposure; in the female sex, it reached its maximum during the pubertal age, suggesting that puberty (estrogen) has a cumulative effect on radiation in the development of thyroid antibodies in females. This study was conducted only 6 to 8 years after the Chernobyl accident, and this may be the reason why cases of overt hypothyroidism were not found. However, the possibility of developing thyroid failure later is very likely in this population.

The release of thyroid antigen from damaged thyroid cells is regarded as the possible mechanism triggering the autoimmune reaction after radiation injury, regardless of the modality of thyroid exposure.

## See Also the Following Articles

Iodine, Radioactive • Thyroid Carcinoma

## Further Reading

- Duffy, B. J., and Fitzgerald, P. J. (1950). Thyroid cancer in childhood and adolescence: A report on twenty-eight cases. *Cancer* 3, 1018-1832.
- Elisei, R., Romei, C., Vorontsova, T., Cosci, B., Veremeychik, V., Kuchinskaya, E., Basolo, F., Demidchik, E. P., Miccoli, P., Pinchera, A., and Pacini, F. (2001). Ret/ptc rearrangements in thyroid nodules: Studies in irradiated and not irradiated, malignant, and benign thyroid lesions in children and adults. *J. Clin. Endocrinol. Metab.* 86, 3211-3216.

- Hancock, S. L., Cox, R. S., and McDougall, I. R. (1991). Thyroid disease after treatment of Hodgkin's disease. *N. Engl. J. Med.* **325**, 599–605.
- Pacini, F., Vorontsova, T., Demidchik, E. P., Molinaro, E., Agate, L., Romei, C., Shavrova, E., Cherstvoy, E. D., Ivashkevitch, Y., Kuchinskaya, E., Schlumberger, M., Ronga, G., Filesi, M., and Pinchera, A. (1997). Post-Chernobyl thyroid carcinoma in Belarus children and adolescents: Comparison with naturally occurring thyroid carcinoma in Italy and France. *J. Clin. Endocrinol. Metab.* **82**, 3563–3569.
- Pacini, F., Vorontsova, T., Molinaro, E., Kuchinskaya, E., Agate, L., Shavrova, E., Astachova, L., Chiovato, L., and Pinchera, A. (1998). Prevalence of thyroid autoantibodies in children and adolescents from Belarus exposed to the Chernobyl radioactive fallout. *Lancet* **352**, 763–766.
- Schneider, A. B., Recant, W., Pinsky, S. M., Ryo, U. Y., Bekerman, C., and Shore-Freedman, E. (1986). Radiation-induced thyroid carcinoma: Clinical course and results of therapy in 296 patients. *Ann. Int. Med.* **105**, 405–412.
- Shore, R. E. (1992). Issues and epidemiological evidence regarding radiation-induced thyroid cancer. *Radiation Res.* **131**, 98–111.
- Socolow, E. L., Hashizume, A., Neriishi, S., and Niitani, R. (1963). Thyroid carcinoma in man after exposure to ionizing radiation: A summary of the findings in Hiroshima and Nagasaki. *N. Engl. J. Med.* **268**, 406–410.

typically of the extracellular domains, cytokine receptors are broadly classified into four subfamilies. The type I or hemopoietic cytokine receptors bind IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, GM-CSF, G-CSF, erythropoietin, growth hormone, prolactin, and oncostatin M. Their extracellular domain has two fibronectin type III modules, the hinge region between which is believed to be the site for ligand contact. The WSXWS (Trp-Ser-X-Trp-Ser) motif is also found in type I receptors (except prolactin and growth hormone) and is located close to the plasma membrane. The cytoplasmic tails of these type I receptors contain two special motifs: a well-defined box 1 and a not so well-defined box 2. Reports suggest that the box 1 and box 2 domains are important for association with and activation of Janus kinases (JAKs). The type II (interferon, IFN) family receptors include receptors for IFN- $\alpha/\beta$  (which share a common receptor), IFN- $\gamma$ , and IL-10. Tumor necrosis factor receptors (TNFRs I and II), as well as receptors for Fas, nerve growth factor, and CD40, are examples of type III cytokine receptors. Cytokine receptors bearing immunoglobulin-like structures are classified as type IV receptors, and the IL-1 receptor proteins are a prototype for this family.

How a given cytokine affects cell function depends on the cell type, its state of differentiation, and the stage in the cell cycle. A given cytokine can exert multiple effects on the same cell (pleiotropic effects), many cytokines can perform the same set of actions on a given cell type (redundant effects), and one cytokine can exert antagonistic or synergistic effects on the action of another cytokine (combinatorial effects). Some of these properties can be explained by the similarities and differences in signaling machinery expressed by distinct cell types at different times during the cell cycle. For example, redundancy in cytokine action may be attributed, at least in part, to the sharing of common signal-transducing receptor chains by several cytokine receptor complexes. A given receptor complex recruits only particular JAK species and activates only selected signal transducer and activator of transcription (STAT) proteins. The association of JAKs and STATs with different receptors in the context of signaling by different cytokines permits different cellular effects to be elicited using a limited repertoire of intracellular signaling molecules. It may be speculated that the existence of different associated proteins (e.g., chaperones, adaptors) in different cell types may then permit cross-talk among different pathways in a regulated yet interdependent fashion, perhaps accounting for the pleiotropism of cytokine actions.

## JANUS KINASES: STRUCTURE AND DOMAINS

Downstream signal transduction events from the cell surface can be mediated by protein tyrosine kinases (PTKs) through either intrinsic receptor tyrosine kinases (RTKs) or recruitment of nonreceptor protein tyrosine kinases (non-RTKs). The JAK subfamily is structurally unique among the more than 10 non-RTK subfamilies (e.g., Src, FAK, Syk, Fes, Tec) that are known. JAKs derive their name from the fact that they possess a catalytic tyrosine kinase domain (JH1) juxtaposed to a second inactive tyrosine kinase-like domain (JH2)—in an analogy to Janus, the mythological two-headed “keeper of gates.” The second JH domain (JH2) is nonfunctional and is often referred to as the “pseudokinase” domain. The other unique feature of JAK proteins is the absence of any Src homology domains (e.g., SH2, SH3), a characteristic feature of other nonreceptor PTK proteins. Four members of the JAK family protein are known in mammals and have been designated JAK1, JAK2, JAK3, and Tyk2. JAK1, JAK2, and Tyk2 are more or less expressed ubiquitously, whereas JAK3 is almost exclusively expressed in hematopoietic tissue. A *Drosophila* JAK has been identified and is called Hopscotch. More recently, piscine JAKs from zebrafish and carp have also been cloned. Alternately spliced versions of JAK2 and JAK3 have been identified, and it has been speculated that they may have a role as natural dominant-negative forms of these proteins.

In each JAK protein (size in the range of 1124–1187 amino acid residues in humans), the nonconserved amino terminal is followed by seven highly conserved domains termed JAK homology domains (JH domains), beginning with JH7 at the amino terminus and ending with JH1 at the carboxy terminus. JH1 is the active tyrosine kinase domain and shares considerable homology with the kinase domains in other PTKs. JH2, the adjacent kinase homology domain, is a nonfunctional domain because it lacks specific amino acids required for kinase activity. Saharinen and colleagues described a regulatory role for this domain in JAK2 function, perhaps through its interaction with the catalytic kinase JH1 domain. The function of the other JH domains is still not understood and is the subject of intensive investigations. Several recent reports suggest a role for the JH7–JH6 region in binding to cytokine receptor chains. It is likely that these domains interact with different adaptor or binding proteins in the context of different receptors and so confer specificity on respective signaling pathways.

## JANUS KINASES: ACTIVATION, MECHANISM OF ACTION, AND SUBSTRATES

Typically, JAKs reside in the cell in an inactive form. Cytokine binding to cell surface receptors leads to receptor chain dimerization/oligomerization, followed by recruitment of JAKs to the cytosolic face of the receptor complexes. Exactly how the JAKs are activated is still being investigated, but a popular model is one of tyrosine phosphorylation by JAKs in trans bound to the separate but now adjacent receptor chains that come close together on receptor chain dimerization/oligomerization triggered by ligand binding. JAK activation by tyrosine phosphorylation can include the participation of other receptor or nonreceptor PTKs. Inactivation of protein tyrosine phosphatases (PTPs), such as by the PTP inhibitor orthovanadate, has been used extensively as a mechanism for “ligand-independent” JAK activation.

Cellular ligands that activate JAKs include, but are not limited to, cytokines and growth factors. Specific cytokines signal through specific receptors and activate specific JAKs. Thus, IL-3 activates mainly JAK2, IL-9 activates mainly JAK3, and IFN- $\gamma$  activates JAK1 and JAK2. IL-6, CNTF, OSM, and LIF all activate JAK1, JAK2, and Tyk2. JAKs then phosphorylate specific tyrosine residues on the receptor chains and create docking sites for the binding of SH2 domain-containing proteins, most importantly STAT family members. STATs are then phosphorylated on key tyrosine residues, and they dimerize and make their way to the nucleus, where they up-regulate target genes. Although JAKs are capable of phosphorylating STATs, other PTKs (including Src) also have a role in this event. JAKs do not seem to exhibit any specificity per se regarding which STATs they activate. In light of this, Rane and Reddy proposed the JAK–Src–STAT model, where JAKs have a predominant role in creating docking sites for STATs, and either Src or JAKs can then phosphorylate STATs. It is noteworthy that docking sites created by JAKs on cytokine receptors have also been shown to bind other substrate proteins such as Shc, Src kinases, and PI-3 kinase, thereby mediating cross-talk among different signaling pathways.

## REGULATION OF JAK ACTIVITY: ROLE OF PHOSPHATASES AND THE SOCS PROTEIN FAMILY

As a pathway critical to several key functions performed by the cell, the JAK–STAT pathway

can be expected to be tightly regulated. The kinase–phosphatase theme for phosphorylation–dephosphorylation of signaling molecules is a general approach to regulate many cellular activities. PTPs, which dephosphorylate tyrosine residues on activated JAKs, are emerging as potential candidates regulating JAK activity, and it is very likely that more than one PTP can play a role in regulating the activity of a particular JAK, depending on the cell type and the cytokine context. SH2 domain-containing phosphatases-1 and-2 (SHP-1 and SHP-2) are by far the most well-studied PTPs. Other PTPs, such as TC-PTP, PTP-1b, and PTPeC, have been implicated in the down-regulation of activity of one or the other JAK.

An exciting new development in the field is the discovery of a family of proteins that negatively regulate JAK activity. The suppressor of cytokine signaling (SOCS) proteins are a family of small, cytokine-inducible, SH2 domain-containing proteins. Eight members of this family (SOCS-1 to SOCS-7 and CIS) in mammals, and three members in *Drosophila*, have been identified. Some of them have been shown to bind the JAK kinase domain and so down-regulate catalytic activity. A second mechanism for down-regulation of JAK activity by SOCS proteins is through a C-terminal domain in these proteins, the SOCS box, which has been suggested to target bound proteins for ubiquitination and subsequent proteosomal degradation. The SOCS box has been shown to be present in more than 20 proteins besides those in the SOCS protein family itself. Perhaps these additional proteins may serve as regulators of pathways other than the JAK–STAT pathway. Moreover, cytokines themselves transcriptionally up-regulate the expression of SOCS proteins as part of a negative feedback loop.

## NEW INSIGHTS: JANUS KINASES AND CYTOKINE RECEPTORS IN LIPID RAFTS

Thermodynamic parameters govern complex biochemical phenomena, and this holds true for signal transduction within cells as well. There is now a growing understanding that multiprotein signaling events, such as those outlined previously, take place within minutes in the context of preformed supramolecular assemblies. Such preformed assemblies—at the level of the plasma membrane—of various signaling components in close proximity make for a more thermodynamically efficient model for signal transduction, in contrast to an initial state with free-floating reactants in the cytosol or in the plane of

the plasma membrane. Recent research has led to the elucidation of organized microdomains within the plasma membrane that specialize in signal transduction. Originally described during the 1950s as microscopic invaginations or “caveolae” in the plasma membrane, such regions are now more commonly referred to as lipid rafts. These raft microdomains are enriched in cholesterol and sphingolipids, contain integral raft proteins (caveolins, flotillins, and stomatins), and have an elaborate repertoire of signaling pathway components. The function of caveolae and rafts has now been confirmed in cholesterol trafficking and various signaling pathways (e.g., T-cell antigen receptor, B-cell receptor, platelet-derived growth factor [PDGF], Ha-Ras, insulin receptor, EphrinB1). The raft-signaling hypothesis proposes that most, if not all, signaling events are initiated at the level of plasma membrane rafts that contain the receptors and associated signaling components in a preassembled supramolecular form. Exactly how the lipid partners play a role in these events is being investigated.

During recent years, components of the JAK–STAT pathway have also been identified in preassembled complexes in lipid rafts in the plasma membrane, and there is growing functional evidence to suggest that cytokine-induced STAT signaling is initiated at the level of such lipid rafts. Various investigators have reported the existence of cytokine receptor chains (gp130, IFN- $\gamma$ R $\alpha$ , IFNAR-1, and IL-2R $\alpha$ ), JAKs (JAK1, JAK2, and Tyk2), and STATs (STAT1 and STAT3) within lipid rafts. That the integrity of the caveolar structure is required in JAK–STAT signaling has been demonstrated in the context of IL-6–STAT3, IFN- $\gamma$ –STAT1, bradykinin–STAT3, and prolactin–STAT5a signaling. We have proposed the “raft–STAT signaling hypothesis” and believe that this is a general mechanism that operates in the broader context of cytokine- and growth factor-induced JAK–STAT signaling.

## JANUS KINASES: PATHOLOGICAL IMPLICATIONS

In mice, JAK1 or JAK2 disruption or deletion invariably results in a phenotype incompatible with life. JAK3 knockout mice exhibit severe hematopoietic anomalies with grossly immunocompromised

phenotypes. In humans, no pathology has been attributed to absent or decreased endogenous activity of JAK1, JAK2, or Tyk2—probably because, as in mice, disruption of their function causes ubiquitous defects that are incompatible with life. JAK3 mutations in humans are expected to cause defects in the immune system, and indeed several patients with severe combined immunodeficiency syndrome (SCID) have been shown to have mutations on the JAK3 locus. Aberrant overactivation of JAK activity has also been implicated in malignancies, including acute lymphoblastic leukemia. Aberrant SOCS activity has been demonstrated to have a role in inflammatory bowel diseases and in gastrointestinal cancers and leukemias, among other diseases.

## See Also the Following Articles

Cytokine Actions, Cellular Mechanism of • Cytokine Receptors • Interferons • Lipid Second Messengers and Receptors • Mitogen-Activated Protein (MAP) Kinases and Receptors • Receptor Serine/Threonine Kinases • Receptor Tyrosine Kinase

## Further Reading

- Kile, B. T., Schulman, B. A., Alexander, W. S., Nicola, N. A., Martin, H. M., and Hilton, D. J. (2002). The SOCS box: A tale of destruction and degradation. *Trends Biochem. Sci.* **27**, 235–241.
- Levy, D. E., and Darnell, J. E., Jr. (2002). STATs: Transcriptional control and biological impact. *Natl. Rev. Mol. Cell Biol.* **3**, 651–662.
- Miyajima, A., Kitamura, T., Harada, N., Yokota, T., and Arai, K. (1992). Cytokine receptors and signal transduction. *Annu. Rev. Immunol.* **10**, 295–331.
- Rane, S. G., and Reddy, E. P. (2000). Janus kinases: Components of multiple signaling pathways. *Oncogene* **19**, 5662–5679.
- Saharinen, P., and Silvennoinen, O. (2002). The pseudokinase domain is required for suppression of basal activity of Jak2 and Jak3 tyrosine kinases and for cytokine-inducible activation of signal transduction. *J. Biol. Chem.* **277**, 47954–47963.
- Sehgal, P. B., Guo, G. G., Shah, M., Kumar, V., and Patel, K. (2002). Cytokine signaling: STATs in plasma membrane rafts. *J. Biol. Chem.* **277**, 12067–12074.
- Shah, M., Patel, K., Fried, V. A., and Sehgal, P. B. (2002). Interactions of STAT3 with caveolin-1 and heat shock protein 90 in plasma membrane raft and cytosolic complexes: Preservation of cytokine signaling during fever. *J. Biol. Chem.* **277**, 45662–45669.
- Taniguchi, T. (1995). Cytokine signaling through nonreceptor protein tyrosine kinases. *Science* **268**, 251–255.

## ETIOLOGY AND PATHOGENESIS

The endocrine regulation of testicular or ovarian functions is performed by the gonadotropic hormones, luteinizing hormone (LH) and follicle-stimulating hormone (FSH), secreted by the gonadotropic cells of the anterior pituitary gland, the function of which is regulated via gonadotropin-releasing hormone (GnRH) released by the hypothalamus. In men, due to LH and FSH deficiency, the testicular functions of testosterone production and spermatogenesis are impaired under conditions of insufficient GnRH release (e.g., in Kallmann's syndrome and IHH). In women, the same applies to estrogen production and ovulation. For the X chromosomal recessive variant of the disease, the underlying mechanism of defective GnRH release has been described as impaired migration of GnRH neurons from the nasal olfactory epithelium to their proper anatomical location in the basal hypothalamus, a process that occurs during normal embryonic development. The gene responsible for this phenomenon is located on the short arm of the X chromosome (Xp22.3) and is called *KAL-1*. The corresponding protein seems to be an extracellular regulator of the directed outgrowth of axons and is called anosmin-1 (after the clinical feature of anosmia found in Kallmann's syndrome). Nevertheless, pedigrees of patients suggest an autosomal dominant and autosomal recessive inheritance as well, and most patients (65%) represent sporadic cases. Well-documented familial case reports demonstrate that Kallmann's syndrome, IHH, and isolated anosmia without concomitant deficient GnRH release can be regarded as manifestations within the spectrum of one underlying disease since these disorders are found in close relatives. Genes currently recognized

to be involved in congenital hypogonadotropic hypogonadism include *KAL-1*, the GnRH receptor, *DAX-1*, and *PROP-1*. Furthermore, sporadic cases of mutations in *ANK-1*, *SF-1*, *FGFR-1*, *LHX-3*, and *HESX-1*, as well as the leptin (*Ob*) gene, the leptin receptor (*Ob-R*) gene, and the prohormone convertase-1 gene, have been reported. However, a genetic basis for IHH has been established in less than 30% of cases; thus, several autosomal and X-linked genes await description.

## CLINICAL FEATURES

The clinical hallmark of Kallmann's syndrome and IHH is the absence or incompleteness of pubertal development accompanied by symptoms of hypogonadism (Table I). Male patients often present with uni- or bilateral maldescendent testes; sometimes, orchidopexy has been performed during childhood. The scrotum is usually hypoplastic, bearing small, soft testes (in contrast to Klinefelter's syndrome, in which the testes are small and firm). Other clinical features of prepubertal hypogonadism are underdevelopment of the penis and prostate; the absence or scarceness of pubic, axillary, or body hair; lack of beard growth; eunuchoid proportions of the body (arm span > body height in males); and a female distribution of fat tissue. Sexual activities are often undeveloped or are reduced, and the patients are infertile. A long-standing deficiency of sex steroids will often cause reduced bone density and slight anemia. Since estrogen production is low as well, gynecomastia is a rare finding. Females with Kallmann's syndrome or IHH display a delay in puberty, primary amenorrhea, and often short stature.

Table I Symptoms of Kallmann's Syndrome and Idiopathic Hypogonadotropic Hypogonadism (Prepubertal Onset)

Organ / function	Males	Females
Larynx	No voice mutation	Infantile
Skin	Female hair pattern, low sebum production	Scarce pubic hair, low sebum production
Bones	Eunuchoidal body proportions, osteoporosis	Often small stature, osteoporosis
Breasts	Rarely gynecomastia	Underdeveloped
Erythropoiesis	Slight anemia	Slight anemia
Muscles	Underdeveloped	Underdeveloped
External genitalia	Infantile	Infantile
Prostata/uterus	Underdeveloped	Underdeveloped
Testes/ovaries	Small volume, often maldescended	Small volume, mostly primordial stage
Spermatogenesis/ovulation	Not initiated	Not initiated
Mood	Often depressed	Often depressed
Libido	Not developed/sexual infantilism	Not developed/sexual infantilism

In patients with Kallmann's syndrome, anosmia is found due to an inability to perceive olfactory stimuli; this is a result of aplasia or hypoplasia of the olfactory bulbs and tracts. The insensitivity relates to aromatic substances, whereas stimulants of the trigeminal nerve, such as ammonia, are perceived. The clinical feature of anosmia is notably absent in IHH, which is otherwise similar to Kallmann's syndrome. Rare anomalies associated with Kallmann's syndrome are impaired hearing, a cleft lip or high arched palate, synkinesis of the extremities, and unilateral renal aplasia. Within the cardiovascular system, malformations such as a double-outlet right ventricle or malposition of the great vessels have been described. There is significant clinical variability with regard to severity and time of onset: Minor forms with partial pubertal development or a late onset during adulthood have been described along with the fully expressed disease. In women, the X-linked form demonstrates penetrance of high variability. It has to be considered that in women a partial development of sexual characteristics can occur due to gonadotropin-independent adrenal steroidogenesis, which does not have a major impact on the development of males.

## ESTABLISHMENT OF DIAGNOSIS

If the clinical picture, physical exam including genitalia, and patient's history lead to the suspected diagnosis of Kallmann's syndrome or IHH, basal serum levels of the gonadotropins, testosterone, and estradiol will yield valuable information. Typically, patients with Kallmann's syndrome or IHH have markedly low levels of the previously mentioned parameters (hypogonadotropic hypogonadism). The responsiveness of gonadotropins to a GnRH stimulus (100 µg intravenously, remeasured after 30 and 45 min) is often poor or absent, which should not immediately be interpreted as a disorder at the pituitary level. The gonadotropic cells in the anterior pituitary often require a certain period of "priming" by means of subcutaneous application of GnRH in a pulsatile manner: For 7 days, 5 µg of GnRH is released every 90–120 min using a portable minipump. If the gonadotropin response (increase in LH and FSH levels  $\geq 3$  IU/liter) to a GnRH bolus is normalized, then the hypothalamic origin of the hypogonadism is indicated. When there is permanent resistance to GnRH, a pituitary disorder or mutation of the GnRH receptor must be suspected.

Putative anosmia can be examined by standardized sets of olfactants that should contain mucosal irritants and gustatory stimuli as controls. If these are not

available, coffee and perfumed soap can be used as aromatic substances and vinegar may be used as the control.

Hypogonadal endocrine disorders require the assessment of other hormone axes that might also be affected: The levels of thyroid hormones and TSH, insulin-like growth factor-1 and growth hormone, cortisol, and prolactin should be determined. The diagnostic workup should include imaging procedures, such as ultrasonography of testes, prostate/ovaries, and uterus as well as magnetic resonance imaging of the hypothalamic-pituitary region, to exclude intracranial neoplasms. Measurement of bone density and bone age (in prepubertal patients) should be performed as well.

Because Kallmann's syndrome is a genetic disorder, a meticulous family history and genetic counseling in cases of desired paternity/maternity are paramount. Sequencing of the *KAL-1* gene and description of polymorphisms/mutations are recommended not only for scientific reasons but also because they may be helpful in counseling the patient and family.

## DIFFERENTIAL DIAGNOSIS

The most important clinical entity related phenotypically to Kallmann's syndrome and IHH is the constitutional delay of puberty (CDP). It presents with hormone profiles similar to those of the previously discussed diseases, and often there is a familial history. To differentiate CDP from Kallmann's syndrome or IHH, a 36-h GnRH pump test followed by a GnRH stimulus is a feasible method. In cases of CDP, but not Kallmann's syndrome or IHH, short-term "priming" is sufficient to provoke a significant LH release ( $\geq 3$  IU/liter). When response is lacking, the test should be performed for 7 days. Moreover, CDP is not linked to anosmia.

Secondary GnRH deficiency may occur in cases of tumors in the region of the diencephalon, in granulomatous illnesses, hemochromatosis, histiocytosis, after trauma/fracture of the skull base, in ischemic or hemorrhagic lesions, and after radiotherapy. Severe cases of malnutrition, such as in Crohn's disease, malabsorption syndrome, anorexia nervosa, and chronic diseases, can cause hypothalamic malfunction. This may also be the case with social deprivation.

Congenital disorders concomitant with hypogonadotropic hypogonadism due to inadequate GnRH release are the Prader-Labhart-Willi, Bardet-Biedl, and Laurence-Moon syndromes, some entities within the category of cerebellar ataxias, and congenital adrenal hypoplasia (the latter often caused by mutations in

the *DAX-1* and *SF-1* genes, which are found in IHH as well).

Hypogonadotropic hypogonadism may also occur in pituitary disorders: Although hypothalamic GnRH release is intact, gonadotropins are not released. This can be due to tumors (e.g., prolactinoma), isolated LH or FSH deficiency, aplasia of the pituitary, or chronic diseases/externally caused damage. In rare cases, GnRH receptor mutations can cause insensitivity to this stimulating hormone. In some families with IHH, this has been determined to be the underlying cause.

Often presenting with a clinical picture similar to that of Kallmann's syndrome or IHH but with hypergonadotropic hypogonadism is Klinefelter's syndrome, which is found in males and caused by one or more additional X chromosomes. The Ullrich–Turner syndrome (single X chromosome) in females is associated with hypergonadotropic hypogonadism as well and may, in some cases, morphologically resemble the picture of gonadotropin deficiency.

## THERAPY

### Treatment Modalities

Patients with Kallmann's syndrome or IHH present with a deficiency of sex steroids. Induction of puberty

and rapid achievement of a higher level of general physical well-being and activity can be obtained by substitution of testosterone in males and estrogens/gestagens in females (Table II). Testosterone administered to males will be aromatized to estrogens physiologically, thus initiating the necessary dual influence of both sex steroids (especially on bone tissue). After sexual differentiation and development have been achieved, the induction of reproductive functions should be discussed with the patient. Direct substitution of sex steroids is not helpful in this instance; For therapy, the administration of gonadotropins [LH/human chorionic gonadotropin (hCG) and FSH injections subcutaneously] or GnRH (via the previously described minipump) is essential (Table III). In males, the induction of spermatogenesis and, thus, achievement of fertility can be obtained within several months in most cases. Some cases may require treatment for approximately 24 months. The time to first appearance of sperm in the ejaculate correlates negatively with testis volume at the onset of therapy. Cryptorchidism or markedly subnormal testicular size represent no contraindication for this therapy. It is advisable to initiate spermatogenesis soon after or during puberty since reinitiation of fertility in later years will occur more rapidly. Regarding testis

Table II Modalities of Sex Steroid Substitution in Kallmann's Syndrome/IHH

Substitution in males			
Application	Preparation	Dosage	
Intramuscular injection	Testosterone enanthate	200–250 mg every 2–3 weeks; for induction of puberty, a 4-week interval may be appropriate	
Intramuscular injection	Testosterone cypionate	200 mg every 2 weeks; for induction of puberty, a 4-week interval may be appropriate	
Transdermal application	Testosterone gel	50–100 mg per day	
Scrotal skin (i.e., Testoderm)	Transdermal testosterone patch	One membrane per day may be inappropriate due to small scrotum	
Nonscrotal skin (i.e., Androderm)	Transdermal testosterone patch	One or two patches per day	
Orally, with meals	Testosterone undecanoate	Two to four capsules of 40 mg per day; not recommended for induction of puberty	
Subcutaneous implantation	Testosterone implants	Three to six implants (200 mg per rod) per 6-month period	
Substitution in females			
Preparation	Application	Dosage	Time frame
Estradiol valerate	Oral	0.2 mg per day	0–6 months
Estradiol valerate	Oral	0.5 mg per day	6–12 months
Estradiol valerate + chlormadinonacetate	Oral	1.0 mg per day (days 1–25); 2.0 mg per day (days 14–25)	Second year of therapy
Estradiol valerate + chlormadinonacetate	Oral	1.5 mg per day (days 1–25); 2.0 mg per day (days 14–25)	Third year of therapy
Common bi- or triphasic oral ovulation inhibitor	Oral	Depending on preparation	Fourth year of therapy and onward

Table III Treatment Modalities to Achieve Fertility in Kallmann's Syndrome / IHH

Substance	Modality	Dosage	Gender
GnRH or Human choriongonadotropin in combination with Human menopausal gonadotropin or in combination with Purified or recombinant FSH	Pulsatile subcutaneously by minipump Subcutaneously or intramuscularly Subcutaneously or intramuscularly Subcutaneously or intramuscularly	5–20 µg/pulse every 2 h 1000–2500 IE two or three times per week 150 IE three times per week 150 IE three times per week	Females and males Males Males Males

volume, a longer duration of treatment is required in patients who have not undergone spontaneous partial puberty or in those who have been diagnosed with cryptorchidism. After successful induction, spermatogenesis can be maintained at a lower level by administration of hCG alone. This regimen represents an alternative to switching to testosterone substitution.

Because healthy boys have a significant increase in reproductive hormones during the early postnatal period, it has been debated whether postnatally diagnosed hypogonadotropic hypogonadism should be treated during early infancy to induce later penile and testicular growth. Thus, Leydig cell proliferation and germ cell differentiation could be improved and lead to better treatment results in later life.

In women, pulsatile GnRH administration to induce follicle growth and ovulation is preferred to injections of gonadotropins since the risk of ovarian overstimulation and pregnancies with multiple fetuses is reduced. Pulsatile treatment in females should be preceded by substitution of steroids to provide normally developed reproductive organs. Thus, the induction of ovulation and a later pregnancy are supported. Once the desire for offspring has been fulfilled, therapy may be changed to sex steroid substitution. During pregnancy, no substitution therapy is needed since during the first 7 to 8 weeks after implantation the functional capacity of the corpus luteum is preserved by hCG secreted by the trophoblast. After the luteal-placental shift, sex steroids are secreted by the placental trophoblast and the decidua. Additional hormone substitution is contraindicated since fetal development may be irrevocably disturbed.

Genetic counseling should be given to patients who desire to have children since there are reports of transmission of the disease to the offspring.

## Monitoring Substitution Therapy

### Men

In men, regular surveillance is required during androgen substitution therapy to ensure that the patient is well adjusted and to avoid side effects. This applies to behavioral aspects, somatic properties, and laboratory parameters. Because testosterone is metabolized by  $5\alpha$  reduction to  $5\alpha$ -dihydrotestosterone or by aromatization to estrogens, the metabolic activity of these steroids should be considered as well. As with other hypogonadal patients receiving androgens, patients with Kallmann's syndrome or IHH report increased beard growth, indicated by a higher frequency of shaving. A male pattern of pubic hair growth will develop. Muscles and physical strength are gained and as body weight increases in lean body mass, the fraction of fat will diminish. Gynecomastia can be caused by increased aromatization to estrogens, especially during therapy modalities leading to high peak levels (i.e., intramuscular injections of short-acting testosterone esters). In this event, the testosterone dose must be adjusted. Sufficient androgen substitution will lead to increased physical and mental activity, alertness, and vigor. Conversely, low levels can be accompanied by lethargy, inactivity, and depressed mood. These symptoms are often described by patients receiving intramuscular testosterone esters. Especially in older patients, in whom induction of puberty is achieved after a psychosocial adaptation has occurred, major psychological changes can lead to restructuring of family and partner relationships. This often requires intensive counseling not only for the patient but also for relatives.

When determining the effects of substitution therapy, the intrinsic pharmacokinetic profiles of different testosterone preparations must be considered. The time at which a sample is obtained is as important

as the time period since the last administration. If injection therapy is chosen, intervals rather than doses should be adjusted accordingly.

Serum estradiol should be assessed when serum levels of testosterone are high, especially when a patient is being treated with enanthate or cypionate ester, and intervals should be prolonged if the serum estradiol level is too high. Bioavailable fractions of testosterone can be estimated from total serum testosterone and SHBG levels.

Because testosterone is a well-known stimulator of the erythropoietic system, the red blood cell count will increase. Supraphysiological levels of testosterone can lead to polycythemia, requiring dose reduction to prevent patients from experiencing embolic or thrombotic events. Because replacing testosterone in hypogonadal patients will increase their bone density, this parameter should be assessed prior to treatment and every 2 years thereafter.

The incidence of prostate carcinoma and benign prostate hyperplasia increases with age. Hence, every patient, especially males older than age 45, should be examined carefully. Rectal palpation, serum prostate specific antigen (PSA) levels, and a transrectal ultrasonographic examination (a transpelvic ultrasonography does not provide exact volume determination) provide the experienced physician with the means for diagnostic accuracy. In the course of treatment, these parameters should be assessed at least yearly. Additional information concerning a functional parameter can be obtained by uroflow measurement. Testosterone increases prostate volume in hypogonadal men, but only to the extent of prostate size in age-matched controls. These parameters must be monitored as well in patients receiving GnRH or gonadotropin substitution. It is likely that prostate growth is pharmacogenetically influenced by the CAG repeat polymorphism of the androgen receptor gene. Since shorter alleles facilitate enhanced androgen action, one might consider determining the polymorphism and, consequently, adapting surveillance and testosterone dose.

### Women

When pregnancy is not desired, in most cases hormone substitution is provided by orally active agents (Table II). Metabolic side effects are well-known from oral contraceptives and should be considered when monitoring therapy. Side effects include impairment of glucose tolerance, an increase in low-density lipoprotein cholesterol and triglycerides, and an increase in thromboembolic events (especially in smoking women). Endometrial effects and induction of menstrual bleeding should be carefully observed.

The development of benign and malignant changes in breast tissue has to be taken into account. Hormone substitution in women will have a positive effect on bone tissue, leading to formation of higher bone mass and preventing osteoporosis. Regarding psychosocial effects, the same phenomena as described for men can occur in women. Cessation of therapy can be discussed after reaching the age of menopause.

In cases of GnRH or gonadotropin substitution, follicular development has to be closely monitored. Overstimulation should be avoided and contraceptive measures should be taken if multiovation occurs.

### See Also the Following Articles

Agonadism, Male and Female • Bardet-Biedl Syndrome • Constitutional Delay of Growth and Puberty (CDGP) • Delayed Puberty and Hypogonadism, Female • Delayed Puberty and Hypogonadism, Male • FSH (Follicle-Stimulating Hormone) • Hypothalamic Disease • Hypergonadotropic Hypogonadism • Hypothalamic Hypogonadism • LH (Luteinizing Hormone) • Menstrual Cycle: An Integrative View • Prader-Willi Syndrome

### Further Reading

- Becker, K. L. (Ed.) (2001). *Principles and Practice of Endocrinology and Metabolism*, 3rd Ed. Lippincott Williams & Wilkins, Philadelphia.
- Büchter, D., Behre, H. M., Kliesch, S., and Nieschlag, E. (1998). Pulsatile GnRH or human chorionic gonadotropin/human menopausal gonadotropin as effective treatment for men with hypogonadotropic hypogonadism: A review of 42 cases. *Eur. J. Endocrinol.* **139**, 298–303.
- Depenbusch, M., von Eckardstein, S., Simoni, M., and Nieschlag, E. (2002). Maintenance of spermatogenesis in hypogonadotropic hypogonadal men with hCG alone. *Eur. J. Endocrinol.* **147**, 617–624.
- Nieschlag, E. and Behre, H. M., (eds.) (2000). *Andrology—Male Reproductive Health and Dysfunction*, 2nd Ed. Springer, New York.
- Pitteloud, N., Hayes, F. J., Boepple, P. A., de Cruz, S., Seminara, S. B., MacLaughlin, D. T., and Crowley, W. F. (2002). The role of prior pubertal development, biochemical markers of testicular maturation, and genetics in elucidating the phenotypic heterogeneity of idiopathic hypogonadotropic hypogonadism. *J. Clin. Endocrinol. Metab.* **87**, 152–160.
- Quinton, R., Duke, V. M., Robertson, A., Kirk, J. M. W., Matfin, G., de Zoysa, P. A., Azcona, C., MacColl, G. S., Jacobs, H. S., Conway, G. S., Besser, M., Stanhope, R. G., and Bouloux, P. M. G. (2001). Idiopathic gonadotropin deficiency: Genetic questions addressed through phenotypic characterization. *Clin. Endocrinol.* **55**, 163–174.
- Seminara, S. B., Hayes, F. J., and Crowley, W. F. (1998). Gonadotropin-releasing hormone deficiency in the human (idiopathic hypogonadotropic hypogonadism and Kallmann syndrome): Pathophysiological and genetic considerations. *Endocr. Rev.* **19**, 521–539.
- Wood, D. F., and Franks, S. (1998). Hypogonadism in women. In *Clinical Endocrinology* (A. Grossman, ed.), pp. 702–717. Blackwell, Oxford.

the near steady-state input of fuels in the blood during starvation. They first reassessed the rate of urinary excretion of nitrogenous compounds and ketone bodies in humans undergoing prolonged starvation and then combined these studies with measurements of arteriovenous concentration differences with regional blood flow rates and calculated exchange rates of substrates, oxygen, and carbon dioxide. These studies led to the demonstration in 1967 by Owen *et al.* that  $\beta$ -OHB and AcAc could replace glucose as the predominant fuel for brain metabolism during prolonged starvation. The extent and importance of ketone body metabolism in sparing glucose utilization by the brain during starvation were not previously appreciated. The dominant role that ketone body metabolism plays in fuel homeostasis was thereby established.

Much of human history involves a search for food, with long periods of prolonged hunger when food was not available. Human metabolism thus developed mechanisms to cope with starvation. The use of stored fat reserves and their conversion to ketone bodies as a source of energy surely played a dominant role in the survival of the human species. The ability of ketone bodies to maintain life through their capacity to supply fuels, maintain electrolyte conservation, promote gluconeogenesis, and maintain acid–base balance during starvation has diminished in significance since humans are usually in the fed or early postprandial states. It is ironic, considering human history, that in the United States and many technologically advanced nations, obesity, and not starvation, has become the major health problem. Public health officials warn that obesity is reaching epidemic proportions in many developing countries and is likely to lead to an equally dire epidemic of diabetes, hypertension, and related heart disease and stroke. With the ready availability of food, the significance of ketone bodies as a source of fuel in normal human metabolism has thus diminished. In fact, to most individuals not familiar with the nuances of whole body metabolism, ketone bodies are unhealthy reminders of diabetes and not an important metabolic fuel required for survival during starvation. It is of note that there is a renewed interest in ketone bodies due to the popularity of diets that are high in fat and protein and nearly devoid of carbohydrate, being recommended by some for weight reduction for obese individuals. In fact, the key to one popular diet, *Dr. Atkins Diet Revolution*, is the maintenance of high levels of ketosis both before and after meals! However, clinical studies of obese people consuming a very-low-carbohydrate but high-fat and

high-protein diet show that they develop only trivial elevations of blood  $\beta$ -OHB concentrations.

## BIOCHEMISTRY OF KETOGENESIS AND KETOLYSIS

After CoA was discovered and accurate methods for measuring  $\beta$ -OHB, AcAc, and acetone became available, knowledge regarding ketogenesis quickly grew. It was soon recognized that ketone bodies were not just intermediates in the oxidation of fatty acids but rather end-products that require their own enzymatic machinery for synthesis. The formation of ketone bodies occurs primarily in the liver via the following enzymatic reactions, all of which are present in the mitochondrial matrix, except for the spontaneous decarboxylation of AcAc to acetone, which occurs in the blood.

The reactions of ketone body synthesis are as follows:

1.  $2 \text{ Acetyl CoA} \rightarrow \text{Acetoacetyl CoA} + \text{coASH}$
2.  $\text{Acetoacetyl CoA} + \text{Acetyl CoA} \rightarrow \beta\text{-Hydroxy-}\beta\text{-methylglutaryl CoA}$
3.  $\beta\text{-Hydroxy-}\beta\text{-methylglutaryl CoA} \rightarrow \text{Acetoacetate} + \text{acetyl CoA}$
4.  $\text{Acetoacetate} + \text{NADH} \leftrightarrow \beta\text{-Hydroxybutyrate} + \text{NAD}$
5.  $\text{Acetoacetate} \rightarrow \text{Acetone} + \text{CO}_2$

The enzymes of ketone body synthesis are as follows: (1) Acetoacetyl CoA thiolase; (2)  $\beta$ -hydroxy- $\beta$ -methylglutaryl CoA synthase; (3)  $\beta$ -hydroxy- $\beta$ -methylglutaryl CoA lyase; (4)  $\beta$ -hydroxybutyrate dehydrogenase; and (5) Spontaneous decarboxylation.

The major source of ketone bodies is the oxidation of fatty acids in the liver. The kidney can synthesize, oxidize, and excrete ketone bodies. The pathway of ketogenesis in the renal cortex is different from that in the liver. A small quantity of AcAc can be synthesized from ketogenic amino acids during starvation. In addition, minute quantities of  $\beta$ -OHB can be synthesized by the central nervous system. The enzymes involved in the hepatic synthesis of ketone bodies are listed above. Although small quantities of acetoacetyl CoA may arise from the last four carbons of long-chain fatty acids during  $\beta$ -oxidation, the bulk of acetoacetyl CoA is formed from a head-to-tail condensation of two molecules of acetyl CoA via reversal of the acetoacetyl CoA thiolase reaction. Another acetyl CoA molecule combines with AcAc-CoA to form  $\beta$ -hydroxybutyrate- $\beta$ -methylglutaryl CoA (HMG-CoA) via the action of HMG-CoA synthetase. Generation of HMG-CoA simultaneously generates a proton; for each molecule of HMG-CoA committed to AcAc formation, one proton is released into the body fluids. HMG-CoA synthetase is nearly exclusively localized to the liver. In the next step in ketogenesis, HMG-CoA

lyase cleaves HMG-CoA to form free AcAc and acetyl CoA. AcAc, the parent ketone body, can be converted to  $\beta$ -OHB by mitochondrial  $\beta$ -hydroxybutyrate dehydrogenase.

Ketone bodies can be used for energy by most tissues. The exceptions are the liver as well as tissues that have no mitochondria, such as red blood cells and the lens of the eye. The liver produces ketone bodies but has an undetectable or low activity of succinyl CoA:acetoacetyl CoA transferase, the major enzyme involved in ketone body degradation. Thus, hepatic activation of AcAc for hepatic oxidation is minimal and ketone bodies are released into the blood. The enzymes involved in the degradation of the major ketone bodies, AcAc and  $\beta$ -OHB, in this sequence are listed below. The catabolism of acetone is more complex and is discussed below.

The reactions of ketone body degradation are as follows:

1.  $\beta$ -Hydroxybutyrate + NAD  $\leftrightarrow$  Acetoacetate + NADH + H<sup>+</sup>
2. Acetoacetate + succinyl CoA  $\leftrightarrow$  Acetoacetyl CoA + succinate
3. Acetoacetyl CoA + CoASH  $\leftrightarrow$  2 Acetyl CoA

The enzymes of ketone body degradation are as follows: (1)  $\beta$ -hydroxybutyrate dehydrogenase; (2) Succinyl CoA:acetoacetyl CoA transferase; and (3) Acetoacetyl CoA thiolase.

There are several differences in the metabolism of ketone bodies in the kidney and in the liver. There is negligible activity of HMG-CoA synthase in the kidney cortex. In addition, the kidney possesses considerable activity of succinyl CoA:acetoacetyl CoA transferase, an enzyme that is absent in the liver. Both organs have acetoacetyl CoA thiolase. The liver can synthesize and release both AcAc and  $\beta$ -OHB into the blood (but it cannot metabolize the ketone bodies further), whereas the kidney synthesizes these compounds by reversing the thiolase reaction. However, net renal ketogenesis is low compared to the rate of ketone body synthesis in the liver.

One important factor in the oxidation of ketone bodies is the concentrations of AcAc and  $\beta$ -OHB in biological fluids. AcAc enters the cell from the extracellular fluid or can be produced in the mitochondria from  $\beta$ -OHB by  $\beta$ -hydroxybutyrate dehydrogenase. Since this reaction reduces NAD to NADH,  $\beta$ -OHB has a greater caloric value than AcAc (4.5 kcal/g versus 4.0 kcal/g). The overall rate of conversion of  $\beta$ -OHB to AcAc is also determined by the oxidation–reduction state (NAD/NADH) of cells in the tissue using the ketone bodies. This ratio can be influenced by a number of factors, including the rate of energy utilization by the tissue and the extent of fatty acid oxidation

that generates both NADH and FADH<sub>2</sub> directly in the mitochondrial matrix. Mitochondrial activation of AcAc in the brain, skeletal muscle, and heart is catalyzed by succinyl CoA transferase, which uses the CoA from succinyl CoA. Acetoacetyl CoA is then cleaved by acetoacetyl CoA thiolase into two molecules of acetyl CoA and the acetyl CoA then enters the tricarboxylic acid (TCA) cycle for oxidation to CO<sub>2</sub>.

Succinyl CoA:acetoacetyl CoA transferase is detected in all tissues with mitochondria except the liver. The synthesis of acetoacetyl CoA via this enzyme occurs at an energy cost since the succinyl CoA that is the CoA donor in this reaction would normally be converted to succinate via succinyl CoA synthase in the TCA cycle, generating a molecule of GTP. The highest activity of succinyl CoA:acetoacetyl CoA transferase is in the myocardium > brain > kidney > other tissues. Acetoacetyl CoA thiolase is present in both the mitochondria and the cytosol. In the mitochondria, it promotes the cleavage of AcAc-CoA into acetyl CoA for oxidation in the TCA cycle.

A significant proportion of the AcAc that is produced is converted to acetone by nonenzymatic decarboxylation. Furthermore, an acid environment promotes AcAc decomposition to acetone. For each molecule of AcAc decomposed to form acetone, one H<sup>+</sup> is consumed in the formation of CO<sub>2</sub>. Acetone production is directly related to the concentration of AcAc that is present in the blood; during diabetic ketoacidosis, approximately 50% of the AcAc produced is converted into acetone. Acetone has a relatively slow rate of turnover, in part due to the large pool of acetone in the body during protracted hyperketonemia. Acetone is converted irreversibly to acetol and then to propyleneglycol (1,2-propanediol). Propanediol is converted to pyruvate, a gluconeogenic precursor. Thus, acetone metabolism violates the general rule that fatty acids (other than propionate) cannot support glucose synthesis in mammals.

The potential importance of acetone metabolism during starvation and diabetes has largely been ignored. Several reports demonstrated that acetone was present in blood, breath, and urine. Although it has been long held that the blood of patients suffering from diabetic ketoacidosis contained little free acetone, Sulway and Malins reported strikingly elevated concentrations of free acetone in the bloodstream of 27 diabetic patients admitted to the hospital in diabetic ketoacidosis. These observations have been neglected by most investigators studying ketone body metabolism. However, there is a general consensus that the overall importance of acetone as a precursor (or fuel) is minor compared to AcAc and  $\beta$ -OHB.

## PHYSIOLOGY OF KETONE BODY METABOLISM

It became evident early in the medical history of metabolic research that both starvation and diabetes mellitus generated states of hyperketonemic acidosis. However, the diabetic state was much more severe and resulted in death. Furthermore, the degree of acidosis and vulnerability of the diabetic patients varied widely. Starving humans usually develop a steady state of ketone body synthesis and catabolism, whereas diabetic patients usually develop severe hyperketonemia (ketosis) and progressively deteriorate to death. Since insulin therapy did not become available until 1921, there is an extensive and sad literature documenting this problem. Ketosis develops when glucose oxidation is suppressed and fat catabolism is accelerated. Diet and physical activity also influence the rapidity and degree of ketosis. Strenuous exercise can also cause a rapid rise in the concentration of ketone bodies in the blood; however, rest and relatively small quantities of glucose rapidly decrease blood ketone body concentrations to normal in nondiabetic individuals.

Occasionally, an obese individual with type 2 diabetes will develop ketoacidosis that is indistinguishable from a lean, insulin-dependent, type 1 diabetic patient. Hydration may diminish the degree of ketonemia and ketonuria in normal patients and in diabetic patients. However, once severe diabetic ketoacidosis develops, rehydration, though it decreases the concentration of blood glucose, has only a marginal effect on hyperketonemia and acidosis.

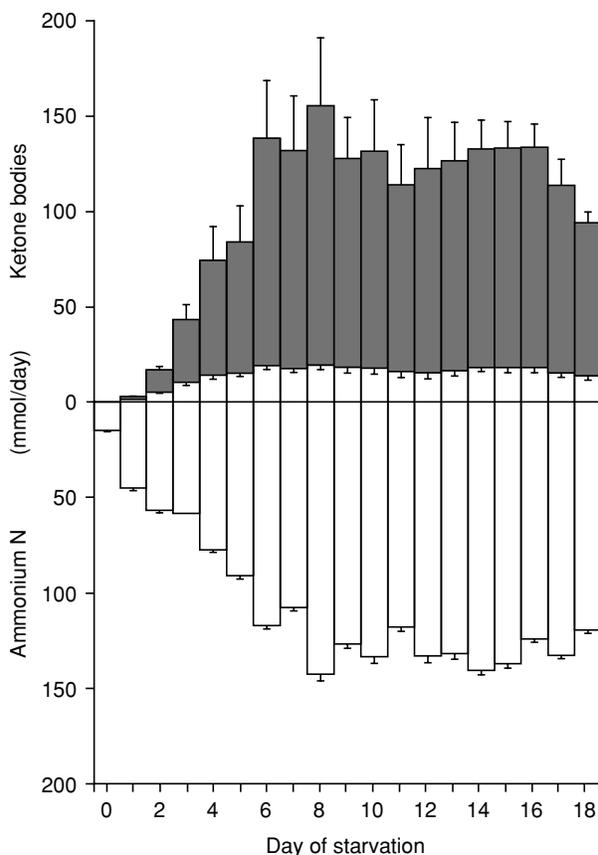
There are different degrees of hyperketonemia so it is important to distinguish the ketosis that occurs during starvation from that which occurs due to catastrophic diabetes. During starvation, ketone bodies are synthesized to provide fuels. Their production rates are equal to their oxidation and excretion rates and a steady state for ketone body metabolism occurs. This is "physiologic hyperketonemia." In contrast, during progressive diabetic hyperketonemia, there is an uncontrolled mobilization of stored fuels in excess of tissue needs due to inadequate insulin action and an overabundance of catecholamine, glucagon, growth hormone, cortisol, etc. The blood is then flooded with an overabundance of fuels. In addition to the hyperketonemia, there is hyperglycemia, acidosis, azotemia, and dehydration. This form of catastrophic catabolism is "pathologic hyperketonemia." It is clearly distinct from physiologic ketosis of starvation. In diabetic ketoacidosis, death ensues if not interrupted by the administration of insulin, fluids, electrolytes, and eventually glucose.

During starvation, when the blood insulin concentration is low and the hypoglycemic counterregulatory hormonal concentrations are high, lipolysis occurs and fatty acids are mobilized from triglycerides stored largely in adipose tissue. Long-chain fatty acids are not water-soluble and are bound to albumin for transport to the liver. After crossing the hepatic cellular membranes, some of the long-chain fatty acids become activated to fatty acyl CoA and undergo repeated steps of  $\beta$ -oxidation, generating acetyl CoA. In the process of  $\beta$ -oxidation, fatty acids yield approximately one-half of their chemical energy, with the remainder of their initial chemical energy in ketone bodies. In the liver during fasting, considerable concentrations of fatty acids undergo  $\beta$ -oxidation, supplying all the energy needs of this tissue. This takes place because  $\beta$ -oxidation of fatty acids occurs in the mitochondrial matrix and generates both NADH and FADH<sub>2</sub> during the process. The change in the mitochondrial redox state decreases the TCA cycle flux, resulting in a transient rise in the concentration of acetyl CoA. The major metabolic fate of this acetyl CoA is to form ketone bodies. Thus, ketone bodies act as a relief valve for the product of the oxidation of fatty acids by providing an alternative route for the disposal of the acetyl CoA. In muscle and other tissues with mitochondria, acetyl CoA derived from the cleavage of fatty acids is oxidized to CO<sub>2</sub> for energy. Ketone bodies have another important metabolic role that is not well understood; like fatty acids, all three ketone bodies induce insulin resistance.

## URINARY KETONE BODY AND AMMONIUM EXCRETION

Urinary ammonium excretion during starvation is directly related to ketonuria (Fig. 1). In humans, the maximum rate of ammonium excretion occurs after 7 to 8 days of starvation. After approximately 18 days of starvation in well-hydrated volunteers, urinary ammonium becomes the principal urinary nitrogenous product. It is of interest to note that the caloric value of protein is greatest when amino acid nitrogen is lost in the urine as ammonium since the synthesis of urea in the urea cycle in the liver requires 4 ATP molecules for each molecule of urea produced. Thus, the most energetically efficient way to excrete urinary nitrogen is as ammonium.

Ketonuria generally peaks after 7 days of starvation in humans and thereafter plateaus, with the maximum rate of excretion of AcAc and  $\beta$ -OHB of 110–120



**Figure 1** Daily urinary excretion of  $\beta$ -OHB (shaded bars) and AcAc (white bars) versus ammonium nitrogen (mean  $\pm$  SEM) in five obese men and women during 18 days of total starvation when only water, salts, and multivitamins were given. Relationship between total ketonuria and urinary  $\text{NH}_4^+$ :  $r > 0.95$ .

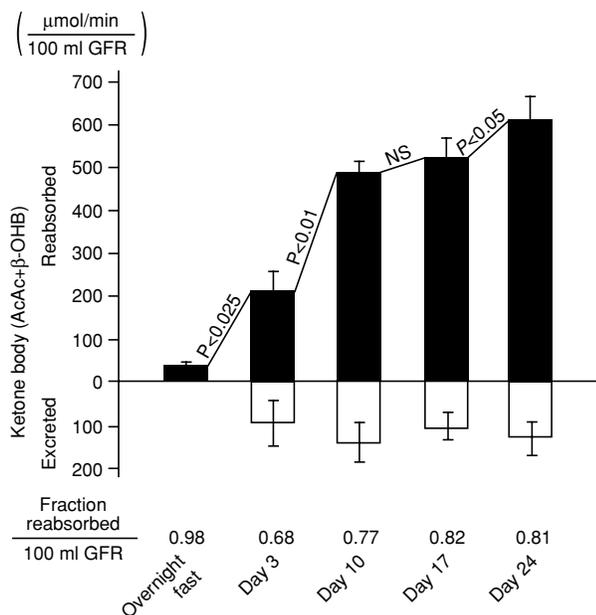
mmol/day. This measurement is complicated by the fact that the accumulation of AcAc in the urinary bladder and/or after collection and storage of the urine may cause some decomposition of AcAc to acetone (acetone is generally not routinely determined in the analysis of ketone bodies). When the daily rates of ketone body excretion were measured very close to the time that the urine collection was completed, the rate of  $\beta$ -OHB excretion was approximately 100 mmol/day (this value peaked at 7–8 days). The rate of excretion of AcAc (19 mmol/day) peaked at approximately the same time during starvation; there are slight decreases in the excretory values of AcAc,  $\beta$ -OHB, and ammonium in the urine as starvation progresses.

As noted above,  $\beta$ -OHB is the predominant ketone body excreted in the urine of humans (a ratio of 5:1 in the urine after prolonged fasting). Depending on

the urinary volume, the concentration of AcAc and  $\beta$ -OHB in the urine is usually approximately 10–20 times that of blood. Whereas AcAc is the minor component of these two ketone bodies, only AcAc reacts with the semiquantitative reagent nitroprusside, so that the routine analysis of ketone bodies using the convenient test strip severely underestimates the actual concentration of ketone bodies in biological fluids since it does not react with  $\beta$ -OHB, the major ketone body in biological fluids.

Early studies performed using dogs and humans suggested that there was a maximal rate of renal tubular reabsorption of ketone bodies. However, when more exacting analytical methods were used to measure the concentration of AcAc and  $\beta$ -OHB in blood and urine, no maximal renal reabsorptive rates for AcAc and  $\beta$ -OHB were found during starvation or diabetic ketoacidosis. The rate of urinary excretion of these freely filtered compounds was limited by the linear increase in their renal reabsorption (Fig. 2). The fractions of both AcAc and  $\beta$ -OHB reabsorbed and excreted per 100 ml glomerular filtrate by starving, obese patients at the height of ketonemia are approximately 80 and 20%, respectively. The same results were found during diabetic ketoacidosis. The excretion of acetone is uniquely low because acetone is soluble in both aqueous and lipid media and thus there is no concentration gradient between the renal parenchyma and the urine.

Since there is no renal threshold for ketone bodies, their loss in the urine was puzzling. Furthermore, these compounds are lost as anions and need to be neutralized by the excretion of  $\text{NH}_4^+$ . Thus, the loss of urinary nitrogen as ammonium should cause wasting of the lean body mass. Subsequently, it was shown that humans can derive only 93% of their energy requirements from fat but need a minimal amount of glucose to maintain (or spark) fuel homeostasis. The kidneys provided approximately one-half of the newly synthesized glucose during prolonged starvation; this glucose is used to supply part of the fuel for the central nervous system. The source of the amino and amide groups used to titrate the acidity of the tubular urine caused by the excretion of ketone bodies is glutamine, mobilized to the kidney from the muscle. The removal of the two nitrogen groups from glutamine results in the generation of  $\alpha$ -ketoglutarate, which enters the citric acid cycle and is used to synthesize glucose. Thus, renal gluconeogenesis and ammoniogenesis are coupled, since ketonuria requires the excretion of ammonium. Ketonuria appears to be a critical biological phenomenon rather than a waste of water-soluble fuels.



**Figure 2** Rates of ketone body (AcAc +  $\beta$ -OHB) reabsorption and excretion in 12 obese (1 male and 11 female) volunteers during prolonged total starvation. GFR, glomerular filtration rate. Data are expressed as means  $\pm$  SEM. Reprinted from Sapir, D. G. and Owen, O. E. Renal conservation of ketone bodies during starvation. *Metabolism* 24, 23–33, 1975, with permission.

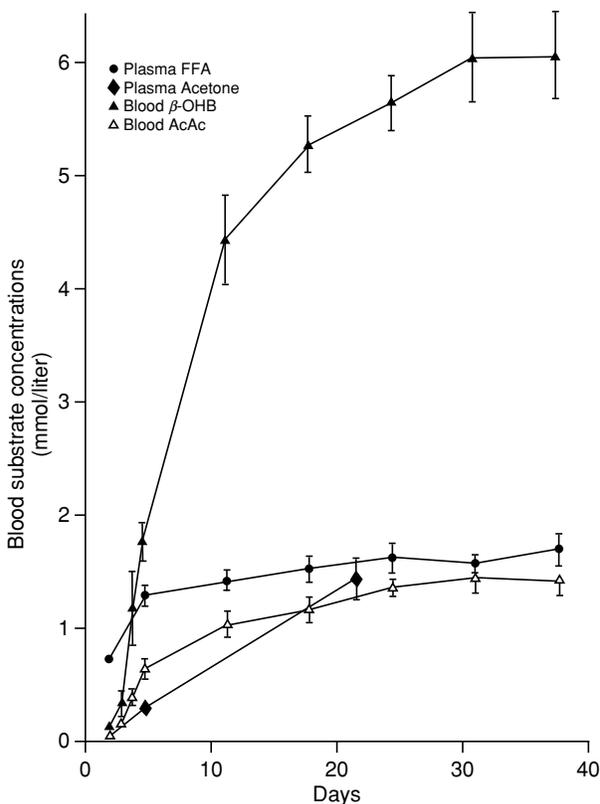
After a few days of starvation, the minimal rate of fat oxidation amounts to approximately 3 g/kg fat-free lean body mass. Thus, a male human with a body weight of 80 kg and a fat-free mass of 64 kg mobilizes approximately 192 g of stored triglyceride daily. Urinary ketone body excretion should amount to approximately 12 g/day. Approximately 47% of the chemical composition of ketone bodies is oxygen; only approximately 6–7 g ( $100\% - 47\% = 53\%$ ;  $12 \text{ g} \times 0.53 = 6.4 \text{ g}$ ) of the ketone body is derived from the highly reduced fatty acids contained in triglycerides. Thus, each day during starvation, only 3% of the fatty acid carbon mobilized from the triglycerides of adipose tissue for energy metabolism is lost in the urine as ketone bodies. In an 80 kg male, the fat mass is approximately 16 kg so that the daily loss of 0.04% of the total fat mass is trivial when there is a constitutional need for gluconeogenesis. Renal function during starvation is thus an extraordinary form of fuel conservation that ensures that other essential functions, including gluconeogenesis, are balanced at a trivial cost to total body energy reserves.

During the first hours of rehydration therapy for severe diabetic ketoacidosis, the quantity of ketone bodies in the urine can be extrapolated to a maximum of approximately 1920–2400 mmol/day. This is 16–20

times greater than that observed during starvation and would be impossible to maintain for more than a few hours. Nonetheless, urinary excretion of these freely filtered anions is limited by the increase in their renal tubular reabsorptive rates. Heightened excretory rates of AcAc and  $\beta$ -OHB during diabetic ketoacidosis are due to the increased filtration load of these ketone bodies. The excretion of the ammonium needed to maintain the electroneutrality required among patients suffering from moderate ketoacidosis has been extrapolated to be approximately 109 to 773 mmol/day ( $\sim 1.5$ – $10.8 \text{ g N per day}$ ) and is inversely related to arterial pH. This high level of ammonium excretion is not coupled to a contribution of glucose to blood by the kidneys because glucosuria induces a net loss of glucose.

## BLOOD KETONE BODIES

There are no other substrates present in the human blood that can change so drastically as ketone bodies and still be compatible with life. After an overnight fast, the concentration of AcAc in the blood is approximately 0.05 mmol/liter. Two hours after breakfast, the concentration of AcAc can fall on occasion to approximately 0.01 mmol/liter. The relative change in the concentration of  $\beta$ -OHB in the blood is comparable. Acetone is practically absent from the blood during brief fasting or postprandial periods unless an individual has been eating a high-fat, carbohydrate-free diet. Prolonged starvation causes an exponential rise in AcAc to approximately 0.60 mmol/liter during the first 3 days of starvation and in  $\beta$ -OHB to approximately 4.50 mmol/liter during the first 10 days of starvation (Fig. 3). An additional mild increase in the concentrations of AcAc and  $\beta$ -OHB occurs as fasting progresses. The concentration of total ketone bodies in the blood plateaus at approximately 6–8 mmol/liter after approximately 18 days of starvation. The ratio of AcAc to  $\beta$ -OHB after this duration of starvation is approximately 1:3 to 1:5. Occasionally, the blood AcAc concentration of a fasting patient will be greater than 2.00 mmol/liter and the  $\beta$ -OHB will be greater than 12.00 mmol/liter. At the highest concentrations of ketone bodies in the blood during starvation, the arterial pH is usually approximately 7.34. The concentration of acetone changes from approximately 0.25 mmol/liter after 3 days of starvation to approximately 1.37 mmol/liter after 21 days of starvation. The range of total ketone body concentration from the postprandial period to prolonged starvation can vary approximately 770-fold (0.02 to 15.4 mmol/



**Figure 3** Blood substrate concentrations in 37 obese subjects during prolonged total starvation. Values are expressed as means  $\pm$  SEM. Plasma acetone was measured in 15 different lean and obese volunteers. The 6 lean individuals had greater venous plasma acetone, AcAc, and  $\beta$ -OHB concentrations than the obese subjects after 3 days of fasting. The 3 obese volunteers had venous plasma acetone concentrations of  $0.29 \pm 0.041$  mmol/liter and  $1.37 \pm 0.19$  mmol/liter after 3 and 21 days of fasting, respectively. FFA, Free fatty acids.

liter). Greater variation occurs in severe diabetic ketoacidosis with blood AcAc concentrations  $>6$  mmol/liter,  $\beta$ -OHB concentrations  $>20$  mmol/liter, and acetone concentrations  $>12$  mmol/liter. The arterial pH in decompensated diabetic patients can be as low as 6.5. Nonetheless, patients in this catastrophic and morbid state of diabetic ketoacidosis can be resuscitated. Thus, the range for total ketone bodies in the blood from the postprandial state to diseased states can vary at least 1900-fold (0.02 to 38.00 mmol/liter) and life can be maintained.

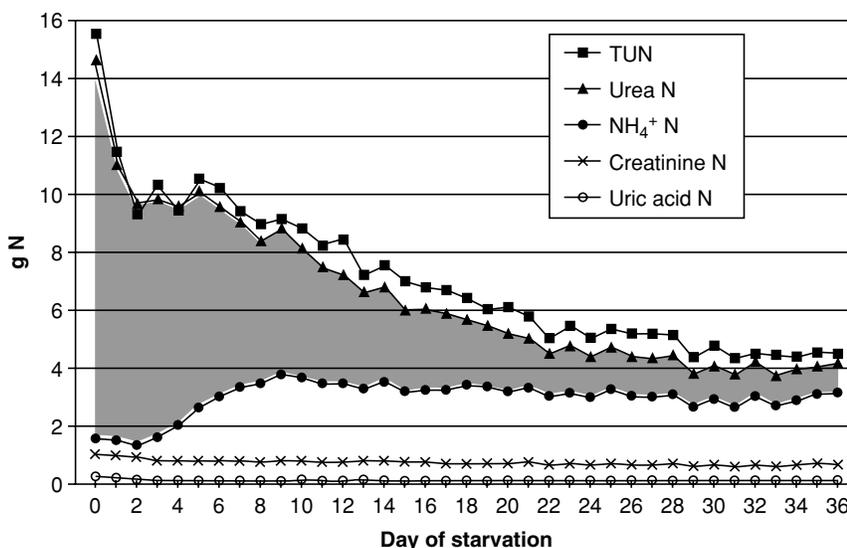
It was not ketone body concentrations in the blood or urinary excretion rates that refocused the understanding of hepatic metabolism during starvation; it was urinary nitrogen excretion. There was simply not enough nitrogen from amino acid catabolism excreted to account for its share of the assumed 100–125 g/day

of glucose needed to be synthesized via hepatic gluconeogenesis during starvation to provide the needed fuel for the brain (Fig. 4). This discrepancy in the level of urinary nitrogen excretion during starvation eventually caused investigators to speculate that the brain may be able to use alternative fuels, ketone bodies, as a source of energy.

## KETONE BODY PRODUCTION AND UTILIZATION RATES

The rate of production of ketone bodies by the liver can be measured by catheterization, isotopic dilution, or magnetic resonance spectroscopic techniques. These methods measure different things. With catheterization techniques, the arteriovenous concentration differences are combined with regional blood flow rates to provide an estimate of the net release or uptake rate of ketone bodies (the Fick principle). Isotopic dilution techniques measure total body equilibration exchange and oxidation rates of these compounds. Usually the values obtained for production rates using tracer techniques are greater than those derived from catheterization studies because tracer techniques measure the total body exchange rates, whereas catheterization techniques determine the net balance of the substrate measured across an organ or vascular bed. Nuclear magnetic resonance detects the labeling of metabolic end-products present in tissues. An additional problem with the measurement of the production and utilization rates of ketone bodies involves the manner in which the data are expressed. When hepatic ketone body production is high and peripheral tissue consumption of ketone bodies is high (especially in the brain), correcting the data to standard body surface areas or body weights introduces potential inaccuracies. This is because a starving person with a body weight of 140 kg and a height of 188 cm does not have a brain or liver that is proportionally enlarged. Also, diet and exercise may have influences on hepatic ketone body production rates. An individual consuming a ketogenic (high-fat, low-carbohydrate) diet is more likely to have higher production rates of AcAc and  $\beta$ -OHB than someone eating the usual balanced diet when both individuals are studied after an overnight 12 h fast. Obviously, the time since the last meal and the presence of metabolic diseases influence these measurements. With the preceding caveats, the following production rates are presented.

When a balanced diet is consumed, total ketone body production after an overnight fast ranges from 0.06 to 0.47 mmol/min/1.73 m<sup>2</sup> (70 kg body weight



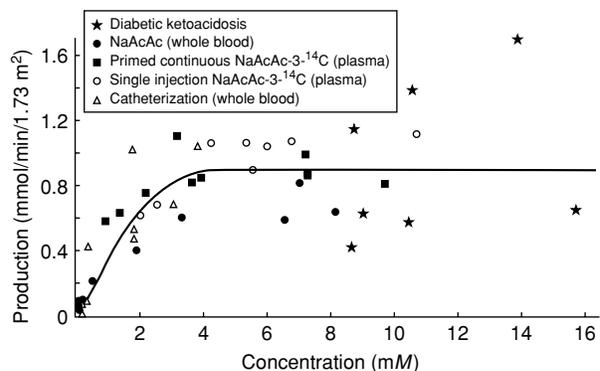
**Figure 4** The mean quantities of nitrogenous compounds excreted daily in five obese men and five obese women undergoing total starvation for 36 days. The components are “stacked” so that their sum can be visualized. TUN, total urinary nitrogen. Reprinted from Owen, O. E., Smalley, K. J., and Jungas, R. L. (2001). Starvation. *In* Handbook of Physiology (L. S. Jefferson and A. D. Cherrington, eds.) Chapter 39, Section 7, Vol. 2, pp. 1199–1225. Copyright 1999 by American Physiological Society. Used by permission of Oxford University Press, London, UK.

was equated to 1.73 m<sup>2</sup> body surface area). The production rates slowly rise hourly as food deprivation continues. After 2–4 days of total starvation, the rate increases to approximately 0.75 to 1.77 mmol/min/1.73 m<sup>2</sup> (70 kg). During the first few days of starvation, urinary ketone body excretion is at a low level and there is some accumulation of these compounds in the body fluids. However, the majority of AcAc plus  $\beta$ -OHB is oxidized to CO<sub>2</sub>. Despite progressive fasting for 17–24 days, the rate of production of ketone bodies is constant at 0.75 to 1.62 mmol/min/1.73 m<sup>2</sup> (70 kg). Only approximately 84% of the ketone bodies that are produced are oxidized; the rest are excreted in the urine or exhaled as acetone. After a few days of starvation, ketone body oxidation can account for 37–44% of the CO<sub>2</sub> production rate or greater than 30% of the energy requirements of a human based on O<sub>2</sub> consumption.

There is a linear relationship, up to 2–4 mmol/liter, between the rate of ketone body production and the concentration of AcAc plus  $\beta$ -OHB in the blood. However, at higher concentrations, such as those found after more prolonged starvation, the concentrations of ketone bodies in blood do not reflect production rates. Thus, there is a dissociation between ketone body production rates and blood concentrations after approximately 3 days of starvation (Fig. 5). The wide range of discordance between production and concentration is more marked among patients suffering from

diabetic ketoacidosis. The primary route of ketone body removal during starvation is oxidation by peripheral tissues. However, several investigators have reported impaired ketone body oxidation rates in diabetic animals during ketoacidosis.

Several mechanisms are responsible for this dissociation between hepatic ketone body production rates and peripheral blood ketone body concentrations during starvation. First, there must be a high rate of hepatic ketone body production to remove fatty acids from the blood during starvation. Second, ketone bodies are not evenly distributed in the extracellular and intracellular water space; they have restricted



**Figure 5** The relationship between total ketone-body blood concentrations and production rates in lean and obese starving volunteers and in diabetic patients during ketoacidosis.

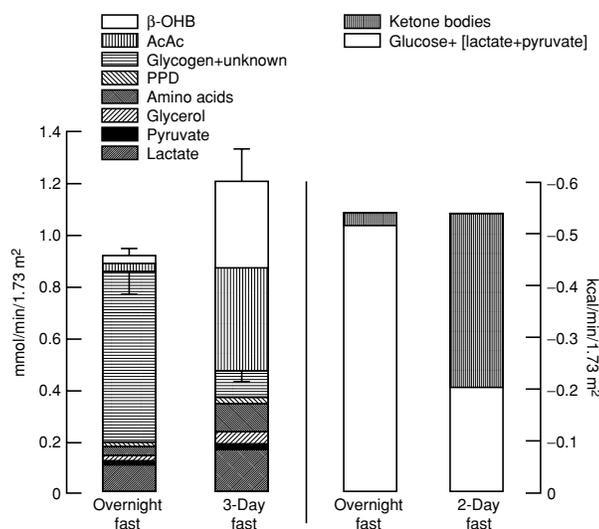
volumes of distribution. Third, there is a limited rate of oxidation of ketone bodies by peripheral tissue. Finally, augmented renal retrieval of ketone bodies contributes to the hyperketonemia of starvation and to diabetic ketoacidosis. However, during the hyperketonemia induced by starvation, the rates of production and removal of AcAc and  $\beta$ -OHB eventually equilibrate, reaching a steady-state level. In uncontrolled diabetic ketoacidosis, ketone body production and removal rates do not equilibrate and death results from progressive ketoacidosis.

One facet of ketone body metabolism that has not been fully appreciated is that hepatic production rates of AcAc and  $\beta$ -OHB are approximately the same after an overnight fast, a 3-day fast, and a 21- to 42-day fast and during diabetic ketoacidosis; i.e., hepatic AcAc production equals hepatic  $\beta$ -OHB production. It is the algebraic sum of preferential extraction and/or release rates of AcAc and  $\beta$ -OHB plus decomposition of AcAc to acetone and the influence of pH that dictate the AcAc to  $\beta$ -OHB ratio in peripheral blood. Another important point regarding normal hepatic ketogenesis relates to hepatic gluconeogenesis and glycogenolysis (Fig. 6). These processes are inversely related in normal hepatic metabolism. There is a reciprocal relationship between the quantity of glucose derived from glycogenolysis and that produced by gluconeogenesis. Thus, when the quantity of glucose derived from glycogenolysis is high, the quantity produced by gluconeogenesis is low and vice versa. In addition, there is a

reciprocal relationship between total hepatic glucose release and ketone body release. After an overnight fast when glucose release is high, ketone body release is low. During prolonged starvation when hepatic ketone body production is at a high level, total glucose production (entirely from gluconeogenesis) is at a low level. These normal reciprocal relationships are absent in patients suffering from diabetic ketoacidosis.

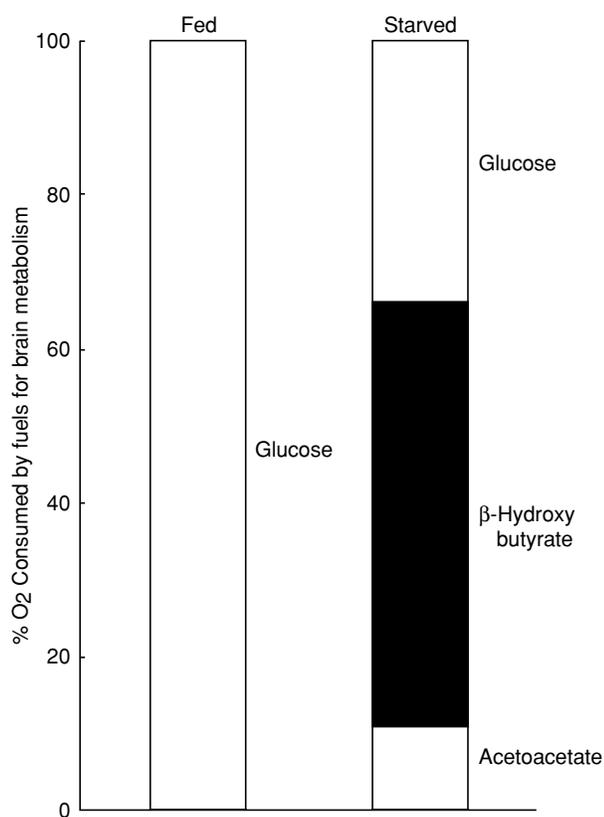
## EXTRAHEPATIC EXCHANGE OF KETONE BODIES

The uptake and (or) release of AcAc and  $\beta$ -OHB across peripheral vascular beds have been estimated by multiplying the arteriovenous concentration differences by the regional blood flow rate or by employing isotopic tracer techniques. However, determining the ketone body exchange or utilization rates using arteriovenous concentration difference measurements has inherent limitations. Blood flow rates are usually difficult to measure with the analytical accuracy needed to exactly quantify exchange rates. Furthermore, the extraction or release of a substrate may vary among individuals. For example, the blood flow rates of lean and obese individuals differ in responses to starvation and diabetes mellitus. Nonetheless, some general statements hold true for the measurement of ketone body metabolism in human tissues during starvation and other physiological and disease states.



**Figure 6** Net splanchnic glucose and ketone body production rates in lean men and women after an overnight fast and after a 3-day fast. Hepatic glucose output excludes that quantity derived from recycled lactate and pyruvate, but this quantity of glucose is extracted from the caloric equivalent expressed as kcal/min/1.73 m<sup>2</sup>. The total caloric equivalents of fuels released into the blood are equal, but the quantities of glucose and ketone bodies have a reciprocal relationship. PPD, propanediol.

The seminal study on ketone body utilization by extrahepatic tissues in humans was reported in 1967, when Owen *et al.* demonstrated that during prolonged starvation  $\beta$ -OHB and AcAc replaced glucose as the predominant fuel for the brain (Fig. 7). Previous studies of brain metabolism showed that the only energy-yielding substrate consistently extracted from the blood by the human brain was glucose. However, the quantity of carbohydrate stored in the human body is limited and can supply glucose for tissues such as the brain and other glucose-utilizing tissues for approximately 1 day. During starvation, the liver produces glucose from amino acids derived from the proteolysis of muscle protein, from glycerol derived from lipolysis, and from acetone that has been derived from AcAc. Recycling lactate and pyruvate to glucose (the Cori cycle) provides no net gain of glucose for



**Figure 7** Fuels consumed by the brain after overnight (fed) and starved states. Arteriovenous concentration differences multiplied by brain flow rates were used to calculate oxygen equivalents in three obese volunteers who underwent 38 to 41 days of total starvation. Extraction of  $\beta$ -OHB plus AcAc by the brain could provide approximately 66% of the energy requirements of the brain after prolonged starvation. The mean quantity of  $\alpha$ -amino nitrogen exchanged by the brain during this study was statistically insignificant.

oxidation by peripheral tissues. The amount of nitrogen excreted in the urine reflects amino acid catabolism. Under these conditions, humans excrete far less nitrogen than would be required if hepatic gluconeogenesis supplied the central nervous system with glucose as the only fuel source.

Cahill first recognized the paradox between the amount of glucose required by the brain if it were the only fuel and the quantity of nitrogen excreted from the body during prolonged starvation. He suggested that it was likely that fuels (ketone bodies) other than glucose were oxidized by the brain during starvation. Owen *et al.* confirmed this hypothesis. After a 38- to 41-day fast, obese volunteers underwent catheterization to measure the exchange rates across the brain and liver of AcAc,  $\beta$ -OHB, glucose, free fatty acids, amino acids, O<sub>2</sub>, CO<sub>2</sub>, and other substrates. This study showed for the first time that during prolonged starvation,  $\beta$ -OHB and AcAc replace glucose as the predominant fuel for brain metabolism (Fig. 7). Subsequently, other researchers have demonstrated that ketone body extraction from arterial blood occurs after brief and intermittent periods of fasting. The initial study performed on three obese volunteers was the key report that modified scientists' concept of the flux of fuels during starvation. It established the textbook concept that the metabolism of ketone bodies is a major adaptation to starvation since their oxidation by the brain spares the need to use glucose in that tissue. This preserves muscle protein since there was less of a requirement for hepatic gluconeogenesis from amino acids.

Striated muscle is the most metabolically active tissue in humans. In normal adults, skeletal muscle weighs approximately 20–35 kg, but in individuals with a large muscle mass, this figure may be increased by a factor of 2 or 3. However, in contrast to the physically active states, resting muscle metabolism (with fat and skin) accounts for only 40% of the energy requirements in the resting state. The total metabolic requirement per unit mass (or volume) of forearm muscle in lean humans during rest is indistinguishable from that in obese humans during rest. Obese humans have a slightly higher concentration of AcAc and  $\beta$ -OHB in the blood after an overnight fast, but the total quantity of ketone bodies extracted in the obese subjects is approximately the same as for lean individuals. After an overnight fast, ketone bodies provide only 5–10% of the energy requirements for muscle. After 3 days of fasting, the arterial ketone body concentration increases 10- to 20-fold. AcAc extraction by muscle rises to a greater extent than  $\beta$ -OHB extraction despite a higher concentration of

$\beta$ -OHB. At this stage of starvation, ketone bodies are the major fuels of respiration, supplying approximately 50–85% of the energy requirements for muscle. The preferential extraction of AcAc over  $\beta$ -OHB becomes more marked as starvation is prolonged. During prolonged starvation, muscle extracts AcAc and releases  $\beta$ -OHB. The sum total of ketone body utilization by striated muscle is grossly diminished after prolonged starvation although arterial concentrations of AcAc and  $\beta$ -OHB are both significantly higher. This is due to a suppression of ketone body oxidation in muscle by fatty acids, which spares AcAc and  $\beta$ -OHB for metabolism by the central nervous system. Thus, ketone bodies are only marginally oxidized by striated muscle during heightened ketonemia; the majority of AcAc extracted by the forearm is converted to  $\beta$ -OHB.

Net renal exchange of AcAc and  $\beta$ -OHB after 21 days of starvation cannot account for ketonuria. This suggests that the kidneys, like the liver, are ketogenic as well as gluconeogenic organs. The quantity of fatty acids extracted from plasma during prolonged starvation can account for the renal excretion of AcAc and  $\beta$ -OHB, as well as the renal synthesis and consumption of ketone bodies.

In 1945, Medes and co-workers noted that ketone body formation was not exclusively a liver function, but may occur in tissues that metabolize acetate. Kidney preparations readily synthesized AcAc and  $\beta$ -OHB from [ $^{13}\text{C}$ ]acetate. However, the concurrent rate of ketone body oxidation was high and they surmised that renal ketogenesis would not yield ketone bodies released to the bloodstream. In 1969, Weidemann and Krebs reported that slices of kidney cortex from starved rats removed more [ $^{14}\text{C}$ ]oleate than could be oxidized, based on the amount of oxygen consumed by the tissue. Oleate increased ketogenesis by six- to sevenfold, but the maximum rates of renal ketogenesis were  $\sim 20\%$  of the maximum rates observed for liver when calculated on a weight-to-weight basis. This suggests that the rate of renal ketogenesis is  $\sim 2\%$  of the hepatic ketogenic rate, which is in agreement with the results of other researchers. In addition, Weidemann and Krebs reported that renal ketogenesis and gluconeogenesis were coupled. Brady *et al.* studied the distribution of  $^{14}\text{C}$  in  $\beta$ -OHB and concluded that a minimum of 11–17% of the ketone bodies formed in diabetic, ketotic rats were derived from extrahepatic tissue, presumed to be the kidneys. Subsequent work showed that ketone bodies presumed to be produced in the kidneys entered the bloodstream before being

excreted in the urine. Owen *et al.* reported that the kidneys of a starving human may extract or release AcAc,  $\beta$ -OHB, or both. However, the mean renal exchange rates for four obese subjects who starved for 21 days and underwent catheterization studies were not significantly different from zero but were different from values reported previously for obese individuals fasting for 35–41 days. Because there was no renal removal of AcAc or  $\beta$ -OHB from the arterial blood in the presence of gross ketonuria, net renal ketogenesis was  $\geq 51 \mu\text{mol}/\text{min}$  or  $\sim 2\%$  of net hepatic ketogenesis. If the kidneys are producing, utilizing, and excreting AcAc and  $\beta$ -OHB, simultaneous net balance and isotopic techniques will have to be combined to quantify these processes in humans in states of augmented ketogenesis.

It is known that hepatic gluconeogenesis and ketogenesis are related. Furthermore, the interplay among ammoniogenesis, ketonuria, and gluconeogenesis has been well documented. However, the interrelationship of renal gluconeogenesis and ketogenesis, which may account for the ketonuria, has not been clearly defined. It is known that the liver synthesizes glucose and ketone bodies and releases these fuels into the blood. It is probable that the kidneys synthesize glucose and release it into the renal venous blood but synthesize ketone bodies and release AcAc and  $\beta$ -OHB primarily into the urine, accompanied by ammonium excretion.

The energy requirement of myocardial muscle is far less than that of striated muscle. The oxidation of ketone bodies by the heart increases after birth when the availability of these compounds increases. Mother's milk has a high fat content and nursing infants have mild ketonemia. However, ketone bodies account for only a small amount of the myocardial oxygen consumption during early childhood. In newborn lambs, carbohydrates provide 69–89% of the energy for the heart and ketone body utilization is negligible. However, fatty acids and ketone bodies become the dominant fuel source for the juvenile lambs. Substrate selection in the isolated heart perfused with physiological concentrations of fuels showed that fatty acids supplied approximately 50% and AcAc contributed approximately 25% of the substrates needed. In addition to serving as a fuel for the myocardium, ketone body utilization may improve the mechanical performance of the heart and the metabolic efficiency of oxygen use. Free fatty acids and ketone bodies compete as fuels but the quantities of AcAc and/or  $\beta$ -OHB oxidized by the heart during prolonged starvation cannot be very great.

## CONCLUSIONS

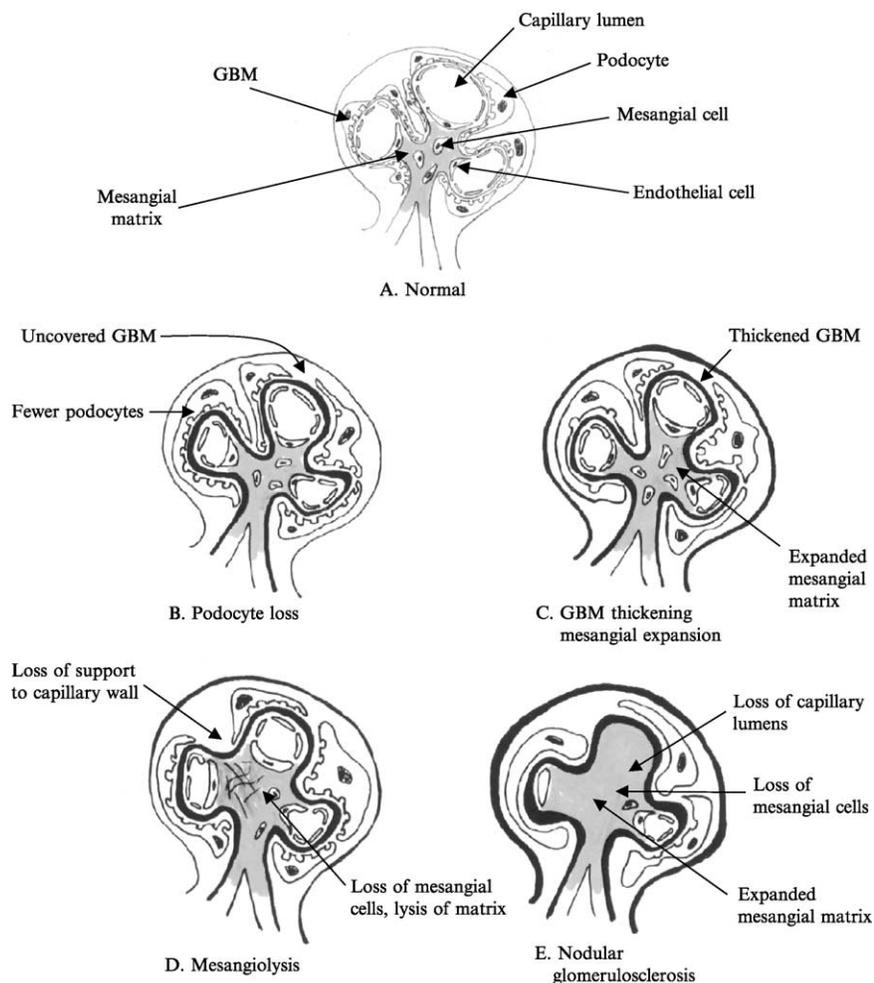
Humans are unique among the animal kingdom for their enlarged brain size to body size ratio. The brain has high energy requirements that must be met from extracranial sources. In the fed and postprandial periods, glucose is the dominant energy-yielding fuel for brain metabolism. During progressive starvation, glucose availability is limited and  $\beta$ -OHB and AcAc become the predominant fuels utilized by the brain. The liver contributes a finite quantity of ketone bodies to the blood. Muscle, an avid consumer of ketone bodies early in starvation, switches to fatty acid oxidation and preferential extraction of AcAc and release of  $\beta$ -OHB during prolonged starvation. The kidney conserves a majority of AcAc and  $\beta$ -OHB that undergoes glomerular filtration, but excretes enough ketone bodies in the urine to promote ammoniogenesis and accompanying gluconeogenesis. Ketone bodies have a wider range of change in concentration than any other fuel transported in the blood. In summary, during starvation, the liver produces, the brain consumes, the kidney mostly conserves, and muscle decreases its use of ketone bodies to maintain fuel homeostasis. The sum of these processes dissociates ketogenesis and ketonemia. The physiologic hyperketonemia of starvation is an important survival mechanism for humans.

## See Also the Following Articles

Anorexia Nervosa • Carbohydrate Metabolism and Hormone-Fuel Interrelationships • Hunger and Satiation • Obesity and Diabetes, Regulation of Food Intake

## Further Reading

- Balasse, E. O. (1979). Kinetics of ketone body metabolism in fasting humans. *Metabolism* **28**, 41–50.
- Brady, P. S., Scofield, R. S., Ohgaku, S., *et al.* (1982). Pathways of acetoacetate's formation in liver and kidney. *J. Biol. Chem.* **257**, 9290–9293.
- Hasselbalch, S. G., Knudsen, G. M., Jakobsen, J., *et al.* (1994). Brain metabolism during short-term starvation in humans. *J. Cereb. Blood Flow Metab.* **14**, 125–131.
- Krebs, H. A., Williamson, D. H., Bates, M., *et al.* (1970). The role of ketone bodies in caloric homeostasis. In "Advances in Enzyme Regulation" (George Weber ed.) Vol. 9, pp. 387–409. Pergamon Press, New York.
- Medes, G., Floyd, N. F., and Weinhouse, S. (1946). Fatty acid metabolism, IV. Ketone bodies as intermediates of acetone oxidation in animal tissues. *J. Biol. Chem.* **162**, 1–9.
- Miles, J. M., Haymond, M. W., and Gerich, J. E. (1984). Control of ketone-body production in humans. In "Recent Advances in Obesity and Diabetes Research" (N. Melchionda, *et al.*, eds.), pp. 283–292. Raven Press, New York.
- Owen, O. E., Kalhan, S. C., and Hanson, R. W. (2002). The key role of anaplerosis and cataplerosis for citric acid cycle function. *J. Biol. Chem.* **277**, 30409–30412.
- Owen, O. E., Licht, J. H., and Sapir, D. G. (1981). Renal function and effects of partial rehydration during diabetic ketoacidosis. *Diabetes* **30**, 510–518.
- Owen, O. E., Mozzoli, M. A., Reichle, F. A., *et al.* (1985). Hepatic and renal metabolism before and after portasystemic shunts in patients with cirrhosis. *J. Clin. Invest.* **76**, 1209–1217.
- Owen, O. E., Reichard, G. A. Jr., Markus, H., *et al.* (1973). Rapid intravenous sodium acetoacetate infusion in man: Metabolic and kinetic responses. *J. Clin. Invest.* **52**, 2606–2616.
- Owen, O. E., Smalley, K. J., D'Alessio, D. A., *et al.* (1998). Protein, fat and carbohydrate requirements during starvation: Anaplerosis and cataplerosis. *Am. J. Clin. Nutr.* **68**, 12–34.
- Owen, O. E., Trapp, V. E., Skutches, C. L., *et al.* (1982). Acetone metabolism during diabetic ketoacidosis. *Diabetes* **31**, 242–248.
- Reichard, G. A. Jr., Haff, A. C., Skutches, C. L., *et al.* (1979). Plasma acetone metabolism in the fasting human. *J. Clin. Invest.* **63**, 619–626.
- Reichard, G. A. Jr., Skutches, C. L., Hoeldtke, R. D., *et al.* (1986). Acetone metabolism in humans during diabetic ketoacidosis. *Diabetes* **35**, 668–674.
- Scofield, R. E., Schumann, W. E., Kumaran, K., *et al.* (1983). Ketone body production in diabetic ketosis by other than liver. *Metabolism* **32**, 1009–1012.
- Skutches, C. L., Owen, O. E., and Reichard, G. A. Jr. (1990). Acetone and acetol inhibition of insulin-stimulated glucose oxidation in adipose tissue and isolated adipocytes. *Diabetes* **39**, 450–455.
- Veech, R. L., Chance, B., Kashiwaya, Y., *et al.* (2001). Ketone bodies, potential therapeutic uses. *Life* **51**, 241–247.



**Figure 1** Evolution of diabetic nephropathy. The figure depicts the components of a normal glomerulus (A) and the changes that develop (B–E) as diabetic nephropathy progresses. GBM, glomerular basement membrane.

metabolic state in which elevated blood sugar concentrations, changes in insulin levels and the response to insulin, other hormones and growth factors, and circulating and intracellular lipids may contribute to the rate of progression of disease. The effects of hyperglycemia are the best studied. Hyperglycemia directly contributes to injury of podocytes and mesangial cells, and it stimulates an increase in the synthesis of matrix proteins that lead to scarring. Clinical studies in patients with both type 1 and type 2 diabetes show that improved glycemic control will slow the rate of progression of diabetic nephropathy. Many patients with diabetes develop high blood pressure. High blood pressure injures the walls of blood vessels, including the delicate capillaries of the glomerulus. Control of high blood pressure reduces this injury and slows the rate of progression of renal disease. Although treatment of high blood pressure with any effective agent is

of benefit, the use of medications that reduce the exposure of the kidney to angiotensin II appears to have the greatest benefit. This implies that angiotensin II directly affects the injury of the glomerulus beyond that which is due to high blood pressure.

## CLINICAL COURSE

Diabetic nephropathy develops slowly over time in patients with diabetes mellitus. Approximately 10 to 15 years after the onset of hyperglycemia, overt proteinuria identifies those patients with the disease. Once significant proteinuria develops, end-stage renal failure develops in 2 to 6 years. Patients with diabetic nephropathy who are treated with hemodialysis have a worse prognosis than do nondiabetic patients. The lipid abnormalities associated with

renal failure, greater difficulties with blood pressure management, and ongoing problems with glycemic control contribute to progression of other complications of diabetes, particularly cardiovascular disease. Recent studies have shown that early attention to glycemic control, smoking cessation, and hypertension management can alter the rate of progression to renal failure, so it is imperative that diabetic patients at risk for future development of nephropathy be identified and treated aggressively. The subsets of patients with diabetes mellitus who are at increased risk for the future development of diabetic nephropathy develop microalbuminuria. A urinary screening test will identify this subgroup approximately 5 years after the onset of hyperglycemia in type 1 diabetic patients. Patients with type 2 diabetes should be screened at the time that the diagnosis of diabetes is made. Although the exact cause of microalbuminuria is not known, some have suggested that it is a surrogate marker for those individuals with an increased genetic risk for developing progressive renal disease when they have diabetes mellitus.

## TREATMENT

Most of the emphasis on treatment is in the identification of individuals at risk for future development of renal disease by measuring urinary albumin excretion rates. In those with microalbuminuria, an angiotensin-converting enzyme inhibitor (ACEI) should be prescribed and the dose should be adjusted to reduce the albumin excretion rate as much as possible. In diabetic patients with hypertension, other medications should be added to the ACEI as needed to reduce blood pressure below 130/85. Aggressive measures should be undertaken to manage weight, encourage daily exercise, and control blood sugar levels as close to the normal range as possible. Efforts should be made to get those patients who

smoke to stop. Similarly, emphasis should be placed on treating lipid disorders because this will reduce the rate of progression to renal failure and will reduce morbidity and mortality from cardiovascular disease. Once renal failure has developed, patients with diabetic nephropathy can be managed by peritoneal dialysis, hemodialysis, or renal transplantation.

## See Also the Following Articles

Cardiovascular Disease in Diabetes • Diabetes, Type 1 • Diabetes, Type 2 • Diabetic Nerve Disease, Neuropathy • Eye Disease in Diabetes • Foot Disease in Diabetes • Kidney Stones • Neurological Disease and Diabetes, Autonomic

## Further Reading

- Deckert, T., Feldt-Rasmussen, B., Borch-Johnsen, K., Jensen, T., and Kofoed-Enevoldsen, A. (1989). Albuminuria reflects widespread vascular damage: The Steno hypothesis. *Diabetologia* **32**, 219–226.
- Gall, M., Borch-Johnsen, K., Hongaard, P., Nielsen, F. S., and Parving, H. H. (1995). Albuminuria and poor glycemic control predict mortality in NIDDM. *Diabetes* **44**, 1303–1309.
- Lewis, E. J., Hunsicker, L. G., Bain, R. P., Rohde, R. D., and Collaborative Study Group. (1993). The effect of angiotensin-converting enzyme inhibition on diabetic nephropathy. *N. Engl. J. Med.* **329**, 1456–1462.
- Osterby, R. (1992). Glomerular structural changes in type I (insulin-dependent) diabetes mellitus: Causes, consequences, and prevention. *Diabetologia* **35**, 803–812.
- Pagtalunan, M. E., Miller, P. L., Jumping-Eagle, S., Nelson, R. G., Myers, B. D., Rennke, H. G., Coplson, N. S., Sun, L., and Meyer, T. W. (1997). Podocyte loss and progressive glomerular injury in type II diabetes. *J. Clin. Invest.* **99**, 342–348.
- Ravid, M., Neumann, L., and Lishner, M. (1995). Plasma lipids and the progression of nephropathy in diabetes mellitus II: Effect of ACE inhibitors. *Kidney Int.* **47**, 907–910.
- Steffes, M. W., Osterby, R., Chavers, B., and Mauer, S. M. (1989). Mesangial expansion as a central mechanism for loss of kidney function in diabetic patients. *Diabetes* **38**, 1077–1081.
- Stout, L. C., Kumar, S., and Whorton, E. B. (1993). Focal mesangiolytic and the pathogenesis of the Kimmelstiel–Wilson nodule. *Hum. Pathol.* **24**, 77–89.

**Table I** Chemical Composition of Kidney Stones (Percentages)

Composition	Approximate frequency
Calcium oxalate	60
Calcium oxalate and calcium phosphate	10
Calcium monohydrogen phosphate (brushite)	10
Struvite	5–10
Uric acid	5–10
Cystine	1
Other	1

conditions associated with an increase risk of renal stone disease.

### Calcium-Containing Stones

Calcium-containing stones account for more than 70% of all renal tract stones. From a physicochemical point of view, the fact that they form is hardly surprising given that urine is almost always in a state of supersaturation for calcium salts. Factors that predispose to the development of idiopathic calcium stones include increased urinary concentrations of calcium and oxalate, alterations in inhibitors of stone formation, and nidus for stone formation.

The concentration of urinary calcium is increased by low urine volume and hypercalciuria. Low urine volume promotes stone formation by increasing the concentration of urinary calcium and by decreasing the urine flow rate. Hypercalciuria is commonly transmitted in an autosomal dominant fashion. Absorptive hypercalciuria is the most common cause of hypercalciuria and is postulated to be due to hyperresponsiveness to vitamin D in the jejunal mucosa. Resorptive hypercalciuria due to primary hyperparathyroidism occurs in approximately 5% of individuals with recurrent stone formation. Primary renal hypercalciuria due to a defect in renal tubular calcium reabsorption occurs in 5% of individuals with recurrent stone formation. A diet high in sodium is associated with increased urinary calcium excretion.

The sources of oxalate are the metabolism of amino acids (e.g., serine, glycine, hydroxyproline), metabolism of ascorbate, and dietary intake (10–20%). Oxalate is excreted renally. The urinary concentration of oxalate is much lower than that of calcium. Increases in absolute excretion of oxalate result in larger proportional changes in urinary oxalate concentration when compared with calcium. Increased

absorption of oxalate is seen with a low-calcium diet and enteric hyperoxaluria. The latter occurs with malabsorption syndromes associated with small bowel resection or Crohn's disease. Calcium is bound by

**Table II** Factors Associated with Stone Formation

Genetic condition
Metabolic disorders
Cystinuria
Primary hyperoxalurias (1 and 2)
Xanthinuria
2,8-dihydroxyadeninuria
Orotic aciduria
Disorders of uric acid metabolism
Lesch–Nyhan syndrome
Phosphoribosyl pyrophosphate synthase hyperactivity
Glucose-6-phosphatase deficiency
Adenine phosphoribosyl transferase deficiency
Renal chloride channel defects
Dent's disease
X-linked hypophosphatemic rickets
X-linked recessive nephrolithiasis
Renal tubular acidosis
Disease process
Medical conditions
Primary hyperparathyroidism
Renal tubular acidosis
Autoimmune
Malabsorption syndromes
Crohn's disease
Hyperthyroidism
Sarcoidosis
Hypercalcemia of malignancy
Surgical conditions
Intestinal resection
Jejunioileal bypass
Anatomic abnormalities
Medullary sponge kidney
Horseshoe kidney
Solitary kidney
Ureteric obstruction
Caliceal diverticulum
Medication
Calcium supplements
Vitamin D
Vitamin C
Sulphonamides
Triamterene
Indinavir
Ephedrine
Ciprofloxacin

free fatty acids allowing increased colonic absorption of unbound oxalate. Inherited hepatic enzyme defects, primarily hyperoxaluria types I and II, are associated with severe hyperoxaluria that commonly presents during childhood with multiple kidney stones and may lead to renal failure.

Citrate is an important inhibitor of stone formation. Urinary citrate forms a soluble complex with calcium and decreases the concentration of free calcium available for crystal formation. Citrate also inhibits crystal propagation by unknown mechanisms. Hypocitraturia can occur as an isolated finding or may be due to chronic systemic metabolic acidosis, renal tubular acidosis, hypokalemia, a high-protein diet, and possibly a high-sodium diet.

Urinary glycoproteins inhibit stone formation at various stages. These proteins include Tamm–Horsfall protein, bikunin, nephrocalcin, and a urinary form of prothrombin fragment 1. The possible role of abnormal glycoproteins in recurrent stone formation is under investigation.

Uric acid is the product of purine metabolism. A diet high in purines can lead to hyperuricosuria and uric acid crystals that act as nidus for calcium oxalate precipitation and promote calcium stone formation.

### Uric Acid Stones

Pure uric acid stones are uncommon. They are radiolucent. The factors that predispose to the development of uric acid stones are the urine concentration of uric acid and low urine (pH < 5.5). Low urine volume increases the risk for uric acid stones by the same mechanisms discussed previously for calcium stones. Chronic hyperuricosuria is associated with gout, a high-purine diet, and inborn errors of metabolism. Acute hyperuricosuria may occur during treatment of myeloproliferative disorders.

### Struvite Stones

Struvite stones are composed of magnesium ammonium phosphate. These stones develop in the context of urinary tract infection with organisms that produce urease (e.g., *Proteus*, *Klebsiella*, *Serratia*, *Mycoplasma*). The urinary pH and concentration of ammonia are increased by the breakdown of urea. Struvite stones grow rapidly, particularly in areas of the pelvicalyceal system that do not drain adequately, and may form stag-horn calculi. These stones usually present with symptoms of urinary tract infection. Struvite stones are difficult to treat because they are themselves infected.

### Cystine Stones

Cystine stones develop as a result of cystinuria, an autosomal recessive trait associated with decreased proximal tubular reabsorption of cystine and other amino acids. Low urine volume and acidic urine promote stone formation.

## INVESTIGATION OF PATIENTS WITH KIDNEY STONES

There is some controversy as to how extensively to investigate patients who have passed a single kidney stone. The controversy is centered around the low mortality and relatively low morbidity of the condition as well as the lack of specific therapy for most forms of renal stone disease. Many clinicians favor the institution of general therapeutic maneuvers without investigation. The European Association of Urology has suggested a classification for stone formers based on frequency of stone passage and relative risk of recurrence. This classification can be adapted to direct investigation.

Analysis of the calculus, where possible, is the cornerstone of the investigation and rational therapy of nephrolithiasis. Patients with infection-related (struvite) stones and cystine stones can be readily identified and treatment can be planned. Patients who make uric acid stones are at high risk for recurrence and require detailed investigations as to the cause of their hyperuricosuria. For calcium stone formers, the European Association of Urology classification divides patients according to whether they are first-time or recurrent stone formers, whether or not there are residual stone fragments, and (for those recurrent stone formers) whether their disease results in rare (mild) or frequent (severe) passage of stones. Also highlighted for aggressive management is a group of patients who are at a predetermined specific risk for recurrence (e.g., primary hyperparathyroidism).

Table III suggests a rational approach to investigation in patients who have passed a single stone. This will detect patients with hyperuricemia and hypercalcemia as well as those with infected urine. The presence of a low bicarbonate level with a urine pH that is not maximally acidified would suggest the presence of renal tubular acidosis.

For patients who have frequent stone passage or a significant number of residual stones visible on radiographs, an attempt to establish the lithogenicity of the urine is a reasonable strategy. Commercially available software can calculate the relative supersaturation for calcium oxalate and calcium phosphate if required.

**Table III Basic Investigations Following Passage of a Single Stone**

Serum analysis
Complete blood count
Biochemical profile
Electrolytes, bicarbonate, creatinine, urea
Ionized calcium, <sup>a</sup> phosphate
Parathyroid hormone
Urate
Urine analysis
Urine culture
Urine pH

<sup>a</sup>Or calcium corrected for albumin concentration.

**Table IV** outlines those additional 24-h urinary investigations that may be useful for recurrent stone formers. Optimal values for these parameters for minimal lithogenicity are shown in parentheses.

## RATIONAL THERAPY FOR KIDNEY STONE DISEASE

The treatment of kidney stones can be divided into general and specific measures. The power of general advice in terms of fluid intake and diet should not be underestimated. It has been suggested that the “stone clinic effect” can result in a 5-year reduction in new stone passage of 60%.

**Table IV Suggested Investigations for Frequent Stone Formers**

Acidified urine (6 mol/L HCl)	
Calcium	(<7.5 mmol male) (<6 mmol female)
Phosphate	(<35 mmol)
Oxalate	(<350 μmol)
Standard urine	
Volume	(>2.5 L)
Sodium <sup>a</sup>	(50–100 mmol)
Creatinine <sup>b</sup>	(7–16 mmol male) (5–14 mmol female)
Urate	(<4.5 mmol)
Citrate	(>2.5 mmol)
Magnesium	(>3.0 mmol)

*Note.* Number in parentheses indicate desirable values for low urine lithogenicity.

<sup>a</sup>To assess compliance with dietary recommendations.

<sup>b</sup>To establish accuracy of 24-h collection.

## Dietary Intervention

Given that stone formation is the end result of the process of crystal formation, aggregation, and growth from a supersaturated solution, the most rational therapy would be to modify the solution such that the potential energy for crystal formation within the solution is lowered. The logical way in which to do this is to increase fluid intake. Consumption of enough fluid such that the daily urine volume is greater than 2 L has been associated with significantly reduced rates of stone formation in large observational studies over prolonged periods of time. Furthermore, a randomized prospective study performed by Borghi and colleagues at the University of Parma, Italy, demonstrated that kidney stone patients randomized to drink enough fluid to generate a urine volume greater than 2 L/day had a significantly lower rate of recurrence of their stones than did the group of patients randomized to receive no strict fluid prescription. In addition, the time to recurrence was significantly longer in the fluid-treated group.

From large observational studies, it appears that many different beverages provide some protection against stone formation. Data exist to support the beneficial effects of coffee (both caffeinated and decaffeinated), tea, beer, and wine in this regard. In contrast, the ingestion of grapefruit juice and apple juice was shown to increase the risk of stone formation in men followed in the Health Professionals Follow-up Study. Grapefruit juice also increased the risk of stone formation in women observed as part of the Nurses' Health Study; however, the mechanism of this effect is unclear.

Although it would seem to be intuitive that a reduction of calcium in the diet would reduce the frequency of kidney stones, that is not the case. The most compelling evidence of the failure of a low-calcium diet to ameliorate renal stone disease came from a prospective randomized study in men performed by Borghi and colleagues. This landmark study compared a diet containing 100 mmol/day of calcium with a diet containing 30 mmol/day of calcium. Members of the latter group were also required to lower their sodium intake to approximately 50 mmol/day and their animal protein intake to 52 g/day. The group on the normal-calcium, low-sodium, low-protein diet had a significantly reduced risk of recurrent stone formation, although the effect took more than 3 years to become manifest.

Few well-designed studies have addressed other dietary components. The combination of a low-protein, high-fiber diet was not shown to be preventive

of recurrent calcium oxalate stones in a prospective study performed under the auspices of the Kaiser Permanente Medical Care Program.

## Pharmacological Intervention

### Calcium Stones

The optimal initial treatment for idiopathic hypercalciuria is the administration of a thiazide diuretic such as hydrochlorothiazide or chlorthalidone (12.5–50.0 mg/day). Thiazide diuretics indirectly and directly increase calcium reabsorption in the proximal and distal tubules. Randomized trials have shown a 20% risk reduction for the development of new calcium oxalate stones. Routine monitoring for serum electrolyte abnormalities is necessary. Hypokalemia can lead to hypocitraturia and so must be avoided. If hypercalciuria persists, amiloride (5 mg/day) may be added to further reduce calcium excretion and correct hypokalemia. Potassium citrate (30–80 mEq/day) is indicated for patients with hypocitraturia. Urinary citrate is increased by decreased reabsorption induced by alkalinization. Patients with hyperuricosuria are at risk for the development of uric acid crystals that serve as a nidus for calcium oxalate crystal formation and have been shown to benefit from administration of allopurinol. Oral calcium and cholestyramine have been considered to bind oxalate and diminish urinary oxalate; however, there is no strong evidence for their use in idiopathic calcium oxalate stones. Oral calcium to bind oxalate and potassium citrate to correct acidosis may be beneficial in enteric hyperoxaluria.

### Uric Acid Stones

The most important aspect of the pharmacological management of uric acid stones is alkalinization of the urine to a range of 6.0 to 6.5. Potassium citrate (30–80 mEq/day) is preferable to sodium bicarbonate because it avoids the sodium load. During potassium citrate therapy, the urine pH must be monitored and not allowed to rise above 7.0 because this will promote the formation of calcium phosphate stones.

### Struvite Stones

Struvite stones are chronically infected. Treatment of struvite stones with antibiotics may help to reduce the rate of growth of the stones; however, definitive therapy for struvite stones usually involves surgical excision of the stones.

### Cystine Stones

Cystine stones are treated with the usual general measures of increased fluid intake and restriction of

sodium intake. Alkalinization of the urine to a pH above 7.0 using potassium citrate increases the solubility of cystine. If these measures are not effective in reducing the rate of stone formation, penicillamine, tiopronin, and captopril may be effective. These medications cleave the disulfide bond and increase the solubility of cystine.

## Surgical Intervention

The surgical management of renal stone disease has been revolutionized by the advent of extracorporeal shock wave lithotripsy (ESWL) and the growth of endourological procedures. Invasive surgery is rarely required except in cases of complicated infected stones. The success rate of ESWL is close to 95% for individual cases. Concerns about the procedure relate to possible long-term complications (although the procedure has been in use for more than 15 years) and to the fact that failure to clear all of the stone fragments after lithotripsy will inevitably lead to an increased risk of recurrence. For ESWL to be effective, the patient must not be too large or too small for the machine. The stone must be targetable and amenable to fragmentation by the shock wave. Once the stone has been broken, there must be a clear drainage pathway given that the presence of ureteric obstruction is a contraindication to ESWL. Not infrequently, ancillary procedures, such as ureteric stenting and endourological fragment removal, are required after ESWL. [Table V](#) details several other situations where ESWL may be less suitable.

Although the shock wave can be delivered externally (as in ESWL), direct application of the shock to the stone is possible using minimally invasive techniques such as percutaneous nephrolithotripsy (PCNL). Under these circumstances, a scope is passed through a percutaneous nephrostomy and fragmenting energy is applied directly to the stone. For ureteric stones, endourological procedures using rigid or flexible ureteroscopes are more likely to be successful than is ESWL alone.

Acute complications of ESWL are infrequent. Loin pain and ureteric obstruction from stone fragment passage (steinstrasse) occur in approximately 6 and 2.5% of cases, respectively. Rarer complications include renal hematoma and direct shock wave damage to other viscera. Long-term development of hypertension has been reported by several groups and is the subject of ongoing investigations. Similarly, reduction in renal function following bilateral ESWL has led to recommendations that a period of

Table V Situations Where Alternatives to ESWL May Be Required

Situation	Explanation	Alternative
Cystine stones	Resistant to fragmentation	Medical management
Struvite stones	Infected, hardcore	Combined strategy
Ectopic kidney (including renal transplant)	Targeting problem	PCNL
Pregnancy	Safety concerns	Conservative management until after delivery
Large stones (> 2.5 cm)	Risk of steinstrasse or multiple residual fragments	PCNL
Bleeding diathesis	Risk of hemorrhage	Correction of problem, medical management, or direct surgical approach

months be allowed to pass before a second kidney is treated by this method.

The removal of stones by ESWL or other surgical methods should not be considered a cure of the renal stone disease. Close attention to medical management is required to prevent recurrence, either from growth of residual stone particles or from *de novo* stone formation in a susceptible individual.

### Management of Acute Renal Colic Due to Stone Passage

The passage of a kidney stone is usually associated with severe pain, hematuria, and nausea and vomiting. Frequently, the person is unable to find a comfortable position and can be seen to be writhing on the bed. This is in contradistinction to persons with severe peritonitis who usually lie very still and resist movement. Despite being called "colic," the pain of stone passage is usually constant and may be in the flank or radiate to the groin. Radiation to the groin area suggests that the stone is in the middle or lower part of the ureter.

If the stone is passed, it should be sent for analysis. Other important investigations in the acute setting should be designed to rule out urosepsis and obstruction as well as to exclude other diagnoses. Helical computed tomography scanning offers high sensitivity and specificity for the diagnosis of ureteric stones and can provide significant information if nephrolithiasis is not the correct diagnosis. Investigations aimed at determining the cause of the kidney stones can be delayed until the acute episode is over.

Patients with acute renal colic can be managed supportively. Kidney stones smaller than 5 mm can

be expected to pass spontaneously over the course of 48 h with symptomatic relief. It should be remembered that disappearance of symptoms does not always imply stone passage, and a confirmatory test is required if the stone is not voided. Indications for surgical intervention during the acute phase include nonpassage of an obstructing stone, urosepsis with obstruction, and ongoing pain. Acute interventions may include percutaneous nephrostomy followed by ESWL, percutaneous lithotripsy, and cystoscopy with retrieval of the stone. Invasive surgery is rarely required.

### See Also the Following Articles

Hypercalciuria • Hypocalcemia, Therapy • Kidney Disease in Diabetes

### Further Reading

- Andreucci, V. E. (1999). New frontiers in renal stone disease. *Nephron* **81**(Suppl. 1).
- Bihl, G., and Meyers, A. (2001). Recurrent renal stone disease: Advances in pathogenesis and clinical management. *Lancet* **358**, 651–656.
- Khan, S. R. (1999). Second Finlayson Colloquium on Urolithiasis. *J. Am. Soc. Nephrol.* **10**(Suppl. 14).
- Morton, A. R., Iliescu, E. A., and Wilson, J. W. L. (2002). Nephrology: I. Investigation and treatment of recurrent kidney stones. *Can. Med. Assoc. J.* **166**, 213–218.
- Pearle, M. S. (2001). Prevention of nephrolithiasis. *Curr. Opin. Nephrol. Hypertension* **10**, 203–209.
- Tiselius, H. G., Ackerman, D., Alken, P., Buck, C., Conort, P., and Galluci, M. (2001). Guidelines on urolithiasis. *Eur. Urol.* **40**, 362–371.
- Wilkinson, H. (2001). Clinical investigation and management of patients with renal stones. *Ann. Clin. Biochem.* **38**, 180–187.

protects this sensitive protein against inactivation at 37°C.

CPN is also an anaphylatoxin inactivator because it cleaves the C-terminal Arg of C3a, C4a, and C5a, but their rates of hydrolysis differ. The Gly-Arg bond of C5a is cleaved approximately 10 times slower than the Ala-Arg terminus of C3a.

Human CPN is considered to be life-sustaining, since no individuals whose blood completely lacks this enzyme are known. For example, inhibition of CPN activity in guinea pigs resulted in sudden death after complement activation when the generated anaphylatoxins were not inactivated.

In humans, low enzymatic activity may be important in the protamine-reversal syndrome. Protamine, when given to neutralize the effects of heparin after extracorporeal circulation, can trigger a catastrophic reaction: pulmonary vasoconstriction, bronchoconstriction, and systemic hypotension. Protamine is a potent inhibitor of this carboxypeptidase (CP) and can also stimulate the generation of kinins and anaphylatoxins; consequently, the inhibition of anaphylatoxin or kinin inactivation may contribute to the protamine reversal syndrome.

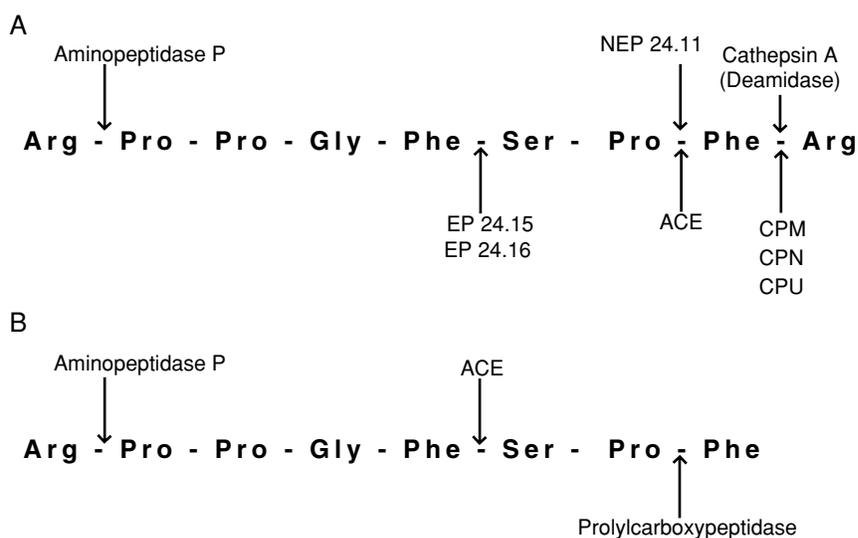
Plasma CPN activity may be an aid in the diagnosis of myocardial infarction where the blood level of creatine kinase (CK) released from heart muscle is measured. CPN cleaves the C-terminal lysine residue from either the M or B subunit of CK to yield des-Lys MM or MB isoforms when separated electrophoretically. A variation in blood carboxypeptidase N activity may affect this CK isoform ratio.

Although CPN cleaves the C-terminal Arg of kinins (see Fig. 1), it preferentially hydrolyzes C-terminal Lys, an important function in the regulation of plasminogen activation. CPN can reduce plasminogen binding to cells because C-terminal lysine residues on cell surface proteins are plasminogen "receptors." As bound plasminogen is more efficiently activated, CPN effectively lowers plasminogen binding to cells and thereby influences plasminogen activation.

No individuals have been reported to completely lack CPN but a 65-year-old patient and his sister had only 20% of normal activity. He suffered from weekly occurrence of angioedema. Other members of the family had angioedema, urticaria, hay fever, and asthma, but only slightly depressed CPN levels. An analysis of the 50 kDa subunit genomic DNA from the 65-year-old patient revealed a frameshift mutation in exon 1 that was not present in any of 128 normal controls.

## CARBOXYPEPTIDASE U

Carboxypeptidase U (CPU; thrombin activatable fibrinolysis inhibitor, TAFI; plasma CPB; CPR; EC 3.4.17.20) is another B-type CP synthesized in the liver and secreted into the blood. In contrast to CPN, it is normally in plasma as an inactive proenzyme of 60 kDa bound to plasminogen. CPU is proteolytically activated during coagulation to the active 35 kDa form by the thrombin-thrombomodulin complex, but it is unstable and rapidly inactivated at 37°C or by further proteolytic cleavage. The primary



**Figure 1** Bonds hydrolyzed in bradykinin (A) and des-Arg<sup>9</sup>-bradykinin (B) by peptidases.

role of CPU is to down-regulate plasminogen binding and activation by cleaving C-terminal Lys residues from cell surface proteins and in fibrin clots.

Regarding the potential function of CPU as a kininase, one of the first groups to identify the enzyme utilized bradykinin as a substrate and studies of the purified enzyme confirmed these findings. However, the reported  $K_m$  of bradykinin is extraordinarily high (10 mM), even though it was lower than that of all other substrates ( $K_m = 63\text{--}290$  mM). This brings into question the physiological relevance of CPU as a kininase, although serum degrades bradykinin fivefold faster than can be accounted for only by CPN. Because CPU is normally a proenzyme and is unstable after activation, CPU may be a significant kininase only in certain circumstances, such as shortly after initiation of the coagulation cascade.

## CARBOXYPEPTIDASE M

Early studies indicated that tissues and urine contained CPN-like activity, but it was assumed that the 50 kDa active subunit would enter organs and attach to plasma membranes or be excreted. When the active CP that cleaved basic C-terminal amino acids was purified from human urine and from placental membrane fractions, it became clear that this was not the case. Purification, cloning, and engineering of recombinant enzyme revealed that this enzyme had a different specificity and structure than those of CPN. Owing to its membrane attachment, it was called carboxypeptidase M (CPM).

CPM is a glycoprotein (23% carbohydrate by weight) with a molecular weight of 62 kDa anchored to the plasma membrane via a glycosylphosphatidylinositol tail. CPM is present in many tissues including kidney, lung, placenta, brain, intestine, peripheral nerves, and blood vessels and in cultured cells such as endothelial cells, fibroblasts, and Madin-Darby canine kidney cells. It is a differentiation-dependent cell surface antigen on white blood cells. The enzyme can be released from the membrane and soluble CPM is found in various body fluids including amniotic fluid, seminal plasma, and urine.

CPM has a pH optimum in the neutral range (6.5–7.5) ideally suited to cleaving peptides at the cell surface. There are no known endogenous inhibitors of the enzyme; thus, it is considered to be constitutively active on the cell surface. Tested with a variety of synthetic and endogenous peptides, CPM cleaved C-terminal arginine in preference over lysine, but the penultimate amino acid residue can also dramatically

affect the rate of hydrolysis. The naturally occurring peptide substrates of CPM include bradykinin, Arg<sup>6</sup>- and Lys<sup>6</sup>-enkephalins, dynorphin A<sub>1–13</sub>, and epidermal growth factor (EGF). Of the substrates tested, bradykinin has the lowest  $K_m$  (16  $\mu$ M) and the second highest specificity constant ( $k_{cat}/K_m$ ).

As a widely distributed ectoenzyme, CPM can participate in a variety of processes, such as control of peptide hormone activity at the cell surface and degradation of extracellular proteins and peptides. It may also carry out the second step in prohormone processing, (i.e., removal of C-terminal Arg or Lys residues from peptides released from prohormones by convertases), if and when incompletely processed peptides are released from secretory granules. Because many peptides work in an autocrine or a paracrine fashion, the location of CPM on the plasma membrane near peptide receptors makes it well suited to control their activities locally.

Within the kinin system, CPM can either inactivate or alter the specificity of kinin peptides by cleaving the C-terminal Arg<sup>9</sup> as mentioned above for CPN. Because CPM is localized on the cell surface, close to the B<sub>1</sub> and B<sub>2</sub> receptors, it is well suited to this important role. Indeed, it has been shown that CPM-like activity on the surface of human lung microvascular endothelial cells was up-regulated approximately twofold by treatment with interleukin-1 $\beta$  and interferon- $\gamma$ ; these conditions increased B<sub>1</sub> receptor responses approximately fourfold. When BK stimulated these cells, the resulting nitric oxide release was prolonged and enhanced because CPM converted the peptide to des-Arg<sup>9</sup>-BK to activate the B<sub>1</sub> receptor. These effects were then blocked by a CP inhibitor or by a B<sub>1</sub> receptor antagonist.

CPM efficiently removes the C-terminal Arg from EGF and may also cleave other growth factors that contain a C-terminal Arg or Lys [e.g., other EGF-like peptides, nerve growth factor (NGF), amphiregulin, hepatocyte growth factor, erythropoietin, macrophage-stimulating protein]. Although the role of the C-terminal basic residue of these factors has not been investigated in detail for most of them, a C-terminal Arg residue is required for association of EGF and NGF with their respective binding proteins.

The role of the up- or down-regulation of CPM in the differentiation of monocytes to macrophages and of B<sub>1</sub> lymphocytes is unknown, but a subject to be investigated further. The released free Arg, as a substrate for the inducible nitric oxide synthase, may provide more nitric oxide in inflammation, possibly just as the related enzyme carboxypeptidase D does in a mouse macrophage cell line.

In the lung, the presence of CPM on type I cells, which constitute 93% of the total alveolar surface, indicates that it may have protective functions here. This enzyme is also readily mobilized from the cell surface. The pulmonary synthesis or release of CPM may be up-regulated in disease states as the enzyme levels in bronchoalveolar lavage fluids were elevated almost fivefold in patients with pneumocystic or bacterial pneumonia or lung cancer.

The functions of CPM in other locations have not been explored. For example, in the placenta it may protect the fetus from maternally derived peptides. CPM in central nervous system (CNS) myelin and in Schwann cells in peripheral nerves may participate in the growth or maintenance of neurons.

## KININASE II

Kininase II (ACE)-type activity, the enzymatic release of Phe<sup>8</sup>-Arg<sup>9</sup> from BK, was detected first with a *Clostridium histolyticum* preparation and then by an enzyme separated from a membrane-enriched fraction (microsomal) of homogenized kidney and also from human plasma. Originally, ACE was discovered by Skeggs and colleagues in the 1950s in horse plasma. It converted the inactive Ang I decapeptide released by renin—called at that time hypertensin or angiotensin—to the active vasoconstrictor octapeptide Ang II.

After the discovery of kininase II in 1965–1967, it was reported in 1970–1971 that the same enzyme cleaves both BK and Ang I by releasing Phe-Arg from BK and His-Leu from Ang I; thus, it is a peptidyl dipeptide hydrolase called peptidyl dipeptidase I. The conclusion that the two activities were due to the same protein was not readily accepted because different amounts of Cl<sup>-</sup> ions are needed to cleave Ang I or BK. The lack of chloride inhibited Ang I conversion by 93%, but the hydrolysis of BK was reduced by only approximately 50%. The discovery of two independent active site domains of ACE led to the finding that some substrates and inhibitors react preferentially with just one of the two active sites, with different Cl<sup>-</sup> requirements.

The  $K_m$  of BK is much lower than that of Ang I (0.1–1  $\mu M$  versus 10–50  $\mu M$ ), which, taken with the turnover number ( $k_{cat}$ ), indicates a higher specificity constant with BK than with Ang I for human ACE; generally, ACE (kininase II) is considered to be the most important human kininase.

In addition to Ang I and BK, ACE cleaves a variety of bioactive peptides *in vitro*. Thus, beyond hydrolyzing C-terminal dipeptides, it released *in vitro* C-terminal tripeptides, e.g., from des-Arg<sup>9</sup>-BK, protected

C-terminal tripeptides [e.g., from substance P, luteinizing hormone-releasing hormone (LHRH)], protected C-terminal dipeptide (e.g., substance P), and even protected N-terminal tripeptide (e.g., of LHRH). Compared to other ACE substrates, it is likely that the Arg<sup>6</sup>-Phe<sup>7</sup> of Met<sup>5</sup>-enkephalin-Arg<sup>6</sup>-Phe<sup>7</sup> is split off with the highest turnover number by the N-domain active center of ACE. The N-domain of ACE hydrolyzes *in vivo* the tetrapeptide Ac-Ser-Asp-Lys-Pro, which is involved in the control of hematopoietic stem cell proliferation by preventing their recruitment into S phase.

After molecular cloning of the somatic form of ACE, two homologous domains were identified, the N- and C-domains; this was taken as an indication for the duplication of an ancestral gene. The germinal, testicular form has only one domain. The highest (89%) sequence similarity of the two domains of human ACE occurs around the active sites. The catalytic sites contain the canonical zinc-binding motif HEXXH. There is a high level of sequence similarity, 80–90%, in both the nucleotide and amino acid sequences, between the somatic mammalian ACE cDNA sequences cloned from different species.

ACE is an ectoenzyme anchored to the plasma membrane by a C-terminal transmembrane domain; most of it is exposed at the extracellular surface of the cell. There are two ACE isoforms; the somatic form of approximately 150–180 kDa is in endothelial, epithelial, and neuronal cells and the smaller isoform (90–110 kDa) is present in germinal cells. Germinal ACE contains 732 amino acids, compared to 1306 in somatic ACE, and corresponds to the C-domain of somatic ACE with a single active site. Germinal ACE has the same hydrophobic transmembrane peptide and cytosolic domains as somatic ACE. The inactivation of the ACE gene in the homozygous male mouse leads to reduced fertility attributed to inactivation of the germinal form of ACE. Somatic ACE is highly expressed, for example, in the plasma membrane of vascular endothelial cells in the lung oriented to metabolize circulating substrates. ACE concentration is probably even higher on the microvilli of the brush borders of absorptive epithelia of the small intestine, placenta, and kidney proximal tubules. T lymphocytes and fibroblasts also contain ACE. In brain, ACE is highly expressed in the choroid plexus, which may be the source of ACE in cerebrospinal fluid, and in ependyma, subfornical organ, basal ganglia (caudate putamen and globus pallidus), substantia nigra, and pituitary. Neuronal ACE and the endothelial and epithelial forms of the enzyme have the same specificities and are not distinct isoenzymes.

Soluble ACE is found in many biological fluids such as serum, seminal, amniotic, and cerebrospinal fluids, released from cell membranes by a class of enzymes called secretases.

The high carbohydrate content of both ACE domains has frustrated attempts to crystallize the enzyme. Finally, in 2003, the crystallization of the human testicular ACE, and thus, the C-domain of ACE without the hydrophobic transmembrane and intracellular domains, was reported. The enzyme was complexed with the inhibitor lisinopril. Although the first orally active ACE inhibitor, captopril, was synthesized on the basis of an assumed similarity to carboxypeptidase A, the structure of the crystallized enzyme resembled instead the other kininase II-type enzyme NEP, despite the lack of sequence homology. The active site is at the bottom of a groove in ACE; this may explain the rather limited size of the peptides that ACE can hydrolyze, consisting usually of less than a dozen amino acids.

Because both CPN and ACE are metallopeptidases, in initial experiments some inhibitors of BK breakdown by CPN, such as metal sequestering agents, inhibited both enzymes. Of other ACE inhibitors, the structures of two snake venom peptides that potentiated BK were reported at a meeting in 1969. The pentapeptide BPP<sub>5a</sub> from *Bothrops jararaca* had a single C-terminal proline, whereas the undecapeptide from *Akistrodon halys blomhoffii* had a C-terminal Pro-Pro sequence. A synthetic nonapeptide, teprotide—found in snake venom originally—also had C-terminal Pro-Pro sequence and it became the first ACE inhibitor to be tested clinically. When this ACE inhibitor was given intravenously to patients, it not only lowered the elevated blood pressure, but was also beneficial in congestive heart failure.

The studies on kininase inhibition and BK potentiation led to the synthesis of ACE inhibitors that are given orally to millions of patients to treat hypertension and a variety of clinical conditions in which the effects are not related to lowering the blood pressure. These conditions range from congestive heart failure to diabetic nephropathy. ACE inhibitors lower the mortality and morbidity after heart failure and they even reduce its incidence. Many laboratory and clinical findings attributed at least part of the results with ACE inhibitors to enhancing the effects of BK. In addition to being involved in BK metabolism and potentiating the effects of B<sub>2</sub> agonists via ACE, some of the ACE inhibitors can directly activate the B<sub>1</sub> receptor of des-Arg-BK followed by the release of nitric oxide, in the absence of the peptide ligand in cultured endothelial cells.

## THE ANGIOTENSIN METABOLITES

Some of the Ang metabolites, enzymatic breakdown products (see Fig. 2), are active on their own. Some of them mimic Ang II, whereas others antagonize the effects of Ang II. Two of them are mentioned here because they potentiate BK, probably by interacting with ACE, but not necessarily acting as inhibitors only.

Ang (1-7) (des-Phe<sup>8</sup>-Ang II) is a substrate of the N-domain, but an inhibitor of the C-domain of ACE. Another enzymatic product of Ang I hydrolysis, Ang (1-9) (des-Leu<sup>10</sup>-Ang I), inhibits both domains of ACE; both Ang (1-7) and Ang (1-9) can enhance BK's effect on the B<sub>2</sub> receptor.

## SERINE CARBOXYPEPTIDASES

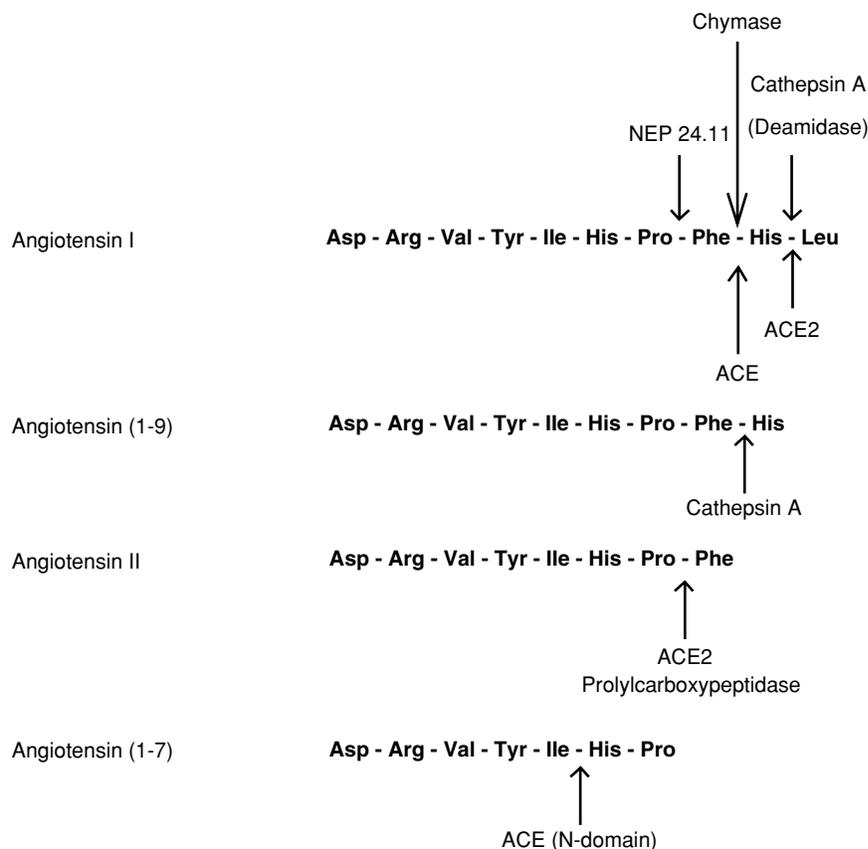
A serine carboxypeptidase-type enzyme, called deamidase, cathepsin A, lysosomal protective protein, or lysosomal CPA, also cleaves BK at the Phe<sup>8</sup>-Arg<sup>9</sup> bond as shown *in vitro*. This enzyme has an acid pH optimum for CP and esterase activities, but deamidates amidated C-terminal amino acids (e.g., substance P) at neutral pH.

Another serine CP, the lysosomal Pro-X CP (prolylcarboxypeptidase or angiotensinase C), inactivates desArg<sup>9</sup>-BK, the ligand of the B<sub>1</sub> BK receptor, at the Pro<sup>7</sup>-Phe<sup>8</sup> bond (see Fig. 1). Prolylcarboxypeptidase occurs in abundance in neutrophils; this may have some bearing on B<sub>1</sub> receptor activity, as it is up-regulated by agents promoting inflammation. Thus, this enzyme may be an inactivator of the B<sub>1</sub> receptor agonists. The same bond is also hydrolyzed in Ang II both by Pro-X-CP and by an enzyme called ACE2 with sequence similarities to ACE, but it is not inhibited by ACE inhibitors (Fig. 2).

## AMINOPEPTIDASES

Aminopeptidases have two different functions in kinin metabolism. A blood-borne aminopeptidase, present also in tissues, converts Lys-BK to BK by cleaving the Lys<sup>0</sup>-Arg<sup>1</sup> bond. A second aminopeptidase, called aminopeptidase P, but originally referred to as "prolidase," is present in erythrocytes, kidney, lung, and other tissues. It inactivates BK by cleaving at Arg<sup>1</sup>-Pro<sup>2</sup>. Rat tissues have the highest concentrations of this enzyme; human tissues apparently express much less aminopeptidase P.

In contrast to ACE and NEP, which are bound by a transmembrane peptide as type I and type II transmembrane enzymes, aminopeptidase P is membrane-bound via a glycosylphosphatidylinositol anchor. Its molecular weight is approximately 90–95 kDa. As with



**Figure 2** Peptide bonds hydrolyzed in angiotensins by peptidases.

some other peptidases, it forms dimers or even oligomers. Aminopeptidase P, a zinc metallopeptidase, can be activated by  $Mn^{2+}$  with some substrates and is inhibited by chelating agents. Other inhibitors include sulfhydryl compounds. Interestingly, many ACE inhibitors also inhibit aminopeptidase P, although with a higher  $K_i$  (in the micromolar range) and  $Mn^{2+}$  enhances this inhibition. This is likely to be due to the presence of Pro or Pro-like structures in the ACE inhibitors and to the effective zinc-binding moieties.

### NEUTRAL ENDOPEPTIDASE 24.11

Neutral endopeptidase 24.11 (neprilysin, NEP) is a zinc metallopeptidase with a single active site containing the HEXXH sequence to bind  $Zn^{2+}$ . The enzyme is a single-chain protein of 742 amino acids and, as a type 2 membrane protein, is bound via an uncleaved N-terminal signal peptide.

NEP is distributed widely; its expression in vascular endothelial cells is low, but epithelial cells, especially in microvillar structures, are rich in NEP. Its presence in the CNS has been investigated in detail

because neuropeptides are among its substrates. The relevance of the high concentration of NEP in the male genital tract, especially in prostate gland, is not known. NEP under the name “common acute lymphoblastic leukemia antigen” (CALLA or CD10) is present in lymphoblasts and also in neutrophils. Of the solid tumors, NEP is highly expressed in malignant liver cells of rats and humans.

NEP cleaves peptides at the N-termini of hydrophobic amino acids and as such it is a second kininase II-type enzyme that releases the C-terminal Phe<sup>8</sup>-Arg<sup>9</sup> of BK. It was first discovered as an endopeptidase that cleaves the B-chain of insulin, but its substrates include enkephalins, endothelin, atrial natriuretic peptide, substance P, and a chemotactic peptide.

Purified human NEP was used to establish the kinetics of hydrolysis of BK. The  $k_{cat}$  for BK is higher with NEP than with ACE but because of the much higher  $K_m$  (120  $\mu M$  versus 0.2  $\mu M$ ) the specificity constant,  $k_{cat}/K_m$ , of NEP is lower (40  $\mu M^{-1} \text{ min}^{-1}$  versus 3667  $\mu M^{-1} \text{ min}^{-1}$ ).

The NEP level in circulating blood plasma is normally low, but increases 60- to 80-fold in adult

respiratory distress syndrome with septic pneumonia. Inhibition of NEP in rat bronchial epithelium markedly enhances bronchoconstriction induced by substance P and BK. BK-induced bronchoconstriction in asthmatic patients is augmented by NEP inhibitors. Indirect actions of BK are also influenced by NEP inhibition, because when BK releases substance P from nerve endings, it is subsequently cleaved by NEP.

Microvascular leakage in guinea pig airways and enhanced myocardial blood flow after sensory nerve stimulation in the rat heart were potentiated by NEP inhibitors, possibly through blocking BK inactivation.

An NEP inhibitor has been used to successfully treat young boys suffering from acute watery diarrhea. The results were attributed to blocking the breakdown of endogenous enkephalins acting on  $\delta$ -opioid receptors.

Because of the increased reactivity to the combined administration of inhibitors of two peptidases, several compounds that inhibit both NEP and ACE, or ACE and aminopeptidase, have become available. Extensive reports deal with the clinical and laboratory experimental results with the dual ACE and NEP inhibitors, for example, with omapatrilat (Vanlev).

Endothelin-converting enzyme, an NEP-related peptidase with some sequence similarity, also cleaved BK at the Pro<sup>7</sup>-Phe<sup>8</sup> bond, albeit with a  $K_m$  approaching 1 mM.

## ENDOPEPTIDASES 24.15 AND 24.16

Endopeptidase 24.15 (EP24.15; thimet oligopeptidase) and endopeptidase 24.16 (EP24.16; neurolysin) are closely related (80% similarity and 63% identity in amino acid sequences) zinc metalloenzymes that contain the HEXXH zinc-binding motif. They are active at neutral pH and cleave a number of biologically active peptides *in vitro*. Both enzymes are abundant in brain and have been purified from rat brain homogenates.

EP24.15 is a 78 kDa protein containing 687 amino acid residues and is expressed at highest levels in brain and testis. When EP24.15 was purified from rat brain, it was determined that the enzyme was identical to two other enzymes named earlier, endopeptidase A, originally described as a kininase, and Pz-peptidase, which cleaved a synthetic substrate of collagenase. EP24.15 is activated by low concentrations of thiols, which convert an inactive multimer, with a blocked catalytic site, to an active monomer by disruption of intermolecular disulfide bridges. EP24.16 does not have these -S-S- bridges.

Tissues such as spleen, liver, kidney, lung, adrenal, and thyroid also contain EP24.15 activity. Although

the enzyme apparently lacks both a signal sequence and a membrane-binding motif, an estimated 10–25% of activity is associated with the plasma membrane fraction of a number of tissues.

EP24.16 is a 704-amino-acid protein that, like EP24.15, is present in both plasma membrane-associated (10–20% of activity) and soluble forms. In addition, EP24.16 is found in mitochondria. After EP24.16 was isolated from rat brain and sequenced, its identity to some other proteins, including a soluble Ang II-binding protein from porcine liver and a microsomal endopeptidase from rabbit liver, was revealed. The enzyme has also been purified from rat ileum and kidney and a variant, called EP24.16B, has been isolated from rat testis.

Both enzymes share most of their natural substrates and preferentially cleave them at the carboxyl side of hydrophobic amino acids. BK is hydrolyzed at the Phe<sup>5</sup>-Ser<sup>6</sup> bond. Other peptides, such as Ang I, opioids, gonadotropin-releasing hormone, substance P, and neurotensin, are also cleaved *in vitro* by both of the enzymes. These substances are hydrolyzed at the same bond by both enzymes, except for neurotensin, which is inactivated at Arg<sup>8</sup>-Arg<sup>9</sup> by EP24.15 and at the Pro<sup>10</sup>-Tyr<sup>11</sup> bond by EP24.16.

Because of the predominantly cytosolic nature of these two enzymes, their roles as kininases are not clear. With the advent of specific, biologically stable enzyme inhibitors, investigators found that the enzymes can be kininases. An inhibitor of both EP24.15 and EP24.16 potentiated the increase in permeability of pial (brain surface) microvasculature induced by BK *in vivo*. This suggests that one or both enzymes can moderate BK's effect on the blood-brain barrier, possibly in cases of inflammation.

The formation of BK 1-5 in the medium of cultured bovine aortic endothelial cells after incubation with BK was inhibited 45% by an EP24.16 inhibitor and 18% by an EP24.15 inhibitor. Thus, even a transiently extracellular location of these enzymes on endothelial cells would allow them to function as kininases in the vasculature.

## EPILOGUE

The very short half-life of kinins *in vivo* (seconds) pointed to the importance of their enzymatic metabolism but the acceptance of this conclusion took decades. It also follows that limited conclusions can be deduced from measuring circulating kinin levels, as kinin's actions are more likely paracrine or autocrine in nature. Only a long time after the discovery of the "original" kininase, CPN, were some of its functions

in the body, for example, on plasminogen activation, clarified.

The induction of so-called Chagas' heart disease by *Trypanosoma cruzi* has been found to be promoted by BK in mice. It was suggested that changes in kininase levels can affect parasite infectivity and thus the outcome of the disease.

ACE inhibitors, which have become widely used, originated from studies performed *in vitro*. The obvious consequence of the seemingly opposite, dual actions of kininase II or ACE on BK and Ang I was that its inhibitors block the release of Ang II while they can enhance the effect of BK. Because related peptidases cleave different active peptides, compounds that inhibit both ACE and NEP when tested clinically could block Ang I, BK, and atrial natriuretic peptide hydrolysis.

## See Also the Following Articles

CCK (Cholecystokinin) • Neurokinins

## Further Reading

- Brosnihan, K. B., Li, P., and Ferrario, C. M. (1996). Angiotensin-(1-7) dilates canine coronary arteries through kinins and nitric oxide. *Hypertension* **27**, 523-528.
- Cao, H., and Hegele, R. A. (2003). DNA polymorphism and mutations in CPN1, including the genomic basis of carboxypeptidase N deficiency. *J. Hum. Genet.* **48**, 20-22.
- Corvol, P., and Williams, T. A. (1998). Peptidyl-dipeptidase A/angiotensin I-converting enzyme. In "Handbook of Proteolytic Enzymes" (A. J. Barrett, N. D. Rawlings, and J. F. Woessner, eds.), pp. 1066-1076. Academic Press, San Diego, CA.
- Erdős E. G. (1970) Bradykinin, kallidin and kallikrein. In "Handb. Exper. Pharmacol." (E. G. Erdős, ed.) Vol. XXV, p. 768; and supplement to Vol. XXV (1979), p. 817. Springer-Verlag, Heidelberg, Germany.
- Erdős, E. G. (2001). Perspectives on the early history of angiotensin-converting enzyme—Recent follow-ups. In "Angiotensin-Converting Enzyme (ACE): Clinical and Experimental Insights" (T. D. Giles, ed.), pp. 3-16. Health Care Communications, Fort Lee, NJ.
- Erdős, E. G., and Deddish, P. A. (2002). The kinin system: Suggestions to broaden some prevailing concepts. *Int. Immunopharmacol.* **2**, 1741-1746.
- Hooper, N. M., and Turner, A. J. (2003). An ACE structure. *Nat. Struct. Biol.* **10**, 155-157.
- Kas, K., Schoenmakers, E. F., and Van de Ven, W. J. (1995). Physical map location of the human carboxypeptidase M gene (CPM) distal to D12S375 and proximal to D12S8 at chromosome 12q15. *Genomics* **30**, 403-405.
- Li, J., Rehli, M., Timblin, B., Tan, F., Krause, S.W., and Skidgel, R. A. (2002). Structure of the human carboxypeptidase M gene. Identification of a proximal GC-rich promoter and a unique distal promoter that consists of repetitive elements. *Gene* **284**, 189-202.
- Norman, M. U., Lew, R. A., Smith, A. I., and Hickey, M. J. (2003). Metalloendopeptidases EC 3.4.24.15/16 regulate bradykinin activity in the cerebral microvasculature. *Am. J. Physiol. Heart Circ. Physiol.* **284**, H1942-H1948.
- Norman, M. U., Reeve, S. B., Dive, V., Smith, A. I., and Lew, R. A. (2003). Endopeptidases 3.4.24.15 and 24.16 in endothelial cells: Potential role in vasoactive peptide metabolism. *Am. J. Physiol. Heart Circ. Physiol.* **284**, H1978-H1984.
- Puleo, P. R., Meyer, D., Wathen, C., Tawa, C. B., Wheeler, S., Hamburg, R. J., Ali, N., Obermueller, S. D., Triana, J. F., Zimmerman, J. L., et al. (1994). Use of a rapid assay of subforms of creatine kinase-MB to diagnose or rule out acute myocardial infarction. *N. Engl. J. Med.* **331**, 561-566.
- Salazar-Lindo, E., Santisteban-Ponce, J., Chea-Woo, E., and Gutierrez, M. (2000). Racecadotril in the treatment of acute watery diarrhea in children. *N. Engl. J. Med.* **343**, 463-467.
- Skidgel, R. A. (1998). Carboxypeptidase M. In "Handbook of Proteolytic Enzymes" (A. J. Barrett, N. D. Rawlings, and J. F. Woessner, eds.), pp. 1347-1349. Academic Press, San Diego, CA.
- Skidgel, R. A., and Erdős, E. G. (1998). Enzymatic breakdown of bradykinin. In "Pro-inflammatory and Anti-inflammatory Peptides" (S. I. Said, ed.), pp. 459-476. Dekker, New York.
- Skidgel, R. A., and Erdős, E. G. (1998). Lysine carboxypeptidase. In "Handbook of Proteolytic Enzymes" (A. J. Barrett, N. D. Rawlings, and J. F. Woessner, eds.), pp. 1344-1347. Academic Press, San Diego, CA.
- Tan, F., and Erdős, E. G. (1998). Lysosomal Pro-X carboxypeptidase. In "Handbook of Proteolytic Enzymes" (A. J. Barrett, N. D. Rawlings, and J. F. Woessner, eds.), pp. 405-407. Academic Press, San Diego, CA.
- Todorov, A. G., Andrade, D., Pesquero, J. B., de Carvalho Araujo, R., Bader, M., Stewart, J., Gera, L., Müller-Esterl, W., Morandi, V., Goldenberg, R. C., et al. (2003). *Trypanosoma cruzi* induces edematogenic responses in mice and invades cardiomyocytes and endothelial cells *in vitro* by activating distinct kinin receptor (B1/B2) subtypes. *FASEB J.* **17**, 73-75.
- Turner, A. J. (1998). Neprilysin. In "Handbook of Proteolytic Enzymes" (A. J. Barrett, N. D. Rawlings, and J. F. Woessner, eds.), pp. 1080-1085. Academic Press, San Diego, CA.
- Yusuf, S., Sleight, P., Dagenais, G., Montague, T., Bosch, J., Pogue, J., Taylor, W., Sardo, L., et al., for the Heart Outcomes Prevention Evaluation (HOPE) Study Investigators. (2000). Effects of an angiotensin-converting enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. *N. Engl. J. Med.* **342**, 145-153.

presentations, a number of patients can achieve significant virilization and go undiagnosed until adulthood. These patients can be discovered incidentally during the workup of a malignancy but are usually diagnosed on the occasion of their evaluation for infertility.

Serum levels of the male hormone testosterone are normally low in prepubertal boys and begin to rise during early puberty. In boys affected with Klinefelter's syndrome, a decreased capacity of the testes to secrete testosterone becomes evident in many cases as puberty continues. Therefore, testosterone levels tend to be either low-normal or frankly decreased in adult patients. This occurs despite an elevation above normal values of the circulating levels of luteinizing hormone (LH), the hormone that normally stimulates testosterone production by the testes. Because of the lack of testosterone, affected patients rarely become normally virilized. On the other hand, the levels of the female hormone estradiol, which is also normally secreted by the human testes, are usually elevated in Klinefelter's patients. These anomalies result in an imbalance between the two sex hormones, with increased estradiol levels and diminished testosterone values producing a hormonal milieu favorable to the development of bilateral gynecomastia. This is observed in more than 90% of Klinefelter's patients.

## DIAGNOSIS AND GENETIC BASES

Postpubertal men with Klinefelter's syndrome exhibit the clinical manifestations listed previously accompanied by a consistent hormonal profile. This profile essentially consists of elevated circulating levels of the two hypophyseal hormones, LH and follicle-stimulating hormone (FSH), together with low levels of the sex hormone testosterone. The typical Klinefelter's patient also exhibits complete azoospermia. However, several variants of the syndrome have been described, with a variable and often less severe degree of both gynecomastia and testosterone insufficiency. Patients with a less severe phenotype are frequently diagnosed at a later age because symptoms of decreased libido and potency, which accompany the fall in testosterone levels seen in adult patients, appear later in life. Therefore, milder forms of the disease can be revealed as late as 45 years of age.

Once suspected on the basis of the clinical and hormonal presentation, confirmation of the diagnosis of Klinefelter's syndrome relies on the demonstration of an abnormal karyotype. Indeed, the genetic basis of Klinefelter's syndrome is an abnormal number of sex chromosomes. In the normal human, a normal karyotype consists of 22 pairs of autosomes (non-sex

chromosomes) and one pair of sex chromosomes. The pair of sex chromosomes should be XX in the female and XY in the male, with the Y chromosome bearing the gene(s) responsible for male sexual differentiation. Therefore, each and every cell of a normal human female should have 46 chromosomes, with two X sex chromosomes (the karyotype being 46,XX), whereas the cell of a normal human male should have 46 chromosomes, with both an X and a Y sex chromosome (the karyotype being 46,XY). As stated earlier in the defining paragraph, there are a number of chromosomal abnormalities associated with Klinefelter's syndrome that have in common the display of at least one supplementary X chromosome. Therefore, a typical Klinefelter's karyotype is 47,XXY. Variants can be 48,XXXY, 48XXYY, or even 49,XXXYY, and they can probably explain the somewhat variable clinical expression of the disease.

The origin of these chromosomal abnormalities is a failure of the sex chromosomes to separate during the formation of the gamete, that is, the spermatozoid in the male or the ovum in the female. When a gamete possessing such an extra X chromosome is fertilized, the resulting zygote will bear an extra X chromosome, either from paternal origin (30% of cases) or from maternal origin (more often). One etiological factor that could potentially account for the nondisjunction of sex chromosomes is the maternal age. However, a genetic predisposition may also exist. Finally, affected patients frequently present 46,XY/47,XXY mosaicism, which is the second most common karyotype seen in Klinefelter's syndrome after 47,XXY. A mosaic state refers to those patients in whom only a subset of cells exhibit the abnormal karyotype. This can be explained by the successful elimination of the extra X chromosome in some cells at the early stages of embryo formation. Of note, mosaicism could account for the milder forms of Klinefelter's syndrome.

## TREATMENT

The treatment of patients suffering from Klinefelter's syndrome is essentially constituted by the replacement of the sex hormone testosterone when there is evidence that an affected individual's own testosterone production is insufficient. This is best achieved by intramuscular injections of depot preparations of testosterone at the average frequency of every 2 to 4 weeks. Frequency of administration depends on the dosage required and the degree of testosterone insufficiency. Oral preparations of testosterone are also available. However, the levels of circulating testosterone achieved in the patients with such preparations

are inconsistent from day to day, rendering the proper administration difficult. In addition, the occurrence of abnormal liver function tests, as well as of hepatic tumors, has been reported with orally administered testosterone. For this reason, and also because parenteral administration of testosterone is more effective in virilizing the patient, oral preparations are generally avoided.

Adequate virilization of affected patients can be achieved almost universally with the appropriate hormonal substitution. When treating adolescent males, special care should be given to avoid premature bone maturation that will occur with testosterone treatment. This is achieved by the administration of low doses of testosterone until the full growth potential of these individuals has been reached. In contrast, testosterone treatment will not correct gynecomastia if the latter is already present. The treatment of choice when gynecomastia is psychologically disturbing to the patient is reduction mammoplasty.

Klinefelter's patients are also usually azoospermic and, hence, infertile. The natural fertility potential of these patients is nonexistent. However, assisted reproductive technologies, and more specifically intracytoplasmic sperm injection (ICSI), offer Klinefelter's patients the possibility of fathering children. In the few successful cases of ICSI reported so far, mature spermatozoa were obtained either directly from testicular aspiration or from the ejaculate. To date, all infants born from Klinefelter's fathers after ICSI have had normal karyotypes because a high percentage of mature spermatozoas retrieved from Klinefelter's patients are chromosomally normal. However, one report showed clearly that in Klinefelter's patients, spermatozoa with the abnormal karyotype can also reach maturity and so could theoretically fertilize the egg. Therefore, the potential to transmit the chromosomal abnormality of the father to the offspring remains a major concern related to these therapeutic approaches. This concern raises the necessity of beginning to evaluate the chromosomal status of sperm before using it for assisted reproductive technologies.

## ASSOCIATED ABNORMALITIES AND COMPLICATIONS

Associated abnormalities include disorders of thyroid function, but clinically significant thyroid insufficiency

is uncommon. However, the early belief that boys affected with Klinefelter's syndrome may exhibit an increased frequency of learning disabilities as well as behavioral problems is erroneous. Adults with Klinefelter's syndrome as a group are not different from normal controls in terms of level of education, employment, socioeconomic status, or criminal behavior.

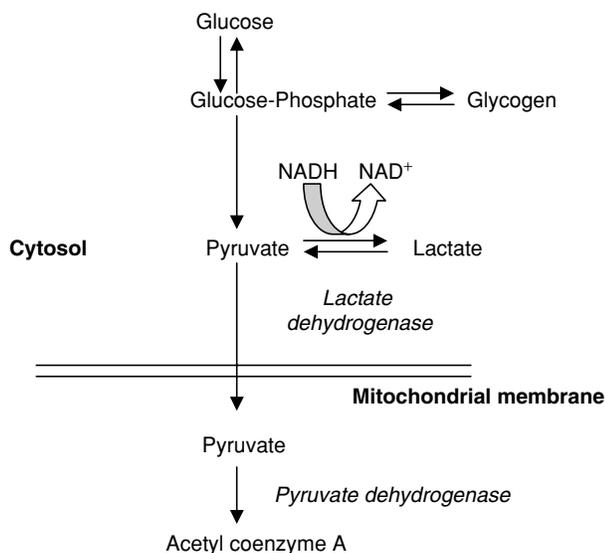
Complications include an increased risk of cerebrovascular disease (stroke) as well as an increased incidence of breast cancer. Indeed, the risk of carcinoma of the breast in Klinefelter's patients is 20 times higher than that in normal men and one-fifth that in women. There is no obvious reason for this increased risk because gynecomastia per se does not represent a premalignancy. A role for the presence of an extra X chromosome has been suggested, as has a role for the hormonal milieu. Other complications encountered in Klinefelter's patients include an increased incidence of diabetes and an increased risk of osteoporosis. This latter complication is probably related to the imbalance in sex hormone concentrations.

## See Also the Following Articles

Assisted Reproductive Technology (ART) • Delayed Puberty and Hypogonadism, Male • Fertility in Men with Spermatogenesis Abnormalities • Gynecomastia • Hypergonadotropic Hypogonadism • Spermatogenesis, Endocrine Control of • Testes, Embryology of • Undescended Testes

## Further Reading

- Foresta, C., Galeazzi, C., Bettella, A., Marin, P., Rossato, M., Garolla, A., and Ferlin, A. (1999). Analysis of meiosis in intratesticular germ cells from subjects affected by classic Klinefelter's syndrome. *J. Clin. Endocrinol. Metab.* **84**, 3807–3810.
- Hargreave, T. B. (2000). Genetics and male infertility. *Curr. Opin. Obstet. Gynecol.* **12**, 207–219.
- Klinefelter, H. F. J., Reifenstein, E. C. J., and Albright, F. (1942). Syndrome characterized by gynecomastia, aspermatogenesis without a Leydigism, and increased excretion of follicle-stimulating hormone. *J. Clin. Endocrinol.* **2**, 615–627.
- Salbenblatt, J. A., Bender, B. G., Puck, M. H., Robinson, A., Faiman, C., and Winter, J. S. (1985). Pituitary-gonadal function in Klinefelter syndrome before and during puberty. *Pediatr. Res.* **19**, 82–86.
- Schwartz, I. D., and Root, A. W. (1991). The Klinefelter syndrome of testicular dysgenesis. *Endocrinol. Metab. Clinics North America* **20**, 153–163.



**Figure 1** Regeneration of  $\text{NAD}^+$  by conversion of pyruvate to lactate permits glycolysis to continue under anaerobic conditions.

explain the higher risk of lactic acidosis than with metformin. Lactic acidosis complicating metformin (dimethylbiguanide) occurs much less commonly, with most cases being reported among patients in whom biguanide therapy is contraindicated (e.g., renal impairment or hypoxic states).

## TREATMENT AND PROGNOSIS

### General Medical Care

Serious initiating or aggravating conditions (e.g., cardiogenic shock) often dictate the poor prognosis of acute severe lactic acidosis. Accordingly, intensive supportive care and correction of precipitating factors are important.

### Bicarbonate

Despite controversy surrounding the benefits of alkali therapy, intravenous sodium bicarbonate remains the mainstay of treatment. High doses of bicarbonate may be required to elevate arterial pH; hemodialysis has been advocated in the management of the resulting sodium and volume overload. Hemodialysis will also remove metformin but the role of dialysis in treating metformin accumulation has been questioned.

**Table 1** Pathological Causes of Lactic Acidosis

Type A
<ul style="list-style-type: none"> <li>• Shock (e.g., septic, cardiogenic)</li> <li>• Respiratory failure</li> <li>• Tonic-clonic epileptic convulsions</li> <li>• Toxins (e.g., cyanide, carbon monoxide)</li> </ul>
Type B
<ul style="list-style-type: none"> <li>• Diabetes mellitus (e.g., in association with diabetic ketoacidosis—rarely severe)</li> <li>• Hepatic failure</li> <li>• Drugs [e.g., biguanides, salicylates, isoniazid, sorbitol (in parenteral feeds)]</li> <li>• Toxins (e.g., ethanol, methanol)</li> <li>• Congenital disorders [e.g., myopathy, encephalopathy, lactic acidosis, and stroke-like episodes (MELAS) syndrome (due to point mutation in mitochondrial DNA at position 3243 in the tRNA<sup>LEU(URR)</sup> gene)]</li> </ul>

## Other Measures

The value of alternative therapeutic approaches (e.g., using the pyruvate dehydrogenase stimulator dichloroacetate or Carbicarb, which contains equimolar amounts of sodium bicarbonate and sodium carbonate) remains unclear.

## See Also the Following Articles

Diabetes, Type 1 • Diabetes, Type 2 • Shock

## Further Reading

- Chan, N. N., Brain, H. P. S., and Feher, M. D. (1999). Metformin-associated lactic acidosis: A rare or very rare clinical entity? *Diabetic Med.* **16**, 273–281.
- DiMauro, S., and Schon, E. A. (2003). Mitochondrial respiratory chain diseases. *N. Engl. J. Med.* **348**, 2656–2668.
- Holt, H. B., and Krentz, A. J. (2003). Metabolic emergencies in type 2 diabetes: Lactic acidosis. In “Text book of Type 2 Diabetes,” (B. J. Goldstein and D. Müller-Wieland, eds.), pp. 183–198. Martin Dunitz, London.
- Howlett, H. C. S., and Bailey, C. J. (2000). A risk–benefit assessment of metformin in type 2 diabetes. In “Drug Treatment of Type 2 Diabetes” (A. J. Krentz, ed.), pp. 61–76. ADIS Books, Auckland, New Zealand.
- Kraut, J. A., and Kurtz, I. (2001). Use of base in the treatment of severe acidemic states. *Am. J. Kid. Dis.* **38**, 703–727.
- Lalau, J.-D., and Race, J.-M. (2000). Metformin and lactic acidosis in diabetic humans. *Diabetes Obes. Metabol.* **2**, 131–137.
- Luft, F. C. (2001). Lactic acidosis update for critical care clinicians. *J. Am. Soc. Nephrol.* **12**, S15–S19.
- Stacpoole, P. W. (1998). Metformin and lactic acidosis: Guilt by association? *Diabetes Care* **21**, 1587–1588.

## ETIOLOGY

LCH is a disease of unknown etiology. Although pulmonary LCH in adults has been reported to be of nonclonal origin, biopsies from lesions in children, with either single-system or multisystem involvement, has demonstrated proliferation of a single LC clone. Several cytokines, such as granulocyte-macrophage colony-stimulating factor, interferon- $\gamma$ , interleukin-1 (IL-1), and IL-10, were also found to be expressed by LCs in LCH lesions. The pattern of cytokine expression induces the recruitment of LC progenitors and in addition favors their maturation and rescue from apoptosis, resulting in the accumulation of pathologic LCH cells. Moreover, ILs, tumor necrosis factor  $\alpha$  (TNF $\alpha$ ), and cluster designation 40 ligand (CD40L) are also produced by T cells inside the lesion, indicating immunologic dysfunction as a possible cause of the disease.

Chromosomal abnormalities of the type showing chromosomal breakage were found in peripheral blood lymphocytes when they were stimulated with phytohemagglutinin. Multisystem involvement was reported to be associated with a higher incidence of chromosomal breakage (13.4%) than single-system involvement (6%). However, it remains unclear whether this difference results as a secondary effect of cytokine production or as an inherent chromosomal instability. Parental solvent exposure, perinatal infections, and a family history of thyroid disease were also reported as being associated with an increased incidence of the disease.

## CLINICAL MANIFESTATIONS

There is a wide spectrum of clinical manifestations of the disease and almost every system can be involved; however, involvement does not necessarily indicate dysfunction. Organ infiltration commonly causes distress and disability, but is of prognostic importance only if vital functions are affected. LCH can be stratified according to system involvement as a single-system disease or as a multisystem disease. A single-system disease may involve only a single site, e.g., a single bone lesion, isolated skin disease, or a solitary lymph node, or may involve multiple sites, e.g., multiple bone lesions and multiple lymph node involvement. Multisystem disease is a multiple-organ disease, with or without dysfunction.

### Bones

LCH frequently affects the bones. The majority of lesions are asymptomatic but can also present as a

painful swelling; lytic skull lesions are the most frequent finding in children, whereas the jaw is more commonly involved in adults. Other bone sites frequently affected include the femur, vertebrae, ribs, pelvis, humerus, and extremities. Clinical examination usually reveals a soft and sensitive protuberance. Skull lesions are not always lytic but they may extend intracranially or impinge on the dura. Patients with such lesions are at greater risk for developing central nervous system (CNS) infiltration and/or diabetes insipidus (DI). Periorbital disease is usually accompanied by proptosis but optic nerve compression is rare. Skeletal lesions can be detected with X rays, which are superior to radioisotope scans, showing well-defined osteolytic areas, a periosteal reaction that resembles malignancy, and a sclerotic areola in healing lesions.

### Skin and Oral Mucosa

Skin rash is very common in infants and may present in any part of the body as brown to purplish papules (Hashimoto-Pritzker disease), resembling healing chickenpox; these lesions heal spontaneously within the first year of life without therapy. Adult patients may develop a red or brown papular rash or crusted dermal areas scattered over the trunk, often healing with depigmentation. Lesions in the perianal region, in the genitalia, in the scalp, or behind the ears are often ulcerative, are associated with great distress, and are usually treated as skin infections before the final diagnosis of LCH is made. Scalp rash can be mistaken for cradle cap, seborrhea, or severe dandruff. In the mouth, LCH often presents as gingival hypertrophy or ulcerative lesions on the soft or hard palate, buccal mucosa, tongue, and lips.

### Lymph Nodes and Thymus

LCH frequently infiltrates the lymph nodes. Involvement of the cervical nodes usually presents as soft, matted nodal groups with accompanying lymphedema. Infiltration of the mediastinal lymph nodes presents as nodular enlargement, making histological examination mandatory for differential diagnosis from infections or neoplasia. The abdominal glands can also be affected, whereas Waldeyer's ring affection can cause obstruction of the upper airways. Additionally, LCH infiltration of regional nodes draining solitary skin or bone lesions is not uncommon. Thymic involvement can be detected either as an enlargement on chest X ray or as morphological changes on biopsy.

## Lungs

Pulmonary LCH runs an unpredictable and variable course, occurring less frequently in children than in adults, in whom it may be related and/or exacerbated by cigarette smoking. Clinical presentation varies from an asymptomatic pattern to a progressive disease that leads to respiratory failure. Children usually present with tachypnea or dyspnea, whereas spontaneous pneumothorax resulting from a bulla rupture is common in adults. The cystic/nodular pattern of lung involvement, with a tendency to form large cavities and widespread fibrosis, leads to lung destruction and appears as “honeycombing” on chest X ray. Diagnosis is confirmed from the findings on biopsy or in bronchial washings.

## Ears

Aural involvement usually presents with discharge either from external otitis due to skin involvement of the aural canal or from polypoid tissue extension into the canal from an adjacent bone infiltration. Differential diagnosis between these two conditions is mandatory as external otitis should be treated with topical regimens, whereas polyps may require curettage. The middle ear can also be affected, sometimes resulting in deafness in persistent cases with ossicle damage. Involvement of the mastoid bones may also occur.

## Liver and Spleen

Liver involvement is associated with hepatomegaly and hepatic dysfunction, leading to hypoalbuminemia, ascites, edema, coagulopathy, and, rarely, hyperbilirubinemia. Splenomegaly, when present, can be massive and may lead to cytopenia due to hypersplenism that may require splenectomy if systemic treatment has failed.

## Bone Marrow and Peripheral Blood

Bone marrow involvement affects mainly children with multisystem disease and is usually associated with involvement of the lymphoreticular system and the skin. However, bone marrow infiltration has been found even in patients with single-site involvement. A mild anemia of chronic disease is the most common finding, although neutropenia and thrombocytopenia might also occur, whereas a normal blood count does not exclude bone marrow infiltration. A reversal of the CD4<sup>+</sup>/CD8<sup>+</sup> ratio is also a very common finding in the acute phase of the disease.

## Gastrointestinal Tract

Gastrointestinal (GI) manifestations from LCH are rare, with the most frequent presenting symptoms being malabsorption and diarrhea. Although endoscopic examination may reveal involvement of the stomach, the diagnostic evaluation of the rest of the GI tract is difficult because of the intermittent pattern of the disease.

## Endocrine System and Central Nervous System

Anterior pituitary dysfunction has been described in up to 20% of patients with LCH, usually associated with DI. However, anterior pituitary function has not been systematically studied in adults; most information has been obtained from studies in children. Anterior pituitary deficiency in LCH is almost always associated with DI; only a few cases of pituitary hormone insufficiency without DI have been described in the literature. Although DI may predate the diagnosis of LCH, it usually develops at approximately 12 months of age, with a range that can extend to many years from diagnosis. Associated features are multisystem disease with skull vault defects and, notably, temporal bone or orbital lesions with intracranial tumor extension. Growth hormone deficiency (GHD) occurs in approximately 40% of affected children and has been related to histiocytic infiltration of the hypothalamus. GHD is frequently the first endocrine defect in addition to DI, with a median latency of approximately 1 year from diagnosis. Growth retardation, although previously described, is thought to be an infrequent presentation of LCH, but GHD should always be considered in children with LCH and DI. As adults with GHD may show an increase in well-being and a favorable metabolic profile in response to GH therapy, assessment of GHD may be an important part of the evaluation of adult patients with LCH.

There are only a few reports on gonadal function in patients with LCH, as most studies have been of prepubertal children. Although the early studies in adults with LCH and DI failed to demonstrate abnormalities of gonadotropin secretion, a few cases of amenorrhea in adults have been described. Similarly, thyroid hormone deficiency can be a major component of anterior pituitary dysfunction in patients with LCH. Adrenocorticotrophic hormone (ACTH) deficiency presents mostly in the context of generalized pituitary involvement, although a few cases of isolated ACTH deficiency have also been described.

Imaging studies of patients with LCH and CNS involvement have shown that more than one type of

lesion may be present in 87% of individual patients, including patients with DI. In addition, abnormalities of the hypothalamo-pituitary axis (HPA) were observed in 68% of patients with CNS lesions and in 81.5% of patients with DI [infundibular thickening, partial or complete empty sella with a lack of posterior pituitary bright spot on T<sub>1</sub>-weighted magnetic resonance imaging (MRI) sequences, or a pituitary mass lesion]. Morphological changes in the HPA are optimally demonstrated by MRI with administration of gadolinium-dimeglumine gadopentetate. A small pituitary or empty sella has also been described in cases of combined anterior and posterior pituitary insufficiency. However, anterior pituitary dysfunction may also occur in the absence of structural changes on imaging and has been attributed to microinjury leading to vascular impairment and scarring. Other possible mechanisms include cytokine modulation from adjacent osseous lesions or an autoimmune effect.

Efforts to identify predictors of late endocrine sequelae in children with LCH showed that dynamic endocrine pituitary testing was not a useful predictor. Neither the site of involvement nor the extent of the disease was associated with further endocrine deterioration. Therefore, it seems that only DI in association with markedly abnormal hypothalamo-pituitary imaging identifies patients with LCH at higher risk for anterior pituitary dysfunction. As DI is associated with multisystem disease and progression that may be greatly delayed, such patients should be receiving regular and prolonged follow-up to identify such dysfunction and provide adequate replacement.

Diabetes insipidus with structural changes in the HPA often heralds the involvement of other parts of the brain with more global neurological or neuropsychological sequelae, depending on the location of the involvement. The signs and symptoms of nonendocrine hypothalamic involvement range from disturbances in social behavior, appetite, and temperature regulation to abnormal sleeping patterns. Further abnormalities, such as disturbances in thermoregulation and adipisia, can make DI difficult to treat and complicate the overall management of these problematic cases. This is particularly significant when there is also memory impairment and problems with compliance. More serious problems from CNS involvement, though of rare incidence, include mass lesions in the brain parenchyma or choroid plexus (resulting in hydrocephalus), space-occupying masses from neighboring bone or meningeal lesions, and diffuse infiltration of the cerebellum, leading to ataxia and visual problems.

Thyroid gland involvement is rare, with only a few reported cases involving goiter and in some cases hypothyroidism. Pancreatic infiltration has also been reported but without apparent dysfunction.

In summary, adult patients with LCH presenting with DI are at high risk for anterior pituitary hormonal deficiency and hypothalamic involvement, especially when accompanied by abnormal HPA imaging. Endocrine abnormalities should be actively sought in patients with LCH, as their recognition and management play important roles in the treatment of this difficult condition.

## EVALUATION OF PATIENTS

Any patient presenting with either single- or multiple-site disease should be evaluated to determine the extent of the disease and subsequent treatment plan. A detailed history and complete clinical examination are always mandatory in a patient suffering from LCH. Imaging studies should include X rays of the chest and long bones (skeletal survey) and, if orbital and/or mastoid involvement is suspected or neurologic dysfunction is present, computed tomography scans of the head or spine and pelvis should be performed. A complete blood count, erythrocyte sedimentation rate, C-reactive protein, liver function tests, serum electrolytes, and urinalysis are also mandatory. A water restriction test is indicated in any patient with symptoms suggestive of DI. The final diagnosis is established with biopsy of lesions and demonstration of positive immunostaining for CD1a and S100 protein. Birbeck granules can be identified only with electron microscopy and should be used in any case where the diagnosis is in doubt.

## TREATMENT

Two prospective studies in children, LCH-I and LCH-II, have established some general treatment guidelines that can also be applied to adults. The selection of treatment chiefly depends on the extent of the disease, which must always be evaluated carefully after a systemic diagnostic approach.

Patients with single-site involvement, such as skin and lymph nodes, can be treated either with prednisone or with a combination of prednisone and velban (vinblastine). The same treatment approach can be used in single, non-CNS bone lesions in which curettage of the infiltrated site is also an efficient therapy. However, in mandibular lesions and with tooth involvement, it is advisable to avoid surgical excision since medication may lead to regression of the lesions. Painful or recurrent bone lesions may also respond to intralesional steroid injection, which for safety reasons are preferably performed under radiographic guidance. Single-site involvement concerning only skin lesions can also be treated with a topical nitrogen mustard and sometimes may respond to topical steroid application. Additionally,

single steroid treatment may decelerate or even halt the process of the disease in adults suffering from the nodular pattern of pulmonary LCH.

Single-medication treatment, such as prednisone or radiation, is not effective in multiple bone lesions or multisystem disease. The treatment protocol of the LCH-III study suggests that these patients should be treated with prednisone and velban for 6 or 12 months.

The LCH-II study showed that VP-16 (etoposide) did not offer any additional therapeutic benefit concerning response, survival, or reactivation frequency, either as monotherapy or in combination with velban and prednisone. Therefore, and in view of its potential leukemogenicity, VP-16 is not included in the standard initial treatment of LCH-III.

According to the LCH-III protocol, patients with multisystem disease and subsequent risk for other organ involvement are randomly divided into treatment groups with either velban–prednisone–6-mercaptopurine or this three-drug combination with methotrexate, initially intravenously and orally during the maintenance phase. If a patient presents recrudescence of the disease after 6 weeks of treatment or does not respond to therapy, it is suggested, according to the LCH-S (salvage) study, that treatment proceed with the use of the purine analogue cladribine (2-CdA) for 4–6 months depending on the response of the patient. Some reports showed that 2-CdA, which can provide both cytotoxic and immunosuppressive effects, is a potent medication in adults with bone, skin, lymph node, pulmonary, and CNS involvement. (The protocols of the LCH-III and LCH-S studies are available from the Histiocytosis Association of America.)

Regarding endocrine involvement, systemic chemotherapy appears to be of little benefit in controlling the progression of the disease over the long term, although focal radiotherapy may halt local disease progression in terms of mass effects. Radiation therapy can also be used in vertebral lesions or lesions of the femoral neck with a risk of fracture or collapse. Other treatments that could be used systemically in the future include anti-TNF $\alpha$  agents, which are

undergoing testing, bone marrow transplantation, and anti-CD1a, which has been used in diagnostic evaluation but also seems promising as a possible therapeutic approach.

## Further Reading

- Arico, M., Girschikovsky, M., Geyereau, T., Klarsy, C., McClain, K., Grois, N., Emile, J.-F., Luking, E., DeJuli, E., and Danesmo, C. (2003). Langerhans cell histiocytosis in adults. Report from the International Registry of the Histiocytic Society. *Eur. J. Cancer*, in press.
- Broadbent, V., Egeler, R. M., and Nesbit, M. E. (1994). Langerhans cell histiocytosis—Clinical and epidemiological aspects. *Br. J. Cancer* **70**, S11–S16.
- Broadbent, V., and Gadner, H. (1998). Current therapy for Langerhans cell histiocytosis. In “Hematology/Oncology Clinics of North America (Langerhans Cell Histiocytosis)” (R. M. Egeler and G. J. D’Angio, eds.), Vol. 12, No. 2, pp. 327–338. W. B. Saunders, Philadelphia, PA.
- Carstensen, H., and Ornvold, K. (1993). The epidemiology of Langerhans cell histiocytosis in children in Denmark 1975–1989. *Med. Pediatr. Oncol.* **21**, 387–388.
- Coppes-Zantiga, A., and Egeler, R. M. (2002). The Langerhans cell histiocytosis X files revealed. *Br. J. Haematol.* **116**, 3–9.
- Gadner, H., Grois, N., Arico, M., Broadbent, V., Ceci, A., Jakobson, A., Komp, D., Michaelis, J., Nicholson, S., Potschger, U., Pritchard, J., and Ladisch, S. (2001). A randomised trial of treatment for multisystem Langerhans’ cell histiocytosis. *J. Pediatr.* **138**, 728–734.
- Kaltsas, G. A., Powles, T. B., Evanson, J., Plowman, P. N., Drinkwater, J. E., Jenkins, P. J., Monson, J. P., Besser, G. M., and Grossman, A. B. (2000). Hypothalamo-pituitary abnormalities in adult patients with Langerhans cell histiocytosis: Clinical, endocrinological, and radiological features and response to treatment. *J. Clin. Endocrinol. Metab.* **85**, 1370–1376.
- Malpas, J. (1998). Langerhans cell histiocytosis in adults. In “Hematology/Oncology Clinics of North America (Langerhans Cell Histiocytosis)” (R. M. Egeler and G. J. D’Angio, eds.), Vol. 12, No. 2, pp. 259–268. W. B. Saunders, Philadelphia, PA.
- Schmitz, L., and Favara, B. E. (1998). Nosology and pathology of Langerhans cell histiocytosis. In “Hematology/Oncology Clinics of North America (Langerhans Cell Histiocytosis)” (R. M. Egeler and G. J. D’Angio, eds.), Vol. 12, No. 2, pp. 221–246. W. B. Saunders, Philadelphia, PA.
- Vassalo, R., Ryu, J. H., Schroeder, D. R., Decker, P. A., and Limper, A. H. (2002). Clinical outcomes of pulmonary Langerhans’ cell histiocytosis in adults. *N. Engl. J. Med.* **346**, 484–490.



## Laron Syndrome

see Growth Hormone Insensitivity

## Fat Mass and Gender Influence Serum Leptin Concentration

The two major factors that determine serum leptin concentrations in humans under conditions of consistent food intake are fat mass and gender. Leptin is positively correlated with fat mass in adults, children, and newborns and is therefore increased in obesity. *LEP* gene expression is greater in larger adipocytes than in smaller adipocytes and leptin secretion is strongly correlated with fat cell volume. Therefore, the elevation in serum leptin results from the release from a greater number of fat cells and increased leptin synthesis in the larger adipocytes of obese subjects.

Serum leptin is significantly greater in women than in men with the same amount of body fat. One explanation for this finding is that women have a significantly greater subcutaneous adipose tissue mass relative to visceral adipose mass than men. *LEP* gene expression/leptin production is greater in subcutaneous than visceral adipocytes. Sex steroids also regulate leptin production. Testosterone reduces serum leptin in hypogonadal men and leptin levels decrease as serum testosterone increases during pubertal development in boys. Estrogen, in combination with anti-androgens, increases leptin in male-to-female transsexuals and testosterone decreases leptin in female-to-male transsexuals, independent of changes in adipose tissue mass. *In vitro*, estradiol stimulates, and dihydrotestosterone inhibits, leptin production in human visceral adipose tissue pieces in culture for 48 h.

## Energy Intake Regulates Serum Leptin Levels

Energy intake can regulate serum leptin concentrations independent of effects on adipose tissue mass. With standard timing for food intake (three meals per day in humans), leptin levels are reasonably constant from day to day, exhibiting a maximal daily variation of approximately 30%. However, serum leptin falls with short-term fasting (24 h) and increases within 4–5 h of refeeding. Maintenance of euglycemia with dextrose infusion prevents the fasting-induced drop in leptin, implicating insulin and/or glucose as the nutritional signal that is recognized by the adipocyte for leptin synthesis. In a prolonged study of energy restriction (moderate and severe), changes in serum leptin were best correlated with changes in glycemia.

Serum leptin exhibits a diurnal profile that is entrained to food intake. The peak in serum leptin occurs at ~0200 h, in both lean and obese subjects under standard living conditions. Day/night reversal shifts

the peak in serum leptin by 12 h. A shift of 6.5 h in meal timing without a change in light or sleep cycles shifts the leptin peak 5–7 h, suggesting that the nocturnal peak in serum leptin is a delayed postprandial response induced by after-meal excursions in glucose and insulin. The macronutrient content of meals also regulates serum leptin levels. High-fat/low-carbohydrate meals (60%/20%) over the course of 1 day reduce leptin levels. High-fat/low-carbohydrate meals induce smaller insulin and glucose excursions than meals of standard fat/carbohydrate content, again implicating insulin and glucose in the nutritional regulation of leptin production.

The regulation of leptin production by glucose and insulin has been directly demonstrated using the euglycemic–hyperinsulinemic clamp. Serum leptin is elevated by the end of prolonged clamps (9 h) at physiologic insulin or within 4–8 h with supraphysiologic insulin concentrations in humans. The synthesis of hexosamines within the adipose tissue is one link between insulin-stimulated glucose uptake/metabolism and leptin production. In rodents, infusion of glucosamine, uridine, or free fatty acids during a 3 h euglycemic–hyperinsulinemic clamp increased muscle uridine diphosphate-N-acetylglucosamine (UDP-GlcNAc), muscle *Lep* gene expression, and serum leptin, compared to saline-infused controls clamped under identical conditions. In transgenic mice overexpressing glutamine:fructose-6-phosphate amidotransferase (GFAT), the rate-limiting enzyme in hexosamine biosynthesis, serum leptin and adipose tissue *Lep* gene expression is significantly increased. In subcutaneous adipose tissue of obese humans, UDP-GlcNAc is increased 3.2-fold compared to lean subjects and there is a significant positive relationship between adipose tissue UDP-GlcNAc and body mass index. An increase in hexosamine biosynthesis in cultured human subcutaneous adipocytes increases leptin release and inhibition of GFAT activity reduces glucose-stimulated leptin release.

## The Sympathetic Nervous System Inhibits Leptin Release from Adipocytes

Leptin-induced activation of the sympathetic nervous system has been postulated to be a negative feedback signal for leptin synthesis. As a mimic of the activation of the sympathetic nervous system, isoproterenol infusion in human subjects significantly reduces serum leptin. This finding is in agreement with animal data indicating that catecholamines inhibit leptin production, although the mechanism through which this occurs has not been elucidated. *In vitro*, catecholamines

inhibit leptin production from 3T3-L1 cells and cultured human adipose tissue pieces.

## LEPTIN AND BODY WEIGHT REGULATION

### The Leptin Receptor

The hypothalamic leptin receptor (Ob-Rb) is a member of the class I cytokine receptor family. As with other cytokine receptors, binding of leptin to its receptor activates an associated janus kinase (JAK), which phosphorylates tyrosines within the leptin receptor itself and signal transducer and activator of transcription (STAT) proteins associated with the receptor. Phosphorylated STAT proteins dimerize and translocate to the nucleus to initiate gene transcription. STAT-3 is the major STAT protein activated by the hypothalamic leptin receptor in mice. Binding of leptin to its receptor has also been reported to activate phosphatidylinositol-3-kinase, phosphodiesterase-3B, and members of the mitogen-activated protein kinase superfamily. Low-level expression of Ob-Rb has been detected in cells other than those in the hypothalamus and it has been shown that leptin activates AMP-activated protein kinase in skeletal muscle.

In addition to the hypothalamic leptin receptor Ob-Rb (long receptor isoform) described above, five other leptin receptors have been identified, all encoded by alternative splicing of the *Lep* gene. The extracellular domain of all leptin receptors is identical; however, the intracellular domains terminate at different lengths. Due to the truncated intracellular domain, the short leptin receptors lack the box 2 motif necessary for STAT protein interaction and these leptin receptors do not phosphorylate STATs. Ob-Ra is the most abundant and widely distributed of the short receptor isoforms and has been found on most cell types in the body. Ob-Ra is highly expressed in the choroid plexus where it functions to transport leptin across the blood–cerebrospinal fluid barrier. The Ob-Re receptor isoform, lacking an intracellular domain and a transmembrane domain, circulates in the blood as a leptin-binding protein, prolonging the half-life of the hormone.

### Leptin-Activated Central Neural Pathways

Leptin receptor is detectable in many areas of the CNS but is highly expressed in the arcuate, paraventricular, ventromedial, and dorsomedial nuclei of the hypothalamus, which is responsible for integration of the numerous neural signals regulating energy intake.

Leptin is detectable in cerebrospinal fluid but leptin likely accesses the arcuate nucleus directly from the blood within the median eminence, which lacks the tight junctions characteristic of the blood–brain barrier. In the arcuate nucleus, leptin reduces the expression of neuropeptides that stimulate food intake and increases expression of neuropeptides that inhibit feeding. As such, leptin suppresses neuropeptide Y and agouti-related peptide expression in one type of neuron and induces the expression of  $\alpha$ -melanocyte-stimulating hormone in proopiomelanocortin neurons. These neurons innervate other nuclei within the hypothalamus and also communicate with other areas of the brain. Leptin-binding neurons in the arcuate nucleus also mediate the effects of leptin on the sympathetic nervous system and hypothalamic pituitary axes.

### Leptin Induces Weight Loss in Mice

The role of leptin in regulation of body weight was first demonstrated in C57BL/6J *ob/ob* mice, which have no detectable serum leptin due to a T-to-C substitution in codon 105 of the *Lep* gene that changes arginine to a premature stop codon. The defective protein is not released by the adipocyte. In *ob/ob* mice, the lack of leptin causes hyperphagia, cold intolerance, and morbid obesity. Daily intraperitoneal injection of recombinant leptin resulted in a dose- and time-dependent decrease in body weight of *ob/ob* mice. Weight loss was mediated by an increase in metabolic rate and locomotor activity, as well as a reduction in food intake. Leptin also reduced body weight in lean C57BL/6J heterozygotes (+/?) and lean wild-type mice, although much higher doses of leptin were needed. In lean mice, leptin administration had no effect on body temperature or locomotor activity but did reduce food intake. Leptin treatment also lowered insulin and glucose in *ob/ob* mice. Since these initial experiments, many investigators have observed that leptin, either peripherally or centrally administered, effectively reduces food intake to stimulate weight loss in most animal models of obesity, with the exception of those with inactivating leptin receptor mutations.

C57BL/KsJ *db/db* mice are hyperphagic, morbidly obese, and diabetic due to a mutation in the hypothalamic leptin receptor gene. A nucleotide substitution (G to T) generates a new splice donor within an exon of the short receptor (Ob-Ra). This new splice donor competes with the upstream splice donor of the long form of the receptor, resulting in alternative splicing and the insertion of a premature termination signal in

the long form of the receptor. The truncated long receptor protein is unable to associate with STAT proteins, interrupting JAK/STAT signaling and resulting in obesity in *db/db* mice. Exogenous leptin administration to C57BL/6J *db/db* mice does not result in weight loss at any concentration tested. A different leptin receptor mutation was identified in Zucker *fatty* (*fa/fa*) rats. An A-to-C substitution changes glutamine-269 to a proline in the extracellular domain of the receptor, reducing the amount of receptor on the cell surface and its capacity to signal. In obese Koletsky rats, a third receptor mutation, T2349A in codon 763 of the leptin receptor gene, creates a premature stop codon just before the transmembrane domain. Leptin treatment does not induce weight loss in either Zucker *fatty* rats or Koletsky rats.

## LEPTIN THERAPY FOR WEIGHT LOSS IN HUMANS

### Leptin and Leptin Receptor Mutations Cause Obesity in Humans

Mutations in the human *LEP* gene were first identified in children of a consanguineous Pakistani family. In these subjects, deletion of a guanine nucleotide in codon 133 results in a frameshift and a truncated leptin protein, which is degraded by the proteasome and not released into the blood. A third child from an unrelated family (at least five generations back) has been found to have the same *LEP* gene mutation. The lack of leptin in these children causes hyperphagia and extreme early onset morbid obesity. A second *LEP* gene mutation, a C-to-T substitution in the first base of codon 105 changing arginine to tryptophan, was identified in a Turkish family. In three subjects homozygous for the mutation, serum leptin was very low but detectable. The low leptin in these subjects results in hyperphagia, morbid obesity, and hypogonadism. One subject also exhibited low sympathetic tone.

As with mutations in the *LEP* gene, mutations in the leptin receptor in humans are rare. A G-to-A substitution in the splice donor site of exon 16 has been identified in three sisters of a Kabilian family. This mutation results in a truncated leptin receptor lacking the transmembrane and intracellular domains. The mutant extracellular domain is present at high concentrations in the blood; acting as a leptin-binding protein to greatly increase serum leptin concentrations. This leptin receptor mutation results in early onset morbid obesity, hyperphagia, and hypothalamic hypogonadism.

### Leptin Induces Weight Loss in Humans with *LEP* Mutations

Farooqi and colleagues have demonstrated that subcutaneous administration of recombinant leptin is an effective long-term therapy for obesity in the three children lacking leptin due to *LEP* gene mutations. With treatment for up to 4 years (0.01 mg/kg lean body weight), leptin therapy induced significant weight loss via a reduction in caloric intake. Weight loss in these children was associated with a reduction in insulin, serum triglycerides, and cholesterol. Leptin treatment also improved T cell proliferation and increased CD4<sup>+</sup> T cell number and function, which was severely impaired in the absence of endogenous leptin production. Adults with congenital leptin deficiency or leptin receptor mutations are prepubertal. Leptin replacement was permissive for the progression through puberty of the oldest child (treatment started at age 10 years) but has not induced precocious puberty in the younger children (3 and 4.5 years old) during the first year of treatment.

### Leptin Resistance

Inactivating mutations in the leptin and leptin receptor gene in humans are rare and not the cause of obesity in the general population. Rather, it has been observed that serum leptin levels are elevated in obese subjects. This observation has led to the hypothesis that obese individuals are “leptin resistant”; i.e., these subjects do not respond to the weight-reducing actions of leptin. Leptin resistance is dependent on the premise that the major function of leptin is to oppose increases in body weight. Flier and colleagues have noted that from an evolutionary perspective, it is difficult to conceive of a mechanism that would limit food intake in times of excess. Therefore, they hypothesize that the major function of leptin is to signal the reduction in energy intake associated with fasting rather than the excess storage of energy in times of food abundance. Evidence for this proposal is derived from the observation that the reduction in leptin with fasting is associated with metabolic, hormonal, and behavioral changes promoting energy conservation and increased energy intake. Leptin administration during fasting prevents these adaptations to food restriction. The concept of leptin resistance is compatible with the role of leptin as a signal of energy deprivation. It is conceivable that the central nervous system in obese individuals does not properly receive or process the leptin signal generated by the adipose tissue, despite elevated serum leptin levels. This would

result in metabolic adaptations to conserve energy and increase energy intake. Several explanations for leptin resistance have been put forth.

High-fat feeding in rodents decreases the effectiveness of intraperitoneal or intravenously injected leptin to induce weight loss, although the response to leptin administered directly to the brain is maintained. Therefore, high-fat feeding appears to cause a defect in access of leptin to the central nervous system. An alternative interpretation is that the palatability of the high-fat diet provides a neural signal for increased consumption that is stronger than the signal for decreased food intake provided by intraperitoneally or intravenously administered leptin. The fact that central leptin administration could overcome this palatability signal may be the result of activation of additional neural pathways that inhibit food intake with central injection.

Suppressors of cytokine signaling (SOCS) are a group of genes rapidly activated following STAT binding in the nucleus. SOCS proteins act in a negative feedback loop to limit intracellular cytokine signaling. SOCS-3 is a potent inhibitor of leptin receptor-initiated JAK/STAT signal transduction in cultured cell lines. Leptin also induces SOCS-3 mRNA expression in the hypothalamus. SOCS-3 message is increased in leptin-binding neurons in the hypothalamus of Agouti mice. Agouti mice are resistant to central leptin administration, suggesting that the increased SOCS-3 expression could cause leptin resistance in these animals. The role of SOCS-3 in leptin resistance in obese humans is not known.

### Leptin Therapy in Obese Adults without Leptin Deficiency

Modest leptin-induced weight loss has been achieved in obese subjects with normal adipose leptin production. In a study conducted by Heymsfield and associates, 47 obese subjects underwent 24 weeks of twice daily injections of leptin (0.1–0.3 mg/kg body weight) or placebo. There was statistically significant weight loss with the highest dose of leptin tested (mean  $\pm$  SD  $7.1 \pm 8.5$  kg,  $n = 8$ ). It is important to note that the dose of leptin needed to induce weight loss in leptin-replete subjects is greater than 10 times that needed in the children with *LEP* gene defects. As discussed above, serum leptin concentrations increase in parallel with body fat in all obese individuals who do not have a *LEP* gene mutation. This apparent “resistance” to the weight-reducing effects of endogenously produced leptin suggests that exogenous leptin administration would be ineffective in decreasing body

weight. The demonstration that leptin therapy can induce weight loss in leptin-replete subjects suggests that leptin resistance is relative and can be overcome with high-dose treatment. However, whether these subjects would become resistant to the weight loss effects of the therapeutic dose of leptin over longer treatment periods is not known.

As discussed above, Flier and colleagues suggest that the major role of leptin is to decrease with prolonged fasting/starvation and initiate processes to conserve energy and maintain body weight. Therefore, the reduction in leptin with caloric restriction during a diet would be counterproductive to the attempt to lose weight. In a controlled study in which subjects lost 10% of body weight, circulating triiodothyronine and thyroxine levels and total daily energy expenditure were reduced. Administration of recombinant leptin to these subjects normalized thyroid hormone levels and increased total energy expenditure. These observations support a role for leptin as opposing weight loss and suggest that leptin therapy may be a useful augmentation to weight loss induced by diet and exercise or other pharmacologic treatment.

### EFFECT OF LEPTIN ON OTHER PHYSIOLOGIC PROCESSES

In addition to its role in body weight regulation, leptin has been implicated as a regulatory signal in a variety of other physiological processes. Many of these functions of leptin are mediated through the central nervous system, although the presence of leptin receptors on numerous nonneural cells suggests that leptin may have direct effects on these cells and tissues.

### Leptin, Glucose Homeostasis, and Lipid Metabolism

Leptin improves insulin sensitivity directly and independently of its effects on food intake and body weight. In rodents, insulin and glucose are reduced shortly after the initiation of leptin treatment, before significant reductions in body fat content are observed. With long-term leptin administration resulting in weight loss, insulin and glucose are reduced to a greater extent than in pair-fed controls. Intracerebroventricular leptin acutely enhances whole body glucose uptake in mice, suggesting that leptin influences insulin sensitivity through central mechanisms. Leptin therapy also reverses insulin resistance and diabetes in two different transgenic models of lipodystrophy through mechanisms not related to reductions in food intake.

Leptin increases lipolysis in white adipose tissue both *in vivo* and *in vitro*. This effect of leptin results from centrally mediated activation of the sympathetic nervous system as well as a direct effect of leptin on adipocytes. Hyperleptinemia in rats does not result in the release of free fatty acids into the blood but rather the oxidation of fatty acids within adipocytes. Leptin has also been observed to stimulate fatty acid oxidation in rodent muscle via activation of AMP-activated protein kinase.

Humans with congenital lipodystrophy have low to undetectable serum leptin levels due to the lack of subcutaneous adipose tissue. Leptin replacement therapy improves glycemic control and decreases triglyceride levels, resting metabolic rate, and self-reported daily food intake.

### Leptin and Reproductive Capacity

Successful reproduction requires an extensive investment of energy. Leptin serves as a permissive signal that energy stores are sufficient for the increased demand for energy during reproduction. *Ob/ob* mice lacking leptin as a signal of energy stores are infertile. Leptin treatment reverses infertility in *ob/ob* mice and accelerates the onset of puberty in normal mice. During starvation, when endogenous leptin levels in mice are low, exogenous leptin maintains LH release and estrous cycle periodicity. In humans, the age of menarche is inversely related to serum leptin concentrations. Humans with *LEP* gene defects have hypothalamic hypogonadism and remain prepubertal. Leptin therapy in these subjects increases gonadotropin secretion and the progression through puberty. Overall, these observations suggest that leptin is a coordinating signal to the central nervous system that energy stores are sufficient to support the higher energy needs associated with reproduction.

Maternal serum leptin increases during pregnancy in both rodents and humans and this increase is not due solely to the increase in fat mass. Human placenta synthesizes leptin, which contributes to the elevation in maternal leptin levels during pregnancy. The role of increased leptin during pregnancy is not yet known. It is possible that the increase in leptin might regulate maternal–fetal energy metabolism or that leptin may act in a paracrine/autocrine manner on placental or fetal growth.

### Leptin and Immune Function

Nutritional deprivation suppresses immune function and malnutrition predisposes to death from infectious

diseases. Leptin appears to be a coordinating signal for activation/inactivation of the immune response during times of energy excess or deprivation. Leptin administration prevents the starvation-induced reduction in immune function in mice. Leptin also stimulates the proliferation of T cells *in vitro*, an effect mediated by direct interaction of leptin with leptin receptor isoform Ob-Rb expressed on the T cell. The number of CD4<sup>+</sup> T cells and their function was severely impaired in children lacking endogenous leptin due to *LEP* gene mutations. Leptin treatment improved T cell proliferation and activity in these children.

### Leptin and Bone Formation

Obesity is protective against bone loss and fractures. A role for leptin as mediator of this protective effect is supported by the observations that leptin administration reduces ovariectomy-induced bone loss in rats and increases bone mineral content and bone density in *ob/ob* mice. That these effects are the result of a direct effect of leptin on bone is inferred from observations that leptin promotes differentiation of bone marrow stromal cells to osteoblasts and promotes mineralization of these cells. Leptin also inhibits osteoclast generation from blood mononuclear cells. In contrast to observations that leptin promotes bone formation, leptin-deficient *ob/ob* and leptin-resistant *db/db* mice have significantly greater bone mass than wild-type littermates. Intracerebroventricular administration of leptin to *ob/ob* mice results in bone loss in these animals. A unifying hypothesis for these disparate findings suggests that the central effects of leptin to decrease bone formation are offset by its peripheral effects to promote bone growth. The importance of this finding is that during periods of starvation, low leptin levels result in suppression of the hypothalamic–pituitary–adrenal and hypothalamic–pituitary–gonadal axes and bone would be lost. However, the low leptin levels in the brain then also activate pathways that protect the skeleton during starvation. The mechanisms by which the hypothalamus regulates bone formation have not yet been elucidated.

### Acknowledgments

Work by the author cited in this article was supported in part by grants from the National Institutes of Health (DK51140), the American Diabetes Association, and the Showalter Trust.

### See Also the Following Articles

Adipocytokines • Alzheimer's Disease and Hormones • Gonadotropin-Releasing Hormone Deficiency, Congenital

Isolated • Obesity and Diabetes, Regulation of Food Intake  
 • Obesity Regulation • Obesity, Treatment of

## Further Reading

- Ahima, R. S., and Flier, J. S. (2000). Leptin. *Annu. Rev. Physiol.* **62**, 413–437.
- Considine, R. V. (2001). Regulation of leptin production. *Rev. Endocr. Metab. Disorders* **2**, 357–363.
- Elmquist, J. K., Elias, C. R., and Saper, C. B. (1999). From lesions to leptin: Hypothalamic control of food intake and body weight. *Neuron* **22**, 221–232.
- Farooqi, I. S., Matarese, G., Lord, G. M., Keogh, J. M., Lawrence, E., Agwu, C., Sanna, V., Jebb, S. A., Perna, F., Fontana, S., Lechler, R. I., DePaoli, A. M., and O’Rahilly, S. (2002). Beneficial effects of leptin on obesity, T cell hyporesponsiveness, and neuroendocrine/metabolic dysfunction of human congenital leptin deficiency. *J. Clin. Invest.* **110**, 1093–1103.
- Flier, J. S. (1998). Clinical Review 94. What’s in a name? In search of leptin’s physiologic role. *J. Clin. Endocrinol. Metab.* **83**, 1407–1413.
- Friedman, J. M. (2002). The function of leptin in nutrition, weight, and physiology. *Nutr. Rev.* **60**, S1–S14.
- Henson, M. C., and Castracane, V. D. (eds.) (2003). “Leptin and Reproduction.” Kluwer Academic/Plenum, New York.
- Heymsfield, S. B., Greenberg, A. S., Fujioka, K., Dixon, R. M., Kusher, R., Hunt, T., Lubina, J. A., Patane, J., Self, B., Hunt, P., and McCamish, M. (1999). Recombinant leptin for weight loss in obese and lean adults. *J. Am. Med. Assoc.* **282**, 1568–1575.
- Khosla, S. (2002). Leptin—Central or peripheral to the regulation of bone metabolism? *Endocrinology* **143**, 4161–4164.
- Lord, G. (2002). Role of leptin in immunology. *Nutr. Rev.* **60**, S35–S38.
- Oral, E. A., Simha, V., Ruiz, E., Andewelt, A., Premkumar, A., Snell, P., Wagner, A. J., DePaoli, A. M., Reitman, M. L., Taylor, S. I., Gorden, P., and Garg, A. (2002). Leptin-replacement therapy for lipodystrophy. *N. Engl. J. Med.* **346**, 570–578.
- Rayner, D. V., and Trayhurn, P. (2001). Regulation of leptin production: Sympathetic nervous system interactions. *J. Mol. Med.* **79**, 8–20.
- Rosenbaum, M., and Leibel, R. L. (1999). Clinical Review 107. Role of gonadal steroids in the sexual dimorphisms in body composition and circulating concentrations of leptin. *J. Clin. Endocrinol. Metab.* **84**, 1784–1789.
- Rosenbaum, M., Murphy, E. M., Heymsfield, S. B., Matthews, D. E., and Leibel, R. L. (2002). Low dose leptin administration reverses effects of sustained weight-reduction on energy expenditure and circulating concentrations of thyroid hormones. *J. Clin. Endocrinol. Metab.* **87**, 2391–2394.
- Tartaglia, L. A. (1997). The leptin receptor. *J. Biol. Chem.* **272**, 6093–6096.
- Zhang, Y., Proenca, R., Maffei, M., Barone, M., Leopold, L., and Friedman, J. M. (1994). Positional cloning of the mouse *obese* gene and its human homologue. *Nature* **372**, 425–432.



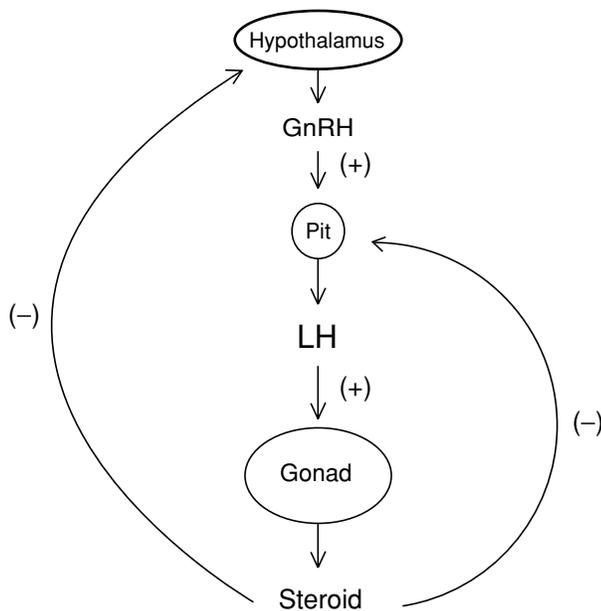
## Léri-Weill Syndrome

see SHOX Disorders



## Leydig Cell Pathology

see Androgen Insensitivity Syndrome



**Figure 1** This schematic is a general introduction to the concept of the regulation of LH secretion and action. The hypothalamus secretes GnRH, which acts on the pituitary gland (Pit) to stimulate (+) the secretion of LH. LH stimulates (+) the gonad to secrete steroids that act on the hypothalamus (right-hand side of figure), pituitary gland (left-hand side of figure), or both to inhibit (-) the secretion of LH.

are composed of cells that exclusively synthesize either FSH or LH. The gonadotropes are stimulated by the hypothalamic polypeptide, GnRH. Only one hypothalamic secretagogue stimulating the synthesis and secretion of both gonadotropins has been identified, although two releasing hormones, one for each gonadotropin, were initially postulated; the term gonadotropin-releasing hormone, rather than luteinizing hormone-releasing hormone, is used throughout this article. The secretion of this hypothalamic secretagogue is episodic and regular. GnRH is released by neurons that terminate on the blood vessels running along the pituitary stalk. These vessels constitute a specialized form of the circulatory system called a portal system. Portal systems are characterized as special veins that arise in one organ system and terminate in another before returning to the heart. They carry the pulses of GnRH to the gonadotropes in the pituitary gland and stimulate the synthesis and release of LH. Accordingly, LH is also released from the pituitary gland in pulses. Changing the pattern of the GnRH pulses results in concurrent changes in the pattern of pulsatile secretion of LH. GnRH binds to a G protein-coupled membrane receptor on the surface of the gonadotrope, activating the  $G_{q/11}$

proteins, which, in turn, stimulate phospholipase C. This enzyme generates inositol 1,4,5-triphosphate and diacylglycerol. The increased levels of these intracellular messengers activate protein kinase C (PKC) and increase intracellular calcium also. PKC and calcium regulate the expression of the  $\beta$ -subunit gene, whereas the secretion of LH is regulated by the increase in intracellular calcium.

During pregnancy, the placenta in equines and primates synthesizes and secretes an LH-like gonadotropin, chorionic gonadotropin (CG), that has more oligosaccharide moieties and a longer amino acid chain in its  $\beta$ -subunit. These modifications confer a longer half-life on CG in the circulation, leading to a more sustained action. Human CG has been used both diagnostically and therapeutically to investigate the action of LH.

## RECEPTOR AND RECEPTOR DEFECTS

The receptor for LH has been described in several mammalian species and is found predominantly in the gonads of each sex. Although LH receptors, identical to those of the gonads, have been found in tumors as well as normal nongonadal tissues, the function of these nongonadal receptors is unclear and a fuller discussion must await future investigation. The LH receptor (85–95 kDa) is a member of the family of G protein-coupled receptors; like all members of this receptor family, it is composed of three domains: a transmembrane domain, an extracellular domain, and an intracellular domain. The transmembrane domain is composed of seven hydrophobic stretches of 20–25 amino acids that form  $\alpha$ -helices traversing the cellular membrane alternating between intracellular and extracellular loops. The extracellular domain of the receptor binds LH, which, in turn, causes the intracellular domain to activate the  $G_s$  protein. This is the first step in an intracellular cascade that continues with the activation of the enzyme adenylyl cyclase and the production of cyclic AMP (cAMP). In turn, protein kinase A (PKA) is activated and phosphorylation of the transcription factor CREB (cAMP regulatory element-binding protein) leads to the expression of genes coding for several factors regulating the cellular functions of proliferation, differentiation, and survival. This pathway has been expanded to include various isoforms of adenylyl cyclase, phosphodiesterase, and anchoring proteins of PKA. Other transcription factors, e.g., stimulatory protein 1, upstream stimulatory factor, and estrogen receptor  $\alpha$  and  $\beta$ , have been shown to be activated by high intracellular

concentrations of cAMP. In addition, CREB can be phosphorylated by kinases other than PKA.

The gene encoding the receptor for LH contains 11 exons, and 13 splice variants have been identified. The last exon encodes the transmembrane and intracellular G protein-activating domains. The extracellular LH-binding domain has much of the heterogeneity of the protein. Several mutations of the LH receptor have been identified and the mutations result either in constitutive activation or in inactivation. Activating mutations of the LH receptor in boys results in precocious puberty observed at 1 to 4 years of age. Activating mutations in girls do not lead to precocious puberty because the gonadotropin FSH is necessary to stimulate folliculogenesis. Mutations that completely inactivate the LH receptor in boys result in genetically male infants displaying female external genitalia but having abdominal testes and lacking Müllerian structures. Mutations that do not completely inactivate the LH receptor in boys, however, result in male infants with micropenis sometimes associated with hypospadias, but the intra-abdominal testes have few, if any, androgen-secreting Leydig cells. In girls, inactivating mutations of the LH receptor, either complete or incomplete, result in normal primary and secondary female characteristics, but also result in amenorrhea because LH is required for folliculogenesis.

## ACTION IN THE OVARY

Before discussing the role of LH in the ovary, a description of folliculogenesis is necessary. During fetal development of the female mammal, the primitive germ cells in the putative ovary enter meiosis and arrest in the prophase of meiosis I. Each oocyte is surrounded by a single layer of granulosa cells, and soon after formation, these primordial follicles enter a resting phase. The mechanisms that stimulate the primordial follicles to initiate folliculogenesis are unknown, but the process begins at once and continues until all follicles are depleted from the ovary. During fetal life in primates and ruminants and within the first 2 weeks after birth in rodents, some of the primordial follicles are activated and become primary follicles. Two layers of cells develop around the oocyte: the outermost are the theca cells that are separated from the innermost layers of granulosa cells by a basement membrane.

The rupture of the follicle, along with the release of the oocyte and some of the granulosa cells surrounding it, is called ovulation. During the periovulatory period, anti-inflammatory factors are produced and appear

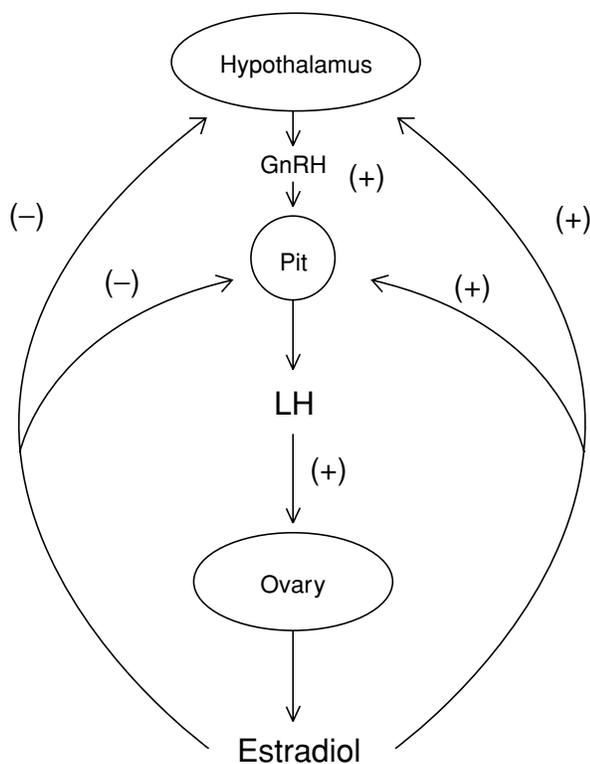
to aid in rapid healing and vascularization of the ruptured follicle. The theca and granulosa cells remaining in the ruptured follicle differentiate into luteal cells, forming the corpus luteum, which characterizes the luteal phase of the ovarian cycle. If no pregnancy occurs, after a finite time specific to each species, the corpus luteum dies, marking the end of the luteal phase. The rapid involution of the corpus luteum is called luteolysis and is regulated by specific factors. Luteolysis is necessary before a subsequent ovarian cycle can begin. The follicular phase, ovulation, the luteal phase, and luteolysis constitute the ovarian cycle and LH plays some role in all phases of the cycle.

LH has three functions in the ovary: (1) stimulation of androgen synthesis during the follicular phase; (2) stimulation of ovulation; and (3), following ovulation, differentiation of the thecal and granulosa cell remnants of the ruptured follicle into luteinized cells. The luteal cells constitute the steroidogenic cells of the corpus luteum, shifting from predominantly estrogen-producing cells to progesterone-producing cells.

LH binds to its receptor on the theca cells of the follicle and stimulates the synthesis of androgens. Steroidogenesis begins with the movement of cholesterol from its storage site on the outer mitochondrial membrane to the inner mitochondrial membrane, where the cytochrome P450-associated enzyme removes the side chain of cholesterol, producing pregnenolone. This mobilization of cholesterol is the rate-limiting step in steroidogenesis and LH regulates the production of a protein called steroidogenic acute regulatory (StAR) protein. The tertiary structure of this 30 kDa protein has a domain containing a cholesterol-binding hydrophobic tunnel that may permit StAR to shuttle cholesterol across the mitochondrial membranes. The steroids are lipid-soluble, unlike cholesterol, and the steroidogenic enzymes are readily accessible for production of the androgen, androstenedione. This androgen diffuses out of the thecal cells and into the granulosa cells, where the enzyme aromatase, stimulated by FSH action, converts androstenedione into estradiol. As the theca and granulosa cells of the maturing follicle proliferate, a greater amount of estradiol is produced. The estradiol, as stated earlier, inhibits the pituitary secretion of LH, but the number of LH receptors on each granulosa cell is also increased under FSH stimulation. An increase in hormonal receptors is generally associated with an increase in sensitivity to the hormone; that is, the cell can detect and respond to lower concentrations of the hormone. With the increased number of LH receptors on the theca and granulosa

cells, only those follicles with sufficient gonadotropin stimulation will continue folliculogenesis. Once the circulating estradiol produced by the antral follicle is elevated, this steroid causes an abrupt release, or surge, of LH from the pituitary (see Fig. 2). The surge of LH causes ovulation by increasing the expression of PKA and increasing the inositol lipid hydrolysis that activates PKC, shifting the pattern of gene expression in the granulosa cells. The shift results in the suppression of cell division of the granulosa cells, utilization of these cells, signaling of the oocyte to reenter meiosis, and rupture of the follicle wall.

After ovulation, the luteinized granulosa and theca cells remaining in the ruptured follicle begin to



**Figure 2** The feedback mechanism regulating LH in the female mammal. The hypothalamus secretes GnRH, which acts on the pituitary gland to stimulate (+) the secretion of LH. This enzyme converts androstenedione, produced by the theca cells of the follicle under LH stimulation, into estradiol. During early folliculogenesis when the growing follicle is small, the low concentrations of estradiol act on the hypothalamus, pituitary gland, or both to inhibit (-) the secretion of LH (left-hand side of figure). Prior to ovulation, the high concentration of estradiol in the blood, due to the increased number of steroidogenic cells in the large antral follicle, causes the sudden release of LH, which causes the follicle to rupture and release the oocyte. This surge of LH is marked by a sudden release of GnRH (+) from the hypothalamus and/or of LH directly (+) from the pituitary (right-hand side of figure). This latter action of estradiol in the female mammal is an example of positive feedback.

synthesize progesterone. Although the luteal cells of some species, e.g., human, are able to produce small amounts of estrogen, the cellular differentiation shifts the steroidogenic pathway in the luteal cell to produce primarily—or in some species only—progesterone. This steroid maintains the endometrium of the uterus in readiness to accept the blastocyst, if fertilization of the oocyte occurs. The corpus luteum under LH regulation continues to survive for a specific period of time, depending on the species. If, on one hand, no blastocyst is formed, then the corpus luteum involutes through programmed cell death. If, on the other hand, implantation of a blastocyst occurs, in the case of primates and horses, chorionic gonadotropin secreted by the blastocyst and later by the placenta binds to the LH receptors on the corpus luteum, rescuing it from luteolysis and maintaining the production of progesterone. This placental gonadotropin in women is called human chorionic gonadotropin (hCG); it has 30 more amino acid residues and is more glycosylated than LH. Because of these differences, hCG has a greater potency and a longer half-life in the circulation than LH itself. The placental CG prevents luteolysis.

In most mammalian species, however, removal of the uterus results in a delay of luteolysis. From studies using the techniques of hysterectomy and vascular anastomosis, the initiation of luteolysis occurs when prostaglandin  $F_{2\alpha}$  ( $PGF_{2\alpha}$ ) from the uterus enters the ovarian artery from the utero-ovarian vein by counter-current exchange. This avoids degradation of  $PGF_{2\alpha}$  by the lungs if it were to be transported by the veins to the heart and then to the lung. Luteolysis is due to an interaction of the uterus and ovary, but the precise trigger for the release of  $PGF_{2\alpha}$  remains unknown.

## ACTION IN THE TESTIS

LH has only one site of action in the adult testis, its receptor on the surface of adult Leydig cells, to which it binds, stimulating the production of the androgenic steroid, testosterone. Before continuing with the effects of testosterone on spermatogenesis, a brief review of the testicular structure and spermatogenesis is necessary. The adult mammalian testis is an ovoid structure covered by a capsule called the tunica albuginea. The parenchyma of the organ is composed of two compartments, the seminiferous tubules and the interstitium between the tubules. The seminiferous tubules contain the epithelium where spermatogenesis takes place. The interstitium contains the Leydig cells, fibroblasts, macrophages, nerves, blood, and lymphatic vessels.

Steroidogenesis begins in the Leydig cell as it does in the theca cell of the ovary with the movement of cholesterol from its storage site on the outer mitochondrial membrane to the inner mitochondrial membrane, where the cytochrome P450-associated enzyme removes the side chain of cholesterol, producing pregnenolone. This mobilization of cholesterol is the rate-limiting step in steroidogenesis and LH regulates the production of StAR protein. The tertiary structure of this 30 kDa protein has a domain containing a cholesterol-binding hydrophobic tunnel that may permit StAR to shuttle cholesterol across the mitochondrial membranes. The steroids are lipid-soluble, unlike cholesterol, and the steroidogenic enzymes are readily accessible for production of the androgen, testosterone. This androgen is a major stimulator of spermatogenesis in the adult male mammal.

Spermatogenesis consists of a series of processes: stem cell renewal, germ cell proliferation, and spermiogenesis. Stem cell renewal is the mechanism that guarantees that a large and undiminishing number of undifferentiated germ cells are continually available for subsequent waves of spermatogenesis throughout the lifetime of the male animal. The spermatogenic stem cells are called type A spermatogonia and, depending on the species, one or more type A spermatogonia have been identified. In addition to renewing their population, the type A spermatogonia also produce more differentiated spermatogonia, often referred to as type B spermatogonia. These latter cells, depending again on species, may produce one or more generations of differentiated spermatogonia. The final generation of differentiated spermatogonia divide and produce the primary spermatocytes. The spermatocytes are the germ cells that undergo meiosis and the products of this specialized cell division are the haploid spermatids. The spermatids enter spermiogenesis, during which the nuclear contents condense, the specialized lysosome, the acrosome, is formed, a tail is produced, and the amount of cytoplasm is reduced. Spermatogenesis requires approximately 2 to 3 weeks to complete, depending on the species of mammal.

Spermatogenesis can be thought of in two ways. First, the series of processes can be viewed as the production of one or more haploid spermatids. In older literature, this is often referred to as the quality of spermatogenesis. Second, the number of spermatids produced is often referred to as quantitative spermatogenesis. This notion is based on the clinical observation that a low number of spermatozoa in the ejaculate of a man is often associated with infertility.

Exactly what that low number is and how it plays a role in fertility is unclear, but it may reflect a need for a large number of spermatozoa to begin the long journey through the female reproductive tract to ensure that enough sperm arrive at the oocyte to effect fertilization.

Before leaving the topic of spermatogenesis and the seminiferous epithelium, the Sertoli cell must be introduced. These cells are the only somatic component of the epithelium and provide the cytoarchitecture of the seminiferous epithelium, nourish the germ cells, and secrete the hormone inhibin B. All hormonal effects on the seminiferous epithelium and on spermatogenesis are mediated by the Sertoli cells. The androgen receptor has been demonstrated conclusively to be in Sertoli cells only and not in germ cells.

Dr. Phillip E. Smith demonstrated the central role of the pituitary to stimulate spermatogenesis in adult rodents and primates nearly 60 years ago by. Surgical removal of the pituitary gland of adult rats led to a decrease in testicular weight and the seminiferous epithelium comprised Sertoli cells and germ cells as mature as spermatids. Hypophysectomy of adult rhesus monkeys resulted in a precipitous decline in testicular size associated with the complete regression of the seminiferous epithelium to the extent that the tissue comprised only Sertoli cells and type A spermatogonia, that is, only stem cells. Further experiments on adult rodents and monkeys have confirmed Dr. Smith's earlier observations. In a study using a GnRH receptor antagonist, which suppresses all gonadotropin secretion, chemical hypophysectomy achieved the same results as observed by Dr. Smith using a surgical hypophysectomy. Namely, the rodent seminiferous epithelium comprised Sertoli cells and germ cells as mature as spermatids, whereas the primate seminiferous epithelium comprised only Sertoli cells and stem spermatogonia.

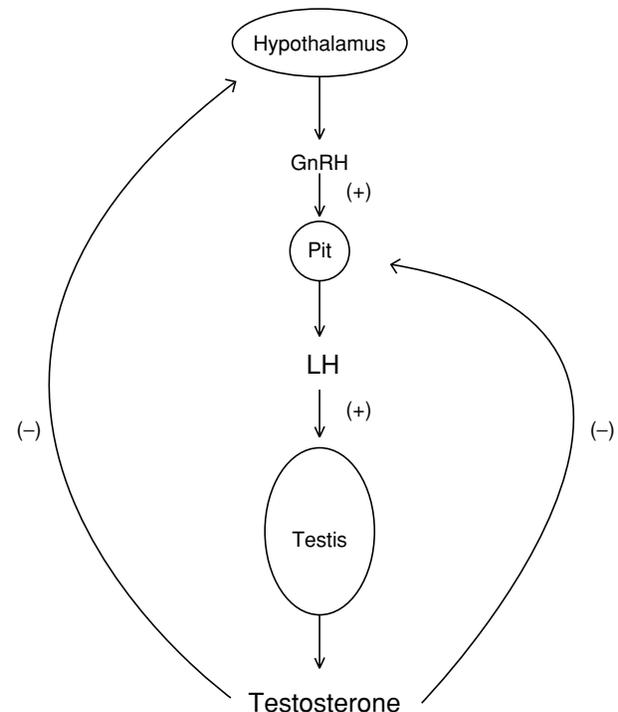
Physiological replacement of testosterone in rats that had been previously rendered hypogonadotropic results in full restoration of the seminiferous epithelium containing the normal number of spermatogenic cells. Testosterone alone is sufficient to stimulate spermatogenesis in this and related species. Replacement of testosterone in hypogonadotropic, hypogonadal rhesus monkeys, however, results in stimulation of testicular growth but to less than normal size. The gonadal growth is due primarily to the stimulation of spermatogenesis by androgen, but morphometric analysis of the seminiferous epithelium of the monkeys revealed that the smaller testicular size was accounted for by a deficit in the numbers of all differentiated, or

type B, spermatogonia. Replacement of FSH in testosterone-treated hypophysectomized adults results in a greater number of all four generations of type B spermatogonia. These observations led to the conclusion that testosterone alone stimulates spermatogenesis, but FSH is necessary to restore spermatogenesis completely in primates. This action of FSH in primates is posited to be the rescue of the differentiated spermatogonia from programmed cell death.

Testosterone is also secreted into the circulatory system and transported to the male reproductive tract, stimulating growth and differentiation. The androgen also causes the increased body size, muscle mass, and behavior characteristic of male mammals. In the case of humans, beard growth and male pattern baldness are also manifestations of testosterone action. Testosterone acts to reduce the secretion of LH; that is, it exerts a negative feedback on LH secretion by the pituitary (see Fig. 3). One distinct pathway for negative feedback has been described in male mammals. In this pathway, found in male primates and sheep, testosterone, or perhaps one or more of its metabolites (estradiol or dihydrotestosterone), acts in the central nervous system to reduce the frequency of GnRH pulses produced by the hypothalamus; this reduction, in turn, reduces the frequency of LH pulses from the pituitary gland. The details of the nervous system pathways involved in the negative feedback of testosterone on LH remain to be elucidated. Studies in male rodents clearly establish the negative feedback system between LH and testosterone, but little consensus exists on whether androgen or its metabolites act on the central nervous system, the pituitary gland, or both.

## FETAL DEVELOPMENT OF THE MALE REPRODUCTIVE SYSTEM

Fetal development of the male reproductive system requires the presence of a testis capable of secreting two hormones, testosterone and Müllerian inhibiting hormone. The formation of a testis requires the action of a gene product, testis determining factor (TDY). This factor causes the undifferentiated gonad to develop into a testis. If this factor is absent or defective, a testis will not form and a female phenotype will be produced. The fetal testis produces testosterone, which stimulates the Wolffian ducts to differentiate into the male reproductive tract consisting of the epididymis, vas deferens, seminal vesicle, and prostate.



**Figure 3** The negative feedback (-) of the androgen testosterone on LH secretion in male mammals. The hypothalamic factor GnRH stimulates (+) the pituitary gland to synthesize and secrete LH. LH, in turn, stimulates the Leydig cells of the testis to synthesize testosterone, which freely diffuses from the cell into the circulation. Testosterone modulates the frequency of the GnRH pulse from the hypothalamus and this latter event reduces the amount of LH secreted by the pituitary gland (left-hand side of figure). This form of negative feedback has been compellingly demonstrated in primates and sheep. Testosterone may act directly on the pituitary gland to inhibit (-) the secretion of LH (right-hand side of figure).

Testosterone also causes the formation of the penis and scrotum into which the testes descend. Testosterone is produced in the fetal testis by Leydig cells and Müllerian inhibiting hormone is produced by the fetal Sertoli cells.

The stimulus for the fetal Leydig cells to produce testosterone cannot be the fetal pituitary since the secretion of testosterone precedes the effective secretion of LH by the fetal pituitary. In the case of higher primates and equines, the placental chorionic gonadotropin may cause the fetal testis to form fetal Leydig cells as well as stimulate those cells to secrete testosterone in response. The situation in the other mammalian genera, however, must be different because these species do not produce a placental gonadotropin. In rodents, the fetal Leydig cells seem to be autonomous; that is, they synthesize testosterone without the stimulus of a gonadotropin. In all mammalian species, the fetal Leydig cells cease producing

androgens soon after birth and seem to disappear in the interstitium of the testes of newborns.

During sexual maturation at the time of puberty, the adult Leydig cells form from undifferentiated cells, called mesenchymal-like cells, in the interstitial spaces between the seminiferous tubules. In rats, these cells actively proliferate from 2 to 4 weeks after birth, but their numbers diminish, suggesting differentiation into a new type of cell. The interstitial cells that give rise to the adult Leydig cell population have been posited to be part of the population of mesenchymal-like cells that during uterine life produced the fetal Leydig cells. Nonetheless, these mesenchymal cells have a very high mitotic index of rapid and multiple cell divisions. The next generation of differentiated cells express the proteins of the adult Leydig cell,  $3\beta$ -hydroxyl dehydrogenase and LH receptors. These latter cells have been called immature Leydig cells and divide once and produce the adult Leydig cell.

### See Also the Following Articles

Androgens, Gender and Brain Differentiation • Fertilization • FSH (Follicle-Stimulating Hormone) • Gonadotropin-Induced Ovulation • Gonadotropin-Releasing Hormone, Family of • Implantation • Ovarian-Follicular Apparatus • Pituitary Gland Anatomy and Embryology • Pregnancy Endocrinology • Puberty: Physical Activity and Growth • Sexual Maturation, Female • Sexual Maturation, Male • Spermatogenesis, Endocrine Control of • Testes, Embryology of

### Further Reading

Achermann, J. C., Weiss, J., Lee, E.-J., and Jameson, J. L. (2001). Inherited disorders of the gonadotropin hormones. *Mol. Cell. Endocrinol.* **179**, 89–96.

Ando, H., Hew, C. L., and Urano, A. (2001). Signal transduction pathways and transcription factors involved in the

gonadotropin-releasing hormone-stimulated gonadotropin subunit gene expression. *Comp. Biochem. Physiol. Part B Biochem. Mol. Biol.* **129**, 525–532.

Ascoli, M., Fanelli, F., and Segaloff, D. L. (2002). The lutopin/choriogonadotropin receptor: A 2002 perspective. *Endocr. Rev.* **23**, 141–174.

Benton, L., Shan, L.-X., and Hardy, M. P. (1995). Differentiation of adult Leydig cells. *J. Steroid Biochem. Mol. Biol.* **53**, 61–68.

De Roux, N., and Milgrom, E. (2001). Inherited disorders of GnRH and gonadotropin receptors. *Mol. Cell. Endocrinol.* **179**, 83–87.

Fimia, G. M., Morlon, A., Macho, B., de Cesare, D., and Sassone-Corsi, P. (2001). Transcriptional cascades during spermatogenesis: Pivotal role of CREM and ACT. *Mol. Cell. Endocrinol.* **179**, 17–23.

Fortune, J. E., Cushman, R. A., Wahl, C. M., and Kito, S. (2000). The primordial to primary follicle transition. *Mol. Cell. Endocrinol.* **163**, 53–60.

Habert, R., Lejune, H., and Saez, J. M. (2001). Origin, differentiation and regulation of fetal and adult Leydig cells. *Mol. Cell. Endocrinol.* **179**, 47–74.

Hillier, S. G. (2001). Gonadotropic control of ovarian follicular growth and development. *Mol. Cell. Endocrinol.* **179**, 39–46.

Kobata, A., and Takeuchi, M. (1999). Structure and function of the N-linked sugar chains of human chorionic gonadotropin. *Biochim. Biophys. Acta* **1455**, 315–326.

Mendis-Handagama, S. M. I. C., and Ariyaratne, H. B. S. (2001). Differentiation of the adult Leydig cell population in the postnatal testis. *Biol. Reprod.* **65**, 660–671.

Niswender, G. D., Juengel, J. L., Silva, P. J., Rollyson, M. K., and McIntush, E. W. (2000). Mechanisms controlling the function and life span of the corpus luteum. *Physiol. Rev.* **80**, 1–29.

Rao, Ch. V. (2001). An overview of the past, present, and future of nongonadal LH/hCG actions in reproductive biology and medicine. *Semin. Reprod. Med.* **19**, 7–17.

Richards, J. S. (2001). New signaling pathways for hormones and cyclic adenosine 3',5'-monophosphate action in endocrine cells. *Mol. Endocrinol.* **15**, 209–218.

Stocco, D. M. (2001). StAR protein and the regulation of steroid hormone synthesis. *Annu. Rev. Physiol.* **63**, 193–213.

Themmen, A. P. N., and Verhoef-Post, M. (2002). LH receptor defects. *Semin. Reprod. Med.* **20**, 199–204.

Tilbrook, A. J., and Clarke, I. J. (2001). Negative feedback regulation of the secretion and actions of gonadotropin-releasing hormone in males. *Biol. Reprod.* **64**, 735–742.



## Liddle's Syndrome

see Mineralocorticoids and Mineralocorticoid Excess Syndromes

reducing arterial tissue cholesterol levels that contribute to such atherosclerosis by facilitating reverse cholesterol transport, HDL inhibits oxidation and aggregation of LDL in the arterial wall. Lower HDL cholesterol may be associated with genetic disorders, nutritional habits (high carbohydrate intake), obesity, hypertriglyceridemia, cigarette smoking, and lack of exercise, whereas HDL levels are increased with alcohol ingestion and estrogens or in familial hyperalphalipoproteinemia, another “longevity” syndrome.

VLDL is a triglyceride-rich lipoprotein synthesized and secreted by the liver. Hypertriglyceridemia is associated with genetic disorders, obesity, diabetes mellitus, renal failure, heavy alcohol intake, and drugs such as estrogens.

The measurement of fasting levels of LDL cholesterol, HDL cholesterol, and triglycerides in the serum is obtained as part of comprehensive assessment of the risk of atherosclerotic vascular disease that is commonly estimated according to the following formula:  $\text{total cholesterol} = \text{LDL cholesterol} + \text{HDL cholesterol} + \text{triglycerides} / 5$ . Hypercholesterolemia is defined as serum total cholesterol of 200 mg/dl or more, according to the National Cholesterol Education Program (NCEP) III guidelines. These guidelines recommend reducing the serum LDL cholesterol to less than 100 mg/dl in persons with coronary artery disease (CAD), peripheral arterial disease (PAD), stroke or transient cerebral ischemic attack, carotid arterial disease, or diabetes mellitus as well as in persons with two or more risk factors (many of whom have the so-called “metabolic syndrome”) who have a 10-year risk of CAD of more than 20% calculated from Framingham Heart Study tables. The NCEP III guidelines further state that persons with two or more risk factors and a 10-year risk of CAD of 10 to 20% should have their serum LDL cholesterol reduced to less than 130 mg/dl. The NCEP III guidelines also state that persons with no risk factors or with one risk factor should have their serum LDL cholesterol reduced to less than 160 mg/dl. A decreased serum HDL cholesterol level of less than 40 mg/dl represents a CAD risk factor. A normal serum triglyceride level, according to NCEP-ATP III guidelines, is less than 150 mg/dl.

The NCEP III guidelines state that other CAD risk factors to tally in a patient’s risk assessment, other than increased serum LDL cholesterol and decreased serum HDL cholesterol, are the following: (1) age (45 years or over for men and 55 years or older for women), (2) cigarette smoking, (3) hypertension (systolic blood pressure of 140 mm Hg or higher, diastolic blood pressure of 90 mm Hg or higher, or treated hypertension),

and (4) a history of premature CAD in a first-degree relative (before 55 years of age in a related man and before 65 years of age in a related woman).

Dyslipidemia includes hypercholesterolemia, increased serum LDL cholesterol, decreased serum HDL cholesterol, and hypertriglyceridemia. Elevated serum total cholesterol, elevated serum LDL cholesterol, and a low-serum HDL cholesterol are risk factors for CAD, PAD, stroke, transient cerebral ischemic attack, carotid arterial disease, heart failure, valvular aortic stenosis, and mitral annular calcium in elderly persons. In general, the higher the serum total cholesterol, the higher the serum LDL cholesterol (not infrequently, however, an elevated LDL may be offset by an increased HDL); the lower the serum HDL cholesterol, the greater the risk of atherosclerotic vascular disease.

Elevated serum triglycerides are associated with an increased risk of atherosclerotic vascular disease. However, except for being a weak independent risk factor for new coronary events in elderly women, hypertriglyceridemia is not an independent risk factor for atherosclerotic vascular disease in elderly or younger men and women (confounded by the inverse relationship between HDL cholesterol and triglyceride levels wherein the low HDL is thought to exert the primary atherogenic effect, a subject of continuing controversy that is further complicated by the high prevalence of the metabolic syndrome in persons at increased CAD risk).

## METABOLIC SYNDROME

The clustering of high serum triglycerides, small dense LDL particles, low serum HDL cholesterol levels, hypertension, insulin resistance (with or without glucose intolerance), and a prothrombotic state is called the metabolic syndrome. This constellation is extremely common in patients with type 2 diabetes and is thought to be a precursor to diabetes in many others who are at increased risk for developing diabetes with aging and the passage of time. Statin drug therapy alone, or in combination with niacin or gemfibrozil, can be used to treat the atherogenic dyslipidemia of the metabolic syndrome.

## DIETARY TREATMENT OF HYPERCHOLESTEROLEMIA

Persons with hypercholesterolemia and CAD or other atherosclerotic vascular disease should be treated with a Step II American Heart Association diet. They should achieve and maintain a desirable weight.

Cholesterol intake should be less than 200 mg/day. Less than 30% of total caloric intake should be fatty acids. Saturated fatty acids should comprise less than 7% of total calories, polyunsaturated fatty acids should account for up to 10% of total calories, and monounsaturated fatty acids should comprise 10 to 15% of total calories. Protein intake should account for 10 to 20% of total calories. Carbohydrates should comprise 50 to 60% of total calories. In addition to loss of weight in obese persons, long-term aerobic exercise, and dietary treatment of hypercholesterolemia, cigarette smoking should be stopped, hypertension should be treated, and diabetes mellitus should be well controlled.

## STATINS

The most effective lipid-lowering drugs in reducing cardiovascular morbidity and mortality in elderly and younger persons are the 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase inhibitors or statin drugs. Statins reduce serum total cholesterol, LDL cholesterol, and triglycerides, and they increase serum HDL cholesterol. The statin drugs include simvastatin, pravastatin, atorvastatin, lovastatin, fluvastatin, and rosuvastatin. Statin drugs suppress cholesterol biosynthesis by competitively inhibiting HMG-CoA reductase, the enzyme that catalyzes the conversion of HMG-CoA to mevalonate, a precursor of sterols (including cholesterol). This action induces up-regulation of LDL receptors in the liver and increased clearance of LDL from the plasma, thereby decreasing plasma cholesterol levels.

Double-blind, randomized, placebo-controlled studies that have included substantial numbers of individuals age 65 years or over have demonstrated that persons with and without CAD treated with statins have a significant reduction in all-cause mortality, nonfatal and fatal coronary events, new or worsening angina pectoris, coronary or noncoronary revascularization, stroke or transient cerebral ischemic attack, intermittent claudication, heart failure, new carotid bruits, one or more new bruits, and hospitalization for cardiovascular events. Individuals both with and without CAD, with PAD, with prior stroke or transient cerebral ischemic attack, with carotid arterial disease, with diabetes mellitus, with heart failure, with treated hypertension, and with valvular aortic stenosis with dyslipidemia have benefited from treatment with statins.

In the Heart Protection Study, where 5806 of the 20,536 men and women randomized to simvastatin or

double-blind placebo were 70 to 80 years of age at study entry and were 75 to 85 years of age at follow-up, 5 years of simvastatin therapy prevented myocardial infarction, stroke, and revascularization in 70 to 100 persons per 1000 treated persons regardless of age, gender, or initial levels of serum lipids. Reduction of serum LDL cholesterol levels by statins to less than 100 mg/dl in elderly persons is more effective at reducing new coronary events and stroke than is reducing serum LDL cholesterol to higher levels in elderly persons.

## OTHER LIPID-LOWERING DRUGS

Other lipid-lowering drugs include the cholesterol absorption inhibitor ezetimibe; bile acid sequestrants such as colestevam, cholestyramine, and colestipol; nicotinic acid; and fibric acid derivatives such as gemfibrozil and fenofibrate. Gemfibrozil may be useful at reducing the incidence of new coronary events and stroke in persons with CAD whose primary serum lipid abnormality is a low-serum HDL cholesterol level. Persons with serum triglycerides of 500 mg/dl or more should be treated with a fibric acid derivative such as gemfibrozil to prevent the development of pancreatitis.

## COMBINATION LIPID-LOWERING DRUG THERAPY

If statin drug therapy at maximal safe and tolerated doses does not lower the LDL cholesterol level to less than 100 mg/dl in persons with CAD, PAD, prior stroke or transient cerebral ischemic attack, carotid arterial disease, abdominal aortic aneurysm, diabetes mellitus, or two or more risk factors with a 10-year risk of CAD of more than 20%, ezetimibe or a bile acid sequestrant such as colestevam should be added to the therapeutic regimen. A combination lovastatin-nicotinic acid formulation also has been approved for clinical use.

## SPECIAL CONSIDERATIONS IN THE MANAGEMENT OF DYSLIPIDEMIA IN THE ELDERLY

The degree to which NCEP III guidelines and the results of clinical trials designed principally to assess the efficacy and safety of hypolipidemic therapy in predominantly younger patients are applicable to the heterogeneous populations of elderly patients (especially those age 75 years or over) remains a subject of controversy and ongoing research. This was noted in

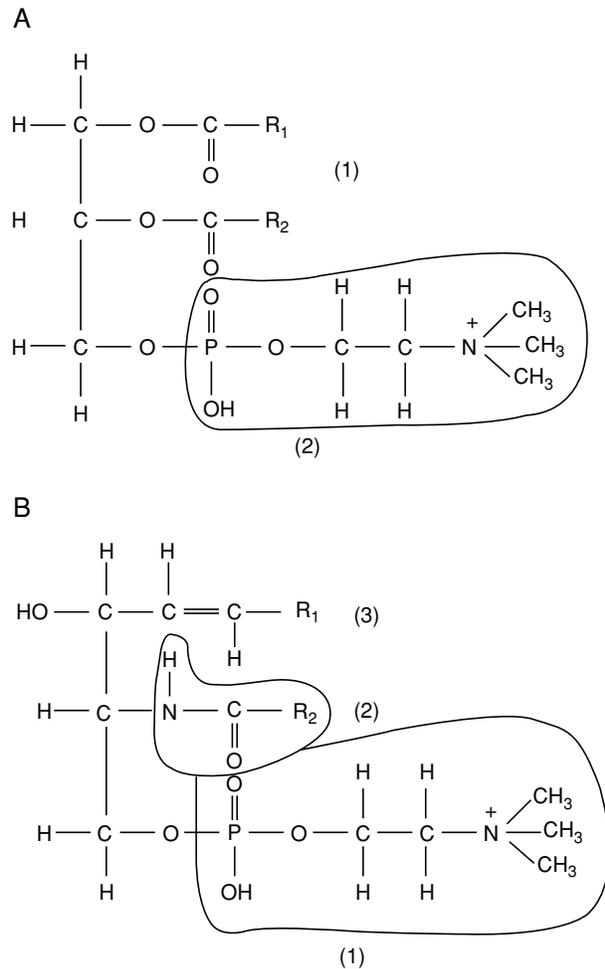
the acknowledgment in the NCEP III guidelines that individualized clinical judgment, with due attention to comorbidities and other medical and social issues, was especially appropriate in the decision-making process in such patients.

### See Also the Following Articles

Atherosclerosis • Familial Low Cholesterol Syndromes, Hypobetalipoproteinemia • Hypercholesterolemias, Familial Defective ApoB (FDB) and LDL Receptor Defects • Hypertriglyceridemia • Low HDL/High HDL Syndromes

### Further Reading

- Aronow, W. S. (2003). Effect of statins on mortality and cardiovascular events in elderly high-risk persons. *J. Am. Geriatr. Soc.* **51**, 717–718.
- Aronow, W. S., and Ahn, C. (2002). Incidence of new coronary events in older persons with prior myocardial infarction and serum low-density lipoprotein cholesterol  $\geq 125$  mg/dl treated with statins versus no lipid-lowering drug. *Am. J. Cardiol.* **89**, 67–69.
- Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. (2001). Executive Summary of the Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). *J. Am. Med. Assoc.* **285**, 2486–2497.
- Heart Protection Study Collaborative Group. (2002). MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20,536 high-risk individuals: A randomised placebo-controlled trial. *Lancet* **360**, 7–22.
- LaRosa, J. C., He, J., and Vupputuri, S. (2002). Effect of statins on risk of coronary disease: A meta-analysis of randomized controlled trials. *J. Am. Med. Assoc.* **282**, 2340–2346.
- Lemaitre, R. N., Psaty, B. M., Heckbert, S. R., *et al.* (2002). Therapy with hydroxymethylglutaryl coenzyme A reductase inhibitors (statins) and associated risk of incident cardiovascular events in older adults: Evidence from the Cardiovascular Health Study. *Arch. Int. Med.* **162**, 1395–1400.
- Sever, P. S., Dahlof, B., Poulter, N. R., *et al.* (2002). Prevention of coronary and stroke events with atorvastatin in hypertensive patients who have average or lower-than-average cholesterol concentrations, in the Anglo-Scandinavian Cardiac Outcomes Trial—Lipid Lowering Arm (ASCOT-LLA): A multicentre randomised controlled trial. *Lancet* **361**, 1149–1158.
- Shepherd, J., Blauw, G. J., Murphy, M. B., *et al.* (2002). Pravastatin in elderly individuals at risk of vascular disease (PROSPER): A randomised controlled trial. *Lancet* **360**, 1623–1630.



**Figure 1** Illustration of phosphatidylcholine and sphingomyelin, the major precursors of lipid second messengers. (A) Phosphatidylcholine: glyceride backbone bound to nonpolar tails containing (1) long chain fatty acyl groups and (2) polar head consisting of negatively charged phosphate group esterified to choline that forms the head alcohol group. (B) Sphingomyelin: (1) polar head group of phosphocholine bound to (2) N-acylated fatty acid linked to (3) sphingosine.  $R_1$  and  $R_2$ , long chain fatty acyl groups.

within the cell (Fig. 2). All of the previously described second messengers derived from the glycerolipid or sphingolipid backbone are discussed individually in the following sections.

## GLYCEROLIPID-DERIVED SECOND MESSENGERS

### Diacylglycerol

Generation of DAG occurs in two phases depending on the source from which it is produced. DAG generated from the breakdown of  $\text{PIP}_2$  by the action of

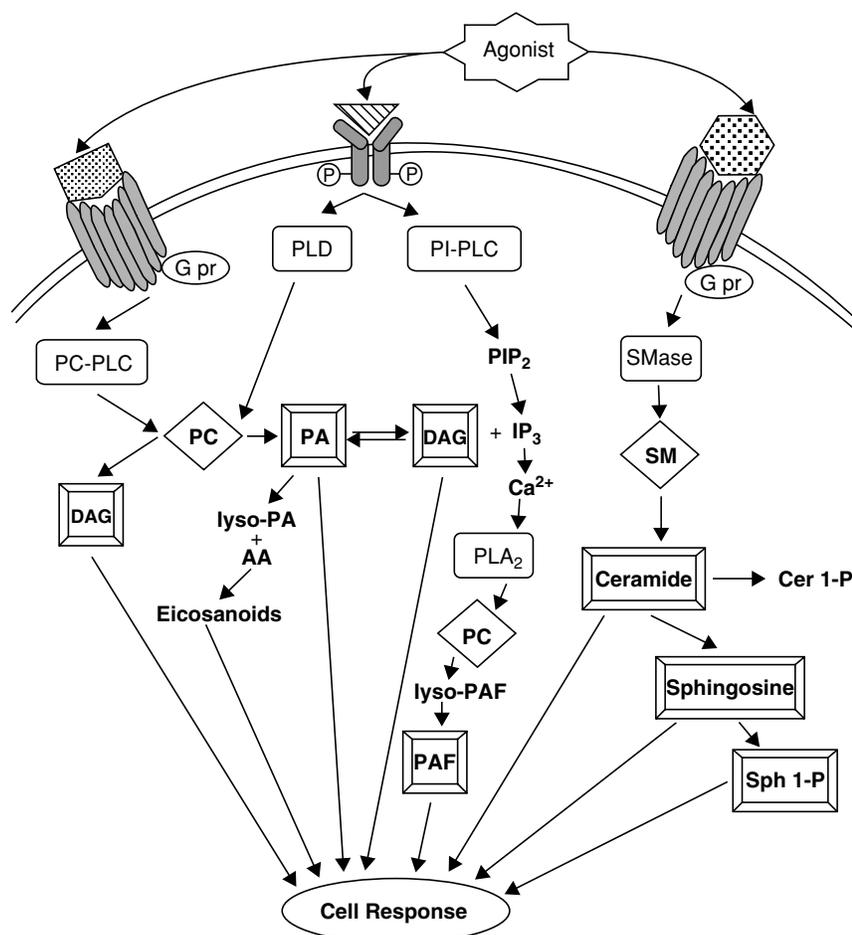
phosphatidylinositol-specific phospholipase C (PI-PLC) is rapid, and the increases in DAG last about 1 min and are accompanied by the release of  $\text{IP}_3$ . The latter causes the release of  $\text{Ca}^{2+}$  from intracellular stores and increases the cytosolic concentration of these cations. As a second messenger, DAG is limited mainly to the activation of protein kinase C. The simultaneous generation of DAG and  $\text{Ca}^{2+}$  leads to the activation of protein kinase C isoforms  $\alpha$ ,  $\beta$ , and  $\gamma$ . The second phase of DAG production is more sustained, takes place without the accompanying release of  $\text{Ca}^{2+}$ , and might be related to the activation of  $\text{Ca}^{2+}$ -independent isoforms of protein kinase C. This increase of DAG results directly from PC by the activation of PC-specific phospholipase C (PC-PLC) or indirectly from PC by the sequential activation of phospholipase D (PLD) and phosphatidate phosphohydrolase to generate PA and DAG, respectively. Recent evidence suggests that there is a distinction in the fatty acid composition between the DAG derived directly or indirectly from PC. PC-PLC activation primarily generates polyunsaturated DAG species, whereas activation of PLD results in the generation of saturated/monounsaturated DAG species.

### Phosphatidic Acid

PA is a bioactive lipid second messenger generated mainly from PC by the stimulation of PLD in response to the activation of receptor tyrosine kinases by protein growth factors. PA has been shown to be involved in the formation of hydrogen peroxide in neutrophils and in activation of the protein kinase that regulates the neutrophil respiratory burst. Regulation of a variety of biological actions of Ras is also mediated by PA as it binds to and inhibits  $\text{p21}^{\text{ras}}$  GTPase-activating protein, effectively increasing the GTP-bound form of Ras and enhancing Ras activity. PA also binds to Raf and translocates it to the membrane where it becomes physiologically active.

### Platelet-Activating Factor

Inflammatory cells rapidly synthesize PAF in response to cell-specific stimuli. PAF is a potent biological second messenger and elicits its significant biological responses at nanomolar concentrations *in vitro* and *in vivo*. PAF-dependent activation of extracellular signal-regulated kinase (ERK) has been demonstrated in Chinese hamster ovary (CHO) cells, human vascular endothelial cells, human neutrophils,



**Figure 2** Illustration of lipid second messengers being generated on agonist-stimulated activation of specific enzymes. Agonist stimulation of G protein-coupled receptors activates PC-PLC, which hydrolyzes PC to DAG. Agonist-mediated activation of receptor tyrosine kinase stimulates PLD, which hydrolyzes PC to PA. PA can be further hydrolyzed by PLA<sub>2</sub> to lyso-PA and AA or can be converted to DAG by phosphatidate phosphohydrolase. PI-PLC that is activated by receptor tyrosine kinase hydrolyzes PIP<sub>2</sub> to IP<sub>3</sub> and DAG. IP<sub>3</sub> induces the release of Ca<sup>2+</sup> ions from internal stores and stimulates PLA<sub>2</sub> to hydrolyze PC to lyso-PAF, which is further acetylated to PAF. Agonist-induced activation of SMase hydrolyzes SM to ceramide, which is further metabolized to Cer 1-P or sphingosine and Sph 1-P. AA, arachidonic acid; DAG, diacylglycerol; G pr, G protein; IP<sub>3</sub>, inositol 1,4,5-triphosphate; lyso-PA, lyso-phosphatidic acid; PA, phosphatidic acid; PAF, platelet-activating factor; PC, phosphatidylcholine; PC-PLC, phosphatidylcholine-specific phospholipase C; PI-PLC, phosphatidylinositol-specific phospholipase C; PIP<sub>2</sub>, phosphatidylinositol 4,5-bisphosphate; PLA<sub>2</sub>, phospholipase A<sub>2</sub>; PLD, phospholipase D; SM, sphingomyelin; SMase, sphingomyelinase; Sph 1-P, sphingosine 1-phosphate; Cer 1-P, ceramide 1-phosphate.

and primary hippocampal neurons. Our laboratory has demonstrated that bFGF-mediated phosphorylation and activation of ERK is regulated by PAF in human umbilical vein endothelial cells (HUVECs). We have also shown that stimulation of the PAF receptor by PAF leads to the time-dependent activation of the signal transducers and activators of transcription-3 (STAT-3) protein via Janus kinase-2 (JAK-2) and Src. Our observation that PAF regulates the synthesis of matrix metalloproteinase in HUVECs suggests its importance in regulating

migration and invasion of endothelial cells during an inflammatory response. Activation of focal adhesion kinase (FAK) by PAF has been observed in endothelial cells, suggesting its involvement in cell motility and adhesion. PAF caused phosphorylation of c-Src and PLC $\gamma$  in rabbit platelets and human epidermoid carcinoma cells. In rat basophilic cells, PAF activated PLC $\beta$  via pertussis toxin-insensitive G proteins. Both PLC $\gamma$  and PLC $\beta$  are important in the further synthesis of PAF within the cell.

## SPHINGOLIPID-DERIVED SECOND MESSENGERS

### Ceramide

Ceramides are generated from SM by the activation of sphingomyelinase, which is itself stimulated by several agonists, including tumor necrosis factor- $\alpha$  (TNF- $\alpha$ ), interleukin-1 (IL-1), nerve growth factor (NGF), 1,25-dihydroxy-vitamin D<sub>3</sub>,  $\gamma$ -interferon, chemotherapeutic agents, ionizing radiation, and ligands that engage Fas or CD28 receptors. Ceramides participate as second messengers in signal transduction by the activation of specific serine/threonine kinases or by the stimulation of protein phosphatases. Ceramide appears to be the primary mediator of inflammation in response to TNF- $\alpha$  by activating a protein kinase that stimulates Raf-1, which initiates signaling through the mitogen-activated protein kinase (MAPK) cascade, leading to the activation of PLA<sub>2</sub> and the release of arachidonic acid. TNF- $\alpha$ -induced ceramide has also been suggested to be required for activation of stress-activated protein kinase (SAPK), leading to NF- $\kappa$ B-dependent gene transcription and apoptosis. Ceramides induce the transcription of cyclooxygenase-1 and -2 as well as a heat shock protein,  $\alpha$ B-crystallin. Ceramides have also been shown to interfere with insulin-induced tyrosine phosphorylation of the insulin receptor substrate-1 and cause insulin resistance by inactivation of MAPK cascade via stimulation of protein phosphatases-2A. Regulation of tyrosine kinase cascades involving FAK and paxillin have been implicated to involve ceramide. PKC- $\xi$ , an atypical PKC that is insensitive to phorbol ester and DAG, may be a direct target for ceramide stimulation, leading to rapid induction of NF- $\kappa$ B. Other potential targets for ceramide-mediated signaling include P23 tyrosine phosphorylation, retinoblastoma gene product (pRb), interleukin-converting enzyme, phospholipase D, and a nuclear phosphoprotein that modulates cell cycle progression.

### Sphingosine and Sphingosine 1-Phosphate

The cellular levels of sphingosine and Sph 1-P are determined by their rate of formation from hydrolysis of ceramide by ceramidase and by phosphorylation involving specific kinases. They can also be generated intracellularly by the action of platelet-derived growth factor, which stimulates cells to release the stored Sph 1-P into the bloodstream, stimulating platelet aggregation and thrombosis. The effects of Sph 1-P may involve a previous interaction with Gi/Go-coupled

receptors, with the subsequent activation of MAPK cascade and the transcription factor activator protein-1 (AP-1). Sphingosine and Sph 1-P have also been shown to decrease the cyclic AMP (cAMP) levels in some cell lines and to mobilize Ca<sup>2+</sup> from internal sources via an inositol-triphosphate independent pathway. A sphingosine-phosphorylcholine-gated Ca<sup>2+</sup> channel has been characterized and may represent the Sph 1-P-gated Ca<sup>2+</sup> channel. Sphingosine has been demonstrated to decrease the formation of DAG by inhibiting both Mg<sup>2+</sup>-dependent and -independent phosphatidate phosphohydrolase activity or by stimulating the activity of DAG-kinase with the subsequent generation of PA. This increase in PA can be further enhanced by sphingosine by rapidly activating phospholipase D in some cell types. Sphingosine and Sph 1-P appear to be involved in cytoskeletal dynamics because they induce actin stress fiber formation and stimulate FAK in Swiss 3T3 cells. One of the most important effects of sphingosine is its ability to inhibit PKC and so attenuate cell growth. Indeed, it even acts as an endogenous mediator of apoptosis.

## CROSS-TALK AMONG DIFFERENT LIPID SECOND MESSENGERS

Glycerolipid- and sphingolipid-derived second messengers can interact with each other at several sites to control signaling within the cell. Ceramides can stimulate the degradation of PA and PAF and the dephosphorylation of Sph 1-P, decreasing the mitogenic signal of these phospholipids within the cell. These changes are related to the increased activity of phosphatidate phosphohydrolase, effectively decreasing the concentration of these mitogenic lipids relative to ceramide. Activation of phosphatidate phosphohydrolase may terminate the signal from PA, PAF, and Sph 1-P but will also generate DAG, ceramide, and sphingosine, which are also bioactive and so are important for cell signaling by modifying the balance of the bioactive lipids that are implicated in cell activation and in controlling cell growth.

Sphingosine can inhibit the phosphatidate phosphohydrolase activities, causing PA to accumulate and thereby decreasing the production of DAG. This would explain the inhibiting effect of sphingosine on PKC, which is the intracellular target of DAG. Sphingosine can also increase the PA:DAG ratio by stimulating DAG kinase and PLD activities. The DNA-binding activity of AP-1 is stimulated by sphingosine and Sph 1-P. Thus, AP-1 may be the convergent point at which there is integration of the

second messengers derived from the sphingolipid or glycerolipid signals.

Control of PLD activity may be critical for the regulation of cellular activities. PLD activity that is stimulated by PA, PAF, and Sph 1-P is inhibited by ceramide. Thus, ceramide inhibits the formation of DAG by the PLD pathway. Senescent cells lack PLD activity, possibly due to the high concentration of ceramide found in these cells.

The apoptotic and antiapoptotic properties of ceramide and DAG, respectively, are well documented. An increase in ceramide levels induces cell death, which can be overcome by the presence of DAG. DAG might abrogate the ceramide-induced signals leading to apoptosis while allowing signals for other events to proceed. Thus, when DAG and ceramide are combined, cell proliferation prevails over cell death. DAG generated from PC by the action of PC-PLC can stimulate acidic sphingomyelinase activity, leading to generation of intracellular ceramides and activation of the transcription factor NF- $\kappa$ B.

PKC isoforms are involved in regulation of diverse cellular functions. Both  $\text{Ca}^{2+}$ -dependent and -independent isoforms of PKC require DAG to be activated. In contrast, the atypical PKC isoforms  $\xi$  and  $\lambda$  are independent of  $\text{Ca}^{2+}$  and not activated by DAG but are regulated by ceramide. Ceramide has also been shown to regulate PKC isoforms  $\alpha$ ,  $\delta$ , and  $\epsilon$  and to induce their translocation from the membrane to the cytosol, suggesting their role in ceramide-mediated apoptosis.

Stimulation of DNA synthesis and cell proliferation by the bioactive lipid phosphates PA, PAF, and Sph 1-P was antagonized by ceramide, which is a potent inhibitor of cell division. It has also been observed that sphingosine, a by-product of ceramide metabolism, increased the DNA synthesis further in the presence of PA and PAF, although the mechanism used is distinct from that used by PA or PAF, suggesting that inhibition of DNA synthesis by ceramide does not depend on its metabolic products.

## RECEPTORS OF LIPID SECOND MESSENGERS

Lipid second messengers generated by agonist stimulation exist as gradients and are regionally localized in the plasma membrane or are enriched in smaller microdomains known as rafts or detergent-insoluble glycoprotein-enriched domains (DIGs). The action of these lipids depends on their binding to discrete

protein domains located within the signaling proteins. Because the generation of lipid second messengers is selectively regulated within a particular membrane on a time scale of seconds to minutes, the binding of these lipids to the protein domains is generally of low affinity and is rapidly reversible, allowing the signaling proteins to undergo random diffusion in the cytosol, interspersed with membrane binding and dissociation.

Hydrolysis of  $\text{PIP}_2$  by PLC generates DAG and  $\text{IP}_3$ , stimulating the release of  $\text{Ca}^{2+}$  ions from internal stores. This induces  $\text{Ca}^{2+}$ -dependent PKC translocation to the membrane and binding of DAG to its unique protein kinase C homology-1 (C1) domain, increasing its affinity for acidic phospholipids. In contrast, increased  $\text{Ca}^{2+}$  ions induce  $\text{cPLA}_2$  translocation to the cytosolic leaflet of the endoplasmic reticulum (and possibly other internal membranes) and binding to the neutral lipids through its protein kinase C homology-2 (C2) domain. It has been shown that PA-mediated translocation of Raf to the plasma and endosomal membrane is mediated by a 4-kDa region of Raf and does not activate kinase activity but facilitates its activation by GTP-Ras.

Many cell surface receptors signal through phosphoinositide 3-kinase (PI3K), which can phosphorylate  $\text{PIP}_2$  and  $\text{PIP}_3$ . These membrane lipids further transmit signals by virtue of their specific interactions with pleckstrin homology (PH) domain-containing signaling proteins. These include the PH domains of kinases Akt, PDK1, and Btk as well as Arf GTP exchange factors ARNO and Grp1, which interact with the inositol phosphate head group.

A lipid that is more important for direction of endosomal traffic than for receptor-linked signaling is phosphatidylinositol 3-phosphate (PI3P). This lipid does not interact with the PH domain-containing proteins but interacts with the early endosomal antigen protein-1 (EEA1) possessing the small  $\text{Zn}^{2+}$ -containing Fab1-YOTP-Vac1-EEA1 (FYVE) domain. This interaction gives a certain degree of directionality to the endocytic membrane traffic.

The intracellular actions of PAF are mediated through a G protein-linked seven-transmembrane receptor that is expressed on the surface of a variety of cell types. PAF can activate its receptor in both an autocrine and a juxtacrine fashion by binding to the PAF-binding pocket within the membrane, effectively amplifying its biological response. The levels of PAF are regulated by both the desensitization of the receptor and rapid degradation of PAF by PAF acetylhydrolase.

## BIOLOGICAL FUNCTIONS OF LIPID SECOND MESSENGERS

Lipid second messengers regulate a variety of biological functions and intracellular signaling events, resulting in diverse cellular functions. Ceramide generated by degradation of sphingomyelin or from sphingosine is known to trigger apoptosis in several cell systems by activation of caspases, down-regulation of PI3K activity, or induction of dephosphorylation and inactivation of antiapoptotic factor Bcl-2, thereby turning off the survival signal leading to apoptosis. Ceramide can also mimic vitamin D3 or TNF $\alpha$ -induced differentiation of HL-60 cells to monocytes. Sphingosine has been shown to act as a mitogen to increase cell proliferation in a PKC-independent manner. Sphingosine can also mediate cell growth arrest by dephosphorylation of the Rb protein, which is essential for a cell to enter the subsequent S phase. Sph 1-P increases intracellular Ca<sup>2+</sup> mobilization and cell proliferation but reduces apoptosis and cell migration.

PAF elicits diverse biological responses in various systems. PAF has a profound effect on airways in asymptomatic individuals by evoking bronchoconstriction, airway hyperresponsiveness, plasma exudation, and mucus hypersecretion. PAF can also mimic anaphylactic symptoms in experimental animals. PAF is involved in inflammatory and immune responses and is capable of eliciting many symptoms associated with endotoxic shock-like hypotension or tissue injury as well as activation of platelets, neutrophils, and macrophages. The elevated plasma PAF levels in patients with psoriasis suggest involvement of PAF. PAF can induce proliferation of human endothelial cells, bone marrow cells, and lung fibroblast cells *in vitro*, whereas hyperplasia of guinea pig and rat airways goblet cells has been observed in response to PAF *in vivo*. The reproductive involvement of PAF has been suggested in sperm motility, ovulation, fetal lung maturation, implantation, and initiation/maintenance of parturition during early pregnancy.

## CONCLUSION

PC and SM are the primary phospholipid precursors of bioactive lipids that take part in regulation of critical signal transduction pathways involved in cell survival and cell death. Existence of lipid gradients in the plasma membrane on agonist stimulation, and translocation of specific cytosolic proteins to enable binding to the lipid metabolites thereby generated, suggests the involvement of lipid-protein complexes

in transferring the agonist-stimulated signal within the cell. Although the PC- and SM-derived bioactive lipids have specific signaling pathways, their ability to tightly regulate each other's levels allows a balance to be maintained, controlling the mitogenic signals of these phospholipids within the cell. Domain-specific interaction of lipid second messengers with specific proteins is reversible and has low affinity, allowing the signaling protein to undergo rapid dissociation and diffusion in the cytosol. The biological functions of these bioactive lipids are diverse, ranging from inducing the cell to undergo apoptosis and cell death to stimulating the cell toward proliferation and cell survival. It can be concluded that a better understanding of the interaction between various lipid second messengers and their downstream signaling events will be essential for elucidation of mechanisms that regulate cell functions.

## See Also the Following Articles

Janus Kinases and Cytokine Receptors • Mitogen-Activated Protein (MAP) Kinases and Receptors • Receptor Serine/Threonine Kinases • Receptor-Regulated Phospholipases • Tumor Necrosis Factor (TNF)

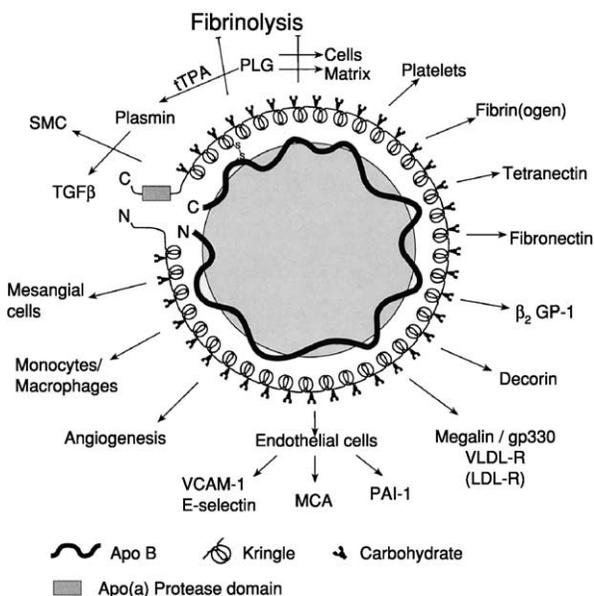
## Further Reading

- Cullen, P. J., Cozier, G. E., Banting, G., and Mellor, H. (2001). Modular phosphoinositide-binding domains: Their role in signaling and membrane trafficking. *Curr. Biol.* **11**, R882–R893.
- Deo, D. D., Axelrad, T. W., Robert, E. G., Marcheselli, V., Bazan, N. G., and Hunt, J. D. (2002). Phosphorylation of STAT-3 in response to basic fibroblast growth factor occurs through a mechanism involving platelet-activating factor, JAK-2, and Src in human umbilical vein endothelial cells: Evidence for a dual kinase mechanism. *J. Biol. Chem.* **277**, 21237–21245.
- Huwiler, A., Kolter, T., Pfrielschifter, J., and Sandhoff, K. (2000). Physiology and pathophysiology of sphingolipid metabolism and signaling. *Biochim. Biophys. Acta* **1485**, 63–99.
- Ishii, S., and Shimizu, T. (2000). Platelet-activating factor (PAF) receptor and genetically engineered PAF receptor mutant mice. *Prog. Lipid Res.* **39**, 41–82.
- Maraldi, N. M., Zini, N., Santi, S., and Manzoli, F. A. (1999). Topology of inositol lipid signal transduction in the nucleus. *J. Cell. Physiol.* **181**, 203–217.
- McPhail, L. C., Waite, K. A., Regier, D. S., Nixon, J. B., Qualliotine-Mann, D., Zhang, W. X., Wallin, R., and Sergeant, S. (1999). A novel protein kinase target for the lipid second messenger phosphatidic acid. *Biochim. Biophys. Acta* **1439**, 277–290.
- Prescott, S. M., Zimmerman, G. A., Stafforini, D. M., and McIntyre, T. M. (2000). Platelet-activating factor and related lipid mediators. *Annu. Rev. Biochem.* **69**, 419–445.
- Venkataraman, K., and Futerman, A. H. (2000). Ceramide as a second messenger: Sticky solutions to sticky problems. *Cell Biol.* **10**, 408–412.

Table I Characteristics of Lp(a) and Apo(a)

- Particle consisting of one LDL and one disulfide-linked apo(a) glycoprotein
- High homology between apo(a) and plasminogen
- Synthesis of apo(a) in the liver, assembly of Lp(a) at cell surface
- Site and mechanism of catabolism discussed controversially
- Size polymorphism of apo(a) gene and protein: > 30 alleles and isoforms
- *LPA* gene located on chromosome 6q26–q27: apo(a) alleles differ by K-IV-2 repeat number
- Lp(a) plasma concentrations: 1000-fold interindividual range (from <0.1 to >300 mg/dl) with skewed distribution (most individuals have concentrations of less than 10 mg/dl)
- Inverse relation between K-IV-2 repeats/molecular weight of apo(a) and Lp(a) concentration
- K-IV-2 polymorphism of apo(a) explaining 30 to 70% of the variability of Lp(a) concentrations; this polymorphism and other sequence variations of the apo(a) gene explain 70 to 90% of the variability of Lp(a) concentrations
- Lp(a) concentrations relatively independent of age and gender

Note. Lp(a), lipoprotein(a); apo(a), apolipoprotein (a); K-IV, kringle-IV.



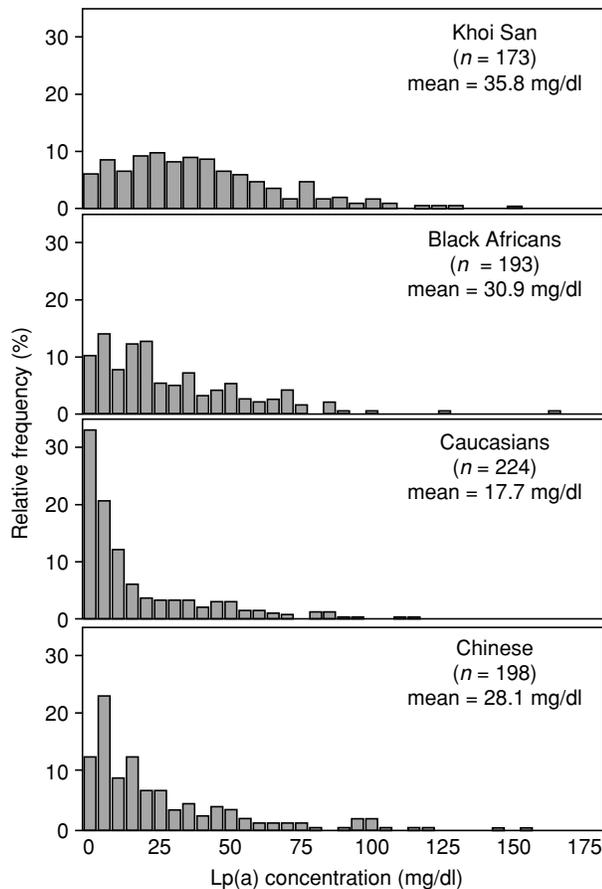
**Figure 1** Model of Lp(a) and reported interactions of Lp(a)/apo(a) with components of the fibrinolytic system, cell receptors, and other binding proteins. The central LDL particle, with its core of neutral lipids and the apo B-100 molecule, is attached to one molecule of apo(a) by a single disulfide bond between one of the K-IV repeats (K-IV-9) in apo(a) and the C-terminal region of apo B. The exact spacial orientation of the components is unknown, and the regularity of carbohydrate chains is a simplification. tTPA, tissue-type plasminogen activator; PLG, plasminogen;  $\beta_2$  GP-1,  $\beta_2$  glycoprotein-1; PAI-1, plasminogen activator inhibitor-1; MCA, monocyte chemotactic activity; VCAM-1, vascular cell adhesion molecule-1; TGF $\beta$ , transforming growth factor  $\beta$ ; SMC, smooth muscle cells. Used with permission from Utermann, G. (2000). Lipoprotein(a). In "The Metabolic and Molecular Bases of Inherited Disease" (C. R. Scriver, A. L. Beaudet, W. S. Sly, and D. Valle, eds.), 8th ed., pp. 2753–2787. McGraw-Hill, New York.

## Lipoprotein(a) Plasma Concentrations

The mean and median concentrations of Lp(a) in Caucasians are about 15 and 8 mg/dl, respectively, with an extremely broad range from less than 0.1 to more than 300 mg/dl. The distribution of Lp(a) plasma concentrations in the general Caucasian population is skewed, and most of the people express low concentrations (Fig. 2). However, there are large differences in concentrations among various ethnic groups. Individuals of African descent have two- to fourfold higher median plasma levels of Lp(a) than do Caucasians, whereas some Asian populations have significantly lower levels. The distribution of Lp(a) levels in some African populations (e.g., Khoi San) is nearly Gaussian (Fig. 2). The reason for these differences among populations is the subject of speculation.

## Genetics of Lipoprotein(a)

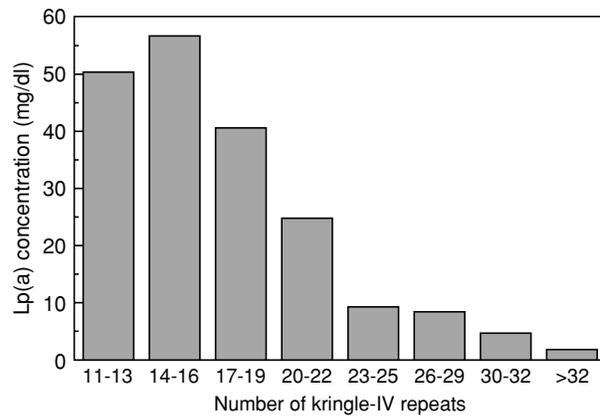
Population and family studies demonstrated that Lp(a) plasma concentrations are highly heritable and are, to a large extent, determined by the apo(a) gene locus. Additionally, two minor loci have recently been identified but are not yet confirmed. Other genes, such as the LDL receptor locus, play a role in patients with rare disease. The apo(a) gene is located in the telomeric region of chromosome 6 (6q26–q27). This gene locus determines a size polymorphism as described in 1987 by Utermann and co-workers. The molecular weight of apo(a) isoforms ranges from 300 to more than 800 kDa. Analyses of genomic DNA have shown that the protein size polymorphism is



**Figure 2** Histograms of the distribution of Lp(a) concentrations in four populations. Used and modified with permission from Kraft, H. G., Lingenhel, A., Pang, R. W. C., Delpont, R., Trommsdorff, M., Vermaak, H., Janus, E. D., and Utermann, G. (1996). Frequency distributions of apolipoprotein(a) kringle IV repeat alleles and their effects on lipoprotein(a) levels in Caucasian, Asian, and African populations: The distribution of null alleles is non-random. *Eur. J. Hum. Genet.* **4**, 74–87. Karger, Basel.

caused by a varying number of K-IV-2 repeats in the apo(a) gene (from 1 to >30).

A negative correlation exists between the molecular weight of apo(a) (corresponding to number of K-IV-2 repeats) and the Lp(a) plasma concentrations. Individuals with low molecular weight (i.e., short) isoforms (up to 22 K-IV repeats) on average have high Lp(a) concentrations, whereas those with high molecular weight (i.e., long) isoforms (>22 K-IV repeats) usually express low Lp(a) values (Fig. 3). Depending on the population, approximately 30 to 70% of the broad variability of Lp(a) plasma concentrations can be explained by this size polymorphism. Several further sequence variations at or near the apo(a) gene exist that increase the within-population variance in Lp(a) levels explained by the apo(a) gene to 70 to 90%. This makes

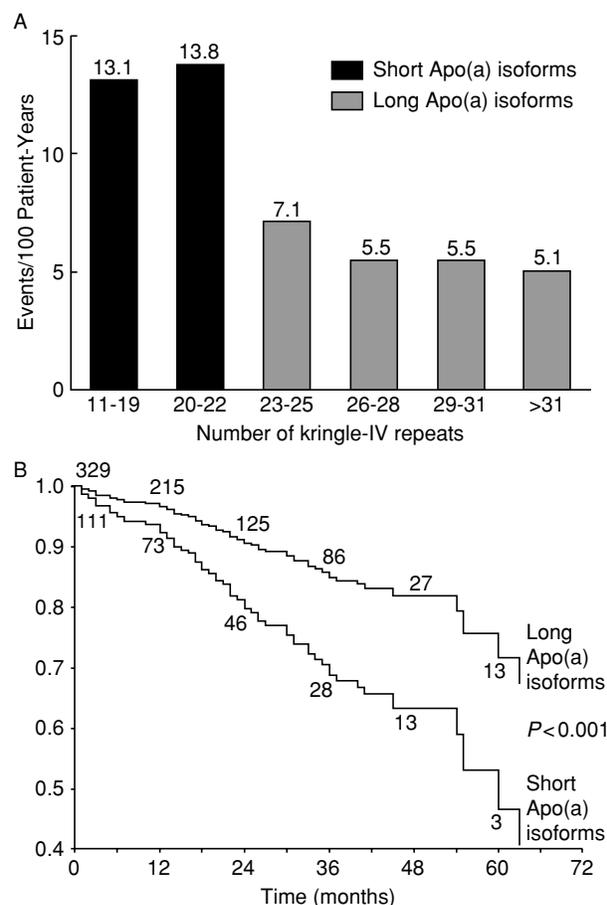


**Figure 3** Graphic demonstration of the association of apo(a) alleles with Lp(a) plasma concentrations in a Caucasian population. Each vertical bar represents the average Lp(a) concentration associated with binned apo(a) size alleles. Used and modified with permission from Kraft, H. G., Lingenhel, A., Pang, R. W. C., Delpont, R., Trommsdorff, M., Vermaak, H., Janus, E. D., and Utermann, G. (1996). Frequency distributions of apolipoprotein(a) kringle IV repeat alleles and their effects on lipoprotein(a) levels in Caucasian, Asian, and African populations: The distribution of null alleles is non-random. *Eur. J. Hum. Genet.* **4**, 74–87. Karger, Basel.

Lp(a) to a lipoprotein with an exceptionally high genetic heritability and control. This high genetic heritability, in combination with the wide interindividual variation of Lp(a), makes it very difficult to investigate nongenetic influences on Lp(a) concentrations. Depending on the study design, it may be mandatory to consider the genetic polymorphism of Lp(a) as a potential confounder in studies aimed to identify environmental and/or other genetic factors.

## LIPOPROTEIN(a) AND ATHEROSCLEROSIS

Numerous retrospective studies, and more recently a remarkable number of prospective case-control studies, have reported significantly higher Lp(a) levels in patients with coronary artery disease, cerebrovascular disease, and peripheral atherosclerosis when compared with those in controls. Many of these studies proposed Lp(a) levels above 30 mg/dl to be associated with an increased risk. However, this threshold depends on the assay used for the measurement of Lp(a). Many studies showed that the risk for atherothrombotic disease associated with Lp(a) increases significantly if other risk factors, such as high LDL or low high-density lipoprotein (HDL) cholesterol concentrations, are also present. Apparently, an interaction of Lp(a) with other lipoproteins changes its atherothrombotic properties.



**Figure 4** Association between the apo(a) size polymorphism and coronary artery disease (CAD) events in a cohort of 440 unselected hemodialysis patients prospectively followed for a period of 5 years. Of the 440 patients, 66 (15%) suffered a CAD event during follow-up. Panel A shows the number of CAD events per 100 patient-years in relation to the number of kringle-IV repeats of apo(a). Panel B shows the CAD event-free survival in patients with only long apo(a) isoforms (>22 kringle-IV repeats) and with at least one short apo(a) isoform (up to 22 kringle-IV repeats). Adjusted results were obtained from a multiple Cox proportional hazards regression analysis. Numbers near the survival curves represent the number of patients at risk within the two apo(a) isoform groups at 0, 12, 24, 36, 48, and 60 months. Used and modified with permission from Kronenberg, F., Neyer, U., Lhotta, K., Trenkwalder, E., Auinger, M., Pribasnik, A., Meisl, T., König, P., and Dieplinger, H. (1999). The low molecular weight apo(a) phenotype is an independent predictor for coronary artery disease in hemodialysis patients: A prospective follow-up. *J. Am. Soc. Nephrol.* **10**, 1027–1036, Lippincott, Williams & Wilkins.

Studies that considered the apo(a) K-IV-2 repeat polymorphism observed that short apo(a) alleles/isoforms were significantly more frequent in patients with coronary artery disease or other forms of atherothrombosis than in controls (Fig. 4). This suggests that

the apo(a) gene locus determines the risk of atherothrombosis through its control of Lp(a) plasma concentrations. Therefore, the association of genetic variation at the apo(a) gene locus with atherothrombosis supports the notion that Lp(a) is a primary genetic risk factor for atherothrombosis. Recently, this association has also been supported by prospective data (e.g., results presented in Fig. 4).

Several mechanisms by which Lp(a) promotes atherosclerosis have been proposed. High Lp(a) concentrations impair activation of transforming growth factor- $\beta$  by down-regulation of plasmin generation, contributing to smooth muscle cell proliferation. Experiments in rabbit arteries indicate that oxidized Lp(a) impairs endothelium-dependent dilation more profoundly than does oxidized LDL. Recent studies demonstrated that Lp(a) induces chemotactic activity to human monocytes in a dose-dependent fashion. Lp(a) enhances the expression of intercellular adhesion molecule-1 (ICAM-1) implicit to the adhesion and transendothelial migration of monocytes. Because Lp(a) accumulates in the subendothelial space of the vessel wall, it may act as a potent chemoattractant for these cells in human atherosclerosis. Others observed that the apo(a) size polymorphism influences the effect of Lp(a) on fibrinolysis in that only Lp(a) with short apo(a) isoforms showed high-affinity binding to fibrin surfaces, thereby acting as a prominent competitive antagonist to plasminogen. These *in vitro* findings suggest that high concentrations of Lp(a) with short apo(a) isoforms should have the most pronounced influence on fibrinolysis. In other words, the same Lp(a) concentrations may be associated with a markedly different atherothrombotic risk, depending on the apo(a) isoform. There is some evidence from population studies that this is indeed the case, although heterogeneity may exist among populations.

## CONDITIONS OF HIGH LIPOPROTEIN(a) LEVELS

### Genetic Influences

The apo(a) gene locus has the strongest influence on Lp(a) levels with usually high Lp(a) levels if the person is a carrier of at least one short apo(a) allele. Approximately 25% of Caucasians are carriers of a short allele. However, there are exceptions to this rule, meaning that some individuals with short apo(a) alleles can express low concentrations and a small fraction of individuals with long alleles may also exhibit high Lp(a) concentrations. In addition, the

relation between apo(a) alleles and Lp(a) concentrations is heterogeneous between populations.

Some conditions are known to influence Lp(a) concentrations (Table II). An inherited disease associated with elevated Lp(a) levels is dominant familial hypercholesterolemia. Heterozygous patients have decreased LDL receptor activities. Some studies did not confirm this observation but might be biased by small sample size combined with the difficulty of detecting a minor gene effect in the presence of a major one. Large family studies demonstrated convincingly that familial hypercholesterolemia affects Lp(a) concentration and that there is a gene dosis effect, with homozygous familial hypercholesterolemia having the highest average Lp(a) concentrations. This conclusion was supported by a large sib-pair analysis in Dutch families with familial defective apolipoprotein B-100 (FDB). FDB-affected sibs revealed higher Lp(a) levels than did nonaffected ones with the same apo(a) genotype, demonstrating an effect of the FDB mutation on Lp(a) concentrations. However, neither study provides evidence that Lp(a) is cleared by the LDL receptor given that the effect of the respective mutations on the Lp(a) concentrations could be an indirect one.

**Table II** Conditions Influencing Lp(a) Concentrations

Conditions increasing Lp(a) levels
<ul style="list-style-type: none"> <li>• Familial hypercholesterolemia</li> <li>• Familial defective apolipoprotein B-100</li> <li>• Renal disease</li> <li>• Systemic lupus erythematoses</li> <li>• Hypothyroidism</li> <li>• Growth hormone therapy</li> <li>• Primary gout</li> <li>• Pregnancy</li> <li>• Postmenopause</li> <li>• Diet: high amounts of elaidic acid or other trans-fatty acids from vegetables and marine sources</li> <li>• Strenuous physical exercise</li> </ul>
Conditions decreasing Lp(a) levels
<ul style="list-style-type: none"> <li>• Abetalipoproteinemia</li> <li>• Lecithin-cholesterol acyltransferase deficiency</li> <li>• Low-density lipoprotein deficiency</li> <li>• Severe hypertriglyceridemia/type I hyperlipidemia</li> <li>• Hepatic disorders associated with decreased liver function</li> <li>• Alcoholism</li> <li>• Hyperthyroidism</li> <li>• Sex hormones (female and male)</li> <li>• Anabolic steroids</li> <li>• Diet: palm oil and n-3 polyunsaturated acids</li> </ul>

## Nongenetic Influences

### Kidney Disease

There is clear evidence that Lp(a) is secondarily elevated in kidney disease. It starts to increase during the earliest stages of disease, already before functional impairment of the kidney can be recognized by an increased serum creatinine concentration. With decreasing kidney function, Lp(a) increases further. The levels are elevated severalfold when the disease is accompanied by severe proteinuria and nephrotic syndrome. The Lp(a) levels during stages of kidney replacement therapy depend strongly on the modality of treatment. Peritoneal dialysis patients have markedly higher levels of Lp(a) than do hemodialysis patients; this might be caused by the tremendous loss of proteins by dialysate and a compensatory increase in protein production, including apo(a) in the liver. Lp(a) levels clearly decrease following kidney transplantation in both treatment groups.

It is an interesting but unexplained phenomenon that the increase of Lp(a) in hemodialysis patients and in patients with mild or moderate (nephrotic) impairment of kidney function is restricted to long apo(a) isoforms when compared with isoform-matched controls. Patients with at least one short apo(a) isoform had nearly identical Lp(a) levels when compared with matched controls. This is in line with the observation that Lp(a) in hemodialysis decreases following kidney transplantation only in patients with long apo(a) isoforms and not in those with at least one short isoform.

### Lifestyle Factors (Diet and Exercise)

Dietary intervention studies have shown that elaidic acid or other trans-fatty acids from vegetables and marine sources elevate Lp(a) levels if they are consumed in great amounts. Trans-fatty acids produced by industrial hydrogenation of edible oils and fats also increase Lp(a). In contrast, palm oil and n-3 polyunsaturated acids lower Lp(a) concentrations.

The effect of exercise on Lp(a) levels is dose dependent. Moderate exercise does not affect Lp(a) levels, strenuous exercise on a daily basis results in higher Lp(a) concentrations.

## INFLUENCE OF THE ENDOCRINE SYSTEM ON LIPOPROTEIN(a) CONCENTRATIONS

### Sex Hormones

On average, there are only small differences in Lp(a) concentrations between men and women. Lp(a) is

Table III Pooled Analysis of Studies of Hormonal Replacement Therapy and Its Influence on Lp(a) Concentrations, Performed by Godsland

Regimen	Number of studies	Percentage change of Lp(a)		
		Mean	95% CI	Range
Conjugate equine estrogens (0.625 mg/day)	13	-23.9	-27.8 to -20.0	-48 to 35
Conjugate equine estrogens (0.625 mg/day plus)	4	-20.7	-25.4 to -16.0	-24 to 16
Medroxyprogesterone acetate (2.5-5.0 mg/day)				
Estradiol 17- $\beta$ (2.0 mg/day)	5	-12.9	-17.6 to -8.2	-20 to -10
Estradiol 17- $\beta$ (2.0 mg/day plus)	3	-34.4	-40.6 to -28.2	-39 to -31
Norethindrone acetate (1.0 mg/day)				
Transdermal estradiol 17- $\beta$ (0.05-0.10 mg/day)	2	5.3	-2.7 to 13.3	-5 to 6
Transdermal estradiol 17- $\beta$ (0.05 mg/day plus)	3	-6.0	-12.4 to 0.4	-13 to 0
Dihydrogesterone (10-20 mg/day)				
Tibolone (2.5 mg/day)	7	-39.2	-44.7 to -33.7	-83 to 3
Raloxifene (60-150 mg/day)	2	-6.9	-10.2 to -3.6	-8 to -5
Placebo controls	6	-1.0	-4.9 to 2.9	-15 to 3
Untreated controls	5	7.1	2.3 to 11.9	-4 to 14

Source. Data according to Godsland, I. F. (2001). Effects of postmenopausal hormone replacement therapy on lipid, lipoprotein, and apolipoprotein(a) concentrations: Analysis of studies published from 1974-2000. *Fertil. Steril.* 75, 898-915, with permission from the American Society for Reproductive Medicine.

elevated during pregnancy and menopause. This increase is clearly counterbalanced by hormonal replacement therapy. Table III shows the percentage changes of Lp(a) levels by various therapy regimens from a pooled analysis of prospective studies performed by Godsland. Conjugated equine estrogens alone were associated with a strong reduction in Lp(a) (-24%). Transdermal administration of estradiol 17- $\beta$  was accompanied by less of a decrease in Lp(a) than was oral administration. The addition of progestogen to estradiol 17- $\beta$  had a further strong decreasing effect on Lp(a) (-35%). The strongest effect, with nearly -40%, was seen for tibolone. The higher the baseline levels of Lp(a), the greater the absolute therapy-induced reduction. This is of special interest in light of the Heart Estrogen/Progestin Replacement Study, which detected a protective effect among those with high baseline Lp(a) levels, although otherwise the study did not reveal a protective effect of estrogen replacement therapy in the secondary prevention group as a whole.

Selective estrogen receptor modulators, such as tamoxifen and structural analogues (e.g., raloxifene), seem to have antiestrogenic effects on the endometrium and breast but estrogenic effects on bone and lipid metabolism. Studies applying this treatment in healthy postmenopausal women or in women with breast cancer showed an Lp(a)-lowering effect that was less pronounced than with hormonal replacement

therapy. Tamoxifen has shown even an Lp(a)-lowering effect in men.

There is strong evidence that testosterone decreases Lp(a) levels. This was shown by exogenous testosterone application as well as indirectly by suppression of endogenous testosterone production either by surgical or pharmacological castration using a gonadotropin-releasing hormone agonist. The effect of endogenous suppression of testosterone production was shown not only in prostatic cancer patients but also in young healthy men. A recent double-blind randomized controlled study using transdermal testosterone application in healthy elderly men over 65 years of age did not affect any of the lipid or apolipoprotein parameters, including Lp(a). Whether this is caused by the transdermal application of testosterone has yet to be investigated.

These data suggest that Lp(a) concentrations are under sex hormone control, but the precise mechanism is not yet defined. Hormone-responsive elements have been identified in the apo(a) 5'-flanking region. The significance of this influence under physiological (nontherapeutic and nonpathological) conditions has yet to be determined.

### Hypothyreosis and Hyperthyreosis

Several studies investigated Lp(a) in hypothyroidisms and hyperthyroidisms. These studies suggest that

Lp(a) is elevated in the hypothyroid state. However, therapy by thyroid hormones did not always show an Lp(a)-lowering effect, a result that might also depend on whether patients showed overt or subclinical hypothyroidism. A recent double-blind placebo-controlled study in 66 women with subclinical hypothyroidism did not show significant changes in Lp(a) by L-thyroxine replacement therapy. Similar observations were made in another placebo-controlled trial.

Studies that investigated hyperthyroid patients generally showed decreased Lp(a) levels in the hyperthyroid state and increased Lp(a) following therapeutic intervention by thyreostatic medications or radioactive iodine therapy.

In conclusion, thyroid hormones seem to have an Lp(a) level-modulating effect. Again, the mechanism of this effect is unknown.

### Influence of Growth Hormone

A large number of studies, including several placebo-controlled double-blind studies, investigated the effect of growth hormone substitution on Lp(a) levels in patients with congenital or acquired growth hormone deficiency. Many of these studies demonstrated a strong Lp(a)-increasing effect of growth hormone. This has raised some concerns about this therapy. However, there is strong evidence from studies in the general population that the atherogenic potential of Lp(a) increases mainly above a threshold of 20 to 30 mg/dl. Because most of the patients have Lp(a) concentrations below 10 mg/dl, the majority of patients receiving growth hormone treatment will not change their atherogenic potential related to Lp(a). Furthermore, the treatment by growth hormone is accompanied by a significant reduction in LDL cholesterol caused by increased LDL receptor activity and by lower triglyceride levels. Therefore, these changes may counterbalance the potential negative effects of an Lp(a) increase.

### Diabetes Mellitus

The epidemiological literature on Lp(a) in diabetes mellitus (DM) is conflicting. Numerous case-control studies are clearly biased by the small number of patients investigated. Lp(a) is not elevated in patients with type 2 DM when only large case-control studies are considered.

The situation is less clear in patients with type 1 DM. Elevated and similar concentrations, as compared with controls, were reported. Most large-scale studies in adult patients described nearly identical

Lp(a) values in patients and controls, whereas some studies in children showed higher concentrations in patients. The Diabetes Control and Complications Trial (DCCT) compared the Lp(a) plasma concentrations of 1299 patients with type 1 DM with those of 2158 controls. The researchers observed nearly identical Lp(a) values in the group with intensive insulin therapy ( $n = 646$ ) as compared with controls (mean 18.5 vs 18.2 mg/dl, n.s.) and slightly but significantly higher values in 653 patients with conventional insulin therapy (22.0 mg/dl,  $P < 0.05$ ). This is in line with the observation in some small but longitudinal studies in type 1 DM patients that showed a decrease in Lp(a) with improvement of glycemic control. A further large prospective study in 105 metabolically poorly controlled patients showed absolutely no changes in Lp(a) concentrations after 3 months of intensive therapy with multiple insulin doses, although the profile of other lipoproteins as well as triglyceride concentrations improved significantly.

A further argument against the conclusion that Lp(a) concentrations are elevated in type 1 diabetic patients came from a study of identical twins that found very similar Lp(a) concentrations in diabetic patients and their healthy co-twins. However, this observation rules out diabetes as a secondary cause of high Lp(a) in type 1 diabetic patients. It cannot exclude that a primary elevation of Lp(a) is associated with type 1 diabetes. Because identical twins share apo(a) alleles identical by descent, a genetic cause for the elevation of Lp(a) in type 1 diabetic patients could not be detected by such a study design. Only few studies were originally designed to investigate whether an elevation of Lp(a) in these patients, if present, is genetically determined. These studies included apo(a) phenotype analysis, and most of them came to the conclusion that apo(a) is not associated with type 1 DM. However, the investigated patient groups were heterogeneous in terms of patients' ages and duration of diabetes. A recent case-control study observed higher Lp(a) concentrations in patients with short diabetes duration and a frequency of short apo(a) isoforms nearly twice as high as in controls matched for sex and age. This preferential association of short apo(a) isoforms with diabetes indicates that the apo(a) gene, or a gene very close to the apo(a) gene, may be a susceptibility gene for type 1 DM. In the latter case, the apo(a) gene must be in linkage disequilibrium with the actual susceptibility gene. This is in line with several studies showing linkage of this chromosomal region with type 1 DM. The difference in apo(a) allele frequencies between patients and controls was found only when the

investigated patients were stratified into groups according to diabetes duration. This might be of considerable importance given that Lp(a) plasma concentrations and the apo(a) size polymorphism are possibly related to survival due to the association with atherosclerotic complications. With increasing duration of diabetes, a pronounced and nearly linear decrease in the frequency of patients with short apo(a) isoforms was observed.

Several studies investigated microalbuminuria, macroalbuminuria, and nephrotic-range proteinuria as confounders that increase Lp(a) in type 1 and type 2 diabetic patients. Studies on microalbuminuria were too inconsistent to draw a final conclusion. The concentrations of Lp(a) increase with increasing albuminuria and proteinuria. A longitudinal study recently showed that Lp(a) decreases significantly with decreasing proteinuria in these patients after starting the dialysis treatment. This is in line with the observation that Lp(a) levels in dialysis patients with diabetes or other causes of kidney disease are similar.

Some small case-control and prospective studies found an association between high Lp(a) plasma concentrations and coronary heart disease or peripheral arterial disease in diabetic patients that was not confirmed by others. A large case-control study ( $n = 500$ ) described significantly higher Lp(a) concentrations in diabetic patients with ischemic heart disease or macroangiopathy as compared with those without these conditions. In this study, a 15 to 30% higher probability of developing macroangiopathy was calculated for patients with hypertension, 10 to 20 years duration of diabetes, and Lp(a) levels greater than 30 mg/dl as compared with those with Lp(a) levels less than 30 mg/dl. Large studies described an association between high Lp(a) levels and diabetic retinopathy.

In conclusion, there is evidence that Lp(a) is an emerging risk factor for atherosclerosis and early death in type 1 and type 2 diabetes but not that Lp(a) is elevated secondarily to the disease. However, the Lp(a) levels are modified when the disease comes along with albuminuria or nephrotic syndrome.

## WHEN SHOULD LIPOPROTEIN (A) BE MEASURED?

Table IV lists groups of individuals in whom routine screening for elevated Lp(a) levels should be considered. These are mainly groups in which other strong atherosclerosis risk factors operate.

It is recommended that Lp(a) be measured for an individualized assessment of atherosclerosis risk for the planning of therapeutic interventions. Because an acceptable Lp(a)-lowering therapy is still lacking, attention should be directed to other risk factors that can be treated in patients with high Lp(a) plasma concentrations. These include high LDL cholesterol, hypertension, DM, and smoking. Besides the measurement of Lp(a), apo(a) phenotyping might be considered in special patient groups such as patients with end-stage kidney disease.

## TREATMENT OF HIGH LIPOPROTEIN(A) LEVELS

Lowering Lp(a) is a frustrating task. No simple and satisfying method or medication is available that can be used in large patient groups and that decreases Lp(a) without major side effects. We do not even know whether therapeutic lowering of high Lp(a) levels alone would be beneficial.

Lp(a) remains unchanged after treatment with widely used lipid-lowering drugs such as hydroxymethylglutaryl coenzyme A (HMG CoA) reductase inhibitors. There is growing evidence that nicotinic acid, and especially its derivative niceritrol, has an interesting Lp(a)-lowering potential. Neomycin has been reported to interfere with the assembly of Lp(a) and to lower Lp(a) moderately. This effect was increased by a combination with nicotinic acid. Lp(a)-lowering steroid hormones are not indicated because of their side effects. Promising specific Lp(a)-lowering effects were described for the antiestrogen tamoxifen. The synthetic steroid tibolone showed a reduction in Lp(a) ranging from 26 to 39% but was accompanied by a decrease in HDL cholesterol between 18 and

**Table IV Patient Groups Suggested for Screening for High Lp(a) Levels**

- 
- Patients with elevated low-density lipoprotein cholesterol concentrations or other forms of dyslipoproteinemia
  - Young patients with atherosclerotic cardiovascular and cerebrovascular disease
  - Individuals from families with a strong history of early cardiovascular or cerebrovascular disease
  - Patients with renal failure
-

23%. Adrenocorticotropin hormone (ACTH) demonstrated a pronounced Lp(a)-lowering effect (up to -65%). Experimental studies observed that tranexamic acid, which blocks the assembly of apo(a), had only a small Lp(a)-lowering effect.

The successful lowering of LDL cholesterol and Lp(a) by LDL apheresis and plasmapheresis in patients with severe hypercholesterolemia and/or high Lp(a) has been demonstrated in several studies, but it is available for only a small number of high-risk patients (e.g., those with drug-resistant hypercholesterolemia). Furthermore, it has to be shown in clinical studies that lowering Lp(a) is beneficial to the patients.

### Acknowledgments

The support by the Austrian Science Fund (P-12819 and P-15480 to G. U.), the Austrian National Bank (Project #9331 to F. K.), and the Austrian Heart Fund (to F. K.) is greatly appreciated.

### See Also the Following Articles

Atherosclerosis • Diabetes, Type 1 • Diabetes, Type 2 • Dysbetalipoproteinemia and Type III Hyperlipidemia • Growth Hormone (GH) • Familial Low Cholesterol Syndromes, Hypobetalipoproteinemia • Hypercholesterolemias, Familial Defective ApoB (FDB) and LDL Receptor Defects • Hypertriglyceridemia

### Further Reading

Craig, W. Y., Neveux, L. M., Palomaki, G. E., Cleveland, M. M., and Haddow, J. E. (1998). Lipoprotein(a) as a risk factor for ischemic heart disease: Meta-analysis of prospective studies. *Clin. Chem.* **44**, 2301-2306.

Danesh, J., Collins, R., and Peto, R. (2000). Lipoprotein(a) and coronary heart disease: Meta-analysis of prospective studies. *Circulation* **102**, 1082-1085.

Geethanjali, F. S., Luthra, K., Lingenhel, A., Kanagasaba-Pathy, A. S., Jacob, J., Srivastava, L. M., Vasisht, S., Kraft, H. G., and Utermann, G. (2003). Analysis of the apo(a) size polymorphism

in Asian Indian populations: Association with Lp(a) concentration and coronary heart disease. *Atherosclerosis* **169**, 121-130.

Godsland, I. F. (2001). Effects of postmenopausal hormone replacement therapy on lipid, lipoprotein, and apolipoprotein(a) concentrations: Analysis of studies published from 1974-2000. *Fertil. Steril.* **75**, 898-915.

Hobbs, H. H., and White, A. L. (1999). Lipoprotein(a): Intrigues and insights. *Curr. Opin. Lipidol.* **10**, 225-236.

Kronenberg, F., Utermann, G., and Dieplinger, H. (1996). Lipoprotein(a) in renal disease. *Am. J. Kidney Dis.* **27**, 1-25.

Kronenberg, F., Kronenberg, M. F., Kiechl, S., Trenkwalder, E., Santer, P., Oberhollenzer, F., Egger, G., Utermann, G., and Willeit, J. (1999). Role of lipoprotein(a) and apolipoprotein(a) phenotype in atherogenesis: Prospective results from the Bruneck Study. *Circulation* **100**, 1154-1160.

Marcovina, S. M., and Koschinsky, M. L. (1999). Lipoprotein(a) concentration and apolipoprotein(a) size: A synergistic role in advanced atherosclerosis? *Circulation* **100**, 1151-1153.

Marcovina, S. M., and Koschinsky, M. L. (2003). Evaluation of lipoprotein(a) as a prothrombotic factor: Progress from bench to bedside. *Curr. Opin. Lipidol.* **14**, 361-366.

McLean, J. W., Tomlinson, J. E., Kuang, W.-J., Eaton, D. L., Chen, E. Y., Fless, G. M., Scanu, A. M., and Lawn, R. M. (1987). cDNA sequence of human apolipoprotein(a) is homologous to plasminogen. *Nature* **330**, 132-137.

Meier, C., Staub, J. J., Roth, C. B., Guglielmetti, M., Kunz, M., Miserez, A. R., Drewe, J., Huber, P., Herzog, R., and Muller, B. (2001). TSH-controlled L-thyroxine therapy reduces cholesterol levels and clinical symptoms in subclinical hypothyroidism: A double-blind, placebo-controlled trial (Basel Thyroid Study). *J. Clin. Endocrinol. Metab.* **86**, 4860-4866.

Mooser, V., and Hobbs, H. H. (1997). Lipoprotein(a) and growth hormone: Is the puzzle solved? *Eur. J. Endocrinol.* **137**, 450-452.

Purnell, J. Q., Marcovina, S. M., Hokanson, J. E., Kennedy, H., Cleary, P. A., Steffes, M. W., and Brunzell, J. D. (1995). Levels of lipoprotein(a), apolipoprotein B, and lipoprotein cholesterol distribution in IDDM: Results from follow-up in the Diabetes Control and Complications Trial. *Diabetes* **44**, 1218-1226.

Utermann, G. (2000). Lipoprotein(a). In "The Metabolic and Molecular Bases of Inherited Disease" (C. R. Scriver, A. L. Beaudet, W. S. Sly, and D. Valle, eds.), 8th ed., pp. 2753-2787. McGraw-Hill, New York.

Utermann, G., Menzel, H. J., Kraft, H. G., Duba, H. C., Kemmler, H. G., and Seitz, C. (1987). Lp(a) glycoprotein phenotypes: Inheritance and relation to Lp(a)-lipoprotein concentrations in plasma. *J. Clin. Invest.* **80**, 458-465.

There is no consensus that lithium causes this hyperthyroidism immunologically or in any other way.

## LITHIUM AND CALCIUM

Lithium alters the calciostat and renders the parathyroid cell less sensitive to calcium. Some patients develop parathyroid adenomas accompanied by an increase in serum calcium and immunoreactive parathyroid hormone.

## OTHER ENDOCRINE EFFECTS

Diabetes insipidus (DI) occurs in up to 40% of patients within a few days of beginning lithium therapy. DI is usually nephrogenic, with high circulating concentrations of arginine vasopressin and an impairment of the renal response to this hormone. Lithium may also inhibit antidiuretic hormone release from the posterior pituitary, leading to central DI.

The clinical effect of lithium on adrenal function is minimal, although urinary aldosterone concentration increases soon after starting the drug. An increase in body weight is commonly seen in patients

receiving lithium. The effects of the drug on glucose metabolism are not thought to be of clinical significance.

## Further Reading

- Birch, N. J. (1988). "Lithium: Inorganic Pharmacology and Psychiatric Use." IRL Press, Oxford, UK.
- Birch N. J., Padgham, C., and Hughes, M. S. (eds.) (1993). "Lithium in Medicine and Biology." Marius, Carnforth, UK.
- Cade, J. F. J. (1949). Lithium salts in the treatment of psychotic excitement. *Med. J. Austr.* **36**, 349-352.
- Fernandez, L., and Fox, R. (1980). Perturbation of the human immune system by lithium. *Clin. Exp. Immunol.* **41**, 527-532.
- Hallcher, L. M., and Sherman, W. R. (1980). The effects of lithium ion and other agents on the activity of myo-inositol-1-phosphatase from bovine brain. *J. Biol. Chem.* **255**, 10896-10901.
- Johnson, F. N. (ed.) (1984). "The History of Lithium Therapy." Memillan, London
- Johnson, F. N. (ed.) (1988) "Lithium and the Endocrine System." Karger, Lancaster, UK.
- Lazarus, J. H. (1986). "Endocrine and Metabolic Effects of Lithium." Plenum, New York.
- Mannisto, P. T. (1980). Endocrine side-effects of lithium. In "Handbook of Lithium Therapy" (F. N. Johnson, ed.), pp. 310-322. MTP Press, Lancaster, UK.
- Schou, M. (1957). Biology and pharmacology of the lithium ion. *Pharmacol. Rev.* **9**, 17-58.

RCT in these conditions are discussed. Due to space limitations, many data on genetically engineered mice, which have altered levels of plasma HDL-C, are not discussed.

## REGULATORY FACTORS FOR PLASMA HDL CHOLESTEROL LEVELS

HDL particles consist of heterogeneous subclasses. They have higher density ( $> 1.063$  g/ml) and smaller in size (Stoke's diameter 5 to 17 nm) and consist of about 50% lipids and 50% proteins. They can be classified by density (HDL2 or HDL3 fraction), electrophoretic mobility ( $\alpha$ - and pre- $\beta$ -electrophoretic mobility), and apolipoprotein composition (Lp A-I or Lp A-II). The quantity and quality of HDLs are regulated by many factors, such as plasma enzymes, transfer proteins, and cell surface receptors (Fig. 1).

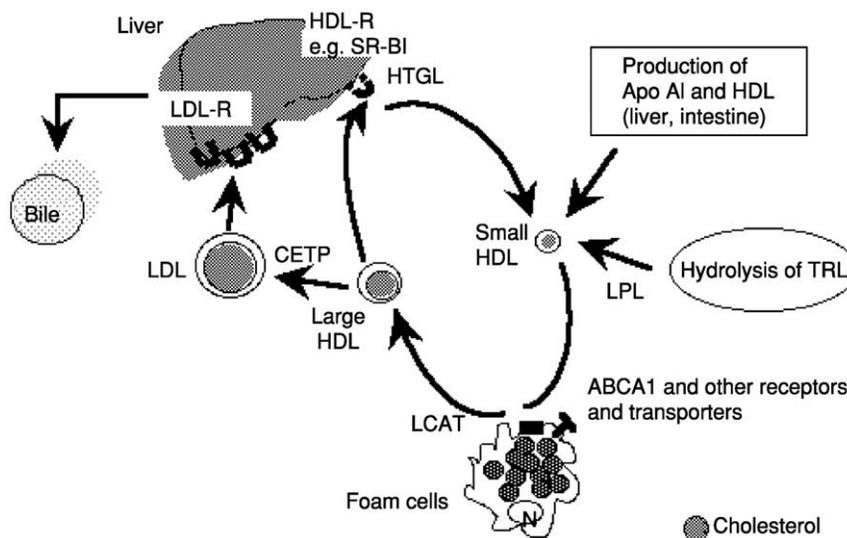
### Synthesis and Secretion of the Components of HDL

One of the major constituents of HDL particles is apo AI, which is synthesized in the liver and small intestine. ABCA1, a defect of which causes some HDL deficiency including Tangier disease, is thought to interact with apo AI and phospholipid on the plasma membrane and generate "nascent HDL" particles, which is a discoidal form. Other receptors and

molecules may be involved, although ABCA1 is requisite for the production of HDL particles. Free cholesterol on the nascent HDL is esterified by the action of lecithin:cholesterol acyltransferase (LCAT). LCAT plays a pivotal role in making HDL particles spherical and mature in form in plasma.

### Modification of HDL Particles in Plasma

In plasma, several enzymes affect the quality and quantity of HDL particles. Intravascular movement of lipids and apolipoproteins is one of the important sources for HDL particles. Two lipases, lipoprotein lipase and hepatic triglyceride lipase (HTGL), are thought to be involved. The hydrolysis of triglyceride-rich lipoproteins (TRLs) by these lipases releases apolipoproteins and phospholipids in plasma, and these constituents can be used for the formation of new HDL particles. HTGL, located on the liver sinusoid, is thought to remodel large and triglyceride-rich HDLs into smaller ones. Cholesteryl ester transfer protein (CETP) is a plasma glycoprotein that facilitates neutral lipids transfer between lipoproteins. CETP transfers cholesteryl ester from HDL to apo B-containing lipoproteins and, conversely, triglyceride from apo B-containing lipoproteins to HDL. Complete lack of CETP causes marked elevation of HDL-C levels in plasma. Phospholipid transfer



**Figure 1** Metabolic map of the generation and regulation of HDL and HDL-mediated cholesterol transport in humans. Arrows denote cholesterol transport in the human body mediated by lipoproteins, plasma lipid transfer proteins, and cell surface receptors. Enzymes and transfer proteins and receptors are shown as red and blue letters, respectively. Cholesterol efflux from peripheral cells is thought to be very diverse. There are some candidates for cell surface receptors, such as ABCA1 and SR-BI.

protein (PLTP) transfers phospholipid and cholesterol from apo B-containing lipoproteins to HDL.

### Hepatic Uptake of HDL and HDL Lipid

In the last step of RCT, there are believed to be at least two distinct pathways available to take up cholesterol from plasma. One is the LDL receptor-mediated pathway, illustrated by human CETP deficiency. The other pathway is the HDL receptor(s)-mediated pathway. Although the impact and significance of this pathway are not completely understood in humans, scavenger receptor class B type I (SR-BI) is the physiologically relevant HDL receptor established in mice. This receptor mediates selective uptake of HDL lipid. There is another possible pathway, in which whole particles of HDL may be taken up and catabolized.

### Degradation of Apo AI or HDL

In the kidney, apo AI is catabolized by a size-dependent filtration process. Cubilin is thought to mediate reabsorption of apo AI from the proximal tubule lumen.

## LOW HDL SYNDROMES

### Primary Low HDL Syndromes

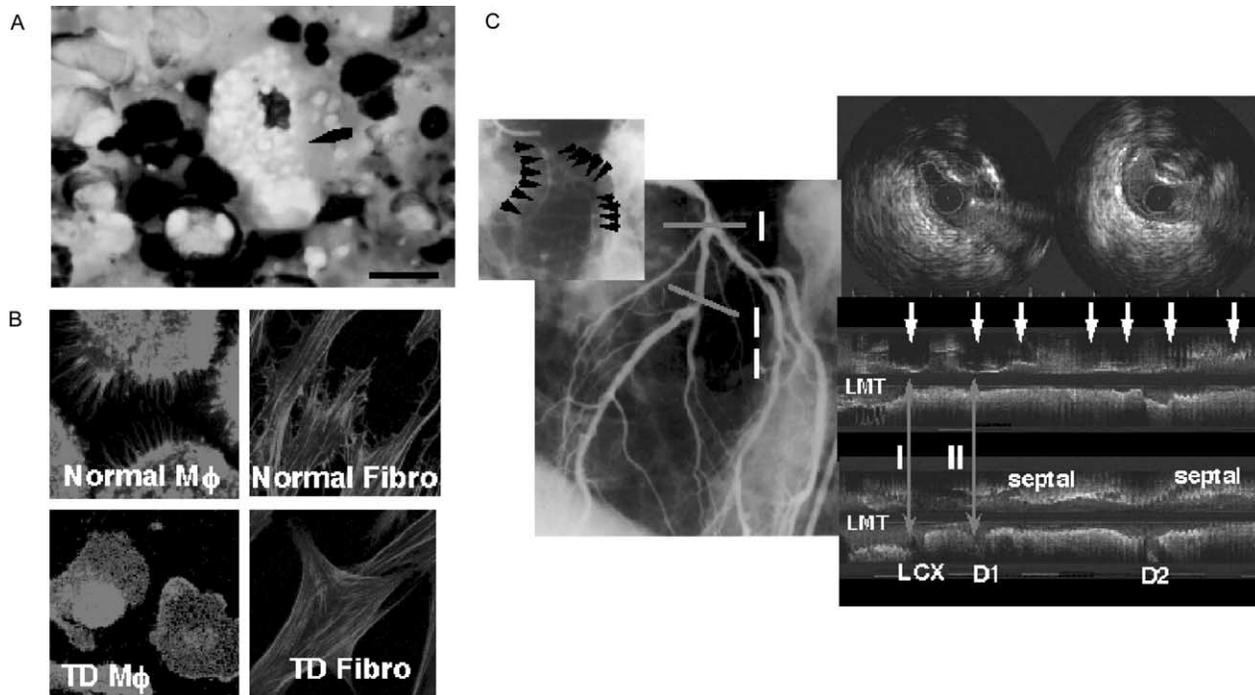
Table I shows the comparison of clinical and biochemical characteristics of primary low HDL syndromes.

#### **Tangier Disease and Familial HDL Deficiency with Mutations in the ATP-Binding Cassette Transporter-1 Gene**

Tangier disease (TD) is a familial HDL deficiency (FHD) originally described by Frederickson *et al.* in 1961 (Table I). Twin brothers living on Tangier Island, Chesapeake Bay, Virginia, suffered from yellow-orange tonsils, hepatosplenomegaly, and neuropathy. In homozygotes, plasma HDL is virtually absent and triglyceride levels are often elevated. Patients suffer from symptoms caused by the accumulation of cholesteryl ester in many tissues throughout the body (Fig. 2). Thrombocytopenia and hemolytic anemia are sometimes observed. Plasma apo AI concentration is extremely low, with a relative increase in precursors of apo AI. Fibroblasts and macrophages from Tangier patients are characterized by defective lipid, cholesterol, and phospholipid efflux. These cells

Table I Comparison of Clinical and Biochemical Characteristics of Primary Low HDL Syndromes

	Tangier disease	Familial HDL deficiency without Tangier disease phenotype	Familial LCAT deficiency	Fish eye disease	ApoAI deficiency
Affected gene	ABCA1	ABCA1 (some cases)	LCAT	LCAT	apoAI
Clinical signs and symptoms					
Typical Tangier phenotype (orange tonsils, hepatosplenomegaly, neuropathy)	+	–	–	–	–
Corneal opacity	+	+	+++	+++	++
Nephropathy	–	–	+	–	–
Risk for coronary heart disease	Moderately increased	Moderately increased	Normal/increased	Normal/increased	Increased
Biochemical data					
Plasma total cholesterol	Low	Low	Low	Low	Normal
LDL cholesterol	Low	Normal	Normal	Normal	Normal
HDL cholesterol	None	None	Very low	Very low	None
Plasma triglyceride	Increased	Normal	Increased	Increased	Normal
% Cholesterol ester	Normal	Normal	Low	Low	Normal
Apolipoprotein (apo)-AI	Very low	Very low	Very low	Very low	None
Relative increase in apoAI precursor	+	+	–	–	–
α-LCAT activity	Normal	Normal	Very low	Very low	Normal
Cell biological data					
Cholesterol efflux from the cells (apoAI mediated)	Absent	Absent	Normal	Normal	Normal



**Figure 2** Clinical manifestations in patients with Tangier disease. (A) Foam cells observed in bone marrow. Reproduced with permission from Komuro, R., *et al.* (2000). *Circulation* **101**, 2446–2448. (B) Altered actin cytoskeletons observed in fibroblasts and macrophages from TD patients. Reproduced with permission from Hirano, K., *et al.* (2000). *FEBS Lett.* **484**, 275–279. (C) Severe coronary atherosclerosis demonstrated by intravascular ultrasonography. Reproduced with permission from Komuro, R., *et al.* (2000). *Circulation* **101**, 2446–2448.

have abnormal actin cytoskeletons in association with decreased expression of Cdc42, a member of the RhoGTPases family. Obligatory heterozygotes have half the normal plasma HDL-C and apo AI levels. This disorder is caused by mutations in the *ABCA1* gene, which is located on the long arm of chromosome 9. In some patients with FHD who do not show the typical Tangier phenotype, such as orange tonsils and hepatosplenomegaly, this is also reported to be associated with mutations in the *ABCA1* gene. The molecular mechanism for the phenotypic differences between TD and FHD remain to be investigated, even though they carry mutations in the same gene. Patients with *ABCA1* mutations are thought to have moderately increased risk for CHD.

#### **Familial LCAT Deficiency and Fish Eye Disease**

Both familial LCAT deficiency (FLD) and fish eye disease (FED) are caused by the mutations in the LCAT gene. Both disorder lead to a marked reduction in plasma HDL-C. The major clinical symptoms in FLD are corneal opacification, anemia, and proteinuria, which may progress to renal failure. Foam cells accumulate in various tissues, including cornea,

kidney, liver, spleen, and arteries. The ratio of esterified to free cholesterol is extremely low. Lipoproteins are morphologically abnormal, with the appearance of multilamellar vesicles, rouleaux, LpX-like particles.

FED is characterized by the presence of corneal opacities and absence of nephropathy. The molecular mechanism for the phenotypic differences between FLD and FED remains to be investigated.

#### **Apo AI Genetic Mutations and Variations Affecting HDL Levels**

Apo AI is the major apolipoprotein in HDL particles. The *apo AI* gene is located on chromosome 11 in a cluster with two other apolipoprotein genes, *apo CIII* and *apo AIV*. Various disruptions and mutations of the *apo AI* gene have been reported. Certain mutations cause complete deficiency of plasma HDL-C levels, also observed for *ABCA1* mutants or LCAT mutants. HDL-deficient patients with *apo AI* mutation do not usually show anemia, proteinuria, orange tonsils, or hepatosplenomegaly. Some mutations or variations in the *apo AI* gene are related to the phenotypic expression of amyloidosis.

## Secondary Low HDL Syndromes

### **Metabolic Syndromes (e.g., Visceral Fat Syndrome, CD36 Deficiency, and Insulin-Resistance Syndromes)**

Low HDL-C is often observed in patients with metabolic syndromes, including visceral fat syndrome, CD36 deficiency, and other insulin-resistance syndromes. Hypertriglyceridemia appears to stimulate the lipid exchange and accelerated catabolism of HDL protein. Reduced activities of plasma lipolytic enzymes cause loss of catabolism of TRL, which leads to a reduction of the source for generating new HDL particles. Patients with low HDL-C have a higher risk for CHD.

### **Inflammations**

It is thought that HDL-C may be a negative marker for systemic or local inflammations. In some pathological conditions with chronic or acute systemic inflammation (e.g., severe infections and some hematological malignancy), HDL-C is reduced in association with an increase in serum amyloid A proteins. Some studies have indicated that low HDL-C may be a marker for cardiac events during short-term follow-up, associated with the increase in C-reactive protein.

### **Cholestatic Disorders**

In obstructive liver diseases, plasma HDL-C levels may be markedly reduced to the same extent as in primary HDL deficiency. Plasma total cholesterol levels are usually increased in these states.

### **Medicines and Drugs**

Some drugs, such as thiazides or probucol, are known to decrease plasma HDL-C levels. ProbucoL is a potent hypolipidemic drug that inhibits xanthoma formation even in patients homozygous for familial hypercholesterolemia and in Watanabe heritable hyperlipidemic rabbits.

## HIGH HDL SYNDROMES

### **Primary High HDL Syndromes**

Etiologies for high HDL syndrome are listed in [Table II](#).

#### **Cholesteryl Ester Transfer Protein Deficiency**

Patients with complete deficiency of CETP activity were found in Japan in the late 1980s. Homozygotes had extremely high plasma HDL-C levels (three to six times normal levels). Plasma levels of apo AI and AII

were increased. Plasma apo B and apo B-containing lipoproteins were relatively decreased. LDL and HDL from the homozygotes are abnormal biochemically and biologically. The HDL particles are large and cholesterol and apo E rich. The LDL particles are small, polydisperse, and apo B rich. The LDLs have reduced affinity for LDL receptors on the fibroblasts, and the HDL have reduced ability to mediate cholesterol efflux from the cells.

The first identified CETP gene mutation was an intron 14 splicing defect (*IN14*), which is a null mutation with a dominant effect on plasma CETP and HDL-C levels. A second CETP gene mutation was a missense mutation in exon 15 (*D442:G*). Both mutations are very common in the Japanese population. The effect of *D442:G* on plasma lipoproteins is less severe than that of *IN14*. Homozygotes for *D442:G* have moderately increased HDL levels. In other countries, some genetic variations have been reported to affect CETP mass and HDL-C levels.

There has been heated discussion about the atherogenicity of the genetic deficiency of CETP. We found a unique area (Omagari, Japan) where *IN14* accumulated. A population-based study in this area showed that there was a U-shaped relationship between HDL-C and the incidence of ischemic electrocardiographic changes. An evaluation of the relationship between CETP gene mutations and CHD in Japanese American men in the Honolulu Heart Program cohort also showed that the men, most of whom had a *D442:G* mutation, had increased risk of CHD. However, the atherogenic effect of CETP deficiency was observed in subjects with moderately

**Table II Primary and Secondary High HDL Syndromes**

#### **Primary high HDL syndromes**

- Cholesteryl ester transfer protein deficiency
- Familial hepatic triglyceride lipase deficiency
- Familial hyper-alpha lipoproteinemia with premature corneal opacity
- Familial hyper-alpha lipoproteinemia with overproduction of apoAI

#### **Secondary high HDL syndromes**

- Chronic massive alcohol intake
- Primary biliary cirrhosis
- Multiple lipomatosis
- Drugs
  - Estrogen
  - Corticosteroid
  - Fibrates
  - Nicotinic acid and its derivatives
  - HMG-CoA reductase inhibitor

increased HDL-C levels but not in subjects with marked hyperalphalipoproteinemia (HALP). Overall, CETP plays a crucial role in the reverse cholesterol transport to transfer cholesterol from HDL to apo B-containing lipoproteins. The transferred cholesterol is taken up by the liver via the LDL receptor-mediated pathway in humans.

#### **Hepatic Triglyceride Lipase Deficiency**

HTGL plays a role in the conversion of very low-density lipoprotein into intermediate-density lipoprotein (IDL) and LDL by its triglyceride lipase activity. This enzyme also has the function for the remodeling of large, triglyceride-rich HDLs into smaller ones. HTGL is reported to enhance the hepatic uptake of HDL lipid. There are several mutations in the HTGL gene. General features of human HTGL deficiency are increased IDL levels and increased large and TG-rich HDL particles. Some patients with HTGL deficiency are reported to have premature atherosclerosis.

#### **Combined Reduction of CETP and HTGL**

Several patients with a combined reduction in CETP and HTGL have been reported to suffer from CHD and corneal arcus. The impact of the combined reduction on atherosclerosis appears to be stronger than that of CETP deficiency alone. One of the possible mechanisms is that both CETP and HTGL play important roles in the remodeling of HDL particles from large to small particles, which are active for cholesterol efflux (Fig. 1). The combined reduction of these two proteins leads to the marked elevation of HDL-C, with the appearance of very large HDL particles, which are not active for cholesterol efflux from the cells.

#### **Hyperalphalipoproteinemia with Increased Production of HDL**

A family with marked hyperalphalipoproteinemia was reported to have overproduction of apo AI. The primary cause(s) is not known.

### **Secondary High HDL Syndromes**

#### **Chronic Alcohol Intake**

It is well-known that chronic alcohol intake increases plasma HDL-C levels. Some enzymes and transfer proteins, such as CETP and HTGL, are altered in patients with massive chronic alcohol intake. The association between alcohol intake and mortality is U-shaped, suggesting that the beneficial effect of

alcohol intake is only observed in mild to moderate drinkers.

#### **Primary Biliary Cirrhosis**

Primary biliary cirrhosis (PBC) is a primary cholestatic liver disease for which the primary defect is not known. In the end stage of this disorder, patients have very low HDL-C with the appearance of Lp X, similar to other obstructive or cholestatic liver diseases. In the early stage, patients with PBC often demonstrate high HDL syndrome. In some, HDL-C is markedly increased to the same extent as that in patients with genetic CETP deficiency. In contrast to CETP deficiency, both activities and protein mass of CETP are markedly increased, whereas HTGL is reduced in patients with PBC.

#### **Medicines and Drugs**

Some drugs, such as estrogen derivatives, fibrates, and statins, are reported to increase plasma HDL-C levels. Prospective studies using the fibrate gemfibrozil demonstrated that increases in HDL-C during treatment were correlated with the prevention of cardiac events.

### **SUMMARY**

Prospective studies using fibrates have provided evidence that HDL-modifying treatments can reduce the incidence of atherosclerotic cardiovascular diseases. The enhancement of RCT has great potential as an antiatherosclerotic treatment and has been further supported by the identification of the novel molecular targets. On the other hand, the association between HDL-C levels and the incidence of cardiovascular disease is U-shaped but not linear, showing that neither an increase nor a decrease in HDL-C always indicate the atherogenicity in individuals. Methods to measure the number of active HDL particles, evaluate the efficiency and dynamics of RCT, and assess the stability of atherosclerotic lesions are required in order to pursue the previously discussed therapeutic approaches.

### **Acknowledgments**

This work was supported by a health science research grant from the Ministry of Health and Welfare of Japan and research grants from the Study Group of Molecular Cardiology (Japan), Japan Heart Foundation, Osaka Heart Club, Japan Heart Foundation/Pfizer Grant for Research on Hypertension and Vascular Metabolism, and Tanabe Frontier Conference (Japan) to K. H. This work was also supported by a grant-in-aid (No. 13671191) to K. H. from the Ministry of Education, Science, and Culture of Japan.

## See Also the Following Articles

ABCA1 Defects • Atherosclerosis • Mixed Lipemias

## Further Reading

- Assmann, G., Von Eckardstein, A. and Brewer, H., Jr. (2001). Familial analphalipoproteinemia: Tangier disease. In "The Metabolic and Molecular Bases of Inherited Disease" (C. R. Scriver, A. L. Beaudet, W. S. Sly, and D. Valle, eds.), 8th ed., Vol. 2, pp. 2937–2960. McGraw-Hill, New York.
- Hayden, M. R., Clee, S. M., Brooks-Wilson, A., *et al.* (2000). Cholesterol efflux regulatory protein, Tangier disease and familial high-density lipoprotein deficiency. *Curr. Opin. Lipidol.* **11**, 117–122.
- Hirano, K., Yamashita, S., Kuga, Y., *et al.* (1995). Atherosclerotic disease in marked hyperalphalipoproteinemia. Combined reduction of cholesteryl ester transfer protein and hepatic triglyceride lipase. *Arterioscler. Thromb. Vasc. Biol.* **15**, 1849–1856.
- Hirano, K., Yamashita, S., Nakajima, N., *et al.* (1997). Genetic cholesteryl ester transfer protein deficiency is extremely frequent in Omagari area of Japan. Marked hyperalphalipoproteinemia caused by cholesteryl ester transfer protein gene mutation is not a longevity syndrome. *Arterioscler. Thromb. Vasc. Biol.* **17**, 1053–1059.
- Hirano, K., Yamashita, S., and Matsuzawa, Y. (2000). Pros and cons of inhibiting cholesteryl ester transfer protein. *Curr. Opin. Lipidol.* **11**, 589–594.
- Hiraoka, H., Yamashita, S., Matsuzawa, Y., *et al.* (1993). Decrease of hepatic triglyceride lipase and increase of cholesteryl ester transfer protein in patients with primary biliary cirrhosis. Relation to abnormalities in high density lipoprotein. *Hepatology* **18**, 103–110.
- Krieger, M. (1999). Charting the fate of the "good cholesterol": Identification and characterization of the high-density lipoprotein receptor SR-BI. *Annu. Rev. Biochem.* **68**, 523–558.
- Matsuzawa, Y., Yamashita, S., Kameda, T., *et al.* (1984). Marked hyper HDL2-cholesterolemia associated with premature corneal opacity: A case report. *Atherosclerosis* **53**, 207–212.
- Miyaoka, K., Kuwasako, T., Hirano, K., *et al.* (2001). CD36 deficiency is associated with insulin resistance. *Lancet* **357**, 686–687.
- Sakai, N., Matsuzawa, Y., Hirano, K., *et al.* (1991). Detection of two species of low density lipoprotein particles in cholesteryl ester transfer protein deficiency. *Arterioscler. Thromb.* **11**, 71–79.
- Tall, A. R. (1995). Plasma cholesteryl ester transfer protein and high density lipoproteins: New insights from molecular genetic studies. *J. Intern. Med.* **237**, 5–12.
- Tall, A. R., Breslow, J. L., and Rubin, E. M. (2001). Genetic disorders affecting plasma high density lipoproteins. In "The Metabolic and Molecular Bases of Inherited Disease" (C. R. Scriver, A. L. Beaudet, W. S. Sly, and D. Valle, eds.), 8th ed., Vol. 2, pp. 2915–2936. McGraw-Hill, New York.
- Von Eckardstein, A., Nofer, J. R., and Assmann, G. (2001). High density lipoproteins and arteriosclerosis. Roles of cholesterol efflux and reverse cholesterol transport. *Arterioscler. Thromb. Vasc. Biol.* **21**, 13–27.
- Yamashita, S., Maruyama, T., Hirano, K., *et al.* (2001). Molecular mechanisms, lipoprotein abnormalities and atherogenicity of hyperalphalipoproteinemia. *Atherosclerosis* **152**, 271–285.



# Luteinizing Hormone

see LH

potentially have significant effects on many processes involved in the pathophysiology of diseases such as diabetes, hypertension, myocardial infarction, and hypercholesterolemia. Magnesium also influences properties of cell membranes; this process is thought to occur by means of calcium channels and ion transport mechanisms. Calcium flux is inhibited by magnesium from sarcolemmal membranes, through competition for binding sites on actin, and via changes in the adenylate cyclase–cyclic AMP system. The calcium-antagonizing effects of magnesium are well known, and it is for this reason that magnesium has been described as the “nature’s calcium antagonist.” Magnesium is also responsible for the maintenance of transmembrane gradients of sodium and potassium. This is why patients with refractory hypokalemia will not respond to potassium supplementation until magnesium deficiency is corrected.

## PATHOPHYSIOLOGY

There are three different biological mechanisms that explain the physiological effects of magnesium in hypertension, diabetes, and hyperlipidemia. First, dysregulation of  $\text{Na}^+$ -dependent  $\text{Mg}^{2+}$  transport is associated with decreased intracellular  $\text{Mg}^{2+}$  and increased intracellular  $\text{Na}^+$ , which may be important in the pathophysiology of hypertension. Second, a relatively low magnesium level creates an intracellular imbalance between calcium and magnesium that results in increased vascular tone in the smooth muscle of the artery and, therefore, in increased blood pressure. Third, magnesium deficiency causes insulin resistance, which in turn causes hyperinsulinemia, resulting in hypertension, diabetes, hyperlipidemia, and atherosclerosis.

## MEASUREMENT OF MAGNESIUM

The easiest and most common way in which to assess magnesium status in the clinic is by the serum magnesium. This measures extracellular magnesium only and so has limited clinical usefulness. If a person has low serum magnesium, also known as hypomagnesemia, the person is definitely magnesium deficient, but many people with normal serum levels may be magnesium deficient. Total serum magnesium concentration in adults is between 0.75 and 0.96 mmol/L. This level, which constitutes a small fraction (0.3%) of body magnesium, is kept remarkably stable, even in the presence of intracellular magnesium depletion or overload. Accordingly, the serum magnesium level is not merely in equilibrium with intracellular

magnesium concentration, and it might not reflect intracellular magnesium status. Most cytosolic magnesium is bound to proteins and nucleotides or is sequestered into intracellular organelles. Only a small fraction is present in the free form, which is responsible for the biological actions of magnesium. Until recently, techniques used to determine intracellular magnesium ( $\text{Mg}^{2+}$ ) included metallochromic indicator dyes, nuclear magnetic resonance, and  $\text{Mg}^{2+}$ -selective microelectrodes. Today, fluorescent probes that accurately measure cellular magnesium are available. Although this fluorescent approach is not used as a routine clinical tool, it has facilitated and enhanced understanding of magnesium regulation and homeostasis in health and disease.

## CAUSES AND PREVALENCE OF MAGNESIUM DEFICIENCY

There are three main causes of magnesium deficiency and hypomagnesemia: decreased dietary intake and lack of magnesium in the water supply, alcoholism, and the use of thiazide diuretics.

The prevalence of hypomagnesemia has been found to vary widely, depending on the patient’s clinical condition. In a general population, 6.9% of patients were shown to be hypomagnesemic. In hospital inpatients on a medical–surgical floor, there was a prevalence of 11%, whereas in the intensive care unit (ICU), it was found to be 20%. In a postoperative ICU setting, the prevalence was 60%. A study of diabetic patients established a prevalence of 25%. Another prevalence study of magnesium levels for 120 patients in an urban minority clinic found that 24% of hypertensive patients and 25% of diabetic patients were hypomagnesemic.

## EPIDEMIOLOGY

Numerous epidemiological studies have found an inverse relationship between magnesium in the drinking water and cardiovascular mortality. This association between magnesium in drinking water and ischemic heart disease was reconfirmed in a major review of the literature done by epidemiologists at Johns Hopkins University.

The largest epidemiological study of magnesium status was the Atherosclerosis Risk in Communities (ARIC) study. This was a 5-year longitudinal study that examined 15,000 patients and compared dietary magnesium, serum magnesium, and race with the prevalence of hypertension, diabetes, and atherosclerosis. The study controlled for the potential

confounding variables of age and body mass index (BMI). The results showed that African Americans had lower dietary magnesium intake along with lower serum magnesium levels, which significantly correlated with a higher prevalence of hypertension, diabetes, and atherosclerosis.

In a 10-year study, 400 high-risk individuals predisposed to coronary heart disease were divided into one group that received a magnesium-rich diet and another group that was placed on a “usual” diet. Increased dietary magnesium was shown to correlate with less sudden death, less total mortality, and lower incidence of hypokalemia and hypomagnesemia.

## MAGNESIUM AND ETHNICITY

An extremely important unanswered question is whether magnesium deficiency, which is more prevalent in the African American community, is a contributing factor to the increased atherosclerosis and mortality rate in this population. Magnesium deficiency has been shown to be a cause of insulin resistance in numerous studies of rigorous methodology. Insulin resistance has been intimately linked to metabolic syndrome X, which is hypertension, diabetes, and hyperlipidemia.

The evidence that there is both an increased incidence of magnesium deficiency and insulin resistance in African Americans comes from various sources. The ARIC study was cited previously. In addition, Ford analyzed the data from the National Health and Nutrition Examination Survey (NHANES) and found that lower dietary magnesium was associated with increased hypertension. Humphries found in young black males that decreased intake of magnesium was associated with insulin resistance even in patients who were not diabetic. In a large study of African Americans, Manolio found an inverse relationship between serum magnesium and fasting insulin levels. As noted previously, African Americans with hypertension or diabetes have a 24 to 25% prevalence of hypomagnesemia.

## CLINICAL SIGNS AND SYMPTOMS

Magnesium deficiency is almost always asymptomatic. There are no pathognomonic signs and symptoms of the magnesium-deficient state. The situation must be severe if clinical manifestations are to occur. This would also always be accompanied by low serum magnesium. Symptoms, when they do occur, generally fall into the categories of cardiac effects, metabolic effects, and neurological effects (Table I).

## THERAPEUTIC USES OF MAGNESIUM

Intravenous magnesium remains the mainstay of the treatment of preeclampsia and eclampsia. It is also the drug of choice in the treatment of torsades de pointes. It is part of the Advanced Cardiac Life Support (ACLS) protocol to use magnesium for this condition.

**Table I Clinical Manifestations of Severe Magnesium Deficiency**

---

<b>Cardiac effects</b>
Atrial fibrillation
Atrial flutter
Supraventricular tachycardia
Ventricular tachycardia
Torsades de pointes
Coronary artery spasm
Hypertension
EKG changes:
Prolonged PR interval
Widened QRS complex
Prolonged QT interval
Atherosclerosis
<b>Metabolic effects</b>
Hypokalemia
Hypocalcemia
Increased intracellular calcium
Hyponatremia
Increased intracellular sodium
Hypophosphatemia
Metabolic alkalosis
Hyperglycemia
Hypercholesterolemia
<b>Neurological effects</b>
Grand mal seizures
Focal seizures
Paresthesias
Dizziness
Vertigo
Ataxia
Nystagmus
Tremor
Myopathy
Dysphagia
Esophageal spasm
Delirium
Personality changes
Depression
Headaches
Migraines
Coma

---

There are a number of epidemiological studies showing that low intake of magnesium is associated with decreased lung function. Therefore, it was thought that a bolus of intravenous magnesium would be helpful during moderate to severe asthmatic exacerbations. When this has been tested clinically, it showed only mild and short-lived symptom improvement and did not show a significant effect on pulmonary function.

While some small randomized controlled trials showed that intravenous magnesium was effective in reducing mortality in the immediate postinfarction period, the Magnesium in Coronary Artery Disease or MAGIC trial proved that it was ineffective for this purpose. The MAGIC trial was a large five-year multicenter randomized controlled trial sponsored by the National Institutes of Health in the United States. It showed that there was no difference in mortality or morbidity at 30 days whether or not intravenous magnesium was used.

There are a number of potential benefits of magnesium supplementation for patients with risk factors for atherosclerosis. In patients who are deficient, magnesium could reduce insulin resistance and have a major impact in decreasing morbidity and mortality. Magnesium is both safe and inexpensive. The most common side effect is gastrointestinal distress. The one clinical condition where magnesium should not be used is in patients with renal insufficiency; because magnesium is excreted through the kidney, there is the threat of magnesium overload in these patients.

Although a number of studies show an inverse relationship between magnesium in the water supply and coronary artery disease, the studies that have looked at oral magnesium supplementation for the treatment of various diseases show inconsistent results. When both magnesium and potassium were used together in patients with congestive heart failure and ventricular ectopy, there was a statistically significant reduction in ventricular arrhythmias.

There have been at least 11 randomized controlled trials looking at magnesium supplementation in hypertension. The largest of these studies, the Trial of Hypertension Prevention (TOHP), studied 698 patients for 6 months and found no benefit. This was a primary prevention trial rather than a therapeutic trial. It is possible that many of the patients were not magnesium deficient. Of the randomized controlled treatment trials, 7 showed a positive effect on blood pressure and 3 showed no effect. It was noted that the higher the dose of magnesium, the greater the likelihood of success. No trials that used less than 20 mmol of magnesium showed an effect,

whereas all trials using 40 mmol of magnesium were positive.

There have been two recent randomized controlled trials using magnesium supplementation in the treatment of diabetes. One used 15 mmol of magnesium per day. It found that many diabetics were intracellularly depleted of magnesium supply. Although this was corrected during the trial, there was no improvement in glycemic control. The trial lasted only 8 weeks long. Another trial that used varying doses of magnesium found no effect on glycemic control at the 20-mmol level but found a statistically significant effect at the 40-mmol level.

There has been only one major trial looking at the effect of magnesium on hyperlipidemia. It showed a statistically significant decrease in triglycerides and in the ratio of low-density lipoprotein (LDL) to high-density lipoprotein (HDL).

There have been two major dietary studies looking at high-magnesium diets. One of these was a 10-year cohort study with more than 400 patients comparing a high-magnesium diet to a low-magnesium diet. It showed a marked decrease in overall mortality, a marked decrease in cardiovascular disease, and a 50% reduction in sudden death. The other study was of the Dietary Approaches to Stop Hypertension (DASH) diet. This diet consists mainly of fruits, vegetables, low-fat dairy products, lean meats, fish, and whole grains. It is a diet that is high in potassium, magnesium, and calcium but that is low in sodium. It has been shown to effectively reduce blood pressure in both African Americans and Caucasians.

High magnesium in both the water supply and the diet has shown value in the prevention of cardiovascular disease and in the treatment of hypertension. The results of magnesium supplementation trials have been inconclusive because measuring intracellular magnesium is difficult and because the magnesium dose has been low and the studies have been of short duration.

## CONCLUSIONS

Magnesium is a critical cofactor in numerous enzymatic reactions in the body, including the whole energy production cycle. It is an intracellular ion that is difficult to measure in a clinically useful way. Intravenous magnesium is the treatment of choice for pre-eclampsia and torsades de pointes. Contrary to this, randomized clinical trials have proven that intravenous magnesium is ineffective in the treatment of an acute myocardial infarction, and in an asthmatic

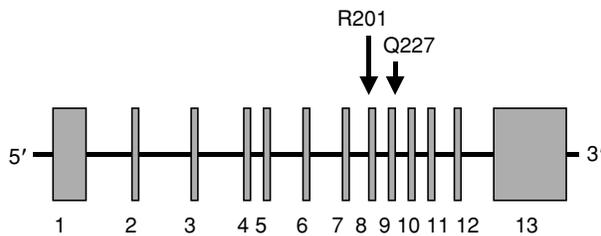
attack. A combination of potassium and magnesium does suppress ventricular arrhythmias in patients with congestive heart failure. Magnesium deficiency has been linked to insulin resistance, which is known to cause hypertension, diabetes, hyperlipidemia, and atherosclerosis. African Americans have a higher prevalence of magnesium deficiency than do Caucasians. High-magnesium diets, such as the DASH diet, should be recommended for the entire population for the prevention of cardiovascular disease. This diet has also been shown to reduce blood pressure regardless of a patient's ethnicity. Clinical trials with magnesium supplements in the treatment of hypertension, diabetes, and hyperlipidemia have been inconclusive. The reason is that intracellular magnesium is difficult to measure in a clinically useful way, and previous studies have had doses of magnesium that were too low and for too short a duration. Further studies with higher doses of magnesium for a longer period of time need to be done before it can be determined definitively whether magnesium is useful in the treatment of hypertension, diabetes, and hyperlipidemia.

### See Also the Following Articles

Atherosclerosis • Hypertension and Diabetes • Hypocalcemia, Therapy • Insulin-Resistant States, Role of Free Fatty Acids (FFA)

### Further Reading

- Altura, B. M., and Altura, B. T. (1995). Magnesium in cardiovascular biology. *Sci. Am.* (May-June), 28-37.
- Fox, C., Ramsoomair, D., and Carter, C. (2001). Magnesium: Its proven and potential clinical significance. *South. Med. J.* **94**, 1195-1201.
- Fox, C., Ramsoomair, D., Mahoney, M., Carter, C., Young, B., and Graham, R. (1999). An investigation of hypomagnesemia among ambulatory urban African-Americans. *J. Fam. Pract.* **48**, 635-639.
- Humphries, S., Kusher, H., and Falkner, B. (1999). Low dietary magnesium is associated with insulin resistance in a sample of young, nondiabetic black Americans. *Am. J. Hypertension* **12**, 747-756.
- Laurant, P., and Touyz, R. M. (2000). Physiological and pathophysiological role of magnesium in the cardiovascular system: Implications in hypertension. *J. Hypertension* **18**, 1177-1191.
- Marx, A., and Neutra, R. (1997). Magnesium in drinking water and ischemic heart disease. *Epidemiol. Rev.* **19**, 258-272.



**Figure 1** Schematic of the GNAS1 gene indicating the location of the Arg201 and Gln227 mutations within exons 8 and 9.

coordinated fashion during intracellular signal transduction. Acting as a second messenger system, the G protein complex is activated when a ligand binds to its receptor, stimulating an exchange of bound guanosine diphosphate (GDP) for guanosine triphosphate (GTP) on the surface of the  $G_{\alpha}$ -subunit. The  $G_{\alpha}$  portion subsequently stimulates adenylate cyclase, leading to intracellular cyclic AMP (cAMP) accumulation and downstream gene transcription. Intrinsic GTPase activity within the  $G_{\alpha}$ -subunit serves to hydrolyze GTP, allowing for a restoration of bound GDP and completion of the signaling cycle. In McCune–Albright syndrome, the completion phase of this cycle is destabilized, resulting in a signaling pathway that becomes and remains “turned on” in an unregulated and autonomous fashion, as shown in Fig. 2. This phenomenon, known as “constitutive activation,” leads to hyperfunction and proliferation of affected cells. G protein-mediated cell signaling is ubiquitous within endocrine glands as well as in many additional cell types, including those found in bone, liver, skin, and heart. Thus, patients with McCune–Albright syndrome may exhibit a myriad of endocrinopathies and other forms of systemic involvement in addition to the classic triad of features that are most commonly present. Therefore, it is important to conceptualize the disorder as existing within a broad spectrum of potential phenotypic manifestations. Both classic and variant forms of McCune–Albright syndrome are reviewed in the next section.

## CLINICAL FEATURES

The precise incidence of McCune–Albright syndrome is unknown, but it is quite rare. At any given time, pediatric endocrine groups in large academic medical centers around the country typically report that they are caring for anywhere from zero to four patients with the disorder. Manifestations of the disease may arise during the neonatal period or at any time during childhood and adolescence. Failure to recognize

features of the disease resulting in a delay in diagnosis is a common problem. Although the true sex ratio is unknown, McCune–Albright syndrome has been reported far more frequently in girls than in boys.

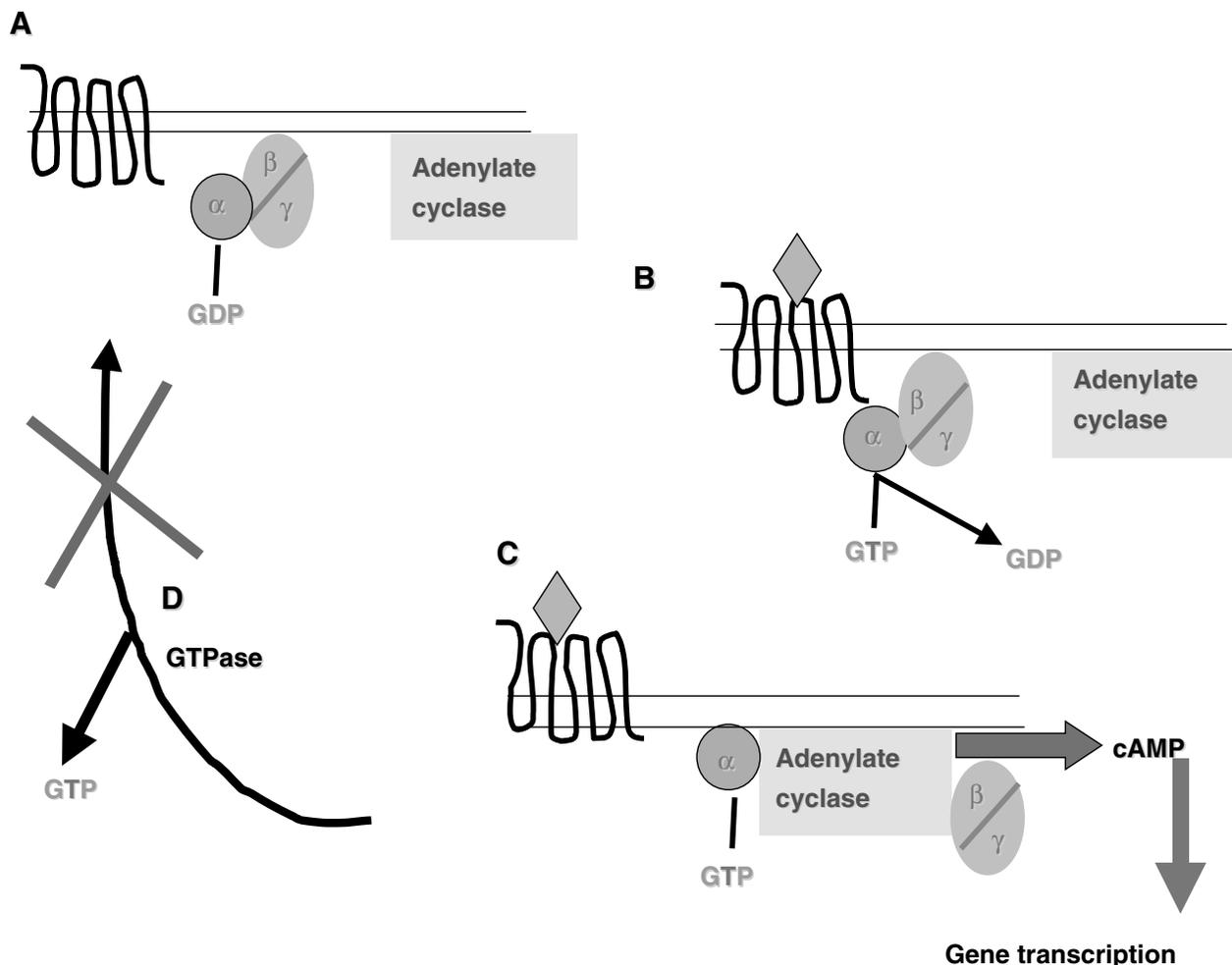
## Classic McCune–Albright Syndrome

The combination of precocious puberty, café au lait pigmentation, and fibrous bone dysplasia constitutes the enduring hallmarks of this disease. However, a significant percentage of patients will develop endocrinopathies above and beyond precocious puberty. Therefore, these additional endocrine problems will be considered in conjunction with the other classic features.

### Precocious Puberty

The diagnosis of McCune–Albright syndrome is most commonly made following the appearance of precocious puberty. Typically described in girls, a sudden onset of vaginal bleeding is heralded or accompanied by acute breast enlargement. These physical changes are due to autonomous ovarian function characterized by the development of large cysts producing high serum levels of estrogen. Once these cysts resolve, the estrogen withdrawal results in a shedding of the endometrial lining and spontaneous bleeding. The first such episode in girls with McCune–Albright syndrome usually occurs between 1 and 6 years of age, although it may become apparent as early as 4 months of age. The frequency with which subsequent episodes arise is unpredictable and quite varied. Whereas some girls experience extended intervals of ovarian quiescence, others go on to develop recurrent episodes of ovarian hyperfunction with frequent menses. In these cases, additional manifestations of precocious puberty include the appearance of pubic and axillary hair, accelerated growth, and skeletal maturation. Ultimately, untreated progressive precocious puberty results in premature closure of growth plates with significant short stature during adulthood. The precocious puberty observed in boys with McCune–Albright syndrome originates from an analogous hyperfunction of the testes, leading to episodic elevations in serum testosterone levels and the early appearance of secondary sexual characteristics. In both sexes, this premature production of sex steroids is categorized as a form of “peripheral precocious puberty,” indicating that it represents a physiological process outside of the central nervous system pathways through which pubertal development is normally mediated.

Reports of ovarian function during the postpubertal period in McCune–Albright syndrome are



**Figure 2** G protein-mediated signal transduction. Panel A depicts the heterotrimeric G protein in the quiescent state, with GDP complexed with the G $\alpha$ -subunit. On ligand binding (B), GDP is exchanged for GTP and the  $\alpha$ -subunit dissociates from  $\beta\gamma$ . Activation of adenylate cyclase (C) results in cAMP formation, leading to downstream gene transcription. In McCune–Albright syndrome, intrinsic GTPase activity within G $\alpha$  is defective (D), rendering the cell cycle turned “on” independent of ligand binding.

inconsistent. Although regular menstrual cycles and normal fertility have been reported in some women, there is also evidence that episodic autonomous ovarian activity persists in many patients. Thus, adolescent and adult women with the disorder may experience irregular periods and occasional prolonged vaginal bleeding.

#### **Other Endocrinopathies**

The propensity to additional endocrine abnormalities is a feature of McCune–Albright syndrome. In each of these cases, the underlying mechanism relates to excessive cellular function from increased intracellular cAMP formation. Affected endocrine glands exhibit a surplus of hormone production and evidence of cell

proliferation. The most frequently encountered endocrinopathy (after precocious puberty) is hyperthyroidism, usually accompanied by an enlargement of the thyroid gland. Somewhat less common is the development of pituitary growth hormone excess, which may also be associated with increased prolactin and causes gigantism if it arises prior to closure of the skeletal growth plates. Rarely, an excess of the adrenal steroid hormone cortisol may occur, resulting in a disorder known as “Cushing syndrome.” Other infrequently reported endocrine problems include hyperparathyroidism and hypophosphatemic rickets. Therefore, periodic testing of blood hormone levels for the early detection of endocrinopathies is a routine part of ongoing follow-up for affected patients. The

Table 1 Additional Endocrine Problems Described in Patients with Classic McCune–Albright Syndrome

Endocrine abnormality	Estimated incidence	Clinical features	Laboratory features
Hyperthyroidism	30–40%	Multinodular goiter	Elevated serum thyroxine and triiodothyronine with suppressed thyroid-stimulating hormone
Growth hormone excess	15–20%	Pituitary hyperplasia or adenoma, resulting in gigantism or acromegaly	Elevated serum growth hormone and insulin-like growth factor I
Cushing syndrome	2–5%	Nodular adrenal hyperplasia	Elevated serum cortisol with low adrenocorticotrophic hormone
Hypophosphatemic rickets	2–5%	Rachitic bone changes	Low serum phosphorous
Hyperparathyroidism	Very rare	Parathyroid adenoma or hyperplasia	Elevated serum and urine calcium

incidence and clinical features of the endocrine manifestations of McCune–Albright syndrome above and beyond precocious puberty are summarized in Table 1.

### Café au Lait Pigmentation

The spectrum of the skin pigmentation exhibited by an affected patient ranges from small discreet birthmarks to dramatic involvement of an entire limb or side of the body. These flat areas are usually faint at birth and become darker and more obvious over time. As implied by the name, the color of these regions is often similar to that of “coffee with milk,” although they may be darker, depending on the baseline skin tone. Often ending abruptly at the midline, the areas of increased pigmentation follow the outgrowth of distinct cell populations during early embryological development, a pattern designated as the “lines of Blaschko.” Another consistent finding is an irregular border configuration reminiscent of a rocky shore and described as the “coast of Maine,” in contrast to the “coast of California” café au lait birthmarks featured in other disorders such as neurofibromatosis. Investigation of biopsied skin cells from these areas has confirmed that the pigmentation is due to increased melanin production within melanocytes containing the activating  $Gs\alpha$  mutation. Figure 3 illustrates the typical appearance of café au lait pigmentation in a patient with McCune–Albright syndrome.

### Fibrous Dysplasia of Bone

Of the triad of clinical features observed in McCune–Albright syndrome, the bone disease carries the greatest potential for morbidity. Fibrous bone dysplasia consists of skeletal lesions in which normal tissue has been replaced by abnormal fibrous bone that may take a variety of forms, depending on the location. These abnormal areas may arise anywhere within the body but are most commonly found in the long bones and skull, with multiple sites of involvement being

typical. The destruction of normal bone that occurs in the limbs leads to an increased rate of fracture, potentially severe deformities, bone pain, and decreased mobility. Also debilitating are areas of bony overgrowth involving the face that may even result in hearing loss or blindness from impingement of nerve fibers that traverse the skull. As expected, the fibrous dysplasia is caused by the presence of the  $Gs\alpha$  mutation within bone-forming cells. With time, both the number and the severity of bone lesions in McCune–Albright syndrome tend to progress, rendering many affected patients confined to wheelchairs by middle adulthood.

### Nonclassic McCune–Albright Syndrome

Nonclassic varieties of the disease may be thought of as those in which the prototypical activating mutation is present in an atypical distribution. This may take the form of an isolated tissue bearing the abnormal  $Gs\alpha$  protein, an unusual combination or subset of the classic triad, or extremely widespread involvement with multisystemic consequences. Each of these possibilities is described in the following subsections.



Figure 3 Characteristic appearance of a café au lait spot in a patient with McCune–Albright syndrome.

### **Isolated $G\alpha$ Mutation**

It has now been recognized that the same mutation that causes McCune–Albright syndrome also has an important role in the genesis of a number of sporadic tumors that may arise within endocrine glands. In this context, the activating  $G\alpha$  mutation is referred to as *Gsp*, an “oncogene,” in view of its action as a genetic switch resulting in abnormal cellular growth and neoplastic transformation. Rare tumors of the thyroid gland, anterior pituitary, testes, and adrenal glands in which this mutation has been found have all been described.

### **Forme Fruste McCune–Albright Syndrome**

Another context in which an isolated  $G\alpha$  mutation is known to occur is in patients exhibiting an apparent isolated feature of the syndrome. Thus, individuals with isolated fibrous bone dysplasia, isolated recurrent ovarian cysts, and isolated café au lait pigmentation all have been demonstrated to harbor the same activating mutation as that found in the classic form of the disease. In a related scenario, cases have been described in which only one or two of the three classic manifestations are present, a situation recognized as a forme fruste variant of the disorder. Similarly, variations on the theme of the classic triad in which a slightly different cell type or endocrine gland is involved are also possible. Given the multitude of tissues in which the mutation may occur, the numbers of possible “faces” of McCune–Albright syndrome are seemingly endless.

### **Multisystemic Involvement**

In a few rare cases, widespread distribution of the  $G\alpha$  mutation has resulted in a devastating, multisystemic form of the disorder. In these patients, the presentation occurs during the neonatal period. Because the classic features are often lacking, however, there is usually a delay in making the diagnosis. Clinical findings in these infants have included severe liver disease, multiple endocrinopathies, heart failure, and even sudden death. It is assumed that these cases represent instances in which the activating mutation occurred very early in embryological life, leading to numerous involved areas within the body. A germline mutation, in which every cell would be expected to have the abnormality, has never been described and would presumably be lethal.

## **TREATMENT OF MCCUNE–ALBRIGHT SYNDROME**

The development of effective and safe therapeutic interventions for the various aspects of McCune–Albright syndrome is an ongoing and evolving

challenge. Only with the remote possibility of gene therapy could a cure for the disease ever be envisioned. A number of established palliative treatment options exist, and several novel approaches that may significantly improve the prognosis and quality of life for patients with McCune–Albright syndrome are under investigation. Historic and potential new therapeutic modalities for each of the classic features of the disease are reviewed in this section.

### **Treatment of Precocious Puberty**

Treatment of precocious puberty is typically reserved for patients with evidence of the progressive form of the condition. Thus, a period of observation is obligatory in all children following the initial episode. In girls with a clinical course for which intervention is indicated, the mainstay of treatment is pharmacological, although surgical removal of the ovaries or ovarian cysts has been performed in some cases. The goals of therapy are to halt progression of pubertal development, prevent vaginal bleeding, and retard the advanced rate of skeletal maturation that will compromise a child’s ability to achieve a normal adult height. Until recently, the greatest success had been obtained with the use of the aromatase inhibitor, testolactone, typically at a dose of 40 mg/kg/day. By decreasing the aromatization of androgens to estrogens, the physiological effect of the autonomous ovarian activity was attenuated in some patients receiving this therapy. However, recent evidence has indicated that improved efficacy may be achieved with the use of tamoxifen (10–30 mg/day), an antiestrogen with tissue-specific stimulatory and inhibitory actions at the level of the estrogen receptor. As experience with the use of this medication for the treatment of precocious puberty in McCune–Albright syndrome continues to grow worldwide, in some cases it is becoming first-line therapy for the disorder. Also under development at this time are a number of additional drugs, including more potent aromatase inhibitors, pure antiestrogens, and pure antiandrogens. It remains to be seen whether any of these agents will ultimately prove to be beneficial in the treatment of children with precocious puberty and McCune–Albright syndrome.

### **Treatment of Café au Lait Pigmentation**

Because the impact of the skin pigmentation in McCune–Albright syndrome is essentially cosmetic, this feature of the disorder is typically the least of

the patient's and parents' concerns. However, in cases where extensive involvement of frequently exposed skin is present, therapeutic options primarily revolve around the use of laser technologies.

### Treatment of Fibrous Bone Dysplasia

Historically, the only therapeutic approach available for the treatment of fibrous bone dysplasia was surgical. Internal rod placement, bone grafting, and curettage have traditionally been reserved for the most severely affected individuals. However, the outlook for patients with McCune–Albright syndrome suffering from symptomatic fibrous dysplasia has brightened considerably during recent years. This is due to the emergence of the bisphosphonates as a promising medical therapy for the disease. In studies conducted previously, short-term administration of the bisphosphonate pamidronate in both children and adults with fibrous dysplasia resulted in a significant decrease in fracture rate and bone pain as well as a concurrent increase in bone mineral density without apparent side effects. If substantiated over time, these results represent a truly remarkable therapeutic breakthrough with the potential to substantially improve the quality of life for many patients with McCune–Albright syndrome.

### CONCLUSION

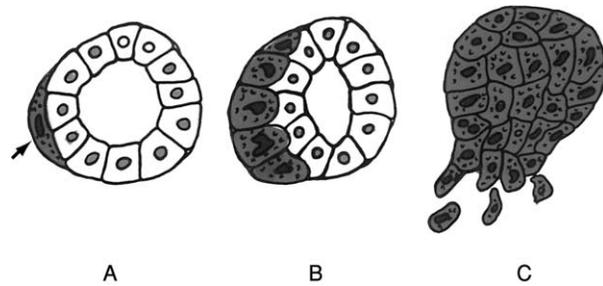
Wonderful strides have been made in the knowledge and understanding of McCune–Albright syndrome. The identification of an activating  $G_{\alpha}$  mutation has afforded valuable insights into normal and abnormal human cellular function, and the development of new therapies continues to improve clinical management. Future collaborative efforts will focus on a determination of the natural history of disease, clarification of remaining enigmas regarding molecular genetics, and designation of optimal treatment strategies for all aspects of McCune–Albright syndrome.

### See Also the Following Articles

Acromegaly, Diagnosis of • Adrenal Tumors, Molecular Pathogenesis • Albright's Fibrous Dysplasia • Autoimmune Polyglandular Syndrome • Gigantism: Excess of Growth Hormone • G Protein-Coupled Receptors • Hyperthyroidism, Childhood and Adolescence

### Further Reading

- Albright, F., Butler, A. M., Hampton, A. O., and Smith, P. (1937). Syndrome characterized by osteitis fibrosa disseminata, areas of pigmentation, and endocrine dysfunction, with precocious puberty in females. *N. Engl. J. Med.* **216**, 727–746.
- De Sanctis, C., Lala, R., Matarazzo, P., Balsamo, A., Bergamaschi, R., Cappa, M., and Cisternino, M. (1999). McCune–Albright syndrome: A longitudinal clinical study of 32 patients. *J. Pediatr. Endocrinol. Metab.* **12**, 817–826.
- Eugster, E. A., and Pescovitz, O. H. (2001). Advances in the treatment of precocious puberty. *Expert Opin. Invest. Drugs* **10**, 1623–1630.
- Eugster, E. A., Rubin, S. D., Reiter, E. O., Plourde, P., Jou, H. C., Pescovitz, O. H., and the McCune–Albright Study Group. (2003). Tamoxifen treatment for precocious puberty in McCune–Albright Syndrome: A multicenter trial. *J. Pediatr.* **143**, 60–66.
- Foster, C. M. (1993). Endocrine manifestations of McCune–Albright syndrome. *The Endocrinologist* **3**, 359–364.
- Foster, C. M., Feuillan, P., Padmanabhan, V., Pescovitz, O. H., Beitins, I. Z., Comite, F., Shawker, T. H., Loriaux, D. L., and Cutler, G. B. (1986). Ovarian function in girls with McCune–Albright syndrome. *Pediatr. Res.* **20**, 859–863.
- Levine, M. A. (1999). Clinical implications of genetic defects in G proteins: Oncogenic mutations in  $G_{\alpha s}$  as the molecular basis for the McCune–Albright syndrome. *Arch. Med. Res.* **30**, 522–531.
- Shenker, A., Weinstein, L. S., Moran, A., Pescovitz, O. H., Charest, N. J., Boney, C. M., Van Wyk, J. J., Merion, M. J., Feuillan, P. P., and Spiegel, A. M. (1993). Severe endocrine and nonendocrine manifestations of the McCune–Albright syndrome associated with activating mutations of stimulatory G protein GS. *J. Pediatr.* **123**, 509–518.
- Weinstein, L. S., Shenker, A., Gehman, P. V., Merino, M. J., Friedman, E., and Spiegel, A. M. (1991). Activating mutations of the stimulatory G protein in the McCune–Albright syndrome. *N. Engl. J. Med.* **325**, 1688–1695.



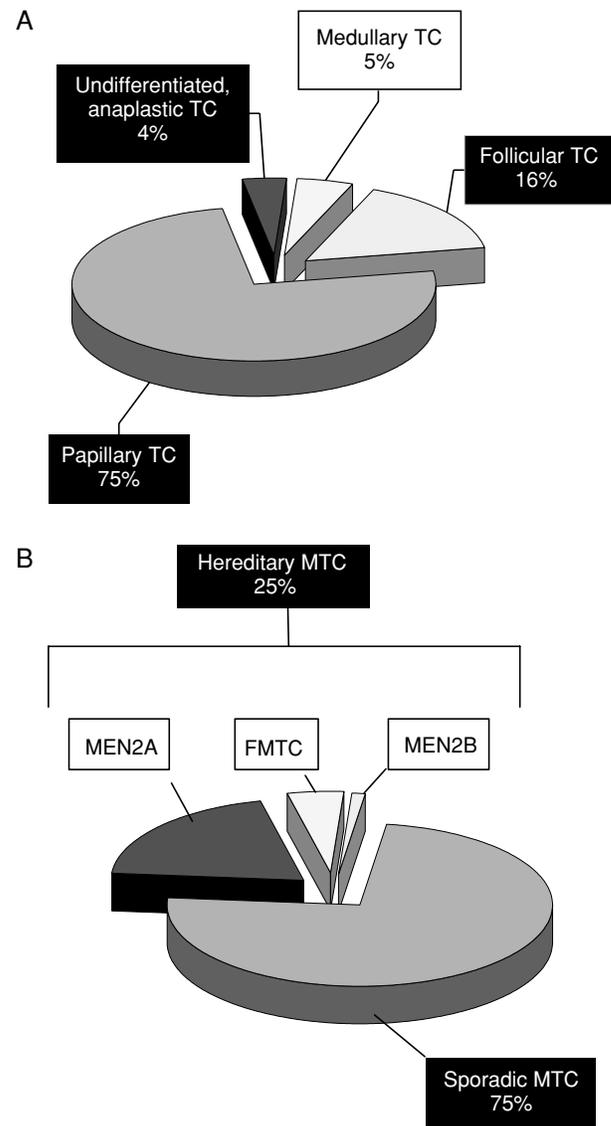
**Figure 1** Schematic presentation of the origin of medullary thyroid carcinoma. (A) C cells are located below the basal membrane (arrow). (B) The diagnosis of C-cell hyperplasia is based on the presence of an increased number of C cells (i.e.,  $\geq 7$  cells per follicle). (C) In medullary thyroid carcinoma, proliferating C cells are growing through the basal membrane.

11 (11p15.3–15.5). The primary transcript (pre-mRNA) of this gene is subject to tissue-specific alternative splicing (see Fig. 3). In the thyroid C cells, CT-mRNA of approximately 1000 nucleotides (exons 1–3 and 4) encodes the CT precursor (prepro-CT) of 141 amino acids. In neuronal cells, exons 1 to 3 are coupled to exons 5 and 6, generating an mRNA of approximately 1100 nucleotides encoding the 128-amino acid precursor for calcitonin gene-related peptide-I (CGRP-I). CGRP-I is widely distributed in the nervous system and particularly in the primary sensory neurons, where it functions as a neurotransmitter.

In 1985, a related gene was identified that encodes a polypeptide highly homologous to CGRP-I, named CGRP-II, but not encoding a CT-like molecule. These two CALC genes (CALC for calcitonin complex) and a pseudogene (CALC-III) all are located in the same region of chromosome 11, suggesting that local duplication of a common ancestral gene is the mechanism of their origin. In 1988, a fourth member of the CALC gene family was identified encoding the constituent protein of amyloid fibrils present in pancreatic islets of patients with diabetes mellitus type 2; therefore, it was named islet amyloid polypeptide (IAPP).

Another member of the CT superfamily was discovered in the medulla of adrenal glands and was named adrenomedullin (ADM). This protein functions as a vasorelaxant, and the ADM gene can be considered as the fifth member of the CALC gene family (see Fig. 4).

The members of the CT superfamily act via heterotrimeric G protein-linked, seven-transmembrane domain receptors coupled to cyclic AMP (cAMP) formation and phospholipase C activation. Two isoforms of the human CT receptor have been identified, named hCTR1 and hCTR2. The CGRP receptor

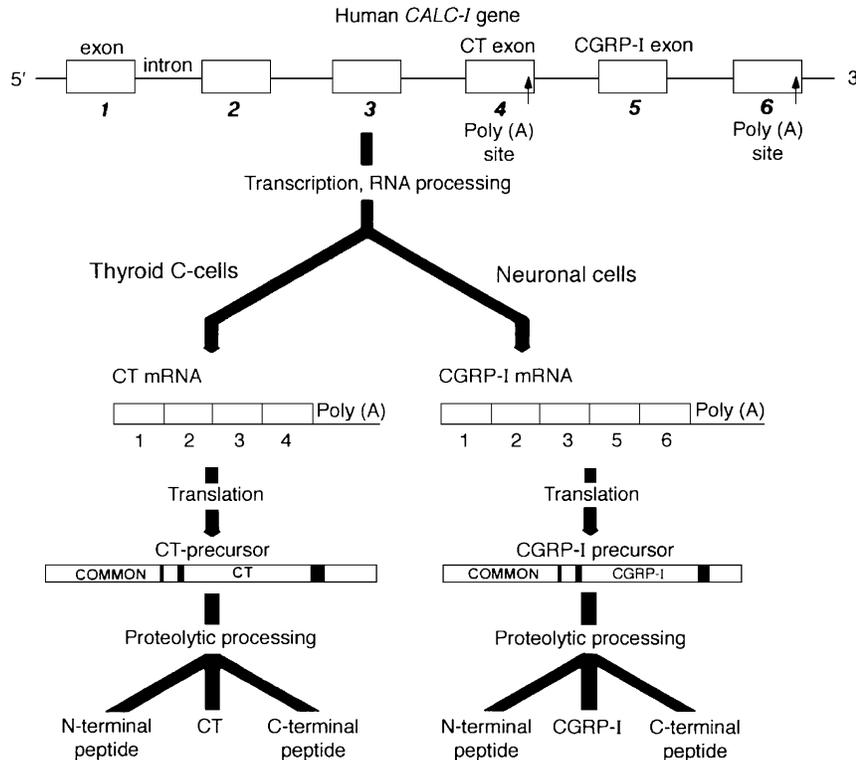


**Figure 2** (A) Classification and relative frequency of thyroid carcinoma (TC). (B) Sporadic versus hereditary medullary thyroid carcinoma (MTC).

has approximately 55% homology with hCTR2 and was named CT-receptor-like receptor (CRLR). Recently, receptor activity-modifying proteins (RAMPs) have been identified; these are single-transmembrane domain proteins acting as coreceptors that specify selectivity of hCTR2 for CT or IAPP and selectivity of CRLR for CGRP or ADM.

## MTC AND THE RET PROTO-ONCOGENE

MEN 2 is an autosomal-dominantly inherited disease caused by activating missense mutations in the RET

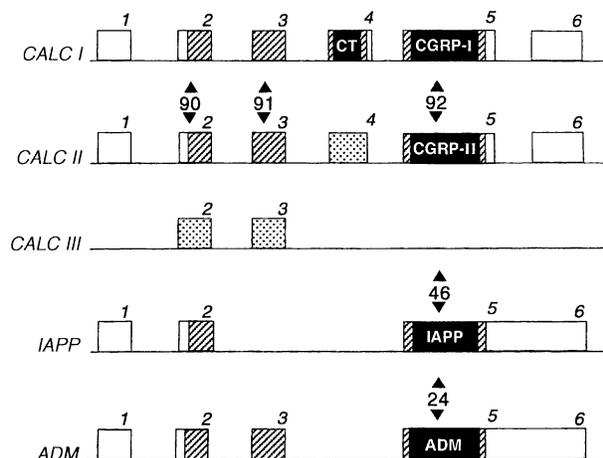


**Figure 3** Alternative expression of the *CALC-I* gene. Schematic representation (not drawn to scale) depicts the human *CALC-I* gene, consisting of six exons and five introns and illustrating the generation of calcitonin (CT) mRNA, calcitonin gene-related peptide-I (CGRP-I) mRNA, and their encoded polypeptides. Exons 1 and 6 are noncoding, exons 2 and 3 encode the amino-terminal polypeptide common to the CT precursor (pro-CT) and CGRP-I precursor, and exons 4 and 5 contain the sequences encoding CT and CGRP-I, respectively, plus the carboxy-terminal polypeptides of the CT- and CGRP-I precursors. In thyroid C cells, exons 1 to 3 are coupled to exon 4, generating CT mRNA, whereas in neuronal cells, exons 1 to 3 are coupled to exons 5 and 6, generating CGRP-I mRNA.

proto-oncogene. The RET protein is a member of the family of receptor tyrosine kinases, which are cell surface glycoprotein receptors that transduce signals for cell growth and differentiation. The RET protein consists of an extracellular part with a ligand-binding domain, a cadherin ( $\text{Ca}^{2+}$ -dependent cell adhesion)-like domain, and a cysteine-rich domain close to the cell membrane. It has a single-transmembrane domain and an intracellular part with two tyrosine kinase subdomains: TK1 and TK2. Activation of RET occurs by the binding of one of its four ligands: glial cell line-derived neurotrophic factor (GDNF), neurturin (NTN), artemin (ARTN), or persephin (PSPN), each of which requires its specific coreceptor GDNF family receptor- $\alpha$ -1,-2,-3, or-4 (GFR- $\alpha$ -1, GFR- $\alpha$ -2, GFR- $\alpha$ -3, or GFR- $\alpha$ -4), respectively (Fig. 5). Interaction of these molecules results in dimerization of RET, cross-autophosphorylation, and subsequent phosphorylation of intracellular substrates. RET is expressed in many cell types, including the C cells of

the thyroid gland and adrenal medulla, and in neurons.

The MEN 2 phenotype is caused by missense mutations affecting cysteine codon 609, 611, 618, or 620 in exon 10 or cysteine codon 630 or 634 in exon 11 of the RET gene (Fig. 6). Mutations in the cysteine codons in exon 10 can also result in FMTC. If cysteine codon 634 in exon 11 is mutated, all organ manifestations of MEN 2A occur, that is, MTC, pheochromocytoma, and hyperparathyroidism. These extracellular MEN 2A/FMTC cysteine mutations lead to a ligand-independent dimerization of receptor molecules and constitutive activation of intracellular signaling pathways. The MEN 2B-associated tumors are caused by mutations in the intracellular TK2 domain (in 95% of the cases involving amino acid 918 and in 5% of those involving amino acid 883 or 922). In more than 50% of cases of MEN 2B with RET codon 918 affected, mutations occur as new (*de novo*) germline mutations.



**Figure 4** The CALC family. Schematic comparison depicts the structure of the CALC-I to -III genes, the islet amyloid polypeptide (IAPP) gene, and the adrenomedullin (ADM) gene. Exons are represented by numbered boxes, with the hatched parts indicating the protein-encoding parts. Shown in black are the parts encoding mature calcitonin (CT), calcitonin gene-related peptide (CGRP), IAPP, and ADM molecules. Spotted boxes indicate "pseudo-exons" with homology to exon 2, 3, or 4 of the CALC-I gene. Percentages of identity in amino acid sequences with those of polypeptides derived from the corresponding regions of the CALC-II gene are indicated.

Approximately 75% of all MTCs are sporadic and solitary in origin (Fig. 2B). In approximately 40% of such tumors, a somatic mutation at codon 918 of the RET gene is present. In 15% of sporadic MTCs, a specific germline variant ( $-193C > G$ ) in the gene encoding the RET coreceptor GFR- $\alpha$ -1 was present, and this was associated with increased expression of GFR- $\alpha$ -1.

#### NATURAL HISTORY OF MTC, GENOTYPE-PHENOTYPE CORRELATION, AND CLINICAL COURSE OF THE DISEASE IN FAMILIES UNDER SURVEILLANCE

C-cell hyperplasia is generally considered as a precursor to MTC in patients with hereditary MTC. The diagnosis of C-cell hyperplasia is based on microscopy criteria: the presence of an increased number of diffusely scattered C cells ( $\geq 7$  per thyroid follicle), clusters of C cells, or 20 or more C cells per (visual) field at a magnification of  $\times 200$ . MTC is diagnosed when nests of C cells appear to extend beyond the basement membrane and to infiltrate and destroy thyroid follicles.

In investigating the natural history of the disease, large MEN 2 families have to be studied over six or

seven generations retrospectively, so that members who were obligate disease gene carriers and who had no intervention can be identified. Extensive studies on large families reveal that there is a clear genotype-phenotype correlation. In four large MEN 2A families with a RET codon 634 mutation, before clinical screening started, the average life span of obligate disease gene carriers (who died of extensive MTC metastases) was 48 years, compared with 60 years in FMTC families with a cysteine codon 618 mutation of the RET gene. Pheochromocytomas and parathyroid gland adenoma do occur in apparent FMTC families, but they occur less often than in MEN 2 families. Thus, although the distinct clinical course of the disease in a family (frequency of pheochromocytoma and parathyroid gland adenoma and life expectancy) depends primarily on the specific RET gene germline mutation that is present, screening for other endocrinopathies is mandatory in all families with MTC.

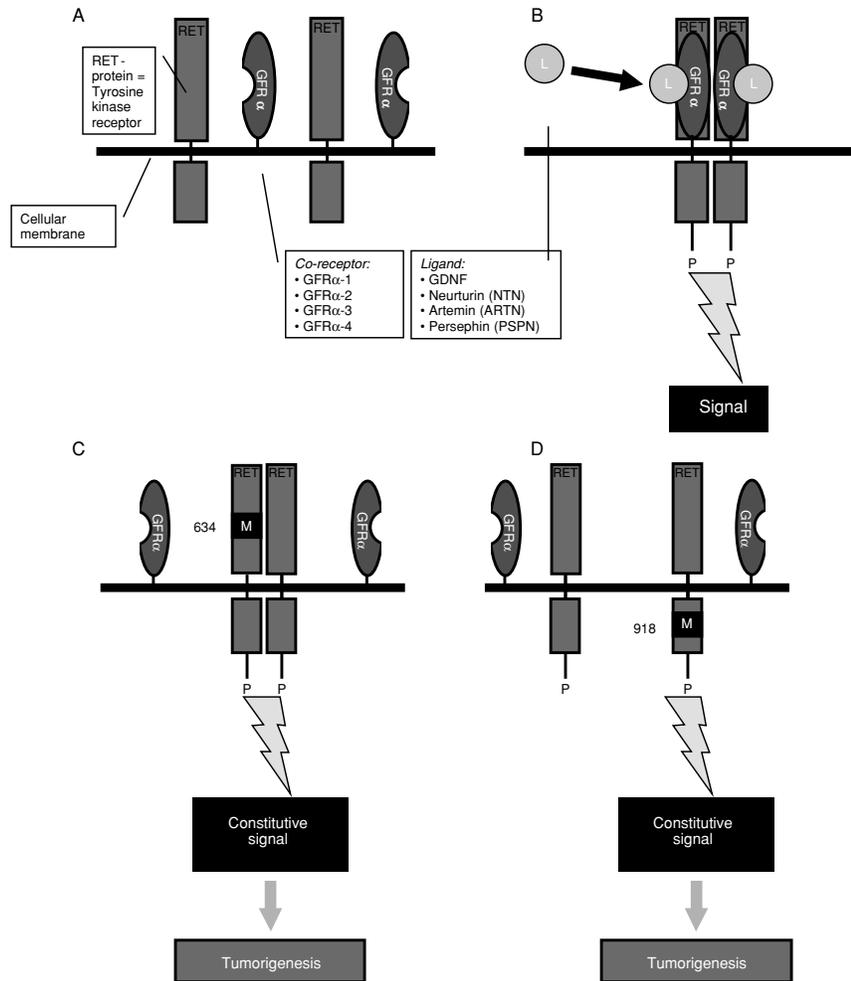
Periodic examination of family members revealed that the age of conversion from normal to elevated plasma CT levels during a C-cell stimulation test was 18 to 31 years (mean 23 years) in FMTC families, compared with 6 to 33 years (mean 16 years) in MEN 2 families.

Although the course of disease in FMTC families with a cysteine codon 618 mutation in the RET proto-oncogene is milder than that in MEN 2A families with a cysteine codon 634 mutation, patients with extensive metastases and an aggressive and unfavorable course of the disease may occur in all such families, irrespective of the nature of the RET mutation. Progression of the disease probably depends on additional somatic mutations in other proto-oncogenes and/or tumor suppressor genes. Individual habits, such as diet and smoking as well as the environment and pollution, may promote the occurrence of these additional mutations.

There is greater morbidity and mortality in MEN 2B than in MEN 2A. The survival curves of patients with MEN 2B are similar to those of patients with sporadic MTC who have somatic RET mutations identical with the most common germline mutations causing MEN 2B.

#### CT MEASUREMENT IN PATIENTS WHERE C-CELL HYPERPLASIA AND/OR MTC ARE SUSPECTED

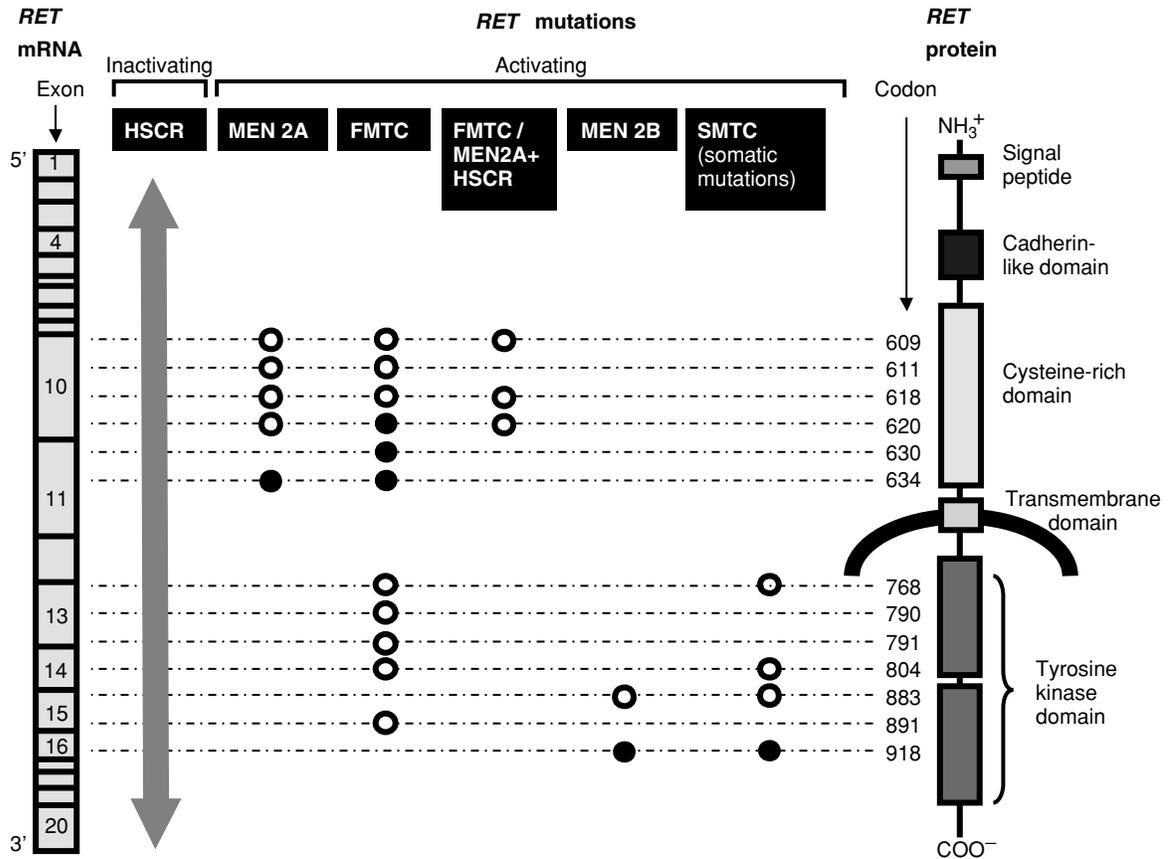
The main secretory product of MTC is CT; therefore, CT can be used as a tumor marker. In all patients with



**Figure 5** Model for the function of the normal (A, B) and MEN 2-mutated (C, D) RET tyrosine kinase receptor-protein complex. (A) In the absence of a ligand (L), both tyrosine kinase receptor RET and coreceptor GFR $\alpha$  are present in a nondimerized form. Activation of the RET protein and of the tyrosine kinase signal transduction pathway does not occur. (B) If a ligand (a member of the GDNF family of neurotropic peptides) is available, interaction with its specific coreceptor will occur. Consequently, the coreceptor will interact with the RET protein. Subsequently dimerization and autophosphorylation of the RET protein occurs, and the tyrosine kinase signal transduction pathway will be activated. As a result, signals for cell growth and differentiation are transduced to the nucleus. (C) Specific mutations (e.g., Cys634Arg in the extracellular cysteine-rich domain of the RET protein) may promote dimerization of the RET protein, independent of the binding of a ligand with its coreceptor. The RET mutation results in continuous autophosphorylation and activation of the tyrosine kinase signal transduction pathway. (D) Other specific mutations (e.g., Met918Thr in the intracellular tyrosine kinase domain of the RET protein) cause autophosphorylation and activation of the tyrosine kinase signal transduction without dimerization of the RET protein. GDNF, glial cell line-derived neurotropic factor; GFR $\alpha$ , GDNF family receptor  $\alpha$ ; P, phosphate residue.

progressive disease, both basal and stimulated plasma CT levels are elevated, whereas in patients with early MTC, CT levels may be elevated only after stimulation of CT release from the C cells. The calcium sensor is expressed in normal C cells; therefore, calcium may regulate secretion of CT. Other CT secretagogues include glucocorticoids, CGRP, glucagon, enteroglucagon, gastrin, pancreaticozym, and beta-adrenergic agents. There is a wide range of C-cell

stimulation test procedures. Results depend on the stimulus used (calcium and/or pentagastrin [PG], the dose, the period of infusion, and the time point of sampling). The standard provocative tests use PG (0.5  $\mu\text{g}/\text{kg}$  body weight intravenously over 5 s) or a rapid infusion of calcium gluconate (2.5 mg/kg over 30 s). Blood samples are obtained at baseline and 1, 2, and 5 min after the stimulus. Some experts believe that the two tests should be combined for maximal



**Figure 6** Genotype–phenotype correlations in MTC associated with mutations in the RET proto-oncogene. Figure depicts an overview of mutations in the RET proto-oncogene and associated diseases. Structure of the RET mRNA (left) and the RET protein (right) are depicted schematically (exons and domains are not to scale). For each disease, the responsible mutations and their locations are depicted. Closed circles represent the most frequently observed mutations, whereas open circles represent mutations that are less frequently observed.

sensitivity. In addition, the test results depend on the age, sex, and condition of the research participant tested (fasting state). All infusion procedures may give false-negative results, and with none of the procedures is it possible to recognize very early MTCs, that is, to discriminate between C-cell hyperplasia and micro-MTC. False-positive results (i.e., of research participants without MTC but only C-cell hyperplasia) also occur. In progressive disease, anaplastic dedifferentiation of MTC may decrease CT production and, as such, the utility of CT as a tumor marker.

**OTHER TUMOR MARKERS FOR MTC**

Besides CT, a number of other substances may be produced by MTC (e.g., carcinoembryonic antigen [CEA], katalcalcin [PDN-21, the carboxy-terminal peptide of the CT precursor], CGRP, IAPP, chromogranin

A, neuron-specific enolase [NSE], somatostatin, adrenocorticotrophic hormone [ACTH], ghrelin), but they are of little value in the diagnosis of MTC. Some MTC metastases do not produce either CT or CEA.

**CT MEASUREMENT COMPARED WITH DNA ANALYSIS IN FAMILIES WITH MTC**

DNA analysis for the detection of mutations in the RET gene (e.g., by direct nucleotide sequence analysis or mutation-scanning techniques such as single-strand conformation polymorphism [SSCP] analysis) is a highly specific method for the identification of individuals with familial MTC. Because of the high specificity and sensitivity of mutation detection in the RET gene, therapeutic decisions can be based on the results of

**Table I** Criteria for DNA Analysis of the RET Gene

• Patients with clinical indication for MEN 2 syndrome or FMTC (bilateral medullary thyroid carcinoma and/or pheochromocytoma)
• Patients with MTC or pheochromocytoma and a family member with MTC or pheochromocytoma
• First-degree family member of a carrier of a MEN 2-specific germline mutation in the RET gene
• Patients with apparently sporadic MTC and an increased risk for MEN 2:
Younger than 35 years of age
Bilateral MTC
MTC and pheochromocytoma

DNA analysis even in asymptomatic family members with negative C-cell stimulation test results. RET mutation analysis is very useful as a means of excluding genetic susceptibility to the development of MTC and, therefore, as a means of ensuring unaffected family members that further screening with biochemical tests can be abandoned safely (see [Table I](#)).

The use of the DNA test has revealed the limitations of the C-cell stimulation test. DNA analysis can identify patients who have MTC but who do not (yet) have elevated plasma CT levels in C-cell stimulation tests. We have identified a number of young MEN 2 gene carriers with MTC, and not just C-cell hyperplasia, who had normal CT levels. On the other hand, some family members who were later proved not to be disease gene carriers had positive C-cell stimulation tests and so underwent total thyroidectomy. They had C-cell hyperplasia, which occurs in normal individuals at a frequency of approximately 5% (i.e., false-positive C-cell test results). Thus, this plasma CT response to C-cell stimulation does not always distinguish physiological C-cell hyperplasia from pathological micro-MTC, a finding that limits the value of the C-cell stimulation test so far as detecting the presence of MTC is concerned (see [Table II](#)).

## PREOPERATIVE SCREENING FOR MTC

In the assessment of at-risk individuals, the DNA test is the gold standard, and a positive result is a sufficient indication for recommending surgery. If an asymptomatic carrier rejects preventive treatment, periodic C-cell stimulation tests are indicated. However, the latter cannot discriminate between C-cell hyperplasia and micro-MTC. Waiting for conversion of stimulation test results from a normal plasma CT level to an elevated level involves the risk of residual carcinoma and/or metastases after surgery and recurrence of the disease.

## SURGERY OF MTC

The treatment for a patient with MTC is surgery. The appropriate procedure is total thyroidectomy and careful lymph node dissection of the central compartment of the neck. Disease gene carriers may be divided into three risk groups, depending on the nature of the predisposition to familial MTC. The highest risk concerns children with a predisposition to MEN 2B, most commonly associated with a germline RET mutation in codon 883, 918, or 922. Total thyroidectomy and central neck dissection during the first year of life are advocated by some because early metastasis of MTC has been reported, but others prefer to delay surgery until these children are older (approximately 5 years of age). The second highest risk group includes carriers in kindreds with a mutation in codon 611, 618, 620, 634, or 891. Such family members should undergo a total thyroidectomy, as well as a central node dissection, before 6 years of age. A lower risk is involved in carriers with a codon 609, 768, 790, 791, or 804 germline mutation. In general, MTCs in these patients develop at a later age, grow more slowly, and behave less aggressively. In this group, surgery may be postponed until an abnormal response to a calcium or PG stimulation test is observed (see [Table III](#)).

**Table II** Dependence of Periodic Clinical Examination Recommended to MEN 2 Disease Gene Carriers on the Specific DNA Mutation in the RET Gene

Type of tumor	Clinical test
<b>MTC</b>	
Prior to and after total thyroidectomy	Annually plasma CT levels
<b>Pheochromocytoma</b>	
RET codon 634 or 918	• From the age of 5 years: Annually metanephrines in 24-h urine
609, 611, 768, 804 or 891	• From the age of 10 years: Every 2 years metanephrines in 24-h urine
<b>Hyperparathyroidism</b>	
RET codon 634	• Annually Ca, chloride, phosphate
609, 611, 618, 620, 790 or 791	• From the age of 10 years: Every 2 years Ca, chloride, phosphate

Table III Dependence of Indication for Total Presymptomatic Thyroidectomy in MEN 2 or Familial MTC on the Specific DNA Mutation in the RET Gene

Degree of malignancy	Mutated RET codon	Age at surgery
1	883, 918, or 922	Early in the first year of life
2	611, 618, 620, 634, or 891	Before 6 years of age
3	609, 768, 790, 791, or 804	After abnormal C-cell stimulation test result

Classification of MTC is based on the pathological tumor, node, metastases system (pTNM). Patients with stage I (tumor restricted to the thyroid gland and diameter <1 cm) or II (tumor locally invasive and >1 cm) and complete tumor resection have the best prognosis and are at minimal risk for death from MTC. However, MTC readily invades the intraglandular lymphatics and spreads to other parts of the gland as well as to pericapsular and regional lymph nodes (stage III). Patients with nodal metastases have a high risk of recurrence of the disease. Patients with stage IV have distant metastases (lungs, liver, and bones) and a high risk of death from MTC (see Table IV).

## POSTOPERATIVE FOLLOW-UP

What degree of increase in plasma CT concentration must be considered as indicating metastases in patients after thyroidectomy? In MTC patients, preoperative CT levels appear to be predictive of tumor size and postoperative CT normalization. When preoperative CT peak levels after C-cell stimulation were less than 50 ng/L, only 1 of 45 patients had increased plasma CT levels after surgery. Interestingly, a correlation also was found between plasma CT peak levels and tumor size, defined as the largest diameter of the largest tumor, in hereditary MTC. Patients with undetectable basal and stimulated plasma CT levels within 24 h after surgery are those having the best prognosis, and they require no further treatment. Even after satisfactory postoperative results, all patients need lifelong follow-up, at least annually, for

monitoring thyroxin replacement therapy and measuring fasting (basal) CT levels. The question arises as to whether C-cell stimulation tests have to be performed lifelong. Despite a normal basal level, there may be an abnormal peak level after stimulation. On the one hand, it is evidence-based medicine that stimulated CT levels are a more sensitive indicator than are basal values. On the other hand, the burden of periodic stimulation tests and the limited clinical consequences of pathological results justify measuring only basal levels. In many patients, it appears that the basal CT levels become positive several years after surgery. Suspicion about residual or metastatic disease arises if basal CT levels exceed 1 ng/L. Over time, during follow-up, patients are their own controls. Patients with persistently elevated or rising plasma CT levels after adequate surgery should be thoroughly evaluated to define the extent of any local or distant disease. Ultrasound computed tomography or magnetic resonance imaging (MRI) can locate metastases. In addition, other imaging methods, such as [<sup>99m</sup>Tc]penta-mercaptosuccinic acid (V-DMSA), [<sup>123</sup>I]meta-iodobenzylguanidine (MIBG), and [<sup>111</sup>In]pentetreotide (related to octreotide) scintigraphy, may be helpful. If local disease in the neck region and/or mediastinum is found and there is no evidence of distant metastases, repeat operation is advocated despite the risk of injury to the nervus recurrence and vocal cord paralysis. Our criteria for reexploration of the neck region are as follows. First, basal plasma CT levels are clearly increased (i.e., >15 ng/L). Second, basal plasma CT levels increase over time. Third, suspected local neck or mediastinal lymph nodes are found, but no distant metastases are detectable by MRI, positron emission tomography (PET), or ultrasonography of the liver, and cytology results of liver specimens are negative.

In patients with evidence of surgically incurable disease (stage IV), monitoring of plasma CT and radiological examination are required. Many such patients show a remarkable stabilization of the disease, and so no further treatment is recommended. A restraint policy (i.e., a “wait and see” approach) is advocated. In patients whose disease shows rapid progress,

Table IV Staging of Medullary Thyroid Carcinoma

Stage	Characteristic features
I	Tumor < 1 cm and restricted to thyroid gland
II	Tumor > 1 cm and/or local invasion
III	Regional lymph node(s) positive
IV	Metastases at distance

*Note.* The highest stage always is applied (e.g., primary tumor <1 cm and regional lymph nodes positive means stage III) Tumors larger than 1 cm are almost always locally invasive.

intervention with cytostatic therapy may be considered. Inhibitors of receptor tyrosine kinase signaling are being developed. These drugs are called “tyrphostins” because they inhibit phosphorylation on tyrosine residues by competition with a tyrosine kinase for substrate and adenosine triphosphate (ATP) binding. Specific inhibition of the RET receptor tyrosine kinase by such drugs might constitute a novel therapeutic approach for MTC that should be investigated.

### GENETIC TESTING OF FIRST-DEGREE RELATIVES OF PATIENTS WITH HEREDITARY MTC

If a RET gene germline mutation is identified in a family, testing of young children is appropriate at any time after birth provided that it is preceded by counseling of the parents. Antenatal testing is technically possible. Would-be mothers should be referred to a clinical genetics center before they are pregnant. Options such as *in vitro* fertilization and preimplantation genetic diagnosis may be considered. However, because early treatment and cure are possible, one may question whether this is advisable and ethically acceptable.

If a RET mutation has not been identified in a family and no affected family member is available to give blood for RET testing, stored histological specimens of affected family members should be sought to provide tissue for DNA extraction and analysis. If suspicion of hereditary disease arises and even this alternative is not available, testing of asymptomatic family members may be considered to reveal a germline RET mutation. If no RET germline mutation can be identified in a MEN 2 family, this negative result cannot be accepted unless all exons of the RET gene are examined by nucleotide sequence analysis. In addition, one has to review the clinical evidence for MEN 2. If this evidence is strong, mutation analysis on a new blood sample has to be repeated. In an extensive family, it makes sense to consider genetic linkage analysis. If positive for the chromosome 10 region harboring the RET gene, all RET exons have to be sequenced again.

### APPARENTLY SPORADIC MTC AND SCREENING OF FAMILY MEMBERS OF PATIENTS WITH APPARENTLY SPORADIC MTC

Indications for testing for hereditary MTC among patients with apparently sporadic MTC are age at diagnosis of less than 35 years, bilateral or multicentric

involvement, and a family history that is positive or suspected. All patients with MTC for whom there is clinical suspicion of familial disease should be tested for germline RET gene mutations in blood DNA. All exons should be sequenced, starting with exons 10, 11, 13, 14, 15, and 16. If no mutation is found, only a small risk of hereditary MTC remains. In this case, tumor DNA can be analyzed. In sporadic MTC, somatic mutations in codon 883 or 918 occur relatively frequently and identification of one of these mutations argues against, but does not exclude, hereditary disease. Conversely, mutations in cysteine codons in exon 10 or 11 are uncommon in sporadic MTC and are suggestive of hereditary disease.

Before blood samples are taken for DNA analysis, detailed information about the consequences of DNA analysis must be provided to the patient. Psychological support may well be needed before DNA test results are disclosed and during follow-up after diagnosis. Written consent must be obtained.

### SUMMARY AND CONCLUSIONS

All individuals with apparently sporadic MTC, but in whom there is suspicion of familial disease, should have mutation analysis of the RET gene. A negative DNA result practically excludes the possibility of hereditary MTC in families where an index case has been investigated and obviates the need for further biochemical evaluation. In families with hereditary MTC, RET gene mutation analysis has superseded measurement of plasma CT in the detection of asymptomatic disease gene carriers. Disease gene carriers may be divided into three distinct risk groups depending on the specific RET gene mutation in the family. The age at which presymptomatic surgery has to be performed depends on the risk group to which the patient belongs. Compared with the results of DNA analysis, the results of CT stimulation tests have become less important in the assessment of the time point of surgery. During follow-up of patients who underwent surgery, measurement of basal plasma CT is still useful. The high sensitivity of measuring stimulated CT levels does not outweigh the burden of lifelong periodic stimulation tests and the limited clinical consequences of slightly elevated levels. Stimulation tests are inevitable for persons at risk who do not want genetic testing.

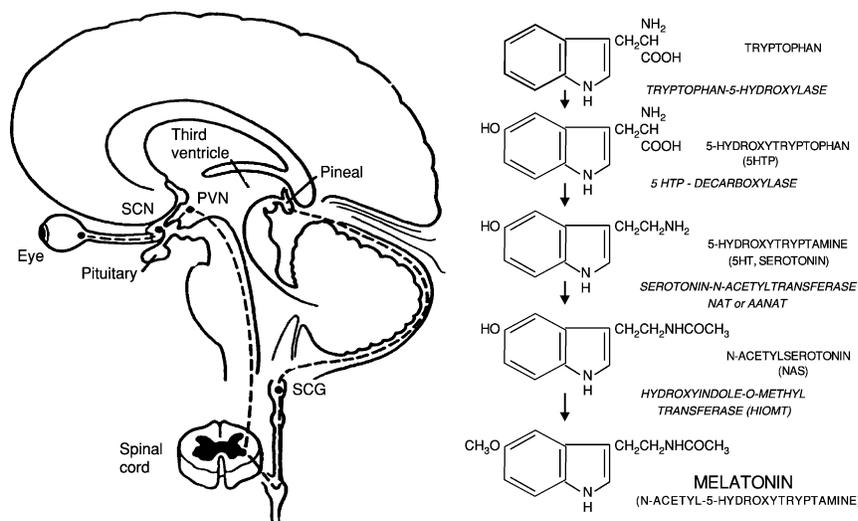
### See Also the Following Articles

Calcitonin, Overview • Parathyroid Cancer • Pituitary Adenomas, TSH-Secreting • Thyroid Carcinoma • Thyroid

Disease, Genetic Factors in • Thyroidectomy • Toxic Adenoma

### Further Reading

- Hansford, J. R., and Mulligan, L. M. (2000). Multiple endocrine neoplasia type 2 and RET: From neoplasia to neurogenesis. *J. Med. Genet.* **37**, 817–827.
- Lips, C. J., Hoepfener, J. W., and Thijssen, J. H. H. (2001). Medullary thyroid carcinoma: Role of genetic testing and calcitonin measurement. *Ann. Clin. Biochem.* **38**, 168–179.
- Lips, C. J. M., Landsvater, R. M., Höppener, J. W. M., Geerdink, R. A., Blijham, G. H., Jansen-Schillhorn van Veen, J., *et al.* (1994). Clinical screening as compared with DNA analysis in families with multiple endocrine neoplasia syndrome type 2A. *N. Engl. J. Med.* **331**, 828–835.
- Moers, A. M. J., Landsvater, R. M., Schaap, C., Jansen-Schillhorn van Veen, J. M., de Valk, I. A. J., Blijham, G. H., *et al.* (1996). Familial medullary thyroid carcinoma: Not a distinct entity? Genotype–phenotype correlation in a large family. *Am. J. Med.* **101**, 635–641.
- Utiger, R. D. (1994). Medullary thyroid carcinoma, genes, and the prevention of cancer [editorial]. *N. Engl. J. Med.* **331**, 870–871.



**Figure 1** Melatonin synthesis and the principal neural pathways innervating the pineal gland. SCN, suprachiasmatic nucleus; PVN, paraventricular nucleus; SCG, superior cervical ganglion. Left panel reproduced from Tamarkin, K., Baird, C. J., and Almeida, O. F. X. (1985). Melatonin: A coordinating signal for mammalian reproduction. *Science* **227**, 714–720.

most circumstances, the rate-limiting step in synthesis (Fig. 1) is the activity of serotonin-N-acetyltransferase (arylalkylamine N-acetyl transferase, AA-NAT), with a major increase (7 to 150-fold) in the activity during the dark phase. Serotonin availability may also play a role. The rhythm of melatonin production is endogenous, being generated by clock genes in the suprachiasmatic nuclei (SCN), the major central rhythm-generating system or “clock” in mammals. (The pineal gland itself and the retina are self-sustaining “clocks” in some if not all lower vertebrates.) The rhythm is synchronized to 24 h primarily by the light–dark cycle acting via the retina and the retinohypothalamic projection to the SCN (Fig. 2).

The cDNAs encoding both AA-NAT and the O-methylating enzyme hydroxyindole-O-methyltransferase (HIOMT) (Fig. 1) have been cloned. It is likely that the human enzyme is regulated primarily at a posttranscriptional level, whereas in rodents the key event appears to be cyclic AMP (cAMP)-dependent phosphorylation of a transcription factor that binds to the AA-NAT promoter. Rapid decline in activity with light treatment at night appears to depend on proteasomal proteolysis.

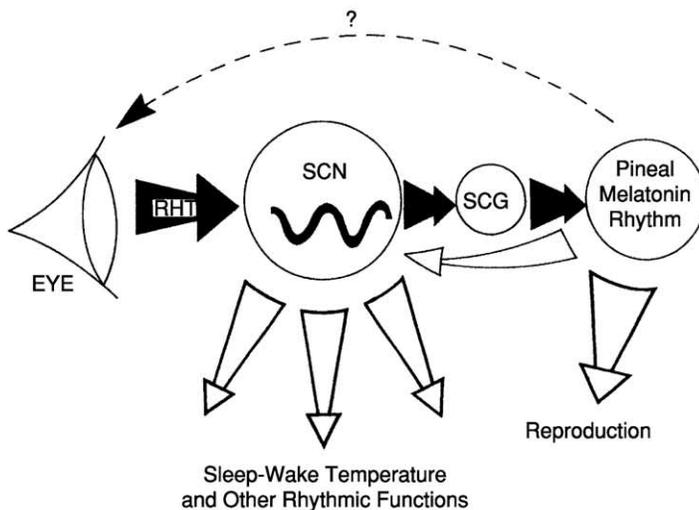
In humans and rodents, melatonin is metabolized to 6-sulphatoxymelatonin (aMT6s), primarily within the liver, by 6-hydroxylation followed by sulphate conjugation. A number of minor metabolites are also formed, including the glucuronide conjugate. Exogenous oral (fast release) or intravenous melatonin

has a short metabolic half-life (20 to 60 min, depending on author and species), with a large hepatic first-pass effect and a biphasic elimination pattern.

Very large interindividual variations in melatonin and aMT6s production are seen with intraindividual stability of both amplitude and timing. This stability leads to the extensive use of melatonin in plasma or saliva and of aMT6s in urine as marker rhythms for circadian phase, for example, in investigating sleep disorders and evaluating adaptation to abrupt phase shifts as in shift work and jet lag. The large individual variations have been ascribed to the size of the pineal gland rather than to variations in enzymic activity.

## PHYSIOLOGICAL ROLE OF THE PINEAL GLAND

The pineal is essentially part of the visual system. Probably the most famous of ancient texts on the pineal is that of the influential French philosopher Descartes. In 1662, Descartes considered that, by movements, the pineal controlled the flow of “animal spirits” into motor nerves and so influenced movements of the body. He thought that the stimulus for pineal function came from visual input to the retina (Fig. 2). The latter is a most remarkable insight given that this is effectively true today. The concept of the pineal as the seat of the soul is usually attributed to Descartes, although this idea probably derives from Herophilos. In lower vertebrates, the pineal is directly



**Figure 2** The French philosopher Descartes considered, during the 17th century, that the pineal was controlled by light input from the eyes, and today we know that light does indeed regulate the timing of the pineal melatonin rhythm by its synchronizing effects on the activity of the central circadian clock in the suprachiasmatic nucleus (SCN). Light (daylength) also regulates the duration of melatonin secretion at night. In many lower vertebrates, the pineal is effectively a “third eye,” registering overall illuminance and acting itself as a central circadian clock.

photoreceptive and often functions as a “third eye” measuring overall illuminance in the environment for the organization of circadian and seasonal functions. Pineal glands of lower vertebrates are innervated by both afferent and efferent fibers. In reptiles and birds, the gland has a mixed photoreceptor and secretory function. Direct neural connections to the brain exist in mammals, but the sympathetic nervous system provides the main input (Fig. 1). Mammalian pineals consist largely of pinealocytes and glia.

The mammalian pineal is secretory and has lost its capacity for direct photoreception. Remnants of its ancient visual role remain in the presence of, for example, opsins in pinealocytes. However, light indirectly controls mammalian melatonin production via the retino-hypothalamic projection (RHT) and the SCN.

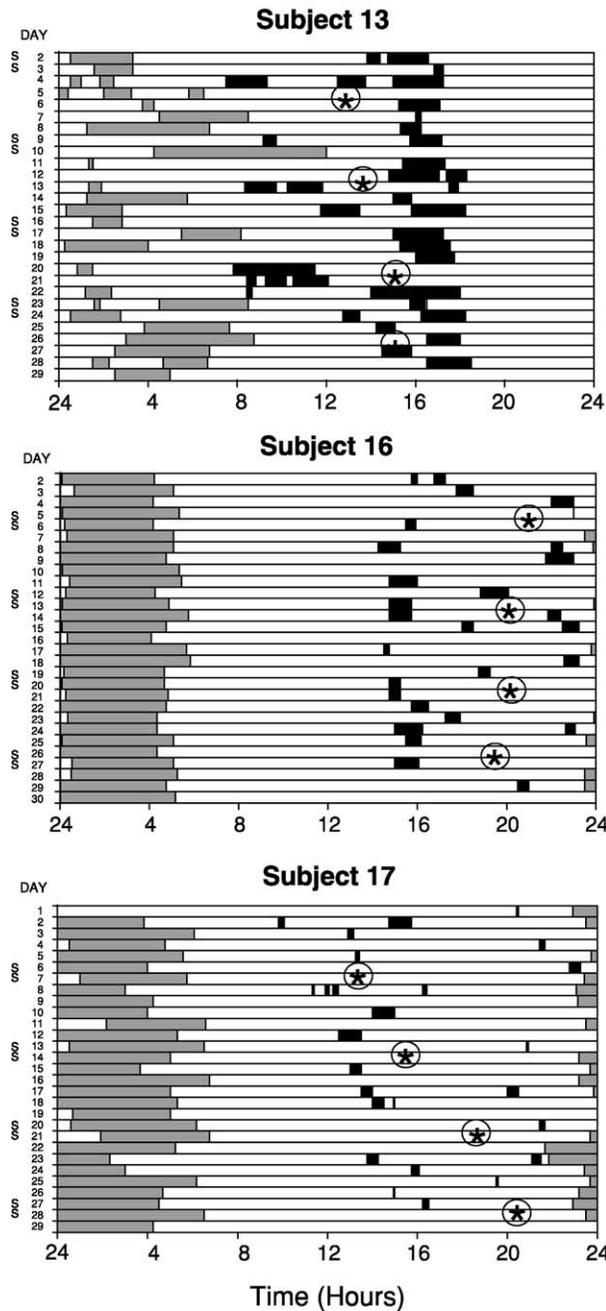
## MELATONIN SECRETION CHARACTERISTICS

In the absence of time cues (e.g., light–dark alternation), the melatonin rhythm, like all circadian rhythms, “free runs”; that is, it assumes a period that

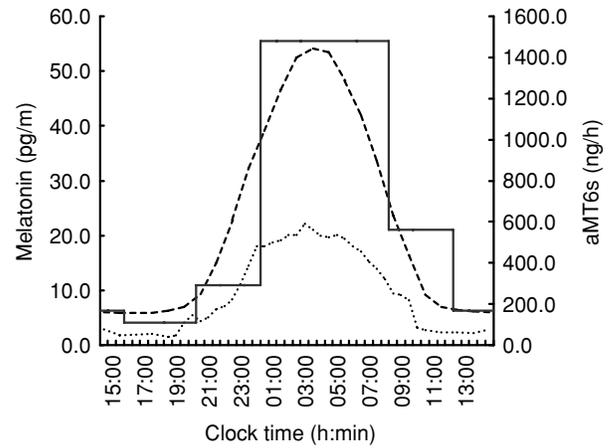
is individually variable and genetically determined, usually somewhat longer than 24 h (on average about 24.3 h). Time cues synchronize or “entrain” the rhythm to 24 h. The most potent time cue is the light–dark cycle, and many blind people with no light perception at all show “free-running” melatonin and other rhythms (e.g., sleep, cortisol, core temperature) in a normal environment (Fig. 3).

In addition to entraining the rhythm, daylength (photoperiod) determines the duration of nighttime secretion both by direct suppression of melatonin and by determining the length of the signal emitted by the SCN. Light of suitable intensity, duration, and spectral quality suppresses melatonin production at night; short wavelengths (approximately 465 nm) are most effective. It is likely that short wavelengths are also most effective for entrainment of the rhythm to 24 h.

In mammals, melatonin is secreted into the bloodstream and also probably into the cerebrospinal fluid (CSF). CSF concentrations are reported to be higher than those of plasma in some species. In humans, the evening melatonin onset is usually around 2100 to 2200 h, it peaks around 0300 to 0500 h, and it declines to daytime values by around 0800 to 0900 h (Fig. 4).



**Figure 3** Three blind subjects (no perception of light) showing abnormal sleep and melatonin rhythms. Subjectively recorded sleep is shown by the gray bars, and naps (classified as such by the subjects) are shown by the black bars. Melatonin was assessed as 6-sulphatoxymelatonin (aMT6s) in sequential urine collections for 48 h weekly, and the asterisks indicate the calculated peak times (acrophase) of the aMT6s rhythms. Note the abnormal timing of aMT6s, which normally peaks between 2400 and 0600 h, and the association of naps with daytime melatonin production. Subject 17 is clearly “free running” (i.e., showing endogenous periodicity), with a steady daily delay of the aMT6s rhythm. All blind subjects with no eyes have shown this phenomenon. Subject 16 is synchronized to 24 h but with an abnormal phase position. Cortisol and core



**Figure 4** Average concentrations of melatonin in plasma (dashed line, average  $N=133$ ), saliva (dotted line, average  $N=28$ ), and 6-sulphatoxymelatonin (aMT6s) in urine (solid line, average  $N=88$ ), with all measurements done by radioimmunoassay. Diagrammatic representation of mean normal values (healthy men and women over 18 years of age) from the author’s laboratory.

The effective cessation of secretion is earlier, according to various models of the secretion characteristics. The melatonin onset and offset correspond to a number of important events related to biological dawn and dusk, including the evening increase in sleep propensity and decline in core body temperature. During winter, particularly in polar regions, the timing of the rhythm may be delayed compared with summer, and this has been attributed to the weaker light–dark cycle during winter.

### MELATONIN RECEPTORS

Melatonin receptors have now been cloned, and three subtypes were initially named Mel-1a, Mel-1b, and Mel-1c. The Mel-1a receptor gene has been mapped to human chromosome 4q35.1. Its primary expression is in the pars tuberalis of the pituitary and the SCN. Mel-1b has been mapped to chromosome 11q21–22, and its expression is in the retina and the brain. Mel-1c is not found in mammals. Two cloned mammalian receptors (Mel-1a and Mel-1b) have recently been

body temperature rhythms show abnormalities similar to those of melatonin. These and other observations in the blind underline the importance of light for synchronizing rhythms and the association of daytime melatonin production with poor nighttime sleep and daytime naps. Reproduced with permission from Lockley, S. A., Tabandeh, H., Skene, D. J., Buttery, R., Bird, A. C., DeFrance, R., and Arendt, J. (1995). Daytime naps and melatonin rhythms in blind people. *Lancet* 346, 1491.

renamed MT1 and MT2. They are a new family of G protein-coupled receptors, have high affinity ( $K_d$  20–160 pmol), and inhibit forskolin-stimulated cAMP formation. Using gene knockout technology and pharmacological manipulations, the results suggest that the phase-shifting receptor is MT2, whereas MT1 is associated with acute suppression of SCN electrical activity in addition to its actions within the pars tuberalis. Several other physiological responses have been ascribed to MT1 and MT2 receptors, including (MT1) melatonin-mediated potentiation of adrenergic vasoconstriction and (MT2) modulation of dopamine release in the retina.

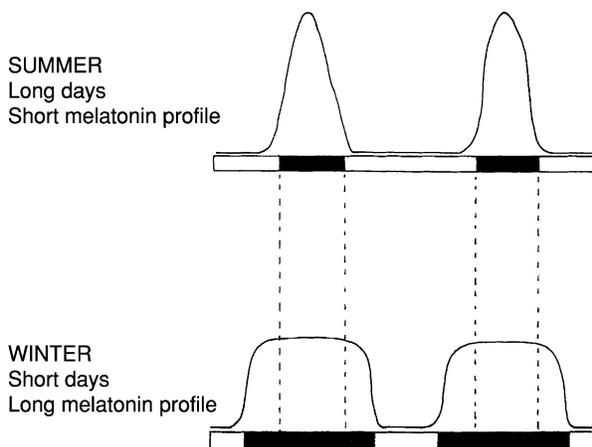
### MELATONIN AGONISTS AND ANTAGONISTS

Large numbers of putative and actual melatonin agonists, together with some antagonists, have now been described. The most interesting have effects similar to those of melatonin on rhythm physiology in both rodents and humans. Moreover, a possible therapeutic role for the agonist S20098 in depression is emerging. A recent review summarized melatonin receptor pharmacology to that point.

### PHYSIOLOGICAL ROLE OF MELATONIN IN MAMMALS

#### Photoperiodism

In all species studied, melatonin appears to act as a timekeeping hormone. It has receptors both within and without the central nervous system and can modify gene expression in certain target organs. Pinealectomy abolishes the ability to respond to changing daylength in photoperiodic seasonal breeders. In mammals, the primary role of melatonin is to provide information about photoperiod for the timing of seasonal functions in daylength-dependent (photoperiodic) seasonal species. These functions include timing of breeding, timing of pubertal development, and seasonal changes in coat growth, behavior, and body weight. The critical signal for the timing of photoperiodic events is the changing duration of melatonin secretion according to daylength, together with previous photoperiodic history (Fig. 5). Artificial generation of winter (long night) melatonin profiles during summer is equipotent with artificial long nights in changing the timing of seasonal functions. Melatonin implants (which are “read” as long nights) are used in agriculture to time the season of birth of,



**Figure 5** The duration of melatonin secretion changes with daylength, providing an internal time cue for the organization of daylength-dependent seasonal functions in photoperiodic species. This changing duration can be seen in humans if they are maintained in different durations of total darkness. However, domestic intensity light at night is sufficient to greatly diminish or eliminate this photoperiodic response.

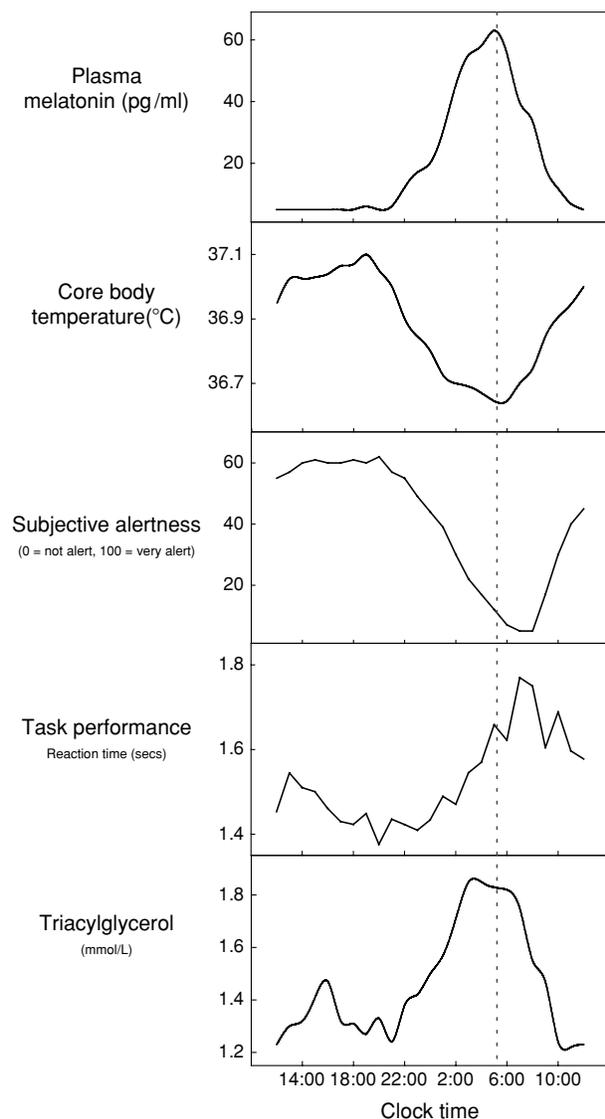
for example, lambs and the growth of winter coat of cashmere goats and mink.

Melatonin influences production of gonadotropins and gonadal hormones via actions within the hypothalamus. The precise receptors have not been fully defined. However, melatonin controls seasonal variations in prolactin by a direct action on the pars tuberalis of the pituitary. This structure is the major site of melatonin receptors (MT1) in most species. Photoperiod-dependent gene expression in the pars tuberalis is directly modified by melatonin.

#### Role in the Mammalian Circadian System

Pinealectomy in mammals leads to subtle effects on circadian rhythms. The speed of adaptation to forced phase shift is enhanced and the activity–rest cycle becomes fragmented in continuous light in rodents. There is good evidence for the importance of the maternal melatonin rhythm in setting circadian phase in neonatal rodents. This may be the most important role for melatonin in the mammalian circadian system given that an essential physiological role in adult mammals remains to be defined. The effects of pinealectomy in humans remain to be clarified; confounding factors are the possible concomitant damage of surrounding structures during surgery. However, suppression of melatonin by the beta-adrenergic antagonist atenolol leads to an enhanced rate of phase shift to applied light.

The peak nighttime levels of melatonin are temporally associated with the nadir in core body temperature (cBT), maximum sleepiness/fatigue, lowest alertness, increased blood lipid (triacylglycerol), glucose, and possibly insulin (Fig. 6). These are correlative associations; however, in the case of cBT and sleep propensity, there is good evidence that melatonin contributes to the nighttime decline and increase, respectively. Light suppression of melatonin at night leads to an increase in cBT and alertness. The increase in alertness correlates to the degree of melatonin



**Figure 6** Relationship of plasma melatonin to other major circadian rhythms. Note the close correspondence between the core temperature nadir and the melatonin peak. Reproduced with permission from Rajaratnam, S. M. W., and Arendt, J. (2001). *Lancet* 358, 999–1005.

suppression, but evidence for the quantitative contribution of endogenous melatonin to these effects is lacking. Exogenous melatonin clearly influences sleep. Possibly, the best evidence for a causal role of melatonin in human sleep comes from studies in free-running blind subjects. When they are out of phase (i.e., when melatonin is secreted during the daytime), the peak levels are strongly associated with an increase in daytime napping (Fig. 3). In general, melatonin reinforces functions associated with darkness. In nocturnal animals, its production is associated with activity rather than sleep.

## MELATONIN AND HUMAN REPRODUCTION

Melatonin declines during development, peaking at 3 to 4 years, with a plateau circa 18 to 35 years, followed by a decline into old age (although it should be noted that very healthy elderly subjects have not shown this decline). Early work suggested that melatonin might have a role in the timing of human puberty, acting in an inhibitory capacity. Certainly, exogenous melatonin is able to delay pubertal development in rats.

Pineal tumors are frequently associated with precocious/delayed puberty, but melatonin production has not been firmly implicated in normal or abnormal pubertal development in humans. Pineal tumors are heterogeneous and may arise from germ cells (teratomas, germinomas, choriocarcinomas, endodermal sinus tumors, and mixed germ cell tumors), pineal parenchymal cells (pineoblastoma and pineocytoma), and the supporting stroma (gliomas). All are rare (less than 1% of intracranial space-occupying lesions) and tend to occur in individuals under 20 years of age with the exception of parenchymal cell tumors, which occur equally in adults and children. Most of the evidence now suggests that precocious puberty is due to the production of human chorionic gonadotropin (beta-hCG) by germ cell tumors of the pineal. Precocious puberty is associated with abnormally low melatonin levels, whereas in delayed puberty and in hypothalamic amenorrhea, melatonin levels are high compared with those in age-matched normal individuals.

There is no doubt that suitable administration of melatonin can modify human reproductive function, generally speaking, in an inhibitory capacity. It is able to suppress the ovulatory luteinizing hormone (LH) peak and to potentiate testosterone-induced LH suppression in very large doses. It can modulate the ultradian properties of LH and follicle-stimulating hormone (FSH) secretion, and it acutely increases

prolactin (although chronic administration lowers prolactin in animals, corresponding to the effects of long nights). A series of studies in males with and without hypogonadism has reinforced the perception that melatonin is essentially inhibitory to human reproductive function. An (unsuccessful) attempt has been made to develop it as an oral contraceptive in combination with a progestin minipill, with very large daily doses (80 mg). These observations fuel worries concerning the possible reproductive side effects of over-the-counter availability of melatonin as a sleep aid in some countries.

### GENERAL HUMAN PATHOLOGY

Numerous reports exist of variations in melatonin secretion as a function of endocrine and other diseases. A problem with many publications is the possible influence of ambient light conditions and posture, together with sleep-wake characteristics of the subjects, especially when a hospitalized group is compared with a nonhospitalized control population. Liver disease such as cirrhosis, which impairs metabolic function, leads to higher than normal plasma concentrations of melatonin. Drugs that stimulate or suppress hydroxylation and conjugation mechanisms, or that compete for metabolic pathways, can be expected to affect circulating melatonin. There is little evidence for a disturbance of melatonin secretion in major sleep disorders such as narcolepsy and Klein-Levine syndrome (intermittent sleeping for days at a time). One very interesting genetic disorder, Smith-Magenis syndrome, is caused by a deletion in chromosome 17p11.2. The syndrome is associated with various physical, developmental, and behavioral disabilities, but it is not fully understood. It is associated with daytime melatonin production and poor nighttime sleep, but with normal cortisol rhythms. Pharmacological suppression of the daytime melatonin leads to improved daytime alertness and nighttime sleep.

Numerous attempts have been made to relate melatonin suppression to exposure to electromagnetic fields, with no consistent data in humans.

Melatonin has been investigated extensively in psychiatric disorders, particularly winter depression and seasonal affective disorder (SAD), in which a causal role has been proposed. No firm conclusions can be drawn yet. Many antidepressant drugs acting on serotonergic or catecholaminergic systems will acutely increase or otherwise modify melatonin production.

There is recent evidence for an involvement of melatonin and the pineal gland in vasopressin production.

### POSSIBLE ROLE OF MELATONIN IN CANCER

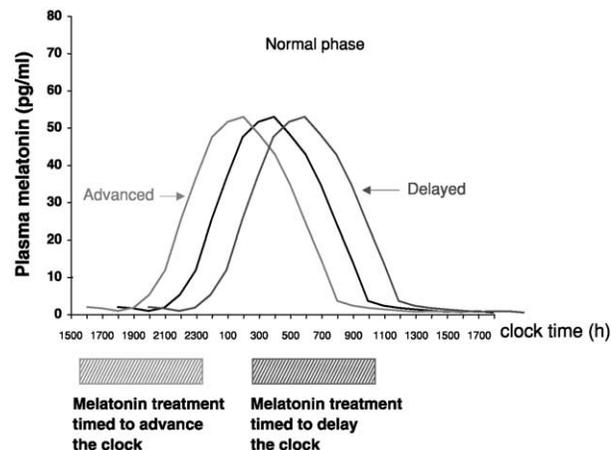
Melatonin suppression has been linked to infertility and increased incidence of certain cancers. There is evidence of low levels of melatonin in hormone-dependent cancer (estrogen receptor-positive breast cancer and prostate cancer). Some cancers are photo-period dependent in rodents, and melatonin may have a role there. Pinealectomy in rats leads to a shortened survival time in dimethylbenzanthracene-induced mammary cancer; melatonin administration reverses this effect. There is good evidence of melatonin inhibition of growth of some hormone-dependent cancer cell lines and also evidence for stimulation of immune function. Combined melatonin therapy with tamoxifen for breast cancer is being investigated in humans.

### MELATONIN AS ANTIOXIDANT/FREE RADICAL SCAVENGER

The melatonin molecule is easily oxidized; therefore, it is not particularly surprising that it demonstrates antioxidant/free radical scavenging effects. The concentrations required are generally orders of magnitude higher than physiological levels. However, there is evidence that endogenous circulating melatonin may be responsible for some of the antioxidant status of blood.

### EFFECTS OF EXOGENOUS MELATONIN IN HUMANS

The sleepiness-inducing properties of melatonin have been known since its discoverer, Lerner, reported this effect after self-administration. A substantial body of evidence indicates that melatonin, when taken during biological daytime (i.e., when endogenous melatonin is very low or absent), induces sleepiness/sleep and lowers alertness, cognitive performance, and cBT. These effects are posture dependent; individuals must be recumbent and preferably in very dim light for maximum responses to be seen. Accompanying these direct effects is the important ability of melatonin to change the timing of the internal circadian clock, as evidenced by shifts in the timing of endogenous melatonin itself, cortisol, cBT, thyroid-stimulating hormone (TSH), and sleep rhythms (Fig. 7). Administration during the (biological) afternoon/early evening leads to phase advances, whereas administration during the early morning can elicit phase delays (to

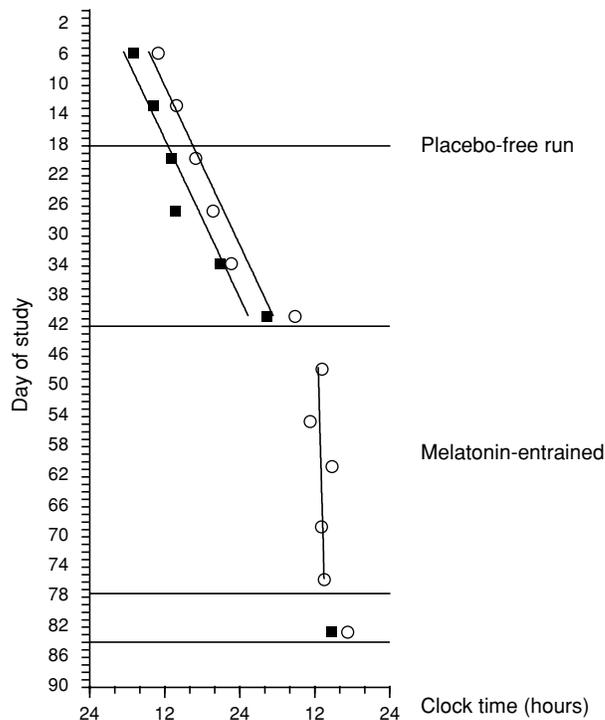


**Figure 7** Diagrammatic representation of phase shifts induced by melatonin. Melatonin treatment, specifically timed according to internal circadian phase (shown here by the timing of the endogenous melatonin rhythm) can advance or delay the circadian clock, as evidenced by shifts in its own endogenous rhythm and that of core body temperature. The shift can be predicted by a “phase response curve” that describes the magnitude and direction of shift that can be obtained with specific circadian timing of treatment. Other rhythms that can be phase shifted by melatonin include sleep, alertness, cortisol, and thyroid-stimulating hormone. The oral doses used are usually in the range of 0.5 to 5.0 mg fast release. The phase shift obtained is dose dependent but individually variable.

this point, shown only for endogenous melatonin core body temperature and sleep). The existing data have been summarized as a “phase response curve” from which specific timing of administration to induce a specific shift can be predicted.

## THERAPEUTIC POTENTIAL

In view of its ability to induce sleepiness, change the timing of sleep, and shift the timing of the internal clock, melatonin has been investigated extensively as a treatment for circadian rhythm-related sleep disorders. It has proved to be successful, when timed correctly, for alleviating jet lag, improving daytime sleep in shift workers, and normalizing sleep time in delayed sleep phase syndrome. It is able to synchronize free-running sleep and other rhythms of some blind subjects with suitable dose and timing of treatment (Fig. 8). It is the latter indication that underlines the importance of such treatment. Circadian sleep disorders can normally be treated with appropriate exposure to bright light to “shift the clock,” inducing a resetting to a normal or near normal pattern of melatonin secretion. In the blind, this is not possible, and so melatonin is the treatment of choice.

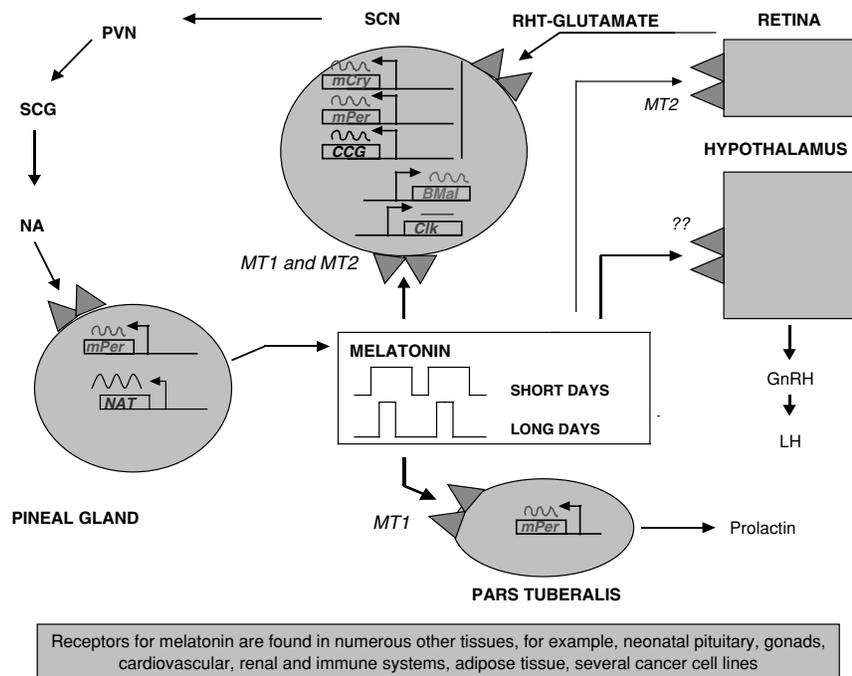


**Figure 8** Melatonin can phase shift and, in some cases, synchronize circadian rhythms in some sighted and blind subjects with suitable timing of treatment and dose. Shown are the times of the calculated peaks (acrophases) of urinary 6-sulphatoxymelatonin (squares) and cortisol (circles) of a free-running blind subject with a period of 24.57 h treated with placebo or 5 mg melatonin daily, timed to phase advance the internal clock. Note that with melatonin, a 24-h period is maintained. Significant beneficial effects on sleep are found using melatonin treatment in non-24-h sleep disorder of the blind. Redrawn from Lockley, S. W., Skene, D. J., James, K., Thapan, K., Wright, J., and Arendt, J. (2000). Melatonin administration can entrain the free running circadian system of blind subjects. *J. Endocrinol.* **164**, R1–R6. Reproduced by permission of the Society for Endocrinology.

It is likely that SCN receptors mediate the circadian effects of melatonin, those in the mediobasal hypothalamus and pars tuberalis influence photoperiodic seasonal reproduction with regard to gonadotropin secretion and prolactin, respectively, and those in the retina mediate the retinal processes influenced by melatonin (Fig. 9). The physiological functions of the multiplicity of melatonin-binding sites in other areas remain to be clarified.

## CONCLUSIONS

Melatonin acts as both an internal clock and an internal calendar to influence the timing of circadian and seasonal rhythms. It is likely that any aspect of



**Figure 9** Diagrammatic representation of the control of production and the functions of melatonin with regard to seasonal and circadian timing mechanisms. RHT, retino-hypothalamic tract; NA, noradrenalin; SCN, suprachiasmatic nucleus; PVN, paraventricular nucleus; SCG, superior cervical ganglion; MT1 and MT2, melatonin receptor subtypes. The melatonin rhythm is generated by a closed-loop negative feedback of clock gene expression in the SCN. *Ctk* and *BMal*, positive stimulatory elements; *Per* and *Cry*, negative elements; *CCG*, clock controlled genes. *Per* and *NAT* mRNA oscillate in the pineal, although posttranscription control is evident in some species. Melatonin influences SCN activity via two or more receptors. MT2 appears to be the phase-shifting receptor in rodents, whereas MT1 is associated with suppression of SCN electrical activity. The MT2 receptor was first characterized in the retina and influences dopamine release. Melatonin conveys photoperiodic information influencing the pattern of *Per* expression in the pars tuberalis for the control of seasonal prolactin variations via an MT1 receptor. Melatonin target sites in the hypothalamus influencing seasonal variations in reproductive hormones have yet to be fully defined. Based (with permission) on an original diagram by Elisabeth Maywood, MRC Laboratory of Molecular Biology, Neurobiology Division, Cambridge, United Kingdom.

physiology that depends on perception of daylength change, and/or that is driven by the master circadian clock in the SCN, is susceptible to the effects of melatonin. These effects may be modulatory and complementary to the effects of light in the case of the adult circadian system. During the perinatal period, maternal melatonin may serve to set the timing of circadian rhythms in the offspring.

Suitable timing of melatonin administration is critical for therapeutic benefit in rhythm disorders. The acute effects of melatonin on sleep, in particular, reinforce the therapeutic potential of melatonin. The possible clinical applications of antioxidant activity remain to be explored.

The effects of melatonin on photoperiod-dependent functions are critically important and include, for example, the timing of pubertal development in photoperiodic species. Melatonin is used commercially in

agriculture to modify the timing of breeding and seasonal changes in pelage.

Current research focuses on the mechanisms of action, with particular emphasis on receptor mechanisms and modification of gene expression in target tissues. Melatonin is the marker rhythm of choice for the investigation of disturbed circadian rhythms in, for example, sleep disorder, shift work, and jet lag.

### Acknowledgment

The author thanks the MRC, the Health and Safety Executive, the National Grid, and the MoD for research support during the writing of this article.

### See Also the Following Articles

Circadian Rhythms: Hormonal Facets • Pineal Gland • Pineal Gland, Evolution of • Prostate Cancer

## Further Reading

- Arendt, J. (1995). "Melatonin and the Mammalian Pineal Gland." Chapman Hall, London.
- Arendt, J. (2000). Melatonin, sleep, and circadian rhythms [editorial]. *N. Engl. J. Med.* **343**, 1114–1116.
- Klein, D. C., Coon, S. L., Roseboom, P. H., Weller, J., Bernard, M., Gastel, J. A., Zatz, M., Iuvone, P. M., Rodriguez, I. R., Begay, V., Falcon, J., Cahill, G. M., Cassone, V. M., and Baler, R. (1997). The melatonin rhythm generating enzyme: Molecular regulation of serotonin-N-acetyl transferase in the pineal gland. *Rec. Prog. Horm. Res.* **52**, 307–358.
- Masana, M. I., and Dubocovich, M. L. (2002). Melatonin receptor signaling: Finding the path through the dark. *Science* **107**, 1–5.
- Reiter, R. J. (1998). Melatonin and human reproduction. *Ann. Med.* **1**, 103–108.
- Reiter, R. J., Acunacastroviejo, D., Tan, D. X., and Burkhardt, S. (2001). Free radical mediated molecular damage: Mechanisms for the protective actions of melatonin in the central nervous system. *Ann. N.Y. Acad. Sci.* **939**, 200–215.
- Shochat, T., Haimov, I., and Lavie, P. (1998). Melatonin: The key to the gate of sleep. *Ann. Med.* **30**, 109–114.
- Vollrath, L. (1981). "The Pineal Organ." Springer-Verlag, Heidelberg, Germany.

GnRH or gonadotropin injections. Development of secondary sexual characteristics can be attained with appropriate hormone replacement.

Pulsatile release of GnRH appears to be an innate characteristic of neurons in the arcuate nucleus. *In vitro* experiments have shown that these neurons have pulsatile activity even when isolated from their normal surrounding structures and microenvironment.

A growing body of evidence has revealed intricate interactions between other regions of the central nervous system and the hypothalamus. Investigators have demonstrated complex neural pathways that have led to a more sophisticated understanding of the cause and effects of hypothalamic function and dysfunction.

Environmental changes, stress, exercise, extremes of weight, and emotional distress may disrupt normal hypothalamic GnRH secretion. Patients with amenorrhea resulting from hypothalamic–pituitary dysfunction are usually hypogonadotropic. In these cases, estrogen replacement should be considered to prevent bone loss and the possibility of osteoporosis. Hypogonadotropic patients with amenorrhea should be screened carefully for eating disorders, which can be fatal without aggressive behavioral and nutritional therapy.

A small number of patients with hypothalamic dysfunction have neoplastic, infiltrative, or infectious lesions of the central nervous system. Structural lesions must be in the differential diagnosis of all patients presenting with primary or secondary amenorrhea of hypothalamic origin.

### The Pituitary and Gonadotropin (Follicle-Stimulating Hormone and Luteinizing Hormone) Release

In response to the pulsatile release of GnRH from the hypothalamus, the anterior pituitary releases the gonadotropins luteinizing hormone (LH) and follicle-stimulating hormone (FSH) in a pulsatile fashion. The frequency and amplitude of GnRH and gonadotropin secretion play pivotal roles in oocyte development, and the alteration of either of these patterns may affect oocyte quality or ovulatory function.

The predominant function of FSH is to stimulate oocyte recruitment and development during the follicular phase of the cycle. The follicular phase is conveniently, but somewhat inaccurately, defined as beginning on the first day of menstruation. Slight but significant increases in FSH secretion can be identified 1 or 2 days before the onset of menstruation. The FSH increment is a response to declining

ovarian hormone production, especially inhibin A, during the late luteal phase.

The initial increase in FSH stimulates the development of a cohort of ovarian follicles containing immature oocytes. This FSH stimulation eventually results in the selection of a single dominant follicle destined for ovulation.

Progressively increasing estradiol secretion during the follicular phase acts to inhibit FSH release prior to the midcycle gonadotropin surge (negative feedback). This allows the dominant follicle a unique advantage over other developing follicles that will ultimately degenerate through the process of atresia. This advantage is predominantly a result of a higher density of FSH receptors, greater steroidogenic capacity, and a more profound angiogenic proliferation within the dominant follicle.

Eventually, estradiol promotes the stimulation of the midcycle gonadotropin surge that initiates the events resulting in ovulation from the dominant follicle (positive feedback). Although a surge in both LH and FSH occurs, LH plays the most significant role in the ensuing ovulation process.

The beginning of the LH surge occurs approximately 35 to 44 h prior to ovulation. The LH surge is critical in regulating the production of progesterone, prostaglandins, and other local factors that contribute to follicle rupture. In response to these conditions, the oocyte is extruded from the ovary into the fimbria of the fallopian tube en route to the uterus.

### The Ovary: Response and Control

The residual follicular tissue undergoes a functional transition. Local growth factors initiate a massive angiogenic process and a drastic increase in the capacity for progesterone production. This process is known as luteinization, and the resultant collection of cells is known as the corpus luteum. Hence, ovulation signals the beginning of the luteal phase.

Pulsatile LH secretion continues to promote the production of estradiol and progesterone during the luteal phase. Progesterone secretion is essential for the induction of secretory changes within the uterine cavity. These changes are a prerequisite for proper embryo implantation.

The corpus luteum appears to have a programmed life span, rapidly regressing 9 to 11 days following ovulation in the absence of pregnancy. During early pregnancy, the conceptus produces human chorionic gonadotropin (hCG), a glycoprotein nearly identical to LH in structure and function. This sustains the steroid-producing capacity of the corpus luteum

until the placenta develops sufficient steroidogenic activity.

There is controversy regarding the effects of dysfunction in the ovary on fertility. Failure of patients to have an appropriate LH surge may inhibit ovulation. These patients can be treated with injections of hCG.

There also appears to be a subset of women who generate an LH surge but do not have rupture of their dominant follicle. The exact etiology of this process, referred to as luteinized unruptured follicle syndrome, is unknown. Because serial ultrasound monitoring is necessary to detect this disorder, the incidence is poorly described in the literature.

It is not clear whether abnormal function of the corpus luteum is a cause of infertility. A possible connection between inadequate progesterone production by the corpus luteum and early pregnancy loss has been termed luteal phase deficiency. Diagnosis and treatment of this disorder are relatively simple, although the efficacy of treatment remains a source of great conflict among experts in the field.

The remarkable coordination necessary to produce oocytes capable of establishing a successful pregnancy has been demonstrated in women receiving gonadotropins for *in vitro* fertilization. Prior to the use of GnRH analogues for pituitary suppression, approximately 20% of patients showed evidence of a premature LH surge. This untimely LH surge often resulted in the retrieval of immature oocytes that fertilized poorly *in vitro*.

Anovulatory disorders represent a large portion of the reproductive endocrinologist's daily practice. The etiology of anovulation is often obscure. Fortunately, the majority of these disorders can be treated effectively with current medical therapies.

Patients with polycystic ovary syndrome (PCOS), often referred to as chronic hyperandrogenemic anovulation, generally present clinically with amenorrhea, hirsutism, and obesity. This syndrome is extremely heterogeneous and may actually represent a pooled collection of various disorders. A strong association between PCOS and insulin resistance, with impaired glucose tolerance, has been recognized. These patients often become ovulatory in response to a strict program of exercise and dietary modification with weight loss. In addition, pharmacological intervention addressing glucose metabolism may be promising in treating anovulation and infertility in these patients.

Certain conditions are known to be associated with accelerated germ cell loss and degeneration of the ovary. Turner syndrome, an abnormality in or absence of one of the X chromosomes, is a classic example of this type of disorder. Characterized by

sexual infantilism, short stature, and streak gonads, these patients have a propensity for the development of concomitant endocrinopathies.

The cessation of ovarian function, which results in menopause, is related to the depletion of functional ovarian follicles. This process usually occurs at approximately 50 years of age. Premature ovarian failure is defined as cessation of ovarian function prior to 40 years of age. Interestingly, this condition can be associated with autoimmune diseases that can be life threatening. In addition, in the subset of women who present with premature ovarian failure prior to 30 years of age, chromosomal abnormalities are occasionally discovered.

In both Turner syndrome and premature ovarian failure, the genetic defect appears to result in an unusually rapid rate of atresia of the follicular pool. This genetic association is particularly interesting as an area for future investigations of physiology and therapeutics.

## THE UTERUS AND MENSTRUATION

Distinct changes in the uterine lining are noted throughout the menstrual cycle. Anatomically, the endometrial lining may be divided into an outer functionalis layer and an inner basalis layer. The role of the functionalis layer is to respond to the alteration of hormones throughout the cycle and to prepare for implantation of the embryo. The functionalis is shed during menstruation. The basalis layer serves as a source of new endometrium after each successive menstruation.

The proliferative phase of the menstrual cycle corresponds with the follicular phase of the ovarian cycle and is marked by estrogen-dependent proliferation of glandular structures and vasculature within the endometrium. Mitotic activity is prominent, and the endometrial lining increases in thickness as much as 10-fold.

After ovulation, the thickness of the endometrium changes little. This is likely a result of inhibition from increasing progesterone production. Tortuosity of the vessels and secretory activity within the glands are the hallmarks of the secretory phase of the cycle. At the time of implantation of the blastocyst, marked edema can be seen within the stroma of the endometrium. Concomitantly, decidualization occurs. The decidua is derived from the stromal cells of the endometrium and is rich in glycogen and lipid content. This layer of cells displays the ability to produce a large number of autocrine and paracrine regulatory

peptides. The decidua is integrally involved in the processes of implantation and placentation.

In the absence of pregnancy, implantation, and the production of hCG by the trophoblasts, the demise of the corpus luteum results in the withdrawal of estrogen and progesterone. This results in apoptosis of the endometrial cells and contraction of the spiral arteries nourishing the surface of the endometrium. Hemorrhage into the stroma and infiltration of the endometrium with inflammatory mediators follow. Eventually, these combined events result in the sloughing of the endometrium, clinically recognized as menstruation.

Many factors are responsible for aberrant menstrual patterns. Patients presenting with amenorrhea, oligomenorrhea, menorrhagia, metrorrhagia, intermenstrual bleeding, or dysmenorrhea can be challenging from both a diagnostic and a therapeutic standpoint. Structural abnormalities in the uterine cavity and neoplastic conditions of the endometrium can contribute to bleeding abnormalities. In addition, alterations in any of the complex processes within the hypothalamic–pituitary–ovarian axis can result in abnormal or absent menstruation.

## THE FUTURE

The next decade should bring exciting advances in our understanding of the mechanisms responsible for normal gamete function, fertilization, implantation, and early pregnancy development. Essential for the development of this growth in knowledge will be a sound basic understanding of the menstrual cycle and an appreciation of the beautifully orchestrated relationships among the brain, ovary, uterus, and other organs.

## See Also the Following Articles

Fertilization • FSH (Follicle-Stimulating Hormone) • Gonadotropin-Releasing Hormone (GnRH) Actions •

Hypothalamic Anovulation, Functional • Hypothalamic-Pituitary Unit • Implantation • Infertility, Overview • Kallmann's Syndrome and Idiopathic Hypogonadotropic Hypogonadism • LH (Luteinizing Hormone) • Ovarian-Follicular Apparatus • Polycystic Ovary Syndrome (PCOS) • Pregnancy Endocrinology • Premenstrual Syndrome (PMS)

## Further Reading

- Chrousos, G. P., Torpy, D. J., and Gold, P. W. (1998). Interactions between the hypothalamic–pituitary–adrenal axis and the female reproductive system: Clinical implications. *Ann. Int. Med.* **129**, 229.
- Ehrmann, D. A., Cavaghan, M. K., Imperial, J., *et al.* (1997). Effects of metformin on insulin secretion, insulin action, and ovarian steroidogenesis in women with polycystic ovary syndrome. *J. Clin. Endocrinol. Metab.* **82**, 524.
- Iuorno, M. J., and Nestler, J. E. (2001). Insulin-lowering drugs in polycystic ovary syndrome. *Obstetr. Gynecol. Clinics North America* **28**, 153.
- Knobil, E. (1989). The electrophysiology of the GnRH pulse generator in the rhesus monkey. *J. Steroid Biochem.* **33**, 669.
- Legro, R. S. (2001). Diabetes prevalence and risk factors in polycystic ovary syndrome. *Obstetr. Gynecol. Clinics North America* **28**, 99.
- McNeely, M. J., and Soules, M. R. (1988). The diagnosis of the luteal phase deficiency: A critical review. *Fertil. Steril.* **50**, 1.
- Petraglia, F., Zanin, E., Faletti, A., and Reis, F. M. (1999). Inhibins: Paracrine and endocrine effects in female reproductive function. *Curr. Opin. Obstetr. Gynecol.* **11**, 241.
- Rasmussen, D. D., Gambacciani, M., Swartz, W., *et al.* (1989). Pulsatile gonadotropin-releasing hormone release from the human mediobasal hypothalamus *in vitro*: Opiate receptor-mediated suppression. *Neuroendocrinology* **49**, 150.
- Schwanzel-Fukuda, M., Bick, D., and Pfaff, D. W. (1989). Luteinizing hormone-releasing-hormone expressing cells do not migrate normally in an inherited hypogonadal (Kallman) syndrome. *Mol. Brain Res.* **6**, 311.
- Tilly, J. L., Kowalski, K. I., Schomberg, D. W., and Hsueh, A. J. (1992). Apoptosis in atretic ovarian follicles is associated with selected decreases in messenger ribonucleic acid transcripts for gonadotropin receptors and cytochrome p450 aromatase. *Endocrinology* **131**, 1670.
- Yen, S. C. (1993). Female hypogonadotropic hypothalamic amenorrhea syndrome. *Endocrinol. Metab. Clinics North America* **22**, 29.

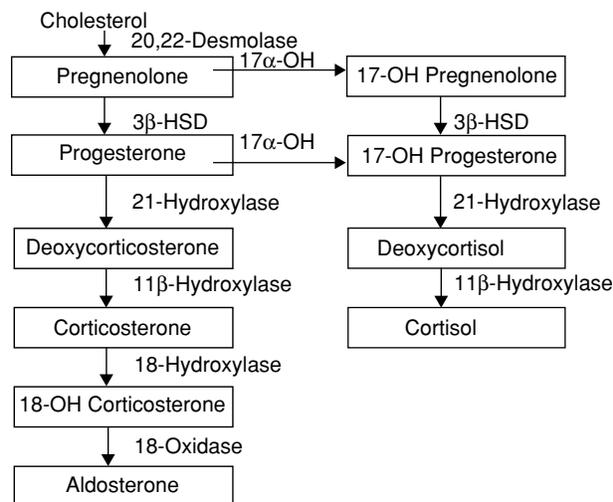


Figure 1 Steroid biosynthetic pathway.

is low-affinity binding to plasma albumin and cortisol-binding globulin. Although a specific high-affinity, aldosterone-binding protein has been reported, it is of doubtful physiological significance.

The liver is the major site of aldosterone metabolism. Aldosterone undergoes reduction by a cytosolic  $5\beta$  reductase and  $3\alpha$  hydroxysteroid dehydrogenase to form tetrahydroaldosterone, which accounts for 40% of aldosterone urinary metabolites. Tetrahydroaldosterone is then conjugated in the liver to glucuronide to yield the major urinary metabolite of aldosterone.

## REGULATION OF MINERALOCORTICOID PRODUCTION

### Renin–Angiotensin System

Angiotensin II is the principal regulator of aldosterone production. The classical renin–angiotensin system (RAS) is outlined in Fig. 2. This consists of a precursor protein, angiotensinogen (produced by the liver), and two enzymes, renin (first described in the juxtaglomerular cells of the kidney) and angiotensin I-converting enzyme (ACE) (widely distributed but found predominantly in the pulmonary vascular endothelium).

Angiotensin II is an octapeptide and acts on the adrenal cortex to stimulate aldosterone secretion and also produce a direct pressor response through increased resistance vessel tone. Secretion of renin (and hence activation of the RAS) is increased under conditions of reduced sodium intake, sodium loss, or reduced extracellular fluid volume because this leads to increased angiotensin II production. Hence, the

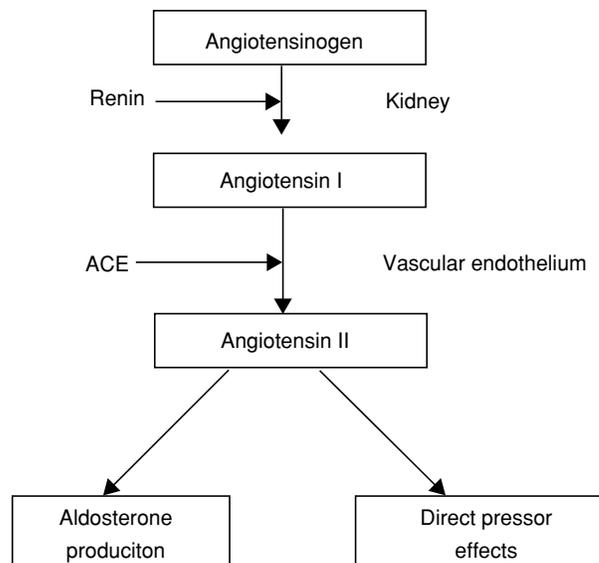


Figure 2 Classical renin–angiotensin–aldosterone system.

RAS is important in acute cardiovascular homeostasis (i.e., the response to head-up tilt or blood loss) and in the regulation of aldosterone secretion in response to salt depletion. It is noteworthy that the aldosterone response to angiotensin II is enhanced in states of classic sodium depletion and is diminished by sodium loading.

Angiotensin II produces its effect on aldosterone secretion via the type I angiotensin II receptor (ATI). Binding to this G protein-coupled receptor results in activation of phospholipase C. In turn, this leads to hydrolysis of phosphatidyl inositol biphosphate, producing inositol triphosphate (IP<sub>3</sub>) and diacylglycerol (DAG), causing a rapid rise in intracellular calcium and activation of both early stages (transport of cholesterol into the mitochondrium) and late stages (aldosterone synthase) of aldosterone synthesis. This rise in intracellular calcium correlates well with aldosterone production.

### Plasma Potassium

Potassium is a powerful direct stimulus to aldosterone secretion; very small increments, which do not alter plasma levels perceptibly, raise the aldosterone secretion rate. Potassium acts at the plasma membrane of the zona glomerulosa by modifying the membrane potential. Increasing potassium concentration leads to depolarization of the cell membrane and activation of voltage-dependent calcium channels. The resultant calcium influx leads to a sustained rise in cytosolic

calcium present for the duration of the stimulus that correlates with aldosterone production.

### Adrenocorticotropin

Infusion of adrenocorticotropin (ACTH) causes a rise in circulating aldosterone levels by direct action on zona glomerulosa cells. The effect of pharmacological doses of ACTH is only temporary and is not sustained. It is hypothesized that this is due to rapid retention of sodium, with suppression of angiotensin II formation caused by the ACTH-mediated elevation of plasma levels of intermediates such as 11-DOC as well as cortisol itself. Hence, the physiological significance of the effect of ACTH on plasma aldosterone remains unclear. Patients with anterior pituitary failure, who lack ACTH, continue to secrete aldosterone responsive to potassium and angiotensin II.

### Other Regulatory Factors

A variety of other peptides and amines influence aldosterone secretion in humans. Of these, atrial natriuretic peptide (ANP) may be one of the more important mechanisms that inhibit aldosterone production. Dopamine also has inhibitory effects. Stimulatory factors include catecholamines (e.g., adrenaline, noradrenaline), acetylcholine, and vasoactive intestinal peptide (VIP).

## EFFECTS OF MINERALOCORTICOIDS

### Mineralocorticoid Receptor

Aldosterone acts by binding to the mineralocorticoid receptor (MR). This is an intracellular receptor belonging to the steroid/thyroid/retinoid/orphan receptor superfamily. It acts as a ligand-dependent transcription factor, so that once it is bound to aldosterone (or cortisol), it will interact directly with hormone-responsive elements (HREs) in the promoter regions of regulated genes. HREs are present in many genes and fall into three groups: the glucocorticoid/progesterone/mineralocorticoid-responsive element to which aldosterone binds, estrogen-responsive element (ERE), and thyroid-responsive element (TRE).

MRs are found in epithelial tissues (renal distal nephron, colon, ducts of sweat, and salivary glands) and nonepithelial tissues (heart, brain, vascular smooth muscle, liver, and peripheral blood leucocytes). There is also evidence of a membrane-associated aldosterone receptor that subserves rapid nongenomic effects of aldosterone on vascular homeostasis.

### 11 $\beta$ Hydroxysteroid Dehydrogenase System

The MR binds both aldosterone and cortisol with similar affinities (0.5–1 nmol/L). Because cortisol is much more abundant in plasma (100- to 1000-fold), it is clear that a mechanism is required to protect the MR against permanent occupancy by glucocorticoid hormones. This role is performed by the 11 $\beta$  hydroxysteroid dehydrogenase (11 $\beta$  OHSD) enzyme system. This converts cortisol into its 11-keto inactive metabolite, cortisone, and also catalyzes the opposite oxoreductase step to reform cortisol from cortisone (Fig. 3). Cortisone has no affinity for the MR.

There are two isoforms of this enzyme encoded by separate genes. The type 1 isozyme is found mainly in liver and adipose tissue; *in vivo*, it generates cortisol from cortisone. The type 2 isozyme inactivates cortisol to cortisone and so is expressed primarily in mineralocorticoid target tissues (renal distal nephron, salivary and sweat glands, and colon). Functional anomalies of the type 2 enzyme result in the clinical entity of apparent mineralocorticoid excess.

### Cellular Response to Aldosterone Binding

When the MR binds aldosterone, it migrates to the cell nucleus, where it binds to an HRE. Following this, transcriptional activation occurs, and during this latent period, DNA transcription and protein

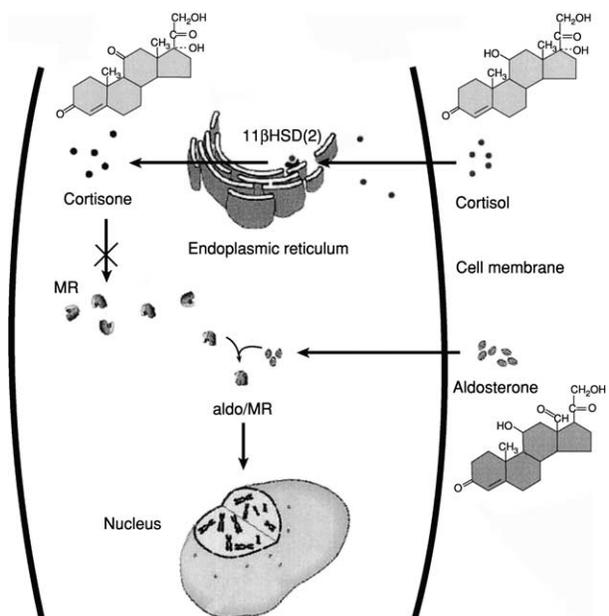


Figure 3 11 $\beta$  hydroxysteroid dehydrogenase system.

translation occur. The first unequivocally aldosterone-induced protein in kidney was demonstrated in 1999. Sgk-1 (serum glucocorticoid-induced kinase, the first of three known isoforms) is regulated transcriptionally by aldosterone and is activated by insulin-regulated phosphorylation. In turn, activated Sgk-1 results in phosphorylation and activation of the epithelial sodium channel. Aldosterone activation of the MR undoubtedly leads to transcriptional regulation of other molecules, but the precise identities and functions of these are beyond the scope of this article.

### Physiological Response to Aldosterone

The best-characterized physiological effect of aldosterone is to increase the reabsorption of sodium in the kidney and at other secretory epithelial sites at the expense of potassium and hydrogen ions. The major sites of aldosterone-induced sodium and potassium transport are luminal cells of the cortical-collecting tubules and the distal convoluted tubule. Aldosterone acts to increase activity of the Na/K ATPase pump on the basolateral membrane of these cells.

Aldosterone also regulates hydrogen ion excretion by the kidney in the distal nephron. Aldosterone-mediated hydrogen ion secretion occurs in the intercalated cells of the collecting tubule. This effect appears to be mediated via an increase in activity of the adenosine triphosphate (ATP)-dependent apical hydrogen ion pump and parallel regulation of the basolateral membrane Cl/HCO exchanger. Thus, aldosterone-induced natriuresis and hydrogen ion secretion appear to be independent events.

Therefore, the overall effect of aldosterone on the renal tubule is to promote sodium retention at the expense of potassium and also to promote hydrogen ion excretion. In turn, this leads to expansion of the extracellular volume as well as plasma volume.

It is likely that the rise in blood pressure reflects mechanisms other than, or in addition to, simple plasma volume expansion and associated increase in cardiac output. For example, activation of the MR in vascular smooth muscle results in alteration in pressor responsiveness to adrenergic stimulation. Moreover, evidence suggests that aldosterone binding by the MR in cardiac tissue regulates collagen formation. Therefore, it is feasible that similar action in peripheral blood vessels might result in remodeling that could sustain an elevated blood pressure. This is supported by evidence suggesting that aldosterone levels are inversely related to arterial compliance in essential hypertension.

Mineralocorticoid receptors are also present in the central nervous system (CNS) and may regulate

central sympathetic outflow as well as thirst and salt intake. Intracerebroventricular administration of aldosterone raises blood pressure without altering peripheral concentration and is independent of its effects on sodium retention. In short, sustained excess of aldosterone is likely to elevate blood pressure through a variety of mechanisms, all of which require activation of the MR.

### Role of Aldosterone in Disease

Secondary hyperaldosteronism occurs due to activation of the RAS and is seen in conditions such as congestive cardiac failure (CCF), malignant hypertension, renal artery stenosis, and hepatic cirrhosis. It has generally been accepted that the pathophysiological effects of aldosterone in these conditions are mediated through sodium retention and expansion of plasma volume. This is supported by the finding that hypertension and edema do not occur in renal salt-wasting forms of secondary aldosteronism such as Bartter's syndrome.

However, recent experimental and clinical evidence suggests that excessive circulating levels of aldosterone can bring about adverse cardiovascular sequelae, such as cardiac fibrosis and left ventricular hypertrophy, independent of its effects on blood pressure or plasma volume. Several investigators have demonstrated this using rat models of renovascular hypertension. In one such experiment, uninephrectomized rats infused with aldosterone over an 8-week period developed progressive elevation of blood pressure, cardiac hypertrophy, and fibrosis. This was also seen in rats that remained normotensive due to concomitant intracerebroventricular infusion of an MR antagonist, demonstrating that these effects are humoral and not hemodynamic. The high prevalence of CCF, which is commonly associated with hyperaldosteronism, suggests a significant role of aldosterone excess as a cause of cardiovascular injury. For example, more recent studies have confirmed that aldosterone excess causes severe cardiac injury associated with perivascular inflammation (in coronary vessels), cerebral tissue damage, and nephrosclerosis. These effects are exacerbated by concomitant sodium loading and are prevented by aldosterone receptor antagonism.

### SYNDROMES OF MINERALOCORTICOID EXCESS

The most common clinical form of mineralocorticoid excess reflects excess production of aldosterone;

Table 1 Classification of Mineralocorticoid Excess Syndromes

Mechanism	Classification	Ligand
Post receptor	Liddle's syndrome	—
Abnormal receptor	Progesterone-induced hypertension	Progesterone
	Glucocorticoid resistance	Cortisol
Abnormal ligand	Syndrome of apparent mineralocorticoid excess	Cortisol
	Congenital adrenal hyperplasia	Deoxycorticosterone
	DOC-producing tumors	Deoxycorticosterone
	Ectopic ACTH syndrome	Cortisol
Normal ligand	Primary aldosteronism	Aldosterone
	GRA	

however, there are numerous other causes and other mineralocorticoids that can be produced in excess (Table I).

### Post Receptor

Liddle's syndrome was first described by Liddle in 1963 when he reported a family in which the siblings appeared to have features of aldosterone excess (early-onset hypertension and hypokalemia) but with suppressed plasma renin and aldosterone levels. It is now known that this syndrome is inherited as an autosomal dominant trait and occurs due to mutations of the genes encoding the  $\beta$ - or  $\gamma$ -subunit of the epithelial sodium channel of the distal nephron. This results in constitutive activation of this receptor (sodium retention and potassium excretion), irrespective of circulating mineralocorticoid levels (usually suppressed). Moreover, in the proband of one of Liddle's original cases, renal transplantation resulted in normalization of blood pressure and electrolyte abnormalities. However, in practice, using antagonists of the sodium channel, such as amiloride and triamterene, treats this condition effectively.

### Adrenal Receptor Causes

#### Glucocorticoid Receptor Resistance

This rare disorder is characterized by decreased sensitivity to cortisol signaling, elevated plasma cortisol, and an absence of any stigmata of Cushing's disease. The compensatory increase in corticotropin-releasing hormone (CRH), ACTH, and cortisol production is also accompanied by increased secretion of mineralocorticoid precursors (11-DOC and corticosterone) and aldosterone as well as adrenal androgens (dehydroepiandrosterone [DHEA], androstenedione, and testosterone). Elevated plasma cortisol levels may

overwhelm the  $11\beta$  OHSD system, allowing cortisol to bind at the MR.

Phenotypically, this disorder has a wide clinical spectrum ranging from fatigue, hirsutism, oligomenorrhea, infertility, and obesity to severe hypertension with hypokalemic alkalosis. In some patients, point mutations in the steroid-binding domain of the glucocorticoid receptor have been identified. However, other cases of cortisol resistance may represent abnormalities in the signaling pathway rather than in the receptor. Treatment is with sufficient dexamethasone to suppress ACTH and to lower blood pressure and adrenal androgens.

#### Progesterone-Induced Hypertension

Progesterone usually acts as an antagonist at the mineralocorticoid receptor. However, a mutation in the ligand-binding domain of the receptor was recently identified. This mutation allows progesterone to act as an agonist at the mineralocorticoid receptor, causing a clinical and biochemical picture indistinguishable from aldosterone excess, although aldosterone levels are suppressed. Because progesterone levels typically increase during pregnancy, affected female members of identified kindred had severe hypertension with associated hypokalemia and hyperkaliuria when pregnant.

### Abnormal Ligand

#### Syndrome of Apparent Mineralocorticoid Excess

In this condition, the type 2 isoform of  $11\beta$  OHSD is inactivated. This enzyme is important in mineralocorticoid target tissues to inactivate cortisol and prevent binding to the MR. If activity of the enzyme is reduced, cortisol binds to the MR, causing classical

mineralocorticoid-type hypertension. Clinically, this presents as hypertension, hypokalemia, and suppression of the renin–angiotensin–aldosterone (RAA) axis. Diagnosis is made by demonstration of an elevated ratio of urinary cortisol to cortisone metabolites or of urinary-free cortisol to urinary-free cortisone. This disorder is inherited as an autosomal recessive trait, and several inactivating point mutations in the type 2 gene have been identified. Dexamethasone can be used to lower plasma cortisol, and patients will also respond to spironolactone.

The same clinical and biochemical phenotype can be seen when the type 2 isoform of 11 $\beta$  OHSD is inhibited by large amounts of liquorice or related drugs such as carbenoxolone (the hemisuccinate derivative of the active component of liquorice, glycyrrhetic acid). Carbenoxolone was once used in the treatment of peptic ulcer disease, but mineralocorticoid side effects were seen in up to 50% of patients. Again, analysis of urine collections will reveal elevated cortisol:cortisone ratios (as urinary metabolites tetrahydrocortisol and tetrahydrocortisone), but to a lesser extent than will the congenital cases. Obviously, treatment is to stop excessive intake of liquorice or to stop the carbenoxolone.

#### **Ectopic ACTH Syndrome**

Fully 80% of patients with Cushing's syndrome are hypertensive, and this increases to 95% in the subgroup with excess pituitary ACTH production (Cushing's disease). Moreover, hypokalemic alkalosis, a hallmark of mineralocorticoid excess, is found in 95 to 100% of these cases, as compared with just 10% in other cases of Cushing's syndrome. Recent evidence suggests that saturation of 11 $\beta$  OHSD type 2 by the very high-cortisol concentrations that typify this condition explains the mineralocorticoid excess state. In support of this, the urinary ratios of both tetrahydrocortisol:tetrahydrocortisone and free cortisol:cortisone are elevated.

#### **Deoxycorticosterone and Corticosterone Secretion**

Excessive amounts of these mineralocorticoids are usually found only in subtypes of congenital adrenal hyperplasia. DOC excess usually occurs along with aldosterone excess in adrenal adenoma, and its contribution to the clinical phenotype is unlikely to be significant. Rarely, adenomas may produce DOC in isolation, and in these circumstances, DOC can act as an agonist at the MR, resulting in hypertension. The majority of pure corticosterone-producing adrenal tumors have been carcinomas.

#### **Congenital Adrenal Hyperplasia**

This describes a group of recessively inherited disorders arising due to mutations of enzymes involved in adrenal corticosteroidogenesis. They are described in detail elsewhere. Only two forms of congenital adrenal hyperplasia are associated with mineralocorticoid excess: 17 $\alpha$ -hydroxylase (excess of DOC and corticosterone) and 11 $\beta$ -hydroxylase (excess DOC due to increased ACTH drive to the adrenal) deficiencies.

#### **Normal Ligand: Primary Aldosteronism**

Primary aldosteronism (PA) can be defined as overproduction of aldosterone independent of its normal chronic regulator, angiotensin II. Recent studies have led to a reevaluation of the apparent prevalence of PA. In particular, wider use of the aldosterone:renin ratio (ARR) as a screening and diagnostic tool has led to a substantial increase in the PA detection rate. Indeed, several groups worldwide now estimate the prevalence of PA in unselected hypertensives to be 5 to 15%. The exact prevalence remains unclear, but a reliable and reproducible study of unselected hypertensives in Dundee, Scotland, suggested a prevalence of 10% both in primary and secondary care settings. Thus, aldosterone excess may be a very important cause of cardiovascular morbidity.

#### **Adrenal Adenoma**

First described by Jerome Conn in 1955, this was previously thought to be the most common cause of PA. Aldosteronomas are usually small (<2 cm in diameter) and benign and have a golden yellow color on their cut surfaces with different adrenal cell types visible on microscopy.

In most cases, control of aldosterone secretion is altered so that the tumor responds to ACTH and not to its usual regulator angiotensin II. In this circumstance, aldosterone adopts a circadian rhythm similar to that of cortisol. However, a number of adenomas remain responsive to angiotensin II.

#### **Bilateral Adrenal Hyperplasia**

The true frequency of this is currently uncertain, but recent studies identifying patients via the ARR suggest that this is now the most common cause of PA. Histologically, it is usually associated with bilateral macro- or micronodular hyperplasia, although rarely it can occur unilaterally. Idiopathic aldosteronism is associated with enhanced adrenal responsiveness to angiotensin II. The reason for this is still unclear.

It has been suggested that this syndrome is part of the spectrum of low-renin hypertension (occurring in up to 30% of elderly hypertensives) and does not in itself constitute a distinct diagnostic entity. This is supported by the findings of adrenal hyperplasia and nodularity in postmortem series of patients with essential hypertension.

Adrenal carcinoma is a rare cause of PA.

### **Glucocorticoid Remediable Adosteronism**

This rare autosomal dominant condition is characterized by aldosterone excess that is regulated by ACTH rather than by angiotensin II. The resulting interruption of negative feedback (aldosterone does not suppress ACTH) causes adosteronism that is suppressible by exogenous glucocorticoid, a useful therapeutic strategy.

The molecular basis of this condition was first described in 1992. The terminal steps in cortisol and aldosterone biosynthesis are catalyzed by the enzymes 11 $\beta$ -hydroxylase and aldosterone synthase, respectively. These are encoded by two genes, CYP11B1 and CYP11B2, that lie in tandem on chromosome 8 and share more than 90% homology at the amino acid level. In glucocorticoid remediable adosteronism (GRA), a chimeric gene is created containing the 5' promoter sequence of CYP11B1 and functional elements of CYP11B2, resulting in aldosterone production under control of ACTH. The hybrid gene can be detected by a polymer-based chain reaction (PCR) that provides a screening test for this type of adosteronism.

### **Clinical Features of Primary Adosteronism**

In the majority of cases, patients are asymptomatic and present simply with refractory hypertension. A few may present with symptoms related to hypokalemia with muscle weakness, tiredness, and polyuria and polydipsia related to nephrogenic diabetes insipidus. However, it should be emphasized that it is now recognized that up to 60 to 70% of patients with PA, as diagnosed by an elevated ARR, are normokalemic.

Nevertheless, spontaneous hypokalemia is rare in untreated hypertension, and its presence should raise suspicion of PA. Patients with hypokalemia on diuretic treatment should stop treatment for 2 weeks and take potassium supplements before repeat serum potassium measurements. As discussed earlier, aldosterone excess may contribute to specific humorally mediated cardiac, renal, and CNS damage. In this regard, there is evidence that patients with PA have more marked left ventricular hypertrophy than do patients with equivalent degrees of blood pressure elevation due to other causes.

### **Diagnosis**

This remains controversial, and no consensus exists regarding an optimal diagnostic protocol. Before performing specific diagnostic tests, serum potassium should be normalized, if necessary by oral supplementation. Both aldosterone and plasma renin activity (PRA) are also affected by posture and sodium intake. Ideally, antihypertensive medications should be stopped several weeks prior to testing due to their effects on the RAA axis (Table II). In practice, this is not always possible or even advisable due to the severity of hypertension. In these circumstances, alpha blockers have been shown to have only negligible effects on the ARR and can be continued. Theoretically, calcium channel blockers are also often continued, but the possibility of false negative results should be remembered. It is simpler to consider the diagnosis of PA in two stages: confirmation of aldosterone excess initially and then investigation of the underlying cause.

### **Aldosterone:Renin Ratio**

Given the problems of controlling posture, sodium intake, and potassium levels, many groups now advocate use of the ARR as a first step in the diagnosis of PA. This ratio appears to be more robust because it is less affected by drug administration, day-to-day and diurnal variation, and posture. However, use of this ratio is not without its problems. One concern is that

**Table II** Effect of Common anti Hypertensive Drugs on the RAA Axis

Drug	Effect	Duration of washout required (weeks)
ACE inhibitors and angiotensin II antagonists	PRA $\uparrow$	2
Diuretics	PRA $\uparrow$	2
Spironolactone	PRA $\uparrow$	6
Beta blockers	PRA $\downarrow$	2
Calcium channel blockers	Aldosterone $\downarrow$	2

whereas most laboratories around the world express plasma aldosterone as picomoles/liter, in the United States it is expressed as nanograms/deciliter. Nearly all investigators report PRA in nanograms/milliliter/hour. This means that the cutoff for a raised ARR can be approximately 25 (plasma aldosterone as ng/dl) or 750 (plasma aldosterone as pmol/L). Although this value is rather arbitrary, most groups have successfully confirmed abnormal regulation of aldosterone in patients with values equal to or greater than 750 by confirmatory testing. It is suggested that the ARR may be used as an initial screening test in individuals with resistant hypertension, hypertension with hypokalemia, and a family history of hypertension.

#### *Postural Testing*

When an abnormal ARR has been identified, confirmatory testing should be performed under more stringent conditions. Antihypertensive drug treatment (except alpha blockers) should be stopped if possible, and ideally, individuals should be on a normal to high-sodium diet. Resting and ambulant plasma aldosterone and renin (as PRA) should be measured. The rationale behind this is that in normal individuals, plasma aldosterone rises with standing. In GRA and the majority of aldosterone-producing adenomas, aldosterone is under ACTH control, resulting in diurnal variation of aldosterone levels and a loss of the normal postural response. For this reason, plasma cortisol should also be measured. In idiopathic aldosteronism due to adrenal hyperplasia and angiotensin II-responsive adrenal adenoma, aldosterone levels continue to rise on standing.

#### *Saline Infusion/Fludrocortisone Suppression*

In both tests, the rationale is that in normal individuals expansion of plasma volume will suppress plasma aldosterone concentrations, whereas in primary aldosteronism further volume expansion does not have the same suppressive effect.

#### *Intermediary Metabolites*

Both 18-hydroxy and 18-oxo cortisol are produced by the adrenal cortex by the action of aldosterone synthase on cortisol. Urinary excretion of these metabolites is increased in patients with aldosterone-producing adenomas, particularly in patients with GRA.

#### *Adrenal Imaging*

This should be performed only when aldosterone excess has been confirmed biochemically because a significant number of nonfunctioning adenomas (“incidentalomas”) can be found when the adrenals

are imaged. Although there have been no recent comparison studies, it appears that computed tomography (CT) is the preferred imaging modality. Most authors report that CT affords better spatial resolution and so is more sensitive in detecting smaller adenomas than is magnetic resonance imaging (MRI), although MRI may give more information on fat content of tumors. Nonetheless, it is estimated that CT detects only approximately 50 to 75% of aldosterone-producing adenomas and is of value in the identification of large (> 2.5 cm) adrenal masses that warrant discussion for excision based on their malignant potential.

#### *Adrenal Vein Sampling*

Although technically difficult and potentially hazardous, this remains the gold standard for diagnosis of adrenal adenomas. Given its risks, it is generally reserved for those in whom adrenal surgery is being considered, but it may be performed in others in whom there is diagnostic doubt. Diagnosis of adrenal adenoma can be made if aldosterone levels are elevated in one adrenal vein compared with the other. Simultaneous cortisol measurements ensure correct positioning of the cannula in the adrenal vein; a gradient of at least 300% between peripheral and central cortisol measurements confirms that this is the case.

In summary, the diagnosis of primary aldosteronism is problematic in that there are a number of potential diagnostic tests, many of which are affected by posture, electrolyte balance, and medications. [Figure 4](#) outlines a possible approach to the initial investigation of mineralocorticoid excess and PA.

#### **Management**

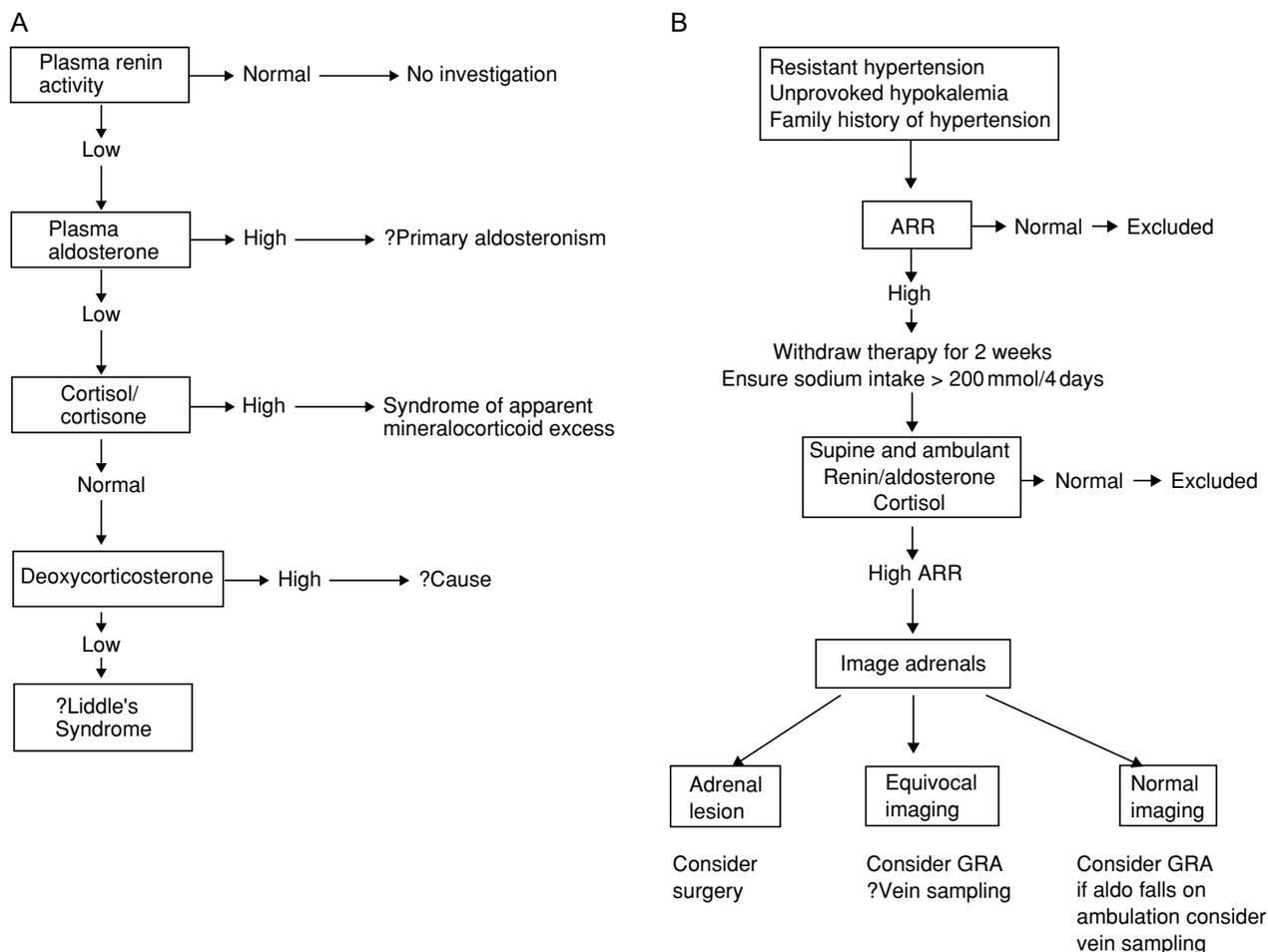
##### *Aldosterone-Producing Adenoma*

Surgical excision in suitably fit individuals can be curative. Preoperatively, patients should be treated with spironolactone to lower blood pressure and normalize plasma potassium. Spironolactone also reduces the risk of postoperative hypoaldosteronism because it allows recovery of the RAA axis, resulting in stimulation of the previously atrophic contralateral zona glomerulosa. After removal of an aldosterone-producing adenoma, serum potassium returns to normal in 100% of cases and blood pressure is normalized in 60% within a month and 75% within a year. Laparoscopic techniques keep morbidity to a minimum.

In patients who are unfit for or decline surgery, medical treatment is preferred.

##### *Idiopathic Aldosteronism*

The most appropriate treatment for idiopathic aldosteronism is blockade of the MR with aldosterone



**Figure 4** (A) Approach to mineralocorticoid excess. (B) Investigation of primary aldosteronism.

receptor antagonists such as spironolactone. Patients may require relatively high doses (occasionally up to 400 mg/day), although many do show a good response to much lower doses. There is evidence that patients with high ARRs, who do not have distinct adenomas, respond well to a relatively low dose of spironolactone. Dose titration is limited by dose-dependent side effects, including nausea, painful gynecomastia, and menstrual irregularities. The selective aldosterone receptor antagonist (SARA) eplerenone may offer considerable advantages, although results comparing the blood pressure lowering efficacy of this with spironolactone in PA are not yet available. Alternatives include the potassium-sparing diuretics, amiloride and triamterene, which act on the distal tubule of the nephron, inhibiting sodium-potassium exchange.

In practice, combination drug therapy using spironolactone or amiloride with one or more agents is

usually required for optimal blood pressure control. Additional agents include calcium channel blockers, angiotensin-converting enzyme (ACE) inhibitors, and angiotensin II antagonists.

#### *Glucocorticoid Remediable Aldosteronism*

Suppression of ACTH with dexamethasone provides a possible long-term therapeutic option. Patients can often be well controlled using small doses (e.g., 0.25 mg/day). Spironolactone is an alternative treatment.

#### *Other Uses of Selective Aldosterone Receptor Antagonists*

The occurrence of secondary aldosteronism in other circumstances, such as CCF, was discussed earlier. The concept of aldosterone-mediated cardiac injury has led to studies that explored the use of SARAs in patients with cardiac disease. The Randomized

Aldactone Evaluation Study (RALES) reported substantial benefit (30% reduction in mortality) in patients with advanced cardiac failure who were given spironolactone in addition to conventional treatment. More recently, the Eplerenone Post-Acute Myocardial Infarction Heart Failure Efficacy and Survival Study (EPHESUS) showed that eplerenone could provide substantial benefit (15% reduction in mortality) in patients following acute myocardial infarction. These are new and exciting data that illustrate the key importance of MR antagonism in cardiovascular pathophysiology and identify important therapeutic opportunities.

## CONCLUSION

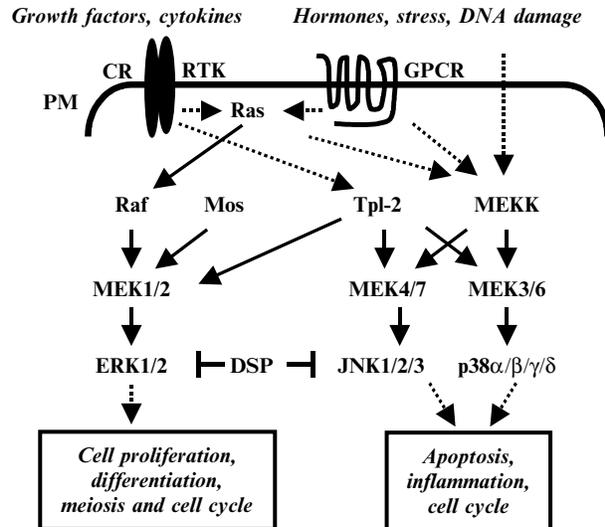
PA is now recognized as the most common cause of secondary hypertension and is thought to exist in up to 10% of unselected hypertensives. Aldosterone-producing adenomas are no longer the most common cause of PA but remain an important diagnosis as surgical removal offers the prospect of cure of hypertension. The majority of patients with PA are now thought to have bilateral adrenal hyperplasia. Development of new, more selective aldosterone receptor antagonists is the major therapeutic challenge to optimize blood pressure control and minimize side effects in this increasing number of patients.

## See Also the Following Articles

ACTH (Adrenocorticotrophic Hormone) • Aldosterone in Congestive Heart Failure • Angiotensin, Evolution of • Atrial Natriuretic Factor and Family of Natriuretic Peptides • Catecholamines • Primary Aldosteronism (PAL) • Tissue Renin-Angiotensin-Aldosterone System

## Further Reading

- Bubien, J. K., Ismailov, I. I., Berdiev, B. K., *et al.* (1996). Liddle's syndrome: Abnormal regulation of amiloride-sensitive sodium channels by beta subunit mutation. *Am. J. Physiol.* **270**, C208–C213.
- Farman, N., and Bocchi, B. (2000). Mineralocorticoid selectivity: Molecular and cellular aspects. *Kidney Intl.* **57**, 1364–1369.
- Lim, P. O., Dow, E., Brennan, G., Jung, R. T., and Macdonald, T. M. (2000). High prevalence of primary aldosteronism in the Tayside hypertensive clinic population. *J. Hum. Hypertension* **14**, 311–315.
- Lim, P. O., Young, W. F., and Macdonald, T. M. (2001). A review of the medical treatment of primary aldosteronism. *J. Hypertension* **19**, 353–361.
- Stewart, P. M. (1999). Mineralocorticoid hypertension. *Lancet* **353**, 1341–1347.
- Vallotton, M. B. (1996). Primary aldosteronism: I. Diagnosis. *Clin. Endocrinol.* **45**, 47–52.
- White, P. C., Mune, T., and Agarwal, A. K. (1997). GRA: Diagnosis, variability of phenotype, and regulation of potassium homeostasis. *Steroids* **60**, 48–51.
- Young, M. J., and Funder, J. W. (2002). Mineralocorticoid receptors and pathophysiological roles for aldosterone in the cardiovascular system. *J. Hypertension* **20**, 1465–1468.



**Figure 1** Schematic of the three major MAP kinase signaling pathways. Growth factors, cytokines, and other hormones activate receptor tyrosine kinases (RTK), cytokine receptors (CR), or G protein-coupled receptors (GPCR) on the plasma membrane (PM) and downstream MAP kinase pathways. Cell stress or DNA-damaging agents may activate signaling pathways independent of a receptor. Inactivation of MAP kinases occurs through dephosphorylation by dual-specificity phosphatases (DSP). Solid and dashed lines indicate direct and indirect interactions, respectively.

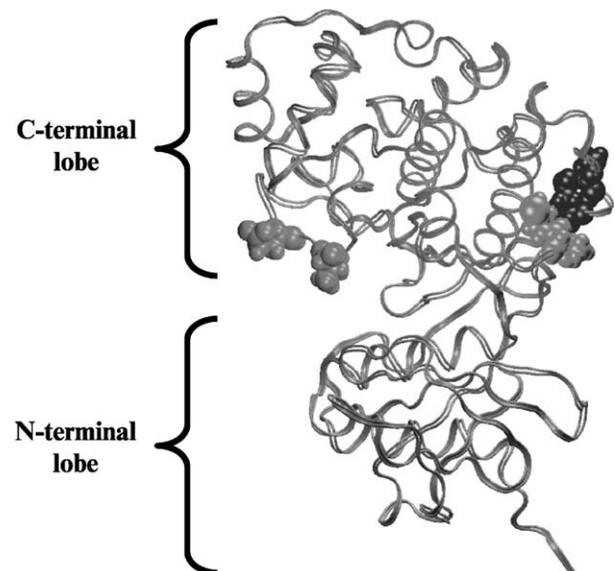
regulate gene expression and protein translation. Following activation, a fraction of ERK1/2 translocates to the nucleus and directly regulates gene expression by phosphorylating several transcription factors, including the p62 ternary complex factor (p62TCF; also called Elk-1), Myc, and proteins belonging to the Ets family. ERK proteins can also regulate the shape and motility of cells by phosphorylating structural proteins associated with microtubules, which are important determinants of the cell shape and architecture. Following activation, ERK proteins may also inhibit their own activity by phosphorylating the signaling protein son of sevenless (SOS), which inhibits further growth factor receptor activation of the ERK signaling pathway.

### ERK Structure and Activation

The ERK proteins contain both N- and C-terminal lobes with a conserved catalytic region containing the regulatory phosphorylation sites in the activation loop (Fig. 2). These regions are important for ATP nucleotide and substrate binding as well as phosphoryl transfer onto substrate proteins. Currently, the upstream kinases, MAP or ERK kinases-1 and-2 (MEK1/2), are the only known activators of ERK1/2.

MEK1 (44 kDa) and MEK2 (45 kDa) are dual-specificity kinases that can phosphorylate both the threonine and the tyrosine residues, which are separated by a glutamate residue within the tripeptide activation motif on ERKs. Although full activation of ERK proteins requires dual phosphorylation of both threonine and tyrosine residues, single phosphorylations on either of these residues may confer partial ERK protein activity. The MEK proteins are also activated through dual phosphorylation of two serine residues primarily by Raf kinases in somatic cells and the Mos kinase in germ cells. However, the MEK proteins may also be activated by other MEK kinases, including the tumor progression locus-2 (Tpl-2) serine/threonine kinase, which is involved in inflammatory responses; the MEK kinase-1, which is involved in apoptotic responses; and mixed lineage kinase-3, which may function as a promoter of proliferation or cell death depending on the extracellular stimulus.

Raf kinases are activated through phosphorylation by a variety of serine/threonine and tyrosine kinases and by recruitment to the cell membrane through interactions with Ras G proteins. Additional phosphorylations of Raf proteins may result in inhibition



**Figure 2** Diagram of the 3D structure of ERK2 showing the spatial relationship of the ERK2 N- and C-terminal lobes. The regulatory threonine and tyrosine residues within the activation loop are shown as spheres on the left. The ATP-binding region resides in the N-terminal lobe. Docking domains important for substrate interactions are represented by aspartate residues as depicted by the darker spheres on the right, and threonine residues as depicted by the lighter spheres on the right are also shown. This bilobed structure is common to many kinases.

of Raf catalytic activity and downstream ERK pathway signaling. For example, protein kinase B (also called Akt) proteins phosphorylate a serine residue in the Raf-1 catalytic domain and inhibit Raf activation of MEK and ERK. Ras proteins are coupled to the activated membrane receptors through adaptor proteins and provide a major link between plasma membrane receptors and activation of the ERK signaling pathway.

The activity of the Raf–MEK–ERK signaling module is also regulated through interactions with other binding proteins. For example, a protein called kinase suppressor of Ras may act as a scaffolding protein that binds to Raf, MEK, and ERK in a complex and determines the degree of ERK activation depending on the stimulus. In addition, a specific MEK-binding protein, MEK partner-1, may be important for directing MEK1 interactions with ERK1 but not ERK2. Lastly, a Raf-interacting protein, the Raf kinase inhibitor protein, may function by preventing Raf activation of MEK1/2. Although these and other ERK pathway binding proteins have been identified, their functional role in regulating ERK signaling events and biological functions remains largely unknown.

## JNK/SAPK MAP KINASES

The c-jun NH<sub>2</sub>-terminal kinase (JNK) proteins are an evolutionarily conserved family of serine/threonine protein kinases. The events leading to the identification and cloning of this family began in 1990, when a mammalian stress-activated protein kinase (SAPK) was discovered. These experiments identified a 54-kDa protein kinase that was activated in cells that were treated with the protein synthesis inhibitor cycloheximide. At the same time, a protein was discovered that had affinity for the transcription factor *c-jun* and phosphorylated *c-jun* on N-terminal residues when cells were exposed to ultraviolet (UV) radiation. In 1994, isolation and cloning experiments confirmed that the JNK and SAPK proteins were identical and belonged to the MAP kinase superfamily. Three isoforms, JNK-1,-2, and -3 (also known as SAPK- $\alpha$ , SAPK- $\beta$  and SAPK- $\gamma$ ), have been shown to exist as at least 10 alternatively spliced variants. Whereas JNK-1 and JNK-2 are ubiquitously expressed, JNK-3 has a restricted expression pattern and is found mainly in the brain, heart, and testis.

The JNK proteins have multifunctional roles and have been proposed to be involved in tumor development, cell growth and differentiation, apoptosis, survival, and cytokine production. JNK proteins may also

mediate cardiac hypertrophic responses during hypertension and ischemia injury in the heart and kidney, and they may be involved in several neurodegenerative diseases. The JNK signaling pathway therefore represents a potential target for therapeutic intervention.

## JNK Structure and Activation

JNKs have a similar core structure as that of ERK proteins but there are differences in the conformation of their activation loop, resulting in differences in the mechanism of regulation. The small N-terminal lobe helps in the orientation and binding of ATP, whereas the large C-terminal lobe aids in substrate recognition. JNKs, like ERK proteins, are activated by phosphorylation on threonine and tyrosine residues, which are separated by a proline within the tripeptide activation loop in the kinase domain.

Mammalian JNK proteins are activated by various extracellular stimuli, including growth factors, cytokines, and cellular stresses such as heat shock, hyperosmolarity, UV radiation, and ischemia/reperfusion. Although the stress-mediated activation mechanism is not clear, it has been hypothesized that stress factors induce receptor clustering and internalization, which lead to JNK activation. The organization of the JNK activation cascade is conserved between the other MAP kinase members, beginning with receptor activation and followed by recruitment of adaptor molecules and activation of small GTP-binding proteins. This is followed by three to four tiers of dual-specificity protein kinases, which culminate in the activation of the JNK proteins. Similar to the ERK proteins, JNK proteins target a variety of substrates and can translocate to the nucleus, where they regulate gene expression by phosphorylating transcription factors.

JNK proteins are activated through dual phosphorylation on threonine and tyrosine residues by MEK4 (also called SEK1 or MKK4) or MEK7 (also called MKK7). Three MEK4 and six MEK7 isoforms have been identified. These different isoforms demonstrate selectivity depending on the extracellular stimulus. For example, MEK7 is primarily activated by cytokines such as tumor necrosis factor (TNF) and interleukins (ILs). In contrast, MEK4 is primarily activated by environmental stress stimuli such as osmotic changes and DNA-damaging agents. Although both MEK4 and MEK7 proteins can activate JNK proteins, MEK4 can also activate the p38 MAP kinases. Another difference is in the specificity of the two MEKs toward the threonine and tyrosine residues within the activation site of JNK; MEK4

preferentially phosphorylates the tyrosine residue, whereas MEK7 has a higher affinity toward phosphorylating the threonine residue on JNK. The JNK-specific MEK proteins are activated by dual phosphorylation of a serine and threonine residue in their activation loop, and they are found in both the nucleus and the cytoplasm. Thus, the JNK proteins may be activated in both the nucleus and the cytoplasm.

A wide variety of upstream kinases phosphorylate the MEK proteins responsible for activating both the JNK and p38 MAP kinases. Some of these kinases include proteins belonging to the MEK kinases (MEKK1–4), the mixed lineage protein kinases (MLK1–3, DLK, and LZK), the apoptotic stimulating kinase (ASK1 and -2), and transforming growth factor- $\beta$  (TGF- $\beta$ )-activating kinase (TAK1) and Tpl2, which also activate MEK1/2. These proteins are in turn activated by the p21-activated kinases (Paks), the germinal center kinases (GCKs), and the hematopoietic progenitor kinases (HPKs). The activity of Paks, GCKs, and HPKs is regulated by G proteins in a manner analogous to Ras G protein activation of Raf kinases in the ERK MAP kinase pathway. In the case of the JNK and p38 MAP kinases, members of the Rho family of G proteins are involved in the initiation of the signaling cascades.

Similar to the ERK MAP kinases, several JNK pathway binding proteins have been identified that function in regulating the assembly and activation of the JNK pathway. Two such proteins have been identified as JNK-interacting proteins (JIPs), which lack enzymatic activity but act as important organizers of JNK pathway complexes. JIP1 and JIP2 are closely related proteins that can bind to JNK proteins, MEK7, and MLK proteins. JIP proteins may be involved in potentiating JNK activation in response to activation by MLK proteins.

## P38 MAP KINASES

The p38 MAP kinase [also known as the cytokine suppressive anti-inflammatory drug-binding protein (CSBP), reactivating kinase (RK), or SAPK] pathway is responsible for mounting a cellular response to many types of stress signals. Stress stimuli that activate p38 MAP kinase include osmotic and temperature shock, proinflammatory cytokines, hypoxia, reactive oxygen species, and irradiation. In some circumstances, p38 may be activated in response to certain growth factors. There are at least six isoforms of p38 that have alternative names, which are given in parentheses: p38  $\alpha_1/\alpha_2$  (CSBP1/2, Mpk2, RK, and

SAPK2a), p38  $\beta_1/\beta_2$  (SAPK2b), p38  $\gamma$  (ERK6 and SAPK3), and p38  $\delta$  (SAPK4). The p38 proteins  $\alpha_1$ ,  $\beta_1$ ,  $\gamma$ , and  $\delta$  are encoded by four separate genes, whereas the  $\alpha_2$  and  $\beta_2$  isoforms are the result of alternative splicing of the messenger RNA of the  $\alpha_1$  and  $\beta_1$  isoforms. Although all of these isoforms are similar enough to be considered members of the p38 family, they are expressed at different levels depending on cell type and have different affinities for p38 substrates. Thus, expression of these isoforms in varying amounts in different cell types may allow cells and tissues to fine-tune their responses to various stimuli.

## p38 Structure and Activation

The topological structure of p38 MAP kinase is similar to that of the ERK proteins, although some differences exist in the activation loop, substrate-binding regions, and the ATP binding site, which helps explain the differences in activation and regulation between the various MAP kinases. The direct activators of p38 MAP kinases, through dual phosphorylation of threonine and tyrosine residues, include primarily MEK3 and -6 and, in some cases, MEK4. Activators of MEK3, -4, and -6 were previously described. Like the ERK and JNK MAP kinase pathways, activation of p38 MAP kinase proteins occurs through a kinase cascade, which is often initiated by activated membrane receptors.

Upon activation, p38 MAPK phosphorylates other kinases in the cytosol and translocates to the nucleus, where it phosphorylates and activates transcription factors as well as proteins that modify the topological structure of DNA. Downstream targets of p38 MAP kinases regulate the expression of genes responsible for the inflammatory response, including cytokines such as TNF- $\alpha$ , IL-1 $\beta$ , and IL-6. Because of its key role as a regulator of the inflammatory response, p38 is currently the target of anti-inflammatory drugs designed to treat acute and chronic inflammatory diseases, such as rheumatoid arthritis and osteoarthritis.

## MAP KINASE ACTIVATION BY RECEPTOR TYROSINE KINASES

The ERK MAP kinase pathway is activated by several growth factor ligands that stimulate RTK activity, including epidermal growth factor (EGF), fibroblast growth factor, hepatocyte growth factor, insulin or insulin-like growth factors, platelet-derived growth factor (PDGF), and vascular endothelial growth factor. Ligand binding induces receptor dimerization

and the activation of intrinsic tyrosine kinase activity, which causes autophosphorylation of tyrosine residues in the cytoplasmic domain. The phosphorylated tyrosine residues of the receptors provide docking sites for Src homology-2 (SH2) domain-containing adaptor proteins, such as SH2-containing collagen-related proteins and growth factor receptor-bound protein-2 (Grb2). Grb2 recruits the guanine nucleotide exchange factor protein SOS, which promotes the active GTP-bound form of Ras and subsequent activation of the MAP kinase signaling pathways.

The JNK and p38 MAP kinases are activated mainly under cellular stress conditions, such as heat shock, irradiation, or hyperosmolarity. Although it is not clear which mechanisms are involved in mediating the stress response and activation of the JNK and p38 pathways, it has been proposed that cell stress may induce membrane receptor clustering and internalization, which could facilitate receptor activation. Nonetheless, similar to the ERK proteins, activation of the JNK and p38 MAP kinases pathways may also be through RTKs or cytokines such as TNF and IL-1 $\beta$ . RTK stimulation of Ras G protein, in addition to activating Raf-1 kinase and the ERK pathway, may activate MEK kinases involved in activating the JNK and p38 pathways.

### ACTIVATION OF MAP KINASES THROUGH G PROTEIN-COUPLED RECEPTORS

G protein-coupled receptors (GPCRs) are integral membrane proteins that contain regions that pass through the membrane seven times; thus, the GPCRs are also referred to as the seven-transmembrane spanning or serpentine receptors. GPCRs are coupled to heterotrimeric G proteins that contain  $\alpha$ ,  $\beta$ , and  $\gamma$  subunits that determine which signaling pathways will be used. The G protein  $\alpha$  subunit binds to GTP and contains the GTP-hydrolyzing activity, whereas the  $\beta$  and  $\gamma$  subunits contain regulatory information. A wide variety of GPCRs, including the adenosine A<sub>1</sub>,  $\alpha$ -adrenergic, and muscarinic acetylcholine receptors, as well as endothelin, angiotensin, histamine, glucagon, and thrombin receptors, initiate physiological responses through activation of MAP kinase pathways.

GPCR activation of MAP kinases may occur through G protein activation of phospholipases, which hydrolyze membrane phospholipids and generate second messengers such as diacylglycerol (DAG) and inositol triphosphate (IP<sub>3</sub>). IP<sub>3</sub> stimulation of

intracellular Ca<sup>2+</sup> and DAG can activate protein kinase C, which phosphorylates and activates Raf kinase. Alternatively, activated GPCRs may transactivate RTKs, which feed into downstream MAP kinase signaling pathways. For example, MAP kinase activation by endothelin or thrombin receptors is coupled to coactivation of the EGF receptor family of RTKs, and the RTK activity is required for endothelin- or thrombin-induced MAP kinase activity.

Other mechanisms for activating MAP kinase pathways may be through the G protein  $\beta$  and  $\gamma$  subunits, which can activate nonreceptor tyrosine kinases belonging to the Src protein family. Src kinases can phosphorylate tyrosine residues on the cytoplasmic domains of receptors and enhance Ras activity or directly phosphorylate and regulate Raf kinase activation. Activation of JNK and p38 MAP kinases by GPCRs has been shown to involve the G $\alpha$  and the G $\beta\gamma$  subunits as well as the activation of the Rho and Ras families of GTPases. Depending on the cell type and the receptor stimulated, G $\alpha$  or G $\beta\gamma$  or both may be involved in the activation of the MAP kinase pathways.

### OTHER MECHANISMS FOR ACTIVATING MAP KINASES

Activation of the MAP kinase signaling pathways may occur through cell-permeable factors such as steroid hormones, which bind to soluble cytoplasmic and nuclear receptors. For example, estradiol, progesterone, and testosterone activate MAP kinase indirectly through steroid receptor interactions with membrane-associated receptors. One proposed mechanism is that estradiol receptors are coupled to membrane-bound receptors such as the EGF receptor, which then facilitate the activation of the MAP kinase signaling pathway in a manner similar to transactivation by GPCRs described previously. Moreover, GPCRs may mediate steroid hormone activation of RTK and MAP kinase signaling.

Some membrane-bound receptors belonging to the receptor serine/threonine kinase (RSTK) family that undergo serine or threonine autophosphorylation on the cytoplasmic domains regulate MAP kinase signaling pathways. For example, the TGF- $\beta$  receptors are RSTKs that regulate a wide variety of biological responses, including cell proliferation, differentiation, extracellular matrix production, and cell death, through the p38 MAP kinases. Activated TGF- $\beta$  receptor stimulates TAK1, which is a novel MEK kinase that stimulates MEK3/6 and p38 and may inhibit cell proliferation.

## MAP KINASE PHOSPHATASES

Inactivation of MAP kinase signaling pathway is accomplished largely through dephosphorylation of threonine and tyrosine residues by unique protein tyrosine phosphatases called dual-specificity phosphatases (DSPs). Of the DSP family, at least six MAP kinase phosphatases (MKPs) have been identified. The MKP proteins show some specificity toward the MAP kinase family members. For example, although MKP1 and MKP2 can target activated ERK, JNK, and p38, MKP3 and MKP4 are specific toward ERK proteins, and MKP5 and MKP6 specifically dephosphorylate JNK and p38. Growth factor activation of the ERK proteins stimulates the expression of MKP1, thus providing a mechanism by which ERK proteins down-regulate their own activity. Other phosphatases belonging to the serine/threonine phosphatases may also play an important role in down-regulating MAP kinases. For example, protein phosphatase 2A (PP2A) is able to inhibit MEK1/2 and ERK1/2 activity in the absence of growth factor-induced DSPs. Similarly, PP2A has been identified as a JNK phosphatase. MAP kinase pathways are also regulated by protein serine/threonine phosphatases of the type 2C (PP2C) family. At least two members of the PP2C family, PP2C $\alpha$  and - $\beta$ , dephosphorylate and inactivate the JNK MAP kinases.

## CONCLUSION

The MAP kinase signaling pathways regulate most physiological functions throughout the life span of a diverse range of organisms, from yeast to humans.

Activation of MAP kinases occurs through plasma membrane receptor-dependent and -independent mechanisms in response to a variety of extracellular stimuli. Importantly, dysregulation of MAP kinase pathways is involved in many human diseases. Therefore, understanding of how MAP kinases are regulated and function is an important goal for the development of new therapies.

## See Also the Following Articles

Janus Kinases and Cytokine Receptors • Lipid Second Messengers and Receptors • Receptor Serine/Threonine Kinases

## Further Reading

- Cobb, M. H., and Goldsmith, E. J. (1995). How MAP kinases are regulated. *J. Biol. Chem.* **270**, 14843–14846.
- Davis, R. J. (2000). Signal transduction by the JNK group of MAP kinases. *Cell* **103**, 239–252.
- Keyse, S. M. (2000). Protein phosphatases and the regulation of mitogen-activated protein kinase signaling. *Curr. Opin. Cell. Biol.* **12**, 186–192.
- Lewis, T. S., Shapiro, P. S., and Ahn, N. G. (1998). Signal transduction through MAP kinase cascades. *Adv. Cancer Res.* **74**, 49–139.
- Lowes, V. L., Ip, N. Y., and Wong, Y. H. (2002). Integration of signals from receptor tyrosine kinases and G protein-coupled receptors. *Neurosignals* **11**, 5–19.
- Luttrell, L. M., van Biesen, T., Hawes, B. E., Koch, W. J., Krueger, K. M., Touhara, K., and Lefkowitz, R. J. (1997). G-protein-coupled receptors and their regulation: Activation of the MAP kinase signaling pathway by G-protein-coupled receptors. *Adv. Second Messenger Phosphoprotein Res.* **31**, 263–277.
- Ray, L. B., and Sturgill, T. W. (1987). Rapid stimulation by insulin of a serine/threonine kinase in 3T3-L1 adipocytes that phosphorylates microtubule-associated protein 2 in vitro. *Proc. Natl. Acad. Sci. USA* **84**, 1502–1506.

particles that contain apo's B100, CII, CIII, and E. After coupling to VLDL receptors in extrahepatic tissues, the triglycerides in VLDL undergo lipolysis by lipoprotein lipase (LPL), which requires apoCII as a cofactor. The liberated free fatty acids (FFA) are taken up by muscle or adipose tissue (through CD36) for oxidation or energy storage, respectively. In this process, VLDL remnants become intermediate density lipoproteins (IDL) that are enriched with apoE and then low-density lipoproteins (LDL). The latter step is also mediated by hepatic lipoprotein lipase (HL), which is attached to the endothelial surface in the liver by heparan sulphate proteoglycans (HSPG) and is facilitated by apoE. VLDL remnants and IDL are removed from the circulation by the low-density lipoprotein receptor (LDLR) and LDLR-related protein (LRP), largely through recognition of apoE. LDL is removed by the LDLR through recognition of apoB100.

In the exogenous pathway, dietary cholesteryl esters and triglycerides are secreted from the intestines in chylomicrons containing apoB48, A-I, and A-IV. They lose A-I and A-IV but acquire apo's CII, CIII, and E. As in VLDL, chylomicron triglycerides are hydrolyzed by LPL in the circulation and the chylomicron remnants are removed from the circulation by LDLR and LRP mainly in the liver through recognition of apoE.

In the reverse cholesterol transport pathway, cholesterol is removed from peripheral tissues by high-density lipoprotein (HDL), which binds to cholesterol-rich cells, at least partly through the SRB1 receptor. The uptake of cholesterol from these cells by HDL is facilitated by the recently characterized ABC1 transporter. HDL-associated cholesterol is then esterified by plasma lecithin cholesterol acyltransferase (LCAT), enlarging the size of the HDL particle. Cholesterol from HDL can be transported into the liver through the action of HL, by the action of the SRB1 receptor, or by the uptake of entire HDL particles. Alternatively, cholesteryl ester transfer protein (CETP) mediates HDL cholesteryl ester to triglyceride-rich particles instead of triglycerides, which are then removed from the circulation by liver LDLR and LRP. In the liver, cholesteryl ester is hydrolyzed and the cholesterol is excreted in the bile as bile acids or free cholesterol.

## LIPID-LOWERING DRUGS

Hydroxy-methylglutaryl coenzyme A (HMG-CoA) reductase inhibitors inhibit the rate-limiting step in hepatic cholesterol synthesis, causing an increase in LDL receptor levels in hepatocytes and enhancement

of remnant and LDL cholesterol removal from the circulation. In addition to lowering LDL cholesterol, they lower triglyceride levels (by decreased hepatic apoB production) and increase HDL cholesterol. Large intervention trials have demonstrated the potency of these drugs in primary or secondary prevention of cardiovascular disease in patients with or without lipid disorders.

Fibric acids stimulate the activity of the liver transcription factor PPAR $\alpha$  that increases LPL activity, thereby enhancing VLDL degradation. PPAR $\alpha$  stimulation also reduces apoCIII, which facilitates VLDL remnant uptake. VLDL production may also be reduced. Stimulation of peroxisomal fatty acid oxidation by fibrates may also contribute to the triglyceride-lowering actions. In addition to reducing triglycerides, fibrates increase HDL cholesterol levels. In several intervention trials, fibrates have been demonstrated to reduce CHD events, but conflicting data on mortality are reported.

Nicotinic acid derivatives decrease both total and LDL cholesterol, reduce VLDL cholesterol levels, and raise HDL cholesterol levels. The mechanism of action is not fully understood, but it appears to inhibit the secretion of apoB-containing lipoproteins from the liver as well as lipolysis in peripheral adipose tissue. In the Coronary Drug Project, niacin significantly reduced CHD events.

## PRIMARY MIXED LIPEMIAS

### Type III Hyperlipoproteinemia (Familial Dysbetalipoproteinemia)

Type III hyperlipoproteinemia is a lipoprotein disorder with a prevalence of 1 to 4 in 10,000. The disorder is usually diagnosed in adults and has a male predominance. Individuals have increased plasma cholesterol and triglyceride levels resulting from increased triglyceride-rich remnant concentrations. Plasma LDL levels are usually reduced, and HDL levels are usually normal. A typical clinical feature is the presence of planar xanthomas at the palmar surface. Patients with familial type III hyperlipoproteinemia have increased susceptibility to CHD and peripheral vascular disease. The typical disorder is caused by a mutation in the apoE gene. The apoE gene has three alleles coding for E4, E3, and E2 that occur in Caucasians with frequencies of 15, 77, and 8%, respectively. The apoE polymorphisms account for approximately 10% of the population variance in LDL cholesterol levels. More than 95% of type III hyperlipoproteinemia individuals have the E2/E2

phenotype. This common E2 mutation, representing an Arg158Cys mutation, causes a nearly total defect of apoE in LDLR binding. The result is a less efficient clearance of VLDL and chylomicron remnants in the liver, leading to increased triglyceride and cholesterol levels. LDL is decreased as a result of the decreased conversion of IDL to LDL. Individuals with heterozygosity for the E2 mutation do not exhibit the phenotype, indicating that the disease is inherited in an autosomal-recessive fashion. Interestingly, only 1 to 4% of individuals with the E2/E2 genotype have the typical phenotype. Hence, type III hyperlipoproteinemia with E2/E2 homozygosity displays a low rate of penetrance. Additional factors that are associated with increased VLDL or chylomicron production or with decreased clearance, such as gender, obesity, hypothyroidism, diabetes mellitus and certain dietary factors, are thought to contribute to manifesting the disease.

The autosomal-dominant inheritance pattern of type III hyperlipoproteinemia has been described for some rare apoE variants. For instance, individuals with heterozygosity for apoE2Lys146Gln or apoE3-Leiden exhibit the phenotype. Nearly all carriers of these rare alleles have the hyperlipoproteinemic phenotype, indicating that this dominant inheritance pattern associates with a high rate of penetrance. In the treatment of this disorder, a thorough search for underlying disorders should be made. Usually, dietary measures (restriction of calories) may normalize lipid levels. If this cannot be achieved, drug therapy may be necessary. HMG-CoA reductase inhibitors have been applied successfully, and fibric acid derivatives may also be useful.

### Familial Combined Hyperlipidemia

Familial combined hyperlipidemia (FCH) is a common lipid disorder affecting 0.5 to 2.0% of the population. The disease occurs in both adults and children. Biochemically, FCH is characterized by familial occurrence of increased cholesterol and triglyceride levels. The FCH lipoprotein pattern includes elevated apoB and LDL cholesterol levels, diminished HDL cholesterol levels, and the presence of small dense LDL. The phenotype may vary within one affected family and even within one individual. Xanthomas are not observed in FCH. Patients with FCH have an increased risk for CHD. FCH is associated with increased body mass index (BMI), waist/hip ratio, and fasting glucose and insulin levels, which are (together with the lipid abnormalities) characteristics of the metabolic syndrome (syndrome X). Thus,

knowledge of the pathogenesis of FCH will provide insight into the metabolic syndrome and vice versa.

Initially, the dominant mode of inheritance was thought to indicate a monogenetic cause. Several candidate genes or chromosomal regions have been suggested. The apoA-I/CIII/A-IV gene cluster, LCAT (chromosome 16q22.1), and 11p were found in Dutch FCH families. However, this could not be confirmed in Finnish FCH families. In contrast, Finnish studies identified candidate regions on chromosomes 1q, 2q, 10p, 10q, and 21q, but in turn, the strong chromosome 1 locus was not found in the Dutch families. These varying results suggest that FCH is a complex disorder where many genes are involved and a large heterogeneity in gene-gene and gene-environment interactions is present.

In the treatment, dietary and lifestyle interventions should be attempted and may be combined with drugs when necessary. Drug therapy should be aimed at the predominating lipid disorder. HMG-CoA reductase inhibitors are a logical choice, and fibric acid derivatives and niacin may be useful as well. Bile acid sequestrants should be avoided because they tend to raise triglyceride levels.

## SECONDARY MIXED LIPEMIAS

### Diabetic Dyslipidemia

Worldwide, diabetes mellitus (DM), and especially type 2 DM, is a rapidly expanding health problem, with estimates suggesting a doubling of its prevalence during the coming years. Cardiovascular disorders are the major cause of mortality in DM. Cardiovascular disorders, in turn, are considered the result of the combined effects of hypertension and alterations in lipid and carbohydrate metabolism, the fibrinolytic system, and inflammatory cascades. The importance of lipid abnormalities in DM is underscored by the fact that in large studies of type 1 and type 2 DM, optimization of glycemic control could prevent microvascular complications but not macrovascular ones. In contrast, in large lipid intervention studies, subgroup analyses revealed favorable effects of lipid-lowering drugs on CHD risk. Lipoprotein disorders in DM are related primarily to abnormalities in triglyceride metabolism, and this may be explained in part by the role of insulin in lipid metabolism. In the absence of insulin action (e.g., insulin deficiency in untreated type 1 DM or insulin resistance), lipolysis from adipose tissue is increased by the action of hormone-sensitive lipase. The increased FFA flux to the liver and the decreased degradation of apoB, also a

result of decreased insulin action, lead to enhanced VLDL production. As insulin facilitates LPL activity, VLDL (and chylomicron) remnant clearance is decreased when insulin action is reduced, thereby further increasing triglyceride (and VLDL-derived) cholesterol levels. Although LDL levels are expected to be lower (due to decreased conversion of VLDL remnants), LDL clearance may also be decreased.

In the presence of triglyceride-rich lipoproteins, enhanced CETP activity may result in a net transfer of cholesteryl ester from LDL and HDL to VLDL remnants. As a result, LDL particles become smaller and denser, and this is thought to be unfavorable with respect to atherosclerosis. The removal of cholesteryl ester from HDL leads to reduced plasma HDL cholesterol measurements. In addition, HDL clearance in the liver is enhanced. Another factor in decreased HDL is the diminished transport of surface lipids from VLDL to nascent HDL caused by decreased LPL activity; this impairs HDL maturation. Decreased HDL cholesterol is associated with increased risk for CHD.

Lipoprotein abnormalities in type 1 DM are the result of insulin deficiency and can be corrected completely by insulin therapy. The situation in type 2 DM is complex; in contrast to type 1 DM, lipid abnormalities usually are not corrected completely with glycemic control. Moreover, this dyslipidemia is often found in patients with insulin resistance without overt diabetes and is one of the features of the metabolic syndrome. Abnormalities in insulin action rather than hyperglycemia are associated with this lipid abnormality, and evidence suggests that a pathological FFA flux from visceral fat to the liver even precedes defective insulin action. Given the high CHD risk in DM, aggressive lipid-lowering therapy is of major importance. Apart from dietary intervention, from a theoretical point of view, HMG-CoA reductase inhibitors, fibrates, and nicotinic acid are rational. Because HMG-CoA reductase inhibitors have been demonstrated to reduce CHD risk in DM patients in large randomized studies, this category of drugs should be the first choice.

### Nephrotic Syndrome

The nephrotic syndrome is almost always accompanied by hyperlipidemia. Plasma VLDL, IDL and LDL cholesterol, and total triglycerides may be increased, whereas HDL cholesterol is decreased. Mortality from CHD is particularly high in the nephrotic syndrome. The increased VLDL and IDL concentrations

may result from decreased clearance due to reduced LPL activity on the vascular endothelium that, in turn, may result from either decreased synthesis or inadequate binding of this enzyme by HSPG to endothelial surfaces. The mechanism of the increased LDL levels is not clear. Synthesis of apoB100 is not related to that of albumin, suggesting a mechanism different from the increased synthesis of nonlipoproteins. HDL concentrations in the nephrotic syndrome are normal, but maturation is impaired, leading to a shift from the larger HDL2 to the smaller HDL3 and resulting in decreased plasma HDL cholesterol concentrations. Because of the high CHD risk, lipid-lowering therapy—with HMG-CoA reductase inhibitors as first choice—must have high priority.

### Hypothyroidism

Because thyroid hormone plays a role in the LDLR expression, clinical or subclinical hypothyroidism is associated with decreased LDL clearance and, consequently, increased LDL cholesterol levels; few patients with hypothyroidism have normal lipid profiles. In addition, hypothyroidism is present in approximately 5% of patients who present primarily with lipid disorders. Hypothyroidism may also be associated with increased triglyceride levels, and this is thought to be the result of decreased LPL activity. Whether hyperlipidemia in hypothyroidism is associated with increased risk for CHD is still being debated. Substitution with thyroid hormone usually corrects the lipid abnormalities.

### Other Conditions

#### **Hypopituitarism**

In patients with hypopituitarism, mixed lipemia may be present. This is likely the result of the combined contribution of hypothyroidism and growth hormone deficiency, the latter of which is accompanied by increased LDL cholesterol levels.

#### **Glucocorticoid Excess**

Excess of glucocorticoids, either in the context of Cushing's syndrome or by steroid hormone therapy, may be associated with increased LDL and VLDL concentrations. Increased VLDL production and increased conversion of VLDL to LDL appear to play a role. Because glucocorticoid excess is also associated with impaired glucose tolerance, lipoprotein abnormalities (as observed in DM) may also be present.

Cushing's syndrome is associated with increased CHD risk; however, the independent contribution of dyslipidemia has not been identified.

### **Human Immunodeficiency Virus**

Patients with human immunodeficiency virus (HIV) infection who are treated with HIV-1 protease inhibitors can develop hyperglycemia, hypertriglyceridemia, or hypercholesterolemia. The mechanism appears to be multifactorial, but data indicate that this effect may be at least partly accounted for by decreased degradation of apo-B and, hence, increased VLDL synthesis. Decreased PPAR $\alpha$  activity may lead to effects opposed to those of fibrates, whereas accumulation of the active portion of sterol regulatory element-binding protein-1c are also involved.

### **See Also the Following Articles**

Diabetes, Type 1 • Diabetes, Type 2 • Dysbetalipoproteinemia and Type III Hyperlipidemia • Familial Low Cholesterol Syndromes, Hypobetalipoproteinemia • Hypertriglyceridemia • Hypopituitarism • Hypothyroidism, Subclinical • Lipoprotein(a) • Low HDL/High HDL Syndromes

### **Further Reading**

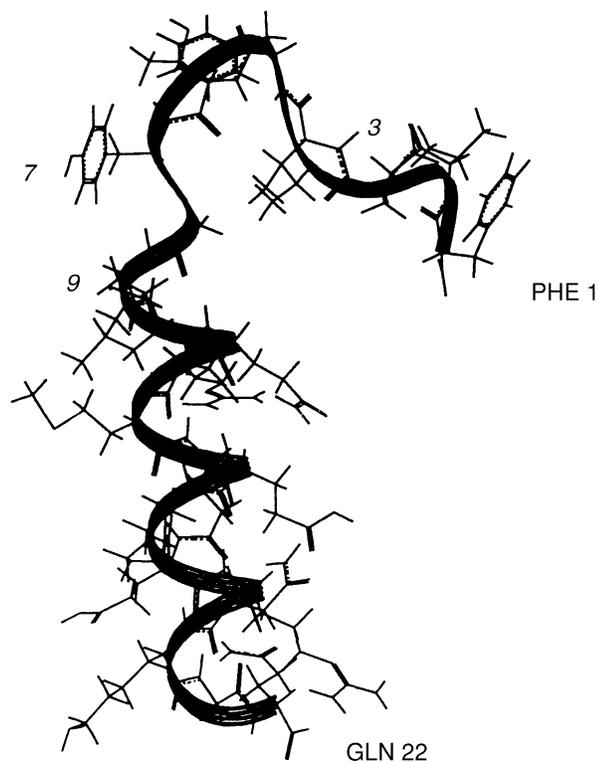
- Aouizerat, B. E., Allayee, H., Bodnar, J., Krass, K. L., Peltonen, L., deBruin, T. W. A., and Rotter, J. I. (1999). Novel genes for familial combined hyperlipidemia. *Curr. Opin. Lipidol.* **10**, 113–122.
- Breslow, J. L. (2000). Genetics of lipoprotein abnormalities associated with coronary heart disease susceptibility. *Annu. Rev. Genet.* **34**, 233–254.
- Faergeman, O. (2000). Hypertriglyceridemia and the fibrate trials. *Curr. Opin. Lipidol.* **11**, 609–614.
- Goldberg, I. J. (2001). Diabetic dyslipidemia: Causes and consequences. *J. Clin. Endocrinol. Metab.* **86**, 965–971.
- Gotto, A., and Pownall, H. (1999). "Manual of Lipid Disorders." Williams & Wilkins, Baltimore, MD.
- Mahley, R., and Rall, S. C. (1995). Type III hyperlipoproteinemia: The role of apolipoprotein E in normal and abnormal lipoprotein metabolism. In "The Metabolic and Molecular Bases of Inherited Disease" (C. Scriver, A. L. Beaudet, W. S. Sly, and D. Valle, eds.), pp. 1953–1980. McGraw-Hill, New York.
- Steiner, G. (2000). Diabetes, lipids, and coronary heart disease: What have we learned from lipid lowering trials? In "Lipids and Vascular Disease" (D. J. Betteridge, ed.), pp. 159–172. Martin Dunitz, London.
- Thompson, G. R., and Barterba, P. J. (2000). Therapeutic approaches to reducing the LDL- and HDL-associated risks of coronary heart disease. *Curr. Opin. Lipidol.* **11**, 567–570.
- Wilson, P. W., Schaefer, E. J., Larson, M. G., and Ordovas, J. M. (1996). Apolipoprotein E alleles and risk of coronary disease: A meta-analysis. *Arterioscl. Thromb. Vascular Biol.* **16**, 1250–1255.

in Merck's laboratories followed a different strategy by first cloning and sequencing a large variety of G protein-coupled receptors and then looking at the nature of ligands for these receptors. Starting from an orphan receptor found in the thyroid with a structure close to that of the receptor for the growth hormone secretagogue (GHS), mass screening of more than 500 peptide and nonpeptide molecules was performed to obtain positive results for motilin. Much of our knowledge of the motilin receptor has been derived from structure-activity studies of peptide analogues and binding of radioactive ligands on various tissue membranes prepared *in vitro*.

Nuclear magnetic resonance showed that the motilin molecule (Fig. 1) is shaped like a golf club or pastoral staff. Its N-terminal portion is responsible for receptor binding and *in vitro* bioactivity. Peptide fragments containing the first 12 amino acids of the molecule have demonstrated full binding capacity *in vitro*, whereas removal of its first amino acid is enough to abolish its activity. Amino acids 4 and 7 also seem to play a major role in this N-terminal sequence. Interestingly, the C-terminal structure, not required for

*in vitro* bioactivity, was found to be essential for *in vivo* bioactivity. N-terminal fragments of 15 amino acids, fully active *in vitro*, were indeed inactive *in vivo*, whereas longer fragments with 19, 20, or 21 amino acids could mimic the motor activity of the 22-amino acid peptide. This is an interesting concept in the field of peptide pharmacology, where the  $\alpha$ -helix-shaped C-terminal structure seems to be important in protecting the whole molecule, probably against circulating degrading enzymes.

Receptors have been studied in membrane solutions prepared from various species (mostly humans or rabbits) and from neural or muscle elements of the intestinal wall. The binding of motilin analogues was different in membranes prepared from the human or rabbit antrum, suggesting that motilin receptors' structure is, as for the motilin substance discussed previously, different among species. Solutions enriched in muscle or in neural elements of the intestinal wall have allowed the identification of specific and distinct responses to various motilin synthetic analogues, indicating the existence of structurally heterogeneous motilin receptor subtypes specific to muscle or neural elements (M and N receptor subtypes).

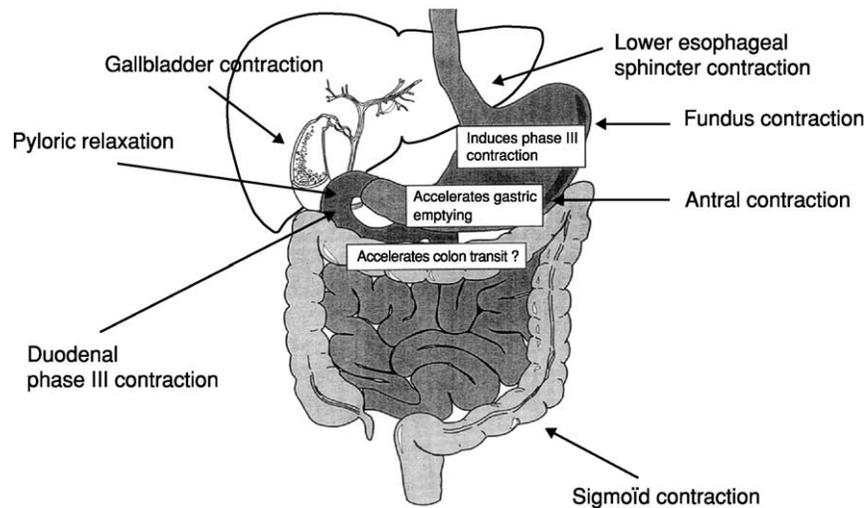


**Figure 1** Structure of the 22-amino acid motilin peptide revealed by nuclear magnetic resonance. The N-terminal curved sequence binds to the receptor, whereas the  $\alpha$ -helix C-terminal structure protects the molecule for *in vivo* bioactivity.

## PHARMACOLOGICAL ACTION

As its name implies, motilin acts on digestive motility. Figure 2 shows the expected effect in humans after the administration of motilin or motilin receptor agonists. Stimulation of antro-duodenal motility remains the dominant action of the peptide.

Initial *in vitro* studies with intestinal tissues from rabbits and humans revealed that motilin can stimulate smooth muscle contraction by a direct effect on muscle cells; the peptide action was seen despite the addition of all neural blockers, including tetrodotoxin. Experiments with isolated muscle cells have confirmed the presence of motilin receptors on muscle cell membranes. However, *in vivo* studies in dogs and humans clearly indicated that the motor action of motilin is mediated by muscarinic transmitters. As discussed previously, binding experiments supported the existence of motilin receptor subtypes (M and N) that are functionally and structurally different. Currently, the most interesting hypothesis proposes that the interdigestive motor action of motilin (induction of phase III contraction) is elicited at low doses through an action via neural cholinergic receptors, whereas postprandial stimulation of antral motility is evoked at higher doses via muscle receptors.



**Figure 2** Biological effects obtained in the human gastrointestinal tract in response to the injection of motilin or a motilin receptor agonist such as erythromycin.

## PHYSIOLOGICAL ROLE

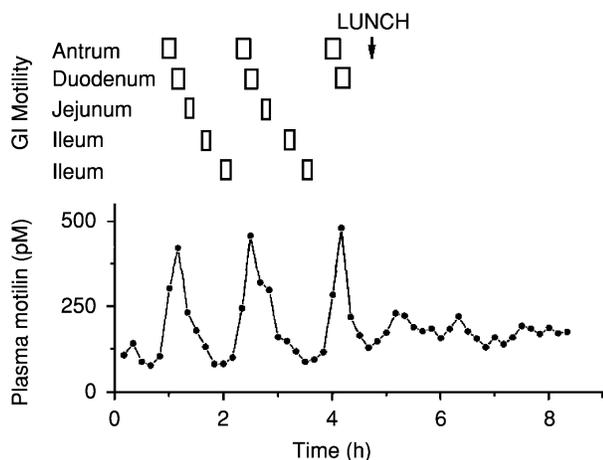
The characteristic action of motilin is the induction of phase III contraction of the migrating motor complex (MMC). The MMC is the basic organization of motor activity of the gut during the fasting interdigestive period. It lasts 80 to 120 min and consists of three successive phases. During phase I, no significant contraction is seen for 20 to 60 min. During phase II, intermittent and irregular contractions start to occur 20 to 60 min before phase III, during which strong peristaltic contractions, lasting 3 to 10 min, start from the stomach and lower esophagus to migrate distally to the duodenum, jejunum, and ileum until the colon. This phase III peristaltic wave has been proposed to clean, from the GI tract, bacteria or nutrients that could accumulate during the digestive period, with deleterious effects on gut (e.g., bacterial overgrowth). Feeding interrupts the intermittent cyclical fasting motility (MMC) and induces a more constant motor activity of moderate amplitude to allow the optimal absorption of ingested nutrients; accelerated transit during phase III would impair nutrient absorption.

Regulatory peptides can exert their influence via different pathways: endocrine (hormonal), neurocrine, paracrine, or autocrine. Motilin is in the unique position, as a GI regulatory peptide, of being an interdigestive hormone. This has been well identified in the dog, where all of Morton Grossman's criteria establishing the endocrine contributions of peptides have been fulfilled. First, regulation of the cyclical pattern of the MMC is under the control of circulating

factors, as shown in various experiments where the MMC persisted in animals when their stomachs had been completely denervated. Second, there is a perfect correlation between circulating peak levels of motilin and the initiation of phase III contractions from the stomach or proximal duodenum. Third, exogenous motilin, given in small doses reproducing physiological plasma variations, induces phase III contraction of the MMC. Fourth, inhibition of circulating motilin by the administration of specific motilin antisera blocked phase III contractions from the upper gut. Although the situation could not be explored under such perfect experimental conditions in humans, most evidence suggests that circulating motilin also plays a key role in the regulation of phase III initiated from the antrum of our species.

## RELEASE MECHANISMS

Most gastrointestinal hormones are released after a meal to allow or facilitate the digestion and absorption of nutrients. Motilin is a unique hormone. It is released periodically during the interdigestive fasting period, and its cyclical release is abolished after a meal, as shown schematically in Fig. 3. Therefore, a "biological clock" somewhere in the organism periodically signals motilin cells to release the peptide into the circulation. *In vitro* preparations of intestinal mucosal cells enriched in motilin cells showed that muscarinic receptors are present on the motilin cell membrane and that protein kinase C activators are the most potent



**Figure 3** Schematic representation of plasma motilin variations in dog. During the fasting interdigestive period, motilin is released cyclically every 80 to 120 min (lower panel) to induce phase III contraction of the migrating motor complex from the stomach to the ileum (indicated by clear boxes in the upper panel). After eating, the motilin cyclical peak increases are abolished for 2 to 8 h (depending on the content and nature of the meal) while the fed pattern motility profile is taking place.

second messengers eliciting motilin release. In the *ex vivo* perfused canine intestine, bombesin has been identified as a direct stimulant of motilin release, whereas the effect of opiates is mediated by acetylcholine. Phenylephrine and somatostatin appear to act on M-cell membrane receptors to block release of the peptide. In humans, meal ingestion is followed by a very early and brief increase in plasma motilin before the interdigestive release cycle is interrupted. This early release can be mimicked by central stimulation with modified sham-feeding and by distension of the fundus with an air-filled balloon. The contribution of this postprandial motilin release (not present in the dog) in the process of nutrient digestion remains to be characterized.

## CONTRIBUTION TO CLINICAL MEDICINE

Some GI hormones, such as gastrin and vasointestinal polypeptide (VIP), are important to clinicians because of the disease symptoms they can generate. Up to now, there is no clinical phenotype attributed to motilin hypersecretion. High levels of circulating motilin have been documented in some patients with pancreatic tumors and Zollinger-Ellison syndrome as well as in patients with carcinoid tumors of the gut. Although it could be tempting to speculate on the

role of motilin in the diarrhea found in these patients, its contribution to the biological alterations remains unknown.

Because motilin hypersecretion could be expected to generate GI hypermotricity and hypersecretion with probable diarrhea, it is logical to expect that hypomotilinemia will induce GI hypomotility. Some investigators have indeed found low levels of plasma motilin in patients with idiopathic intestinal pseudo-obstruction and idiopathic or postoperative gastroparesis. To this point, however, plasma motilin measurement has not been shown to be useful for inclusion in the workup diagnoses of any clinical situations.

On the other hand, motilin was recently of major interest to medical clinicians because of the capacity of motilin receptor agonists to act as powerful stimulants of GI motor activity in patients with hypokinetic disorders. Itoh was the first to observe that erythromycin could mimic the motor effect of motilin when injected in dogs. It was soon established that erythromycin was in fact acting on motilin receptors, and Janssens and colleagues made the capital observation that erythromycin was the most potent gastrokinetic ever tested to stimulate gastric emptying in diabetic patients with gastroparesis. Since then, erythromycin, whether administered intravenously (i.v.) or by mouth (p.o.), is used by many clinicians for the treatment of patients with gastroparesis or intestinal pseudo-obstruction. Motilides, or motilin receptor agonist substances derived from the erythromycin macrolide and with improved gastrokinetic activity but devoid of antibiotic properties, have been developed by many pharmaceutical companies. At least three motilides have been tested in humans, but for various reasons (e.g., rapid tachyphylaxis, no significant clinical benefit, potential side effects), clinical trials with these newly derived molecules have failed to substantiate the impressive pharmacological potential seen with i.v. erythromycin. Whether the gastrokinetic capacity of motilin receptor agonists will be amenable to commercial development and clinical exploitation remains to be seen.

## THE CENTRAL NERVOUS SYSTEM AND THE PEPTIDE FAMILY

Most GI peptides are found in the brain and/or act as neuropeptides, but the situation remains unclear for motilin. Motilin mRNA has been identified in brain tissues, but RIA determination of motilin content brought ambiguous results. Motilin administration in

this organ induced actions (suggesting the existence of motilin receptors in the brain) that were quite unexpected: appetite stimulation, growth hormone (GH) release, anxiety suppression, and so forth. Yet no valid data support its role as a neuropeptide.

Most peptides are members of a “peptide family” that includes structurally related compounds; gastrin-CCK, and secretin-VIP are typical examples. Motilin remained alone; however, recently a new peptide discovered in the gastric mucosa shows 25% similitude to motilin. This new peptide has been called motilin-related peptide (MTL-RP) by some, but it is more often recognized as ghrelin because of its effect on GH release. Interestingly, ghrelin administration induces central actions that are similar to those described previously for motilin (e.g., GH release, appetite stimulation). Peripherally, ghrelin mimics motilin and appears to be the most potent gastrokinetic agent we have ever tested in the rodent. Future studies should

tell us more about the importance of this new family of peptides.

### See Also the Following Articles

CCK (Cholecystokinin) • Gastrin • Ghrelin • GI Hormone Development (Families and Phylogeny) • GI Hormones Outside the Gut: Central and Peripheral Nervous System • GI Tract, General Anatomy (Cells)

### Further Reading

- Brown, J. C., Cook, M. A., and Dryburgh, J. H. (1973). Motilin, a gastric motor activity stimulating polypeptide: The complete amino-acid sequence. *Can. J. Biochem.* **51**, 533–537.
- Itoh, Z. (1997). Motilin and clinical application. *Peptides* **18**, 593–608.
- Peeters, T. L. (1993). Erythromycin and other macrolides as prokinetic agents. *Gastroenterology* **105**, 1886–1899.
- Poitras, P. (1984). Motilin is a digestive hormone in the dog. *Gastroenterology* **87**, 909–913.

female genitalia. Evidence indicates that ALK2 is the MIS type I receptor. Expression of ALK2 in coelomic epithelium as well as in the circumferential mesenchyme of the Müllerian duct may explain the antineoplastic effects of MIS on epithelial ovarian cancers, which are derivatives of the coelomic epithelium.

Receptor or R-SMADs, the co-SMAD SMAD-4, and the inhibitory or I-SMADs function downstream of the TGF- $\beta$  superfamily type I receptors. Once phosphorylated, R-SMADs dimerize with SMAD-4 and translocate to the nucleus, where they modulate SMAD-responsive gene expression either directly or through recruitment of additional DNA-binding proteins, such as CREB, CBP/p300, or FAST-1. The repressors of SMAD signaling such as SNIP1 interfere with the interaction between SMAD-4 and CBP/p300; I-SMADs 6/7 impair phosphorylation/activation of R-SMADs.

## MIS DURING DEVELOPMENT

In males, serum MIS is present at birth, peaks in infancy, and declines at puberty. Excluding a short time period after birth, serum MIS and testosterone are reciprocally related. This is best illustrated in male puberty, when a precipitous decline in MIS occurs during the characteristic increase in testosterone. In contrast, female ovaries produce low levels of MIS at birth, increasing slightly at puberty to levels similar to those of the adult male.

## OBSERVATIONS FROM TRANSGENIC MICE AND HUMAN SYNDROMES WITH MIS DEFECTS

Human Müllerian duct regression is complete 51 days after ovulation. Failure of MIS signaling in males results in persistent Müllerian duct syndrome (PMDS), in which affected individuals show normal male external/internal genitalia development but manifest a retained cervix, uterus, and Fallopian tubes. In PMDS, findings of crossed ectopia of the contralateral gonad and Müllerian structures contained within an inguinal hernia are common. Unilateral or bilateral testicular maldescent is also common and many PMDS patients are infertile. The phenotype seen in MIS and MIS RII knockout mice recapitulates that seen in human PMDS. Approximately half of human PMDS patients have undetectable MIS levels and harbor MIS gene mutations; most of the remaining patients show

normal/elevated serum MIS levels and have MIS RII gene mutations.

Normally virilized MIS-deficient mice develop focal Leydig cell hyperplasia and tumors. In contrast, transgenic mice overexpressing MIS show undervirilization and cryptorchidism secondary to the effect of MIS on Leydig cell function. Interestingly, in addition to the expected Müllerian regression, female overexpressors develop masculinized ovaries with seminiferous tubules, Sertoli cells, and a paucity of germ cells. MIS is also expressed in the granulosa cells surrounding the small and preantral follicles where MIS activity may function in oocyte development.

## EXTRA-MÜLLERIAN EFFECTS

### Antiproliferative Effects

Human ovarian epithelium derives from the coelomic epithelium, which expresses both type I and type II receptors, suggesting MIS as a potential therapeutic for the most common ovarian cancers that originate from these cells. Preclinical *in vitro* and *in vivo* studies of a human ovarian cancer cell line transplanted beneath the renal capsule of immunosuppressed mice showed that delivery of highly purified human recombinant MIS (hrMIS) via intraperitoneal injection or MIS-producing tissue implants could suppress tumor growth. We found that human breast cancer cells also express the MIS type II receptor, as does the normal involuting but not rapidly growing lactating breast. Furthermore, growth of human breast cancer lines is inhibited by hrMIS *in vitro*. Similar observations have been made of MIS growth inhibition against human prostatic cell lines, indicating additional targets for MIS as a cancer therapeutic against other reproductive tumors expressing MIS RII.

### Regulation of Steroidogenesis

The role of MIS as a modulator/regulator of Leydig cell differentiation/function was suggested by the phenotypes of transgenic mice that either over- or under-express MIS or MIS RII. The highest MIS-producing male transgenics were undervirilized, had Leydig cell hypoplasia with low testosterone, and ultimately developed gonadal exhaustion. Conversely, those mice underexpressing MIS or MIS RII showed Leydig cell hyperplasia and high testosterone. A plausible explanation was derived when it was shown that Leydig cells express MIS RII and that

MIS suppresses testosterone production and Cyp17 expression, which catalyzes the committed step of testosterone synthesis. Cyp17 mRNA levels were highest in the gonads of mice with the lowest MIS levels and Cyp17 mRNA decreased after MIS treatment. When added to cAMP-stimulated MA-10 cells, a mouse Leydig cell line, MIS inhibited testosterone biosynthesis 10-fold, coincident with a decrease in the transcriptional activity at the Cyp17 promoter. MIS suppression of androgens as well as the detection of MIS type II receptor in normal prostate and prostatic cancers indicate that testing MIS in the setting of benign prostatic hypertrophy and prostate cancers may be warranted. MIS inhibition of aromatase activity has also been documented in cultured fetal gonads. It is of interest to determine if MIS has a therapeutic role in hyperandrogenic women with polycystic ovarian syndrome.

## DIAGNOSTIC APPLICATIONS

Normal human gender-specific MIS levels have been established from virtually all developmental stages using an enzyme-linked immunoabsorbant assay, and this assay is helpful in the evaluation of gonadal abnormalities and tumors. MIS is an excellent marker of testicular activity, specifically of the Sertoli cell compartment. Accordingly, in prepubertal boys with nonpalpable gonads, MIS can be used to differentiate anorchia from intraabdominal testes. MIS is also a useful diagnostic tool in intersex states, with levels correlating with the mass of testicular tissue. Consequently, MIS is abnormally low in gonadal dysgenesis, in which functional testicular tissue is subnormal. In androgen-insensitivity syndromes, MIS is elevated; thus, MIS levels can differentiate abnormal testicular determination from defects in androgen biosynthesis or sensitivity. MIS is also extremely valuable as a marker of tumor burden or recurrence in patients with granulosa or sex-cord tumors.

## CONCLUSION

MIS plays an important role in normal sexual development and differentiation, and it has proven diagnostic value for intersex disorders, undescended testis, and gonadal tumors. Furthermore, MIS may have important therapeutic applications in a number of common reproductive tumors in both sexes. The success of these applications will likely depend on the efficiency and cost-effectiveness of MIS production. A complete understanding of MIS downstream signaling pathways and the induced gene products essential for its effect may lead to the development of therapeutics capable of modulating this pathway, which could be of benefit.

## See Also the Following Articles

Agonadism, Male and Female • Androgens, Gender and Brain Differentiation • Endocrine Disruptors and Male Sexual Differentiation • In Vitro Fertilization (IVF) • Testes, Embryology of

## Further Reading

- Behringer, R. R., Finegold, M. J., and Cate, R. L. (1994). Müllerian-inhibiting substance function during mammalian sexual development. *Cell* **79**(3), 415–425.
- Gustafson, M. L., *et al.* (1992). Müllerian inhibiting substance as a marker for ovarian sex-cord tumor. *N. Engl. J. Med.* **326**(7), 466–471.
- Lee, M. M., *et al.* (1997). Measurements of serum Müllerian inhibiting substance in the evaluation of children with nonpalpable gonads. *N. Engl. J. Med.* **336**(21), 1480–1486.
- Mishina, Y., *et al.* (1996). Genetic analysis of the Müllerian-inhibiting substance signal transduction pathway in mammalian sexual differentiation. *Genes Dev.* **10**(20), 2577–2587.
- Teixeira, J., *et al.* (1999). Müllerian-inhibiting substance regulates androgen synthesis at the transcriptional level. *Endocrinology* **140**(10), 4732–4738.
- Teixeira, J., Maheswaran, S., and Donahoe, P. K. (2001). Müllerian inhibiting substance: An instructive developmental hormone with diagnostic and possible therapeutic applications. *Endocr. Rev.* **22**(5), 657–674.

## CLINICAL DESCRIPTION OF APS TYPES I AND II

In 1980, Michael Neufeld, Noel Maclaren, and Robert Blizzard presented data from a survey of patients with APSs and proposed a classification for the syndromes observed:

APS type I was defined as present in patients who have at least two of the triad of Addison's disease, hypoparathyroidism, and chronic mucocutaneous candidiasis. Other associated autoimmune disorders were also allowed to be present.

APS type II was defined as Addison's disease with autoimmune thyroid disease and/or type 1 diabetes mellitus. Autoimmune disorders other than hypoparathyroidism and candidiasis were also sometimes present.

APS type III was defined as patients presenting with thyroid autoimmune disease and any other autoimmune disease except Addison's disease or hypoparathyroidism. APS type III could be grouped separately or together with type II.

APS type IV was defined as patients presenting with two or more organ-specific autoimmune diseases that do not otherwise fall into the category of type I, II, or III. This form of APS is frequently present in patients with non-endocrine autoimmune disorders. In this article, we consider APS type IV to be part of APS type III.

### APS TYPE I

#### Prevalence

APS type I is a rare disorder. The highest prevalence, 1 in 9000, has been reported in the Iranian Jewish population. It is also more common in the Finnish and Sardinian populations, with prevalences of 1 in 25,000 and 1 in 14,500, respectively. APS type I may appear in a sporadic or familial form and is now known to be an autosomal recessive disease caused by mutations of the autoimmune regulator (AIRE) gene. According to Michael Neufeld and colleagues, females exceed males at all ages in a ratio of 1.6 to 1.

#### Presentation

Of the classic features, chronic mucocutaneous candidiasis occurs at the youngest age, often during infancy. Hypoparathyroidism may occur shortly thereafter, usually before adolescence. Addison's disease tends to appear after the onset of hypoparathyroidism and as late as the fourth decade of life. Chronic active

hepatitis and malabsorption due to celiac disease tend to occur in APS type I and not in APS type II. Chronic active hepatitis is particularly important because it has been reported as the cause of death in a significant number of patients with APS type I. In addition, gonadal failure and diffuse vitiligo may be found in patients with APS type I, and in 1985, it was recognized that hypopituitarism and diabetes insipidus also occur. In 1990, P. Ahonen and colleagues reviewed their data from 68 patients in 54 families in Finland (coining the term "autoimmune polyendocrinopathy–candidiasis–ectodermal dystrophy [APECED]) and found that all of their patients had candidiasis at some time with hypoparathyroidism (79%), adrenocortical failure (72%), and/or gonadal failure (60% in females and 14% in males). Several of their patients also had dental enamel hypoplasia (77%) and/or keratopathy (35%) that was not attributable to hypoparathyroidism and whose endocrine dysfunction did not manifest until the fifth decade of life. These observations underscored the need for lifelong surveillance for new components of the disease. In 1998, Corrado Betterle and colleagues from Italy published their experience following 41 patients for 20 years (1967–1996). They compared the clinical manifestations in their patients with those of Neufeld and Ahonen. The occurrence of clinical manifestations and the age of onset were similar. They reported calcifications of the basal ganglia in 17 of their patients, most likely secondary to hypoparathyroidism. As in earlier published reports, Betterle and colleagues showed that 20% of their patients had chronic active hepatitis, again emphasizing the importance of continued surveillance of liver function in these patients (Table I).

### Genetics of APS Type I

APS type I is the only autoimmune disease known to be caused by a defect in a single gene. This disease is inherited as a simple recessive Mendelian trait. It occurs primarily in the familial form but can occur sporadically. The defective gene in this disorder was identified by positional cloning and was called AIRE (Fig. 1). The AIRE (autoimmune regulator) gene is located on chromosome 21q22.3. More than 40 mutations have been described. All of the mutations described are localized to the coding portion of the gene and are present in all ethnic populations, with the exception of one mutation that was found only in the homozygous form of APS type I in Iranian Jewish patients. The Iranian Jewish population also seems to have a milder clinical expression of this disorder.

**Table I** Autoimmune Polyglandular Syndrome (Percentages)

Type I	
Endocrine disorders	
Hypoparathyroidism	90
Addison's disease	60
Gonadal failure	45
Autoimmune thyroid disease	10
Type 1 diabetes mellitus	1
Hypopituitarism	
Diabetes insipidus	
Nonendocrine disorders	
Mucocutaneous candidiasis	75
Malabsorption syndromes	25
Alopecia	20
Pernicious anemia	15
Chronic active hepatitis	10
Vitiligo	5
Sjörge's syndrome	
Dystrophy of nails and dental enamel	
Progressive myopathy	
Type II	
Endocrine disorders	
Addison's disease	100
Autoimmune thyroid disease	70
Type 1 diabetes mellitus	30
Gonadal failure	10
Hypopituitarism	<1
Nonendocrine disorders	
Alopecia	
Pernicious anemia	
Chronic active hepatitis	
Vitiligo	
Premature greying	

Source. Adapted with permission from Riley, W. J. (1992). (Autoimmune polyglandular syndromes. *Horm. Res.* 38 (Suppl. 2), 10, Karger, Basel.

This group of patients rarely has any of the mucocutaneous candidiasis or ectodermal disorders common in all of the other affected populations.

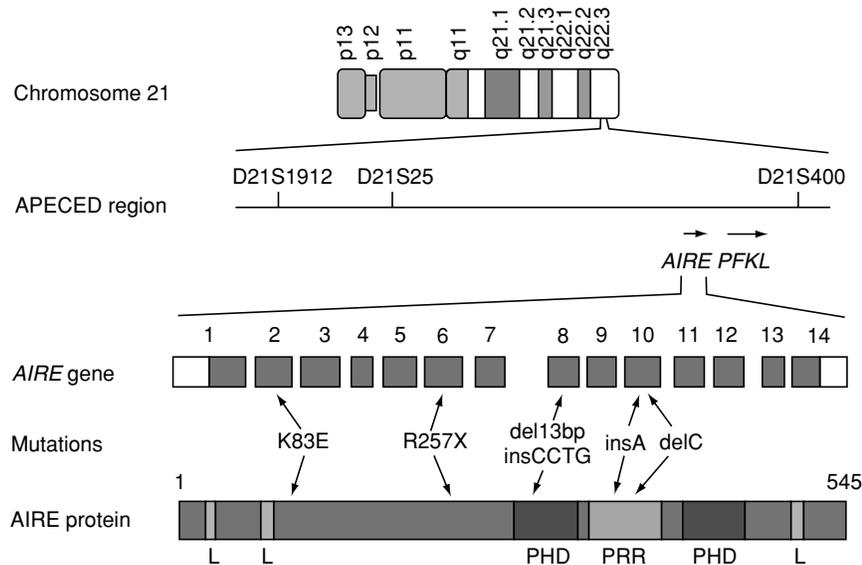
The protein encoded by the AIRE gene is localized to the cell nucleus and acts as a transcription factor that regulates autoimmunity by promoting the ectopic expression of peripheral tissue-restricted antigens in medullary epithelial cells of the thymus. AIRE-deficient thymic medullary epithelial cells showed a specific reduction in ectopic transcription of genes encoding peripheral antigens. Animals lacking a functional AIRE protein in the thymus exhibit a defined profile of autoimmune diseases. AIRE is expressed not only in the thymus but also in the spleen, lymph

nodes, peripheral leukocytes, testes, and adrenal gland. There is no good genotype-phenotype correlation of the mutations with the syndrome. However, Maria Halonen and colleagues reported an association between the most common mutation of the AIRE gene, R257X, and a higher prevalence of candidiasis than with other mutations.

Antibodies that cross-react with steroid-producing cells have been commonly observed in patients with APS type I. Mark Leshin reported that although steroid antibodies were observed in both males and females, a considerably greater number of females manifested clinical hypergonadotropic hypogonadism secondary to gonadal failure. Studies describe adrenal cortex autoantibodies and steroid-producing cell autoantibodies associated with autoimmune Addison's disease and premature ovarian failure. Betterle and colleagues also showed that a number of these autoantibodies, detected by indirect immunofluorescence, reacted with steroidogenic enzymes. The prevalence of these antibodies has varied depending on the techniques used to detect them. The fact that epitope-specific autoantibodies are expressed against a number of tissues in APS type I indicates a polyclonal immune response. In addition, the T-cell dysfunction results in localized infections with *Candida albicans*.

All of the autoantigens in the adrenal cortex and gonads are cytochrome P450 enzymes involved in steroid synthesis. Usually, the appearance of these antibodies signals a high risk of developing failure of the target organs. In Addison's disease, the three major autoantigens are the 21-hydroxylase (P450c21) and 17-hydroxylase (P450c17) enzymes and the cholesterol side chain cleavage enzyme (P450scc). Patients with APS type I might not express antibodies to all of the P450 enzymes. The most commonly expressed antibody is to P450c21. All three P450 enzymes are expressed in the adrenal cortex, and P450c17 and P450scc are also expressed in the gonads. In addition, their expression in the brain and skin has been described (Table II).

In most autoimmune diseases, including APS type II, there is an association with certain human leukocyte antigen (HLA) alleles, incurring higher susceptibility to the autoimmune disease. The association is particularly strong with the occurrence of tissue-specific antibodies. APS type I was not known to have any correlation with HLA class II haplotypes until recently, when Addison's disease was associated with HLA-DRB1\*03, alopecia was associated with HLA-DRB1\*04-DQB1\*0302, and type 1 diabetes correlated negatively with HLA-DRB1\*15-DQB1\*0602. The same HLA associations have been established previously for the isolated endocrine disorders.



**Figure 1** Schematic of chromosome 21 showing the physical map of the APECED region (top) and the genomic structure of the AIRE gene, features of its protein product, and mutations found in APECED patients (bottom). The AIRE gene maps proximal to the 5' end of the PFKL (phosphofruktokinase of liver type) gene between DNA markers D21S25 and D21S400 and has 14 exons. The most common mutation found in APECED patients was the Arg change to a stop codon (R257X) in exon 6. Three other mutations were also detected: deletion of 13 nucleotides (del13bp) and insertion of a 4-nucleotide repeat CCTG (insCCTG) in exon 8, resulting in a truncated AIRE protein product; insertion of A (insA) and deletion of C (delC) nucleotides in exon 10, again resulting in a truncated AIRE protein product; and a missense substitution of amino acid Lys to Glu (K83E) in exon 2. The location of mutations in gene exons and respective sites in the protein are shown by arrows. The AIRE protein (545 amino acids) contains three LXXLL motifs (L), two PHD-type zinc fingers (PHD), and a proline-rich region (PRR), suggesting that it functions as a transcriptional factor or coactivator. Adapted from Peterson, P., Nagamine, K., Scott, H., Heino, M., Kudoh, J., Shimizu, N., Antonarakis, S. E., and Krohn, K. J. E. (1998). APECED: A monogenic autoimmune disease providing new clues to self-tolerance. *Immunol. Today* **19**, 385, with permission from Elsevier Science.

## APS TYPE II

### Prevalence

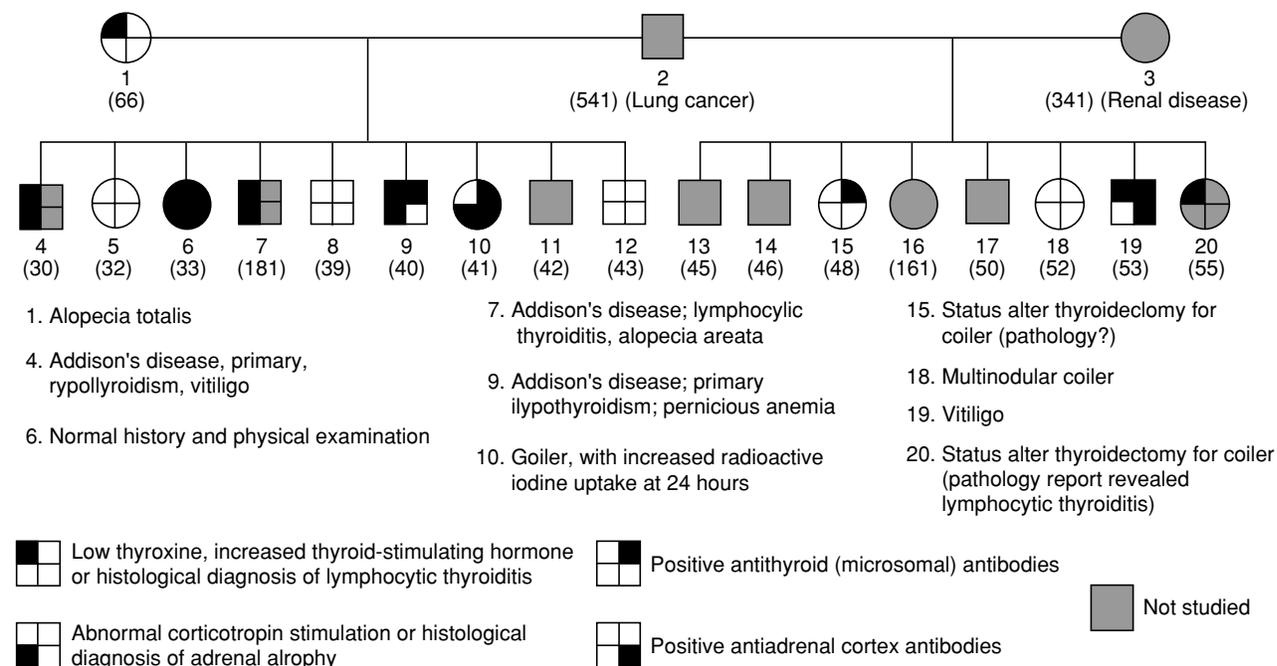
In 1980, a report of familial Schmidt's syndrome was published. The report described 17 offspring of a common father. Endocrine abnormalities were documented in 8 of the 12 siblings studied and one of the two mothers in this kindred. The age of clinical onset ranged from 18 to 38 years. Adrenal and thyroid hypofunction with organ-specific antibodies were found. Pernicious anemia, vitiligo, and alopecia were reported as well. This familial study helped to elucidate the variable presentation of adrenal, thyroid, and other autoimmune abnormalities in Schmidt's syndrome (Fig. 2).

The occurrence of primary Addison's disease in the general population is 40 to 50 per 1 million. Recent studies have suggested an increasing prevalence, or greater recognition, of the disease. In 40% of the patients with idiopathic Addison's disease, a second or third endocrine disease is present. The female-to-male

predominance is approximately 1.8 to 1. Other non-endocrine autoimmune disorders have been described as well, sometimes classified separately as APS type III or IV. Although a clear relationship exists, the current recommendation is that it might not be necessary to evaluate all patients with clinically latent adrenal insufficiency. The prevalence of adrenal insufficiency among patients with Hashimoto's thyroiditis is low. Only if other evidence of APS type II is present, such as type 1 diabetes mellitus either in the patient or in a first-degree family member, should the possibility of adrenal insufficiency be investigated more thoroughly.

### Presentation

The initial description of APS type II included the concurrent appearance of adrenal insufficiency and autoimmune thyroid disease or type 1 diabetes mellitus. For the purpose of this discussion, we include all associated autoimmune disorders as also being variable expressions



**Figure 2** Family pedigree. Numbers in parentheses indicate ages in years; daggers indicate deceased. Adapted with permission from Anderson, P. B., Fein, S. H., and Frey, W. G., III. (1980). Familial Schmidt's syndrome. *J. Am. Med. Assoc.* **244**, 2070.

of APS type II as proposed by Neufeld and colleagues, where not all of their patients with APS type II had Addison's disease. The age of onset was much more varied when compared with their APS type I patients, occurring during childhood to late adulthood with peak incidence during midlife. Chronic mucocutaneous candidiasis and hypoparathyroidism were not present in any of their patients with APS type II. Chronic active hepatitis and malabsorption were also essentially non-existent in their patients, although these were common in the patients with APS type I. Fully 69% of patients had autoimmune thyroid disease, which was overactive as well as underactive. In addition, 52% of patients had type 1 diabetes mellitus. In a group of older patients, Betterle and colleagues demonstrated atrophic gastritis by biopsy in 95% of older patients with vitiligo. They concluded that 95% of all patients with vitiligo would have pernicious anemia and/or gastric atrophy. Vitiligo, alopecia, and premature graying of the hair are relatively common and should alert the physician to the possible presence of other diseases (Table I).

### Genetics of APS Type II

The inheritance of APS type II is poorly defined, and the familial aggregation does not follow Mendelian inheritance. Hence, in contrast to APS type I, APS

type II is a more complex disease that results from the interaction of an even greater variety of genetic and environmental influences. Many members of affected families were noted to have an increased frequency of the HLA B8/DR3 haplotype, indicating that HLA was one susceptibility gene. More recent studies have focused on the cytotoxic T-lymphocyte antigen 4 (CTLA4) gene, which confers susceptibility to autoimmunity. CTLA4 is a negative regulator of T-cell activation, and the association of CTLA4 with APS type II was variable (OR = 1.73) depending on the specific autoimmune disease being studied.

### APS TYPE III

APS type III consists of a diverse group of multiple autoimmune endocrine disorders developing in the same patient over time. There are a few considerations to keep in mind. Approximately 3 to 8% of patients with type 1 diabetes mellitus or autoimmune thyroid disease have celiac disease. Measuring auto-antibodies to tissue transglutaminase is now the typical method for screening for celiac disease. The frequency of frank or latent pernicious anemia is increased among patients with autoimmune thyroid disease who are over 40 years of age. Atrophic gastritis with achlorhydria but normal B12 absorption is

Table II Circulating Antibodies to Endocrine Antigens in Autoimmune Polyglandular Syndromes

	Autoantigen	Tissue/Cells
Addison's disease	21-OH, P450scc 17-OH	Enzymes of the adrenal cortex
Hypogonadism, premature menopause	CYP450scc 17-OH	Ovary: granulosa/theca cells Testis: Leydig cells Placenta: syncytiotrophoblasts
Hashimoto's thyroiditis/ hypothyroidism	Thyroid peroxidase Thyroglobulin TSHr (blocking)	Thyroid enzyme Thyroid-secreted protein Thyrocytes
Graves' disease	TSHr (stimulating)	Thyrocytes and extra ocular fat cells
Hypoparathyroidism	Calcium-sensing receptor	Parathyroid/Other tissues
Type 1 diabetes mellitus	GAD <sub>65</sub> 38 kDa IA-2, IA2B Insulin L-amino acid decarboxylase	Pancreatic beta cells
Vitiligo <i>Alopecia areata</i>	Tyrosinase Tyrosine hydroxylase	Melanocyte scalp cells Gastric parietal cells
Pernicious anemia	H <sup>+</sup> , K <sup>+</sup> ATPase	Gastric parietal cells
Achlorhydria	Intrinsic factor	Gastric mucosa and chief cells
Chronic hepatitis active	P450D6, 2C9 P450 1A2	Hepatocytes
Celiac disease	Antigliadin Antiendomysial Transglutaminase	Small intestine
Malabsorption	Tryptophan hydroxylase	Intestine chromaffin cells

Source. Adapted from Ten, S. New, M. and Maclaren, N. Clinical review 130: Addison's disease 2001. *J. Clin. Endocrinol. Metab.* **86**, 2915, with permission from the Endocrine Society.

also frequent in patients with autoimmune thyroid disease. Approximately 20 to 30% of patients with autoimmune thyroid disease have circulating parietal cell antibodies, compared with a frequency of 60 to 90% of patients with the APS syndrome.

The clinical and epidemiological characteristics of the group of patients with diabetes mellitus and coexistent autoimmune thyroid disease differ from those of patients with diabetes alone. The frequency of insulin dependence is 60%, and the median age of diagnosis of patients with insulin-dependent diabetes and thyroid disease (36 years) is greater than the mean age of patients with insulin-dependent diabetes alone (25 years). The female-to-male predominance of patients with type 1 diabetes mellitus and thyroid disease is much greater (6.4 to 1) than the ratio for patients with diabetes alone (1 to 1).

Over the past 20 to 30 years, an increasing incidence of hypophysitis has also been reported in patients with Hashimoto's thyroiditis. This pituitary autoimmune disorder, as well as diabetes insipidus, has also been recognized in patients with APS type II.

One-fifth (20%) of patients with generalized myasthenia gravis have been reported to have clinical autoimmune thyroid disease, with a higher incidence of clinical endocrine disease described in patients with onset of myasthenia after 40 years of age than before (35 vs 17%).

## TREATMENT

There are some special considerations for patients with APS. For example, thyroxine therapy with untreated adrenal insufficiency may precipitate an adrenal crisis. In addition, a decreased insulin requirement or increased hypoglycemia in a patient with type 1 diabetes mellitus may be one of the many signs of adrenal insufficiency or hypothyroidism. On the other hand, increased hyperglycemia may be an early sign of autoimmune hyperthyroidism. Therefore, it is necessary to be vigilant of the possible development of thyroid, adrenal, and other autoimmune diseases when considering appropriate therapy and management of endocrine autoimmune disorders.

### See Also the Following Articles

Adrenal Insufficiency • Diabetes, Type 1 • Hashimoto's Disease • Hypoparathyroidism • McCune-Albright Syndrome • Thyroid Autoimmunity

### Further Reading

Betterle, C., Dal Pra, C., Mantero, F., and Zanchetta, R. (2002). Autoimmune adrenal insufficiency and autoimmune polyendocrine syndromes: Autoantibodies, autoantigens, and their applicability in diagnosis and disease prediction. *Endocrinol. Rev.* **23**, 327–364.

Betterle, C., Greggio, N. A., and Volpato, M. (1998). Clinical review: Autoimmune polyglandular disease type I. *J. Clin. Endocrinol. Metab.* **83**, 1049–1055.

Betterle, C., Volpato, M., Greggio, A. N., and Presotto, F. (1996). Type 2 polyglandular autoimmune disease (Schmidt's syndrome). *J. Pediatr. Endocrinol. Metab.* **9**(Suppl. 1), 113–123.

Gottlieb, P. A., and Fain, P. R. (2002). Genetic basis of autoimmune adrenal deficiency. *Curr. Opin. Endocrinol. Diab.* **9**, 237–243.

Heino, M., Peterson, P., Kudoh, J., Shimizu, N., Antonarakis, S. E., Scott, H. S., and Krohn, K. (2001). APECED mutations in the autoimmune regulator (AIRE) gene. *Hum. Mutat.* **18**, 205–211.

Leslie, D., Lipsky, P., and Notkins, A. L. (2001). Autoantibodies as predictors of disease. *J. Clin. Invest.* **108**, 1417–1422.

Ten, S., New, M., and Maclaren, N. (2001). Clinical review 130: Addison's disease 2001. *J. Clin. Endocrinol. Metab.* **86**, 2902–2922.

or a pentagastrin-stimulated calcitonin value is  $>100$  pg/ml. A basal calcitonin value of  $>30$  pg/ml or a stimulated value  $>200$  pg/ml is highly predictive of MTC (positive predictive value 93%, sensitivity 80%). Hypercalcitonemia also occurs in other disease states, such as other neuroendocrine tumors, renal failure, chronic lymphocytic thyroiditis, and follicular thyroid adenomas. False-positive test results for abnormal calcitonin values in individuals suspected to have MTC occur in 5–10% of the tests, as determined by retrospective testing for *RET* germ-line mutations in these individuals. Fifty percent of patients with a negative pentagastrin test but a positive test for *RET* mutation had already developed MTC. MTC can also give rise to ectopic hormone production, such as ectopic adrenocorticotrophic hormone syndrome.

Before thyroidectomy is performed, the physician should always screen for pheochromocytoma by measuring fractionated urinary metanephrines and/or plasma free-metanephrines. This is especially important since MEN2-associated pheochromocytomas are “adrenergic”; i.e., they cause symptoms such as palpitations rather than sustained hypertension. Before prophylactic thyroidectomy was routinely performed in individuals at risk for MTC (e.g., carriers of a *RET* germ-line mutation), MTC rarely was at stage 1 but instead was usually at an advanced stage (stages 3 and 4) when it was discovered in a patient. Therefore, prophylactic thyroidectomy with central lymph node dissection has been recommended for individuals with a *RET* germ-line mutation. This has lowered the mortality rate from hereditary MTC to  $<5\%$ . Multifocal hyperplasia of parafollicular C cells has been regarded as a precursor lesion for hereditary MTC. Metastases usually develop in the cervical and mediastinal lymph nodes as well as in lung, liver, or bone. Calcitonin is an excellent tumor marker in postsurgical follow-up. In addition, carcinoembryonic antigen is helpful. Diarrhea from humoral factors may develop and can be improved by reducing the tumor burden. For instance, patients with MTC metastatic to the liver, high calcitonin values, and diarrhea may experience cessation of diarrhea and a fall in calcitonin levels to near normal after partial liver resection. For disseminated metastatic MTC, no good therapeutic measures exist. None of the different chemotherapeutic regimens has proven beneficial nor has radiation therapy has proven beneficial in these relatively insensitive tumors.

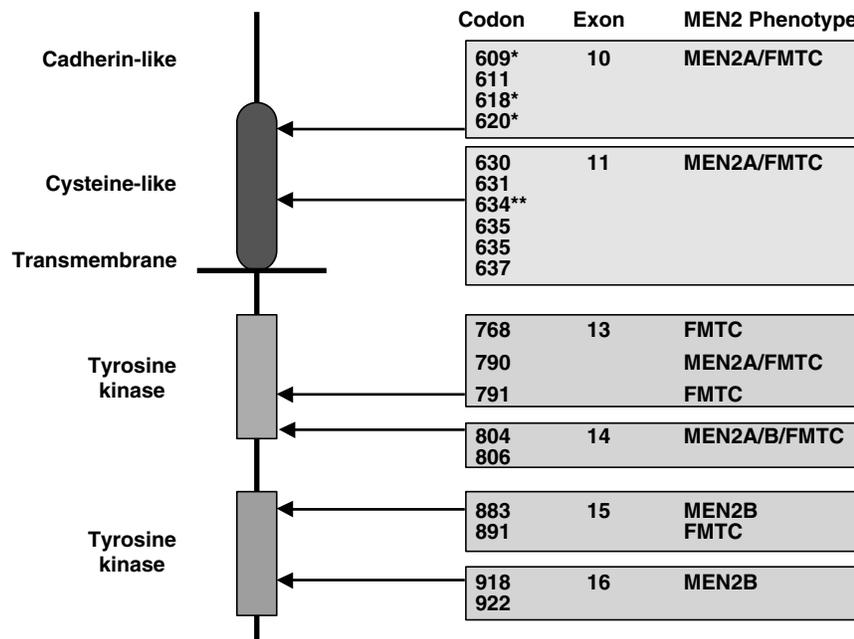
Pheochromocytoma develops in approximately 50% of patients with MEN2A during their lifetime. The true prevalence can be determined by autopsy studies of such patients. Typically, MEN2 pheochromocytomas

predominantly produce epinephrine and metanephrine; i.e., they are adrenergic. Affected patients develop adrenergic symptoms, such as palpitations, sweating, and paroxysmal hypertension. Screening for pheochromocytoma in patients at risk for MEN2 should start at the age of approximately 6 years. MEN2 pheochromocytomas can be unilateral but are often bilateral, typical for a hereditary syndrome. Rarely ( $<5\%$  of cases), MEN2 pheochromocytomas are metastatic. Once the presence of a pheochromocytoma has been confirmed by a more than threefold elevation of metanephrines from baseline values, an imaging study such as high-resolution (thin cuts, 1–2 mm) computed tomography or magnetic resonance imaging of the adrenal glands should be performed. Before surgery, a metaiodobenzylguanidine (MIBG) scan should also be performed to rule out extra-adrenal lesions. The sensitivity of MIBG is approximately 80%, whereas the specificity is 100%. All individuals scheduled for adrenalectomy for a biochemically confirmed pheochromocytoma or an adrenal “incidentaloma” should be treated with an alpha-blocker, such as phenoxybenzamine, and/or a beta-blocker for 7–10 days prior to surgery. The preferred surgical approach is laparoscopic adrenalectomy—if possible, adrenal cortical-sparing laparoscopic adrenalectomy.

Primary hyperparathyroidism in patients with MEN2A develops in approximately 25% of cases and usually is mild; i.e., most affected individuals have no symptoms. Screening should include serum calcium and intact parathyroid hormone. Surgical intervention should be based on the same criteria as for sporadic primary hyperparathyroidism, i.e., serum calcium elevation  $>0.25$  mmol of the upper limit of normal, 24 h urinary calcium  $>400$  mg, creatinine clearance reduced by 30%, *T* score of bone mineral density (radius, femur, spine) minus 2.5 SD, and age less than 50 years. Surgical procedures should be similar to those in patients with other hereditary hyperparathyroidism, such as multiple endocrine neoplasia type 1.

### Familial Medullary Thyroid Carcinoma

In a few families with *RET* germ-line mutations (probably fewer than 20 kindreds), affected individuals are known to develop only medullary thyroid carcinoma during their lifetime. Whether such patients indeed do not have pheochromocytoma can be proven only by autopsy studies. Rigorous criteria for familial medullary thyroid carcinoma (FMTC) patients have been established to avoid incorrect designations and



**Figure 1** Genotype–phenotype correlations for *RET* mutations in patients with MEN2. \*, Associated with Hirschsprung’s disease. \*\*, cutaneous lichen amyloidosis (CLA).

include the following: an adequate medical history, particularly in older family members, multiple carriers, or affected members over the age of 50 years; and more than 10 carriers in the kindred. Late onset of familial medullary thyroid carcinoma commonly occurs in patients with *RET* germ-line mutations in codons 790, 791, and 804 (Fig. 1). Often, basal and stimulated pentagastrin calcitonin values are used in deciding when to perform thyroidectomy. Some authorities recommend waiting no longer than until age 10 in individuals with negative calcitonin testing and *RET* mutations in the aforementioned codons.

### Multiple Endocrine Neoplasia Type 2B

MEN2B is the most aggressive form of all MEN2 subforms. It affects approximately 5% of all MEN2 patients. MTC can occur in the first weeks of life. Therefore, some authorities recommend prophylactic thyroidectomy with central lymph node dissection in patients with *RET* germ-line mutations in exon 16 within the first months of life. Similarly, pheochromocytoma becomes manifest earlier (by approximately 10 years) than in individuals with MEN2A and therefore should be screened for earlier. On the other hand, MEN2B patients usually do not develop hyperparathyroidism. Instead, they have a

characteristic habitus with marfanoid features and intestinal and mucosal ganglioneuromas.

## MOLECULAR ASPECTS

### The RET Receptor Tyrosine Kinase

*RET* denotes “rearranged during transfection” and was discovered in 1985 during transfections assays. It has been classified as a proto-oncogene located at chromosome 10q11.2 with activating mutations leading to MEN2. Interestingly, so-called inactivating *RET* mutations have been found in patients with Hirschsprung’s disease or intestinal aganglionosis. In addition, some kindreds with *RET* germ-line mutations in codons 618 and 620 have MEN2 and Hirschsprung’s disease. *RET* is expressed in neural crest-derived cells, such as the parafollicular C cells in the thyroid gland and the chromaffin cells in the adrenal medulla. *RET* consists of 21 exons and encodes a receptor tyrosine kinase. Six exons, exons 10, 11, 13, 14, 15, and 16, are called hot spots, since germ-line mutations in these exons are found in approximately 97% of patients with MEN2. The most common affected exon is exon 11 including codon 634, which is mutated in approximately 80% of patients with MEN2A. More than 94% of patients with MEN2B have *RET* germ-line mutations in exon 16, most

commonly in codon 918. There are strong genotype–phenotype correlations (see Fig. 1). Ligands for the *RET* receptor include glial cell line-derived neurotrophic factor, neurturin, persephin, and artemin. *In vitro* studies indicate that certain *RET* germ-line mutations can lead to autophosphorylation and/or dimerization with constitutive activation of RET. The RET receptor consists of several domains, a cadherin-like domain, a cysteine-rich domain, a transmembrane domain, and an intracellular tyrosine kinase domain (see Fig. 1).

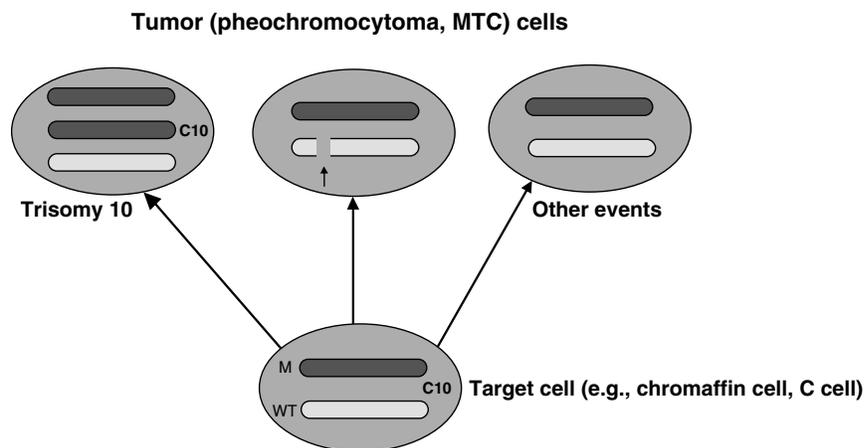
### RET Downstream Signaling

A complex array of downstream interactions mediates RET's functions in normal and tumorigenic cells. For instance, RET stimulates the RAS pathway, including activation of the mitogen-activated protein kinase pathway, which is necessary for differentiation and neuronal survival. RET also can activate phosphatidylinositol 3-kinase (PI3-kinase), which has been implicated in cell proliferation and motility. Furthermore, through PI3-kinase activation, RET can activate protein kinase B, also leading to cell proliferation and survival. Through activation of the c-jun N-terminal kinase pathway, RET may contribute to tumorigenesis.

### Model for Tumorigenesis of MEN2

Transgenic mice heterozygous for the M918T *RET* mutation develop hyperplasia of C cells and

adrenomedullary hyperplasia without progression to medullary thyroid carcinoma or pheochromocytoma. On the other hand, mice homozygous for this *RET* mutation not only have an earlier onset of C cell and adrenomedullary hyperplasias but also progression to medullary thyroid carcinoma and pheochromocytoma. It is puzzling that related and unrelated patients with the same heterozygous germ-line mutation in *RET* develop the respective tumor (MTC or pheochromocytoma) at widely different ages, i.e., at age 1 and at age 81, and that only a few cells in the target organs develop into tumors. In an attempt to answer this question *in vivo*, studies showed evidence of a so-called second hit in tumors of patients with heterozygous *RET* germ-line mutations, leading to tumor formation. In MEN2-associated pheochromocytomas, overrepresentation of mutant RET through duplication of mutant RET in trisomy 10 or loss of wild-type RET was found. Similarly, such an allelic imbalance between mutant and wild-type *RET* could also be demonstrated in MEN2-associated medullary thyroid carcinoma (Fig. 2). In addition, in the stable and commercially available TT cell line, which is derived from medullary thyroid carcinoma with a *RET* germ-line mutation in codon 634, a tandem amplification of mutant RET was found, leading to the overrepresentation of mutant RET. Selected C cells in the thyroid gland and selected chromaffin cells in the adrenal gland may undergo such a second-hit event, giving these cells a growth advantage and making them prone to more replication errors. Somatic *RET* mutations in MEN2-associated MTC



**Figure 2** Model for tumorigenesis in MEN2-associated pheochromocytomas. In an individual with MEN2, each cell of the target organs (e.g., adrenal medulla) carries a *RET* germ-line mutation in the heterozygous state with one intact wild-type *RET* allele. Selected chromaffin cells undergo a second hit through duplication of the mutant *RET* allele in trisomy 10 or loss of the wild-type allele, giving these cells an overrepresentation of mutant RET and a growth advantage with subsequent tumor formation. M, mutant *RET* allele; WT, wild-type *RET* allele; C10, chromosome 10. Modified from Koch *et al.* (2002), *J. Clin. Endocrinol. Metab.* 87, 5367–5384, with permission.

or pheochromocytoma are rare and may represent a tumor progression phenomenon. A few patients have double germ-line mutations in *RET*; the functional significance of this finding has yet to be determined. A subset of MEN2-associated pheochromocytomas have somatic *VHL* gene alterations, possibly leading to tumorigenesis or representing events of tumor progression.

### RET Mutation Testing

All cases of sporadic MTC should be tested for *RET* germ-line mutations, since approximately 7% of patients with apparently sporadic MTC have germ-line mutations in *RET*. Also, all patients presenting with pheochromocytoma should be tested for germ-line mutations in *RET*, since a study including 271 patients found that up to 24% of these patients with apparently nonsyndromic/sporadic pheochromocytoma had germ-line mutations in the *VHL*, *RET*, *SDHD*, or *SDHB* genes. Patients with apparently sporadic primary hyperparathyroidism without other features of MEN2, such as MTC, should not routinely be tested for *RET* germ-line mutations. If one of the so-called hot spot exons does not reveal a *RET* mutation, especially in members of MEN2 families, the other 15 exons of *RET* should be tested. Prophylactic thyroidectomy with central lymph node dissection within the first 6 months should be recommended for children with *RET* mutations in codons 883, 918, or 922. For children with mutations in codons 611, 618, 620, or 634, this surgical procedure should take place before 5 years of age. Children with mutations in codons 609, 768, 790, 791, 804, or 891 also should undergo prophylactic thyroidectomy but the recommendation for the exact age at which the child should undergo surgery varies between the ages of 5 and 10, especially for children with mutations in codons 790 and 791.

### See Also the Following Articles

ACTH (Adrenocorticotrophic Hormone) • Adrenal Tumors, Molecular Pathogenesis • Hyperparathyroidism, Primary • Medullary Thyroid Carcinoma • Parathyroid Glands,

Pathology • Pheochromocytoma • Thyroidectomy • Von Hippel-Lindau Syndrome

### Further Reading

- Bieglmayer, C., Scheuba, C., Niederle, B., Flores, J., and Vierhapper, H. (2002). Screening for medullary thyroid carcinoma: Experience with different immunoassays for human calcitonin. *Wien Klin. Wochenschr.* **114**, 267–273.
- Bilezikian, J. P., Potts, J. T., Fuleihan, G. E. H., *et al.* (2002). Summary statement from a workshop on asymptomatic primary hyperparathyroidism: A perspective for the 21st century. *J. Clin. Endocrinol. Metab.* **87**, 5353–5361.
- Brandi, M. L., Gagel, R. F., Angeli, A., Bilezikian, J. P., *et al.* (2001). Guidelines for diagnosis and therapy of MEN type 1 and type 2. *J. Clin. Endocrinol. Metab.* **86**, 5658–5671.
- Gagel, R. F. (2001). Multiple endocrine neoplasia type 2. In “Endocrinology” (L. J. DeGroot and J. L. Jameson, eds.), 4th ed., pp. 2518–2532. W. B. Saunders, Philadelphia, PA.
- Huang, S. C., Koch, C. A., Vortmeyer, A. O., Pack, S. D., Lichtenauer, U. D., Mannan, P., Lubensky, I. A., Chrousos, G. P., Gagel, R. F., Pacak, K., and Zhuang, Z. (2000). Duplication of the mutant *RET* allele in trisomy 10 or loss of the wild-type allele in multiple endocrine neoplasia type 2-associated pheochromocytoma. *Cancer Res.* **60**, 6223–6226.
- Koch, C. A., Huang, S. C., Azumi, N., Zhuang, Z., Chrousos, G. P., Vortmeyer, A. O., and Pacak, K. (2002). Somatic *VHL* gene deletion and mutation in MEN 2A-associated pheochromocytoma. *Oncogene* **21**, 479–482.
- Koch, C. A., Pacak, K., and Chrousos, G. P. (2002). The molecular pathogenesis of hereditary and sporadic adrenocortical and adrenomedullary tumors. *J. Clin. Endocrinol. Metab.* **87**, 5367–5384.
- Machens, A., Gimm, O., Hinze, R., Hoppner, W., Boehm, B. O., and Dralle, H. (2001). Genotype–phenotype correlations in hereditary medullary thyroid carcinoma: Oncological features and biochemical properties. *J. Clin. Endocrinol. Metab.* **86**, 1104–1109.
- Niccoli-Sire, P., Murat, A., Rohmer, V., Franc, S., Chabrier, G., *et al.* (2001). Familial medullary thyroid carcinoma with noncysteine ret mutations: Phenotype–genotype relationship in a large series of patients. *J. Clin. Endocrinol. Metab.* **86**, 3746–3753.
- Pacak, K., Linehan, W. M., Eisenhofer, G., Walther, M. M., and Goldstein, D. S. (2001). Recent advances in genetics, diagnosis, localization, and treatment of pheochromocytoma. *Ann. Intern. Med.* **134**, 315–329.
- Schuffenecker, I., Virally-Monod, M., Brohet, R., *et al.* (1998). Risk and penetrance of primary hyperparathyroidism in multiple endocrine neoplasia type 2A families with mutations at codon 634 of the *RET* protooncogene. *J. Clin. Endocrinol. Metab.* **83**, 487–491.
- Takahashi, M. (2001). The GDNF/*RET* signaling pathway and human diseases. *Cytokine Growth Factor Rev.* **12**, 361–373.

**Table 1** Precipitating and Exacerbating Factors for Myxedema Coma

- 
- Hypothermia
  - Infections
  - Cerebrovascular accidents
  - Drugs
    - Anesthetics
    - Tranquilizers
    - Amiodarone
    - Sedatives
    - Narcotics
  - Trauma
  - Gastrointestinal bleeding
  - Metabolic disturbances exacerbating myxedema coma
    - Hypoglycemia
    - Hyponatremia
    - Hypoxemia
    - Hypercapnia
    - Acidosis
    - Hypercalcemia
- 

## LABORATORY DIAGNOSIS

As in uncomplicated hypothyroidism, a significantly elevated thyroid-stimulating hormone (TSH) with very low or undetectable thyroid hormone levels is seen. The degree of TSH elevation can vary a lot between patients; a TSH level seen in a patient with mild signs and symptoms of hypothyroidism could be comparable to that seen in another patient with myxedema coma. Similarly, it is not unheard of for patients to be functional with undetectable levels of thyroid hormone (at least for some period of time).

An exception to the rule of an elevated TSH level may be the patient with central hypothyroidism, although even in this case it is possible for the TSH to be in the normal range or even somewhat elevated, possibly reflecting bioinactive TSH.

## TREATMENT

Given the high mortality associated with myxedema coma, treatment must be instituted as soon as diagnosis is strongly suspected. Thyroid hormone therapy alone might not be enough in the presence of the multiple system organ dysfunction associated with myxedema coma. Ventilatory support is often required and helps to prevent respiratory failure, a common cause of death in these cases. An intensive care unit is always the best place in which to care for these patients. Mechanical ventilation and empirical antibiotic therapy are often indicated together with monitoring of blood gases. It is important to realize that prolonged ventilatory support may be required despite adequate therapy of the hypothyroidism.

Correction of hyponatremia is of equal importance. This can easily contribute to the altered mental status of these patients, especially if the serum sodium concentration is less than 120 mEq/L. Intravenous saline and dextrose are used to correct any volume depletion and to provide minimal nutritional support. If serum sodium concentration is less than 120 mEq/L, use of small amounts of hypertonic saline may be indicated; however, this requires very close monitoring of sodium concentration changes to avoid central pontine myelinolysis. The intravenous fluids may be warmed to help correct hypothermia. Thyroid hormone will ultimately restore body temperature, and external heat with warming blankets must be used with great caution because they may act to cause vasodilation and provide too precipitous a fall in peripheral vascular resistance. Rather than external warming, it would be more prudent to use ordinary blankets or increase the ambient room temperature.

If there is any suspicion of adrenal insufficiency, stress dose steroids should be given after a baseline cortisol level is drawn. The presence of signs such as hypoglycemia, hyponatremia, hyperkalemia, and hypotension is highly suggestive. This can be of extreme importance because correction of the hypothyroidism without correction of adrenal insufficiency can precipitate an adrenal crisis. Rarely, these patients require vasopressor support for hemodynamic stability.

The most important part of therapy is the replacement of the thyroid hormone deficiency. Which regimen of thyroid hormone administration to use remains controversial. Administration of levothyroxine (T4) by itself may result in insufficient levels of triiodothyronine (T3) because of inadequate mono-deiodination of T4 to T3 in sick patients. On the other hand, T4 therapy provides a steady smooth rise of the hormone level and is less likely to be associated with any adverse effects. A commonly used dosing regimen involves administration of a high dose (300–600 µg) intravenously the first day to replete the body's stores and then about 50 to 100 µg daily intravenously or orally after that.

Administration of T3 has the advantage that its onset of action is much faster than that of T4, an advantage that can be very important for patients' survival. The drawback is that there may be potential for an increased incidence of complications, especially in those patients who have primary cardiac diseases or who are at risk for arrhythmias or ischemia. Intravenous preparations of T3 are available, and common doses for myxedema coma are 10 to 20 µg intravenously every 4 h for the first day and then 10 µg every 6 h for 1 to 2 days. Oral administration is usually possible after that.

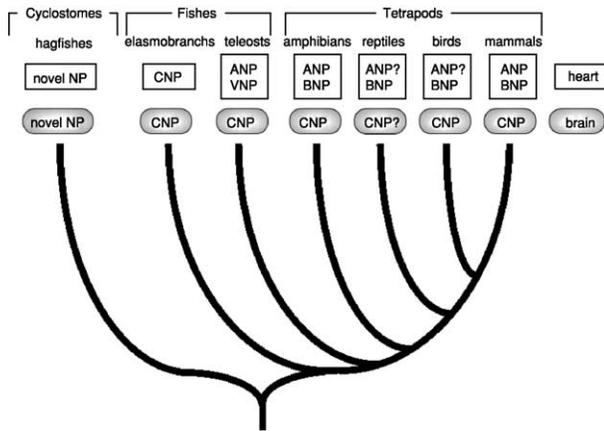
Another approach is to administer both T4 and T3 initially. T4 is given at a dose of 4  $\mu\text{g}/\text{kg}$  lean body weight (approximately 200–300  $\mu\text{g}$ ) followed by 100  $\mu\text{g}$  24 h later and then 50  $\mu\text{g}$  daily either intravenously or orally as tolerated. The initial T3 dose is 10  $\mu\text{g}$  intravenously every 8 to 12 h until the patient is able to tolerate oral intake. The response to therapy can be quite variable, and close monitoring of all such patients is extremely important.

### See Also the Following Articles

Hypothyroidism, Systemic Manifestations of • TSH Function and Secretion

### Further Reading

- DeRubertis, F. R., Jr., Michelis, M. F., and Bloom, M. G. (1971). Impaired water excretion in myxedema. *Am. J. Med.* **51**, 41–53.
- Holvey, D. N., Goodner, C. J., and Nicoloff, J. T. (1964). Treatment of myxedema coma with intravenous thyroxine. *Arch. Int. Med.* **113**, 89.
- Lindberger, K. (1975). Myxedema coma. *Acta Med. Scand.* **198**, 87–90.
- Martinez, F. J., Bermudez-Gomez, M., and Celli, B. R. (1989). Hypothyroidism: A reversible cause of diaphragmatic dysfunction. *Chest* **96**, 1059–1063.
- Pereira, V. G., Haron, E. S., and Lima-Neto, N. (1982). Management of myxedema coma: Report on three successfully treated cases with nasogastric or intravenous administration of triiodothyronine. *J. Endocrinol. Invest.* **5**, 331–334.
- Skowsky, R. W., and Kikuchi, T. A. (1978). The role of vasopressin in the impaired water excretion of myxedema. *Am. J. Med.* **64**, 613–621.
- Wilson, W. R., and Bedell, G. M. (1960). The pulmonary abnormalities in myxedema. *J. Clin. Invest.* **39**, 42.
- Yamamoto, T. (1984). Delayed respiratory failure during the treatment of myxedema coma. *Endocrinol. Jap.* **31**, 769–775.
- Zwillich, C. W., Pierson, D. J., and Hofeldt, F. D. (1975). Ventilatory control in myxedema and hypothyroidism. *N. Engl. J. Med.* **292**, 662–665.



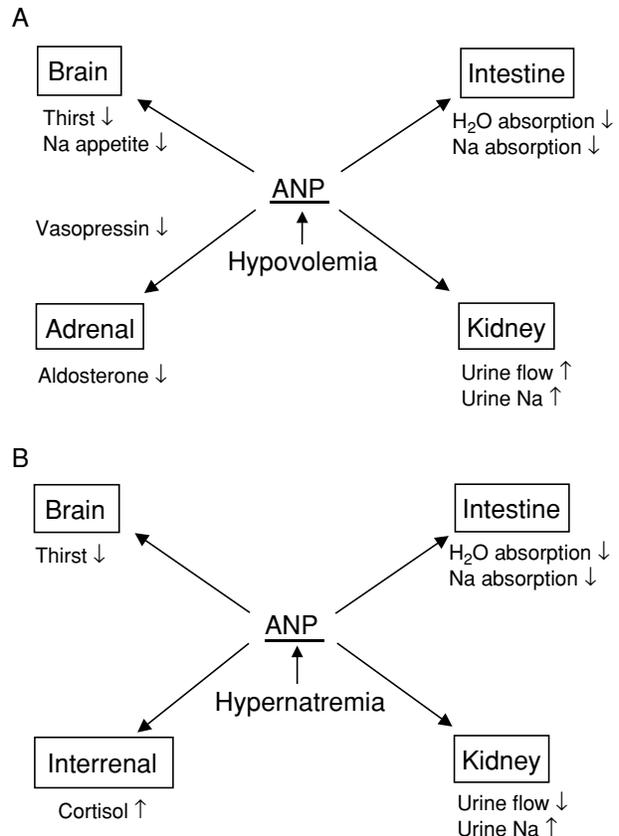
**Figure 1** Phylogenetic distribution of natriuretic peptides in vertebrates. NPs in rectangles are circulating hormones secreted from the heart. NPs in ellipses are paracrine factors synthesized mainly in the brain.

and ventricular natriuretic peptide (VNP) secreted from the heart and paracrine CNP synthesized and stored most abundantly in the brain (Fig. 1). In nonmammalian vertebrates, BNP and CNP are identified in birds, whereas ANP, BNP, and two types of CNP are identified in amphibians (Fig. 3). Thus, ANP, BNP, and CNP appear to be common members of the NP family in tetrapods. However, in fishes, ANP, CNP, and VNP are identified in teleost fish. VNP has a uniquely long C-terminal tail sequence that extends from the intramolecular ring structure (Fig. 3). In elasmobranchs, CNP is the only NP in the heart and brain. In cyclostomes (hagfish), a novel NP that displays a chimerical structure sharing all features of NPs is the sole NP in the heart and brain (Fig. 1). It remains undetermined whether CNP or hagfish NP is an ancestral form of the NP family.

CNP is structurally the most conserved peptide in the NP family (Fig. 3), whereas BNP is highly variable even in mammals (Fig. 3). ANP exhibits high sequence identity in mammals but has low identity among various classes of vertebrates. VNP is well conserved in teleosts; its sequence is more than 80% identical between eel and rainbow trout that are phylogenetically distant and is completely identical between rainbow trout and chum salmon.

## MOLECULAR EVOLUTION OF NATRIURETIC PEPTIDE RECEPTORS

Cloning of NP receptors in nonmammalian species is still scanty. In tetrapods, NPR-A and NPR-B have been cloned in two species of frogs (*Xenopus laevis* and



**Figure 2** Secretory stimulus and biological actions of ANP on water and sodium metabolism in mammals (A) and seawater fish (B).

*Rana catesbeiana*), whereas no NPR has been cloned in birds and reptiles. In fishes, NPR-A and NPR-B have been cloned in eel (*Anguilla japonica*) and medaka (*Oryzias latipes*), and NPR-C has been cloned in eel. In addition, a new receptor that lacks a cytosolic guanylyl-cyclase domain, named NPR-D, has been cloned in eel. Unlike clearance-type NPR-C that is scattered ubiquitously in various tissues in large numbers, however, NPR-D is localized selectively in the brain. Thus, NPR-D may serve as a good model to pursue biological functions of guanylyl cyclase-deficient receptors. NPR-D is a tetramer as NPR-A and NPR-B that also differs from dimeric NPR-C.

## FUNCTIONAL EVOLUTION OF THE NATRIURETIC PEPTIDE SYSTEM

In birds, as in mammals, rat ANP is diuretic and natriuretic in the chicken, indicating that ANP is a volume-regulating hormone. However, chicken BNP stimulates the secretion of hypertonic NaCl solution

## Atrial (A-type) natriuretic peptide (ANP)

Eel	SKSSSPCFGGKLDRIGSYSGLGCNSRK-NH <sub>2</sub>
Trout	SKAVSGCFGARMDRIGTSSGLGCSPKRRS
Bullfrog	SMRRSSDCFGSRIDRIGAQSGMGCGRRF
Clawed frog	SLRRSSDCFGGRIDRIGAQSGMGCNSHRF
Man	SLRRSSCFGGMRDRIGAQSGGLGCNSFRY
Rat	SLRRSSCFGGRIDRIGAQSGGLGCNSFRY

## Brain (B-type) natriuretic peptide (BNP)

Bullfrog	SNCFGRRIDRIDSVSGMGCNGSRNRY
Chicken	MMRDSGCFGRRIDRIGSLSGMGCNSRKN
Pig	SPKTMRDGCFGRRLDRIGSLSGLCNVLRRY
Dog	SPKMMHKGCFGRRLDRIGSLSGLCNVLKRY
Ox	GPKMMRDGCFGRRLDRIGSLSGLCNVLRRY
Man	SPKMVQSGCFGRKMDRISSSSGLGCKVLRH
Hamster	NSKKMHNSGCFGQRIDRIGSFSRLGCNVLKRY
Rat	NSKMAHSSSCFGQKIDRIGAVSRLGCDGLRLF

## C-type natriuretic peptide (CNP)

Spotted dogfish	GPSRGCFGVKLDRIGAMSGLGC
Spiny dogfish	GPSRSCFGLKLDRIGAMSGLGC
Eel	GWNRCFGLKLDRIGSLSGLGC
Killifish	GWNRCFGLKLDRIGSMSGLGC
Bullfrog I	GYSRGCFGVKLDRIGAFSGLGC
Bullfrog II	GTSKGCFLKLDRIGAMSGLGC
Chicken	GLSRSCFGLKLDRIGSMSGLGC
Man	GLSKGCFGLKLDRIGSMSGLGC

## Ventricular natriuretic peptide (VNP)

Eel	KSFNSCFGTRMDRIGSWSGLCNSLKNGTKKKIFGN
Trout	KSFNSCFGNRIERIGSWSGLCNNVKTGNKKRIFGN

**Figure 3** Amino acid sequences of natriuretic peptides in vertebrates. Brackets show disulfide bonds formed by two cysteine residues. Identical amino acid residues in the same group are shaded.

from the nasal salt gland of seawater-adapted ducks (*Anas platyrhynchos*), suggesting a specific role for sodium extrusion in marine birds. In reptiles, mammalian ANP is vasodepressor, but no osmoregulatory actions have been examined. In amphibians, homologous ANP is without effect on NaCl absorption by the skin, but it inhibits vasotocin-stimulated NaCl absorption in the bullfrog. Collectively, NP peptides seem to decrease body sodium in tetrapods in general.

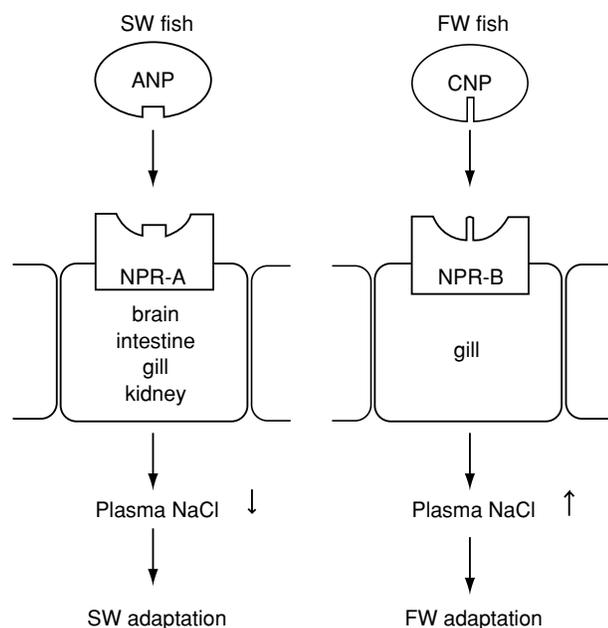
In teleost fish (eel), ANP secretion caused by an increase in plasma osmolality is more prominent than that caused by an increase in blood volume (Fig. 2). ANP secretion occurs after transfer of eels from fresh water to sea water when blood volume decreases and plasma sodium concentration increases. Thus, ANP secretion after seawater transfer is stimulated solely by hypernatremia even with an inhibitory signal

by hypovolemia. The stimulation of ANP secretion occurs *in vitro* from the isolated eel atrial tissue when sodium concentration in the medium is increased.

Consistent with the secretory stimulus, ANP specifically decreases sodium ions from the body, resulting in a decrease in plasma sodium concentration in seawater-adapted eels. This is exemplified by the fact that ANP decreases urine volume but increases urine sodium concentration (Fig. 2). With respect to the intake, ANP inhibits drinking in seawater eels and inhibits intestinal absorption of sodium from the ingested sea water. ANP also seems to increase sodium excretion by the gills of seawater fish. Furthermore, ANP stimulates the secretion of cortisol, which acts as a mineralocorticoid as well as a glucocorticoid in teleost fish to facilitate long-term adaptation to sea water (Fig. 2). Most of these ANP

actions are observed only in seawater fish. Therefore, ANP is an important hormone that promotes seawater adaptation in teleost fish (Fig. 4).

In contrast, CNP appears to be a hormone important for freshwater adaptation (Fig. 4). Unlike mammals in which CNP is principally a paracrine factor, CNP circulates in the blood of freshwater fish. The CNP-specific receptor, NPR-B, is expressed more abundantly in the gills and other osmoregulatory organs of freshwater fish than in those of seawater fish. Furthermore, CNP infused into freshwater fish promote sodium uptake from the environment, resulting in an increase in plasma sodium concentration. This is in contrast to the decrease caused by ANP infusion in seawater fish. These results, together with seawater-adapting effects of ANP, indicate that the NP system is a key hormonal system that governs the diverse environmental adaptability of euryhaline fish (Fig. 4).



**Figure 4** Effects of natriuretic peptides (ANP and CNP) on adaptation to sea water (SW) and fresh water (FW) in euryhaline fish.

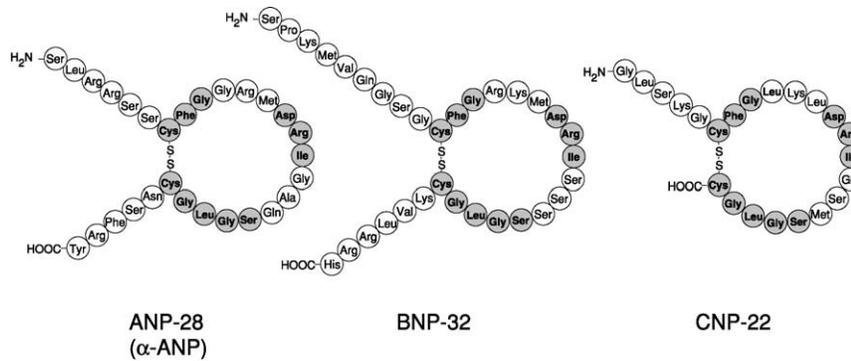
In the shark, where CNP is the only NP in the heart and brain, CNP stimulates hypertonic NaCl secretion from the rectal gland of marine species. Because CNP is a circulating hormone in the shark, as is the case with ANP in teleosts and tetrapods, CNP may take over the role of ANP in marine elasmobranchs. Taken together, it seems that the essential action of ANP is sodium extrusion throughout vertebrate species. This is particularly eminent in marine fishes, but it becomes a volume-regulating hormone in mammals, probably because sodium and water are regulated together in the same direction in terrestrial species. This illustrates the functional evolution of the NP system during the course of landward migration in the vertebrate phylogeny.

### See Also the Following Articles

ACTH,  $\alpha$ -MSH, and POMC, Evolution of • Angiotensin, Evolution of • Atrial Natriuretic Factor and Family of Natriuretic Peptides • Natriuretic Peptides • Neuropeptide Y, Evolution of • Prolactin, Evolution of • Somatostatin, Evolution of • Steroid Receptors, Evolution of

### Further Reading

- Evans, D. H. (1990). An emerging role for a cardiac peptide hormone in fish osmoregulation. *Annu. Rev. Physiol.* **52**, 43–60.
- Farrell, A. P., and Olson, K. R. (2000). Cardiac natriuretic peptides: A physiological lineage of cardioprotective hormone? *Physiol. Biochem. Zool.* **73**, 1–11.
- Inoue, K., Naruse, K., Yamagami, S., Mitani, H., Suzuki, N., and Takei, Y. (2003). Four functionally distinct C-type natriuretic peptides found in fish reveal new evolutionary history of the natriuretic system. *Proc. Natl. Acad. Sci. USA* **100**, 10079–10084.
- Loretz, C. A., and Pollina, C. (2000). Natriuretic peptide in fish physiology. *Comp. Biochem. Physiol. A* **125**, 169–187.
- Takei, Y. (2000). Structural and functional evolution of the natriuretic peptide system in vertebrates. *Int. Rev. Cytol.* **194**, 1–66.
- Takei, Y. (2001). Does the natriuretic peptide system exist throughout the animal and plant kingdom? *Comp. Biochem. Physiol. B* **129**, 559–573.
- Takei, Y., and Hirose, S. (2002). The natriuretic peptide system in eel: A key endocrine system for euryhalinity? *Am. J. Physiol.* **282**, R940–R951.

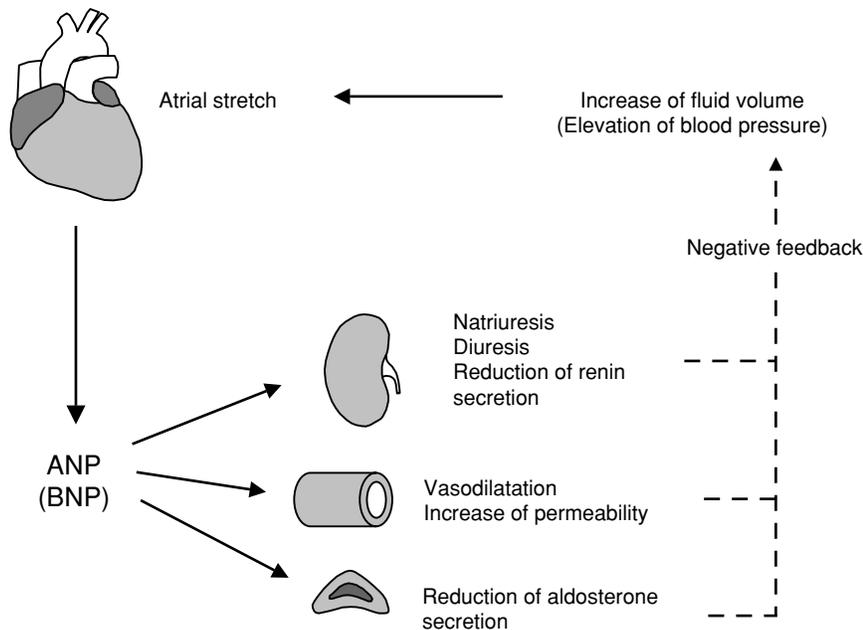


**Figure 1** Amino acid sequences of human ANP-28 ( $\alpha$ -ANP), BNP-32, and CNP-22. Shaded amino acids are common to the three natriuretic peptides.

The third member, CNP, is present in the human brain, but in contrast to ANP and BNP, little CNP is detected in the heart. The two molecular forms, CNP-53 and CNP-22, were processed from the pro-hormone CNP(1–103). The major molecular form circulating in human blood is CNP-22 (Fig. 1), although the level is much lower than those of ANP and BNP. The vascular endothelium also produces CNP, probably contributing to plasma CNP. The endothelial production is augmented by growth-promoting factors of the vascular smooth muscle, such as thrombin and transforming growth factor- $\beta$ , and by factors involved in the pathogenesis of sepsis, such as tumor necrosis factor- $\alpha$  and lipopolysaccharide.

**BIOLOGICAL ACTIONS**

Natriuresis and diuresis occur with a reduction of blood pressure following intravenous infusion of ANP in humans (Fig. 2). The natriuretic peptides exert an endothelium-independent vasodilator effect by acting directly on the vascular smooth muscle cells to elevate the intracellular cyclic guanosine 3',5'-monophosphate (cGMP) level. Despite the reduced blood pressure, the increase in heart rate is relatively small due to an inhibitory effect on sympathetic nerve activity. ANP lowers plasma renin activity and inhibits aldosterone secretion from the adrenal glomerulosa. According to *in vitro* experiments, the natriuretic peptides inhibit



**Figure 2** Schematic representation of the actions of ANP (and BNP) and fluid volume regulation by ANP.

cellular proliferation of the vascular smooth muscle and renal mesangial cells.

The natriuretic peptides induce natriuresis and diuresis largely by acting on the glomeruli and inner medullary collecting ducts of the kidney. ANP dilates the afferent arterioles with little effect or constrictor effect on efferent arterioles, leading to an increase in intraglomerular capillary pressure that elevates the glomerular filtration rate (GFR). ANP relaxes the mesangial cells of the glomerulus in a cGMP-dependent manner, and this relaxation also increases the GFR. Another important site of action appears to be the inner medullary collecting duct, where the natriuretic peptides inhibit sodium and water reabsorption. In addition, increased blood flow to the renal medulla partly contributes to natriuresis and diuresis induced by the natriuretic peptides. Fluid volume is reduced by their actions not only on the kidneys but also on vascular permeability. ANP increases the vascular permeability, thereby shifting fluid volume to the extravascular space.

Both the natriuretic peptides and their receptors are present in the central nervous system (CNS). When infused to the lateral ventricles of the rat brain, ANP inhibits salt appetite and water drinking and lowers blood pressure, suggesting roles of the natriuretic peptides in the CNS. Centrally infused ANP reduces the secretion of antidiuretic hormone (ADH) and adrenocorticotrophic hormone (ACTH). These central actions appear complementary to the systemic roles of the natriuretic peptides in reducing fluid volume and blood pressure.

## RECEPTOR AND INTRACELLULAR MECHANISM OF THE ACTIONS

The biological actions discussed previously are mediated by binding of the natriuretic peptides to specific receptors on the cell surface. Three subtypes of natriuretic peptide receptors (NPRs) have been identified. The two subtypes NPR-A and NPR-B are transmembrane proteins that have similar structures: an extracellular ligand-binding domain and an intracellular domain containing protein kinase-like and guanylate cyclase portions. The third type of receptor is a "clearance" receptor (NPR-C), the main function of which is clearance of the natriuretic peptides from the bloodstream.

Most of the natriuretic peptide actions are mediated by elevation of the intracellular cGMP level. Binding of the natriuretic peptides to the extracellular domain of NPR-A or NPR-B activates guanylate

cyclase of the intracellular domain, increasing the conversion of guanosine triphosphate (GTP) into cGMP. The cellular response occurs by the cGMP-dependent cascades, including activation of protein kinase G, although the precise mechanisms of cGMP actions appear to be complex. In contrast, NPR-C has neither a protein kinase-like nor a guanylate cyclase domain in its short intracellular portion. Following binding of the natriuretic peptides to NPR-C, the receptor-ligand complex is internalized to the cytoplasm, where the peptides are degraded by lysosomal enzymes. Thus, NPR-C participates in the metabolism of natriuretic peptides.

The modes of action of ANP, BNP, and CNP depend on the specificity and distribution of their receptors. ANP activates NPR-A most potently, followed by BNP, whereas CNP has a much lesser affinity. On the other hand, CNP is highly specific for NPR-B, compared with ANP and BNP, in terms of cGMP generation. Both NPR-A and NPR-B are present in a number of tissues, such as the kidney, adrenal gland, and vascular smooth muscle, where the natriuretic peptides exert their effects. The affinity of NPR-C for BNP seems to be less than that for ANP or CNP, suggesting a lower clearance rate of BNP. NPR-C is widely distributed in many tissues, including the vascular endothelium and smooth muscle.

## METABOLISM

Circulating natriuretic peptides are metabolized by two mechanisms: clearance by NPR-C and enzymatic cleavage by neutral endopeptidase (NEP)-24.11. Both mechanisms seem to be important for the metabolism of natriuretic peptides. NEP-24.11 is widely distributed in many tissues and organs, including the kidney and lung. The natriuretic peptides filtered through the glomerulus of the kidney are metabolized by NEP expressed on the brush border membrane of renal tubules. NEP-24.11 cleaves the ringed structure of ANP, an essential portion for the biological action, but is rather a nonspecific enzyme and hydrolyzes not only the natriuretic peptides but also other bioactive peptides such as bradykinin and endothelin.

## PLASMA LEVELS

In healthy adults, the plasma levels of ANP, BNP, and CNP are in the ranges 20 to 40, 2 to 4, and 1 to 2 pg/ml, respectively. Both the ANP and BNP levels increase with aging and are transiently elevated by physical exercise. In patients with congestive heart

failure, the plasma ANP levels progressively increase in relation to severity of the disease. Because atrial stretch is the major stimulus of ANP secretion, the plasma ANP level can be a marker of volume retention or a preload to the heart. Compared with ANP, the plasma level of BNP is lower in healthy individuals but increases to a much greater degree in heart failure as well as in myocardial infarction. BNP is a hormone produced mainly by cardiac ventricles responding to stress to the heart, and its plasma level is clinically used as a sensitive indicator to evaluate cardiac function. Based on the inhibitory actions of ANP and BNP on the renin–angiotensin system and sympathetic nerve activity, the elevated plasma level is thought to be involved in the mechanism counteracting excessive volume retention and vasoconstriction in heart failure (Fig. 2). The plasma levels of ANP and BNP also increase in patients with hypertension and renal failure. These increases appear to be responses against blood pressure elevation and volume retention. In patients with renal failure, reduced clearance of ANP and BNP partly contributes to the increased plasma levels. In contrast to ANP and BNP, little change is observed in the plasma level of CNP in heart failure or hypertension, but the CNP level rises markedly in sepsis, and this increase appears to be due to increased production in the vascular endothelium by various cytokines.

## THERAPEUTIC IMPLICATIONS

The therapeutic potential of the natriuretic peptides has been considered for various diseases in which either fluid volume retention or excessive vasoconstriction occurs such as heart failure, hypertension, and renal failure. When infused intravenously to patients with congestive heart failure, ANP exerted favorable effects—reductions in pulmonary capillary wedge pressure and peripheral vascular resistance and an increase in cardiac output. ANP is currently

used as an intravenous agent for heart failure patients in Japan, and BNP has become available in the United States. On the other hand, because the natriuretic peptides are metabolized by NEP-24.11, substantial efforts have been made to develop orally available drugs that inhibit the enzymatic activity of NEP-24.11. Indeed, the NEP inhibitors elevate the plasma level of endogenous ANP when administered intravenously or orally in patients with congestive heart failure. This elevation leads to natriuresis, diuresis, reductions in pulmonary capillary wedge pressure and right atrial pressure, and a reduction in plasma renin activity. Orally active NEP inhibitors could be useful in the treatment of not only congestive heart failure but also hypertension.

## See Also the Following Articles

Atrial Natriuretic Factor and Family of Natriuretic Peptides  
 • Natriuretic Peptide System, Evolution of

## Further Reading

- Jewicki, J. A., and Protter, A. A. (1995). Physiological studies of the natriuretic peptide family. In "Hypertension: Pathophysiology, Diagnosis, and Management" (J. H. Laragh and B. M. Brenner, eds.), 2nd ed., pp. 1029–1053. Raven Press, New York.
- Levin, E. R., Gardner, D. G., and Samson, W. K. (1998). Natriuretic peptides. *N. Engl. J. Med.* **339**, 321–328.
- Maack, T. (1995). Receptors of natriuretic peptides: Structure, function, and regulation. In "Hypertension: Pathophysiology, Diagnosis, and Management" (J. H. Laragh and B. M. Brenner, eds.), 2nd ed., pp. 1001–1019. Raven Press, New York.
- Potter, L. R., and Hunter, T. (2001). Guanylyl cyclase-linked natriuretic peptide receptors: Structure and regulation. *J. Biol. Chem.* **276**, 6057–6060.
- Rademaker, M. T., and Espiner, E. A. (2001). Hormones of the cardiovascular system. In "Endocrinology" (L. J. DeGroot and J. L. Jameson, eds.), 4th ed., pp. 1791–1809. W. B. Saunders, Philadelphia.
- Thibault, G., Amiri, F., and Garcia, R. (1999). Regulation of natriuretic peptide secretion by the heart. *Annu. Rev. Physiol.* **61**, 193–217.

deriving from bowel, pancreas, kidneys, placenta, thyroid, pituitary, and hypothalamus. Within the stomach, ghrelin is produced by the enteroendocrine cells, probably the X/A-like cells, a major endocrine population in the oxyntic mucosa, the hormonal product of which was previously unknown.

Ghrelin is the first peptide isolated from natural sources in which the hydroxyl group of one of its residues (namely, serine 3) is acylated by *n*-octanoic acid. The acylation seems critical for the peptide to be able to cross the blood-brain barrier and is essential for binding GHS-R type 1a and for endocrine actions. In fact, nonacylated ghrelin, which circulates in far greater amounts than the acylated form, does not bind GHS-R type 1a and has no endocrine activities. However, nonacylated ghrelin exerts some nonendocrine actions including cardiovascular and anti-proliferative effects, probably binding different GHS-R subtypes or other receptor families. It should be noted that it is possible that ghrelin is not the sole natural GHS-R ligand and that GHS-R type 1a is one of a group of receptors for such ligands. Indeed, GHS-R is also bound by other molecules, such as Des-Gln14-ghrelin, which is a gastric homologue of ghrelin, with the exception that one glutamine is missing, and which has the same biological activities as ghrelin; adenosine, which binds but does not activate the receptor; and cortistatin, which is a neuropeptide that is homologous to somatostatin, which is in turn unable to recognize GHS-R type 1a.

Moreover, there is strong evidence suggesting the existence of additional receptor subtypes. Specific binding sites for peptidyl GHS only, with very low binding affinity for ghrelin and nonpeptidyl GHS, have been shown in a wide range of nonendocrine peripheral human tissues.

Circulating ghrelin levels mostly reflect gastric secretion. The regulation of ghrelin secretion is still largely unknown, but it has already been shown that circulating ghrelin levels are increased by fasting and decreased by food intake and glucose but not by gastric distension. In agreement with the critical influence of nutrition on ghrelin secretion, circulating ghrelin levels are increased in anorexia and cachexia but reduced in obesity.

The functional relationship between ghrelin and insulin is still unclear. Indeed, there is a negative association between their circulating levels, and insulin administration has been reported to decrease ghrelin secretion independent of glucose variations.

Though ghrelin drives the GH response to fasting in humans, a feedback regulatory mechanism between ghrelin and GH seems unlikely. On the other hand,

both somatostatin and cortistatin exert a potent inhibitory effect on ghrelin secretion in humans.

Thus, all these data clearly indicate that ghrelin represents a true turning point in understanding the control of the GH-insulin-like growth factor-I (IGF-I) axis but, indeed, in addition to potent GH-releasing actions, ghrelin and synthetic GHS have other remarkable endocrine and nonendocrine actions that are often unrelated to their GH-releasing activity.

## ENDOCRINE ACTIONS OF GHRELIN

### GH-Releasing Action

Ghrelin and synthetic GHS possess a strong and dose-related GH-releasing effect, which synergizes with that of GH-releasing hormone (GHRH), indicating at least partially different mechanisms of action. Nevertheless, GHS need GHRH activity to fully express their GH-releasing effect. In humans, the GH response to GHS is strongly inhibited by GHRH receptor deficiency or antagonists as well as by hypothalamus-pituitary disconnection, accordingly with the assumption that the most important action of GHS takes place at the hypothalamic level.

In both animals and humans, ghrelin and synthetic GHS act as functional somatostatin antagonists at both the pituitary level and the hypothalamic level. In fact, the GH response to ghrelin and GHS is refractory to the modulatory effects of substances enhancing or inhibiting somatostatin secretion or action.

In humans, the GH-releasing effect of ghrelin and GHS undergoes marked age-related variations increasing at puberty, persisting in adulthood, and decreasing with aging. The mechanisms underlying the age-related variations in the GH-releasing action of GHS differ according to age and could involve variations in estrogenic levels, GHRH and/or somatostatin expression, and GHS-R and/or endogenous ghrelin expression.

### PRL- and ACTH-Releasing Action

In humans, the stimulatory effect of ghrelin and GHS on PRL secretion is slight, independent of both gender and age, and probably comes from direct stimulation of somatomammotroph cells.

The acute stimulatory effect of ghrelin and GHS on the hypothalamus-pituitary-adrenal axis in humans is remarkable and similar to that of naloxone, arginine-vasopressin (AVP), and even corticotropin-releasing hormone (CRH) but probably vanishes during prolonged treatment. Under physiological

conditions, the ACTH-releasing action of GHS totally depends on central nervous system-mediated mechanisms including AVP-, neuropeptide Y (NPY)-, and/or  $\gamma$ -aminobutyric acid-mediated actions.

The ACTH response to GHS is generally sensitive to the negative feedback action of cortisol but is surprisingly exaggerated (and higher than the response to CRH) in patients with pituitary ACTH-dependent Cushing's disease as well as in some patients with ectopic ACTH-dependent Cushing's syndrome. Interestingly, ghrelin and GHS-R are expressed in abnormal human pituitary and in other neuroendocrine tumors including ACTH-secreting tumors; GHS also stimulate ACTH release from human ACTH-secreting pituitary adenomas but not from normal human pituitary.

### Influence on Endocrine Pancreatic Function

Ghrelin and GHS-R type 1a are also present in normal and neoplastic endocrine pancreas. In agreement with these data, an increasing number of studies suggest that ghrelin plays a significant role in the fine-tuning of insulin secretion and glucose metabolism, integrating the hormonal and metabolic response to fasting.

In humans, a clear relationship between ghrelin and insulin secretion has been found and ghrelin has been shown to induce a significant increase in plasma glucose levels that is surprisingly followed by a reduction in insulin secretion. Specifically, ghrelin may likely block the inhibitory effects of insulin on gluconeogenesis and have a direct, non-GHS-R type 1a-mediated, stimulatory effect on glycogenolysis, since this activity is not shared by synthetic GHS.

### Influence on Gonadal Function

GHS-R are present in the testis and in the ovary. Leydig cells synthesize ghrelin, which, in turn, induces a significant inhibition of human chorionic gonadotropin (HCG)- and cyclic AMP-stimulated testosterone secretion *in vitro* coupled with a significant decrease in HCG-stimulated expression levels of steroid acute regulatory protein and synthetic enzymes. These data, together with evidence that intracerebroventricularly injected ghrelin inhibits pulsatile luteinizing hormone secretion in rats, provide evidence for a possible action of ghrelin in the regulation of the gonadal axis and of testicular function.

## NONENDOCRINE ACTIONS OF GHRELIN

### Regulation of Energy Balance

Exogenous ghrelin induces weight gain in rodents by increasing food intake and reducing fat utilization. These actions are GH independent and are mediated by a specific central hypothalamic network of NPY- and Agouti-related protein neurons that is also modulated by leptin; ghrelin and leptin might actually be complementary players of one regulatory system that informs the central nervous system about the status of energy balance.

A premeal rise in circulating ghrelin levels suggests its role as a hunger signal triggering meal initiation and this action would be mediated by GHS-R non-type 1a subtypes, as suggested by evidence that GHS analogues devoid of any GH-releasing effect stimulate food intake.

Peripheral ghrelin could reach GHS-R in the hypothalamus through the general circulation to regulate food intake and energy homeostasis. Ghrelin-containing cells are also present in the mediobasal hypothalamus, where GHRH-secreting neurons and the neuroendocrine network regulating energy balance are located.

### Effects on Behavior

In mice, ghrelin induces anxiogenic activities after both intracerebroventricular and intraperitoneal administration, an effect that is significantly inhibited by the administration of a CRH receptor antagonist. These findings suggest that ghrelin may play a role in mediating neuroendocrine and behavioral responses to stressors and that the stomach could play an important role, not only in the regulation of appetite, but also in the regulation of anxiety.

### Gastro-entero-pancreatic Actions

It is not surprising that, as a gastric hormone, ghrelin acts at the gastro-entero-pancreatic level, where GHS-R type 1a have been demonstrated. Interestingly, there is also a close structural and functional similarity between motilin and ghrelin. Also, the gastrointestinal motilin receptor 1A and the GHS-R type 1a show a high degree of structural homology.

Ghrelin stimulates gastric acid secretion and motility. Interestingly, these actions are mediated by the cholinergic system, at least partially, at the central level.

Moreover, ghrelin exerts a potent inhibitory effect on pancreatic cholecystokinin-induced exocrine secretion in rats via modulation of intra-pancreatic neuron activities.

It should be noted that the endocrine cells of the stomach may undergo both hyperplastic and neoplastic phenomena and even silent gastric carcinoids often synthesize ghrelin. These might represent conditions of ghrelin hypersecretion, whereas gastrectomy is associated with a 65% reduction in circulating ghrelin levels. The clinical impact of ghrelin hyper- and hyposecretion is still unknown.

### Cardiovascular Actions

In the cardiovascular system, both GHS-R type 1a and specific binding sites for only peptidyl GHS have been reported, suggesting the coexpression of different GHS-R subtypes.

Peptidyl GHS markedly protects against cardiovascular ischemic damage and improves cardiac performance in normal rats and in different animal models of GH deficiency, ischemic heart disease, or dilated cardiomyopathy. On the other hand, acute administration of high-dose peptidyl GHS induces transient coronary vasoconstriction via activation of the binding glycoprotein CD36.

In humans, hexarelin, a peptidyl GHS, increases the left ventricular ejection fraction in normal subjects, in hypopituitary patients with severe GH deficiency (GHD), and in patients with ischemic dilated cardiomyopathy. Although ghrelin has been reported to be devoid of cardioprotective effects, its chronic administration has been shown to improve cardiac contractility and to reduce systemic vascular resistance in GHD rats and in rats with chronic heart failure.

In humans, independent of GH-IGF-I and nitric oxide systems, ghrelin significantly decreases systemic vascular resistance and mean arterial pressure and increases cardiac index and stroke volume index in normal subjects and in patients with chronic heart failure. The vasodilating activity of ghrelin has been attributed to antagonism of endothelin-1.

### Antiproliferative Actions

GHS-R have been also found in neoplastic endocrine and nonendocrine tissues even from organs that do not express these receptors under physiological conditions, such as the breast.

Both normal and neoplastic thyroid tissues express GHS-R and ghrelin, at least in parafollicular cells.

Both ghrelin and synthetic GHS inhibit cell proliferation of thyroid follicular, papillar, and anaplastic tumor cell lines.

GHS-R was also shown in breast tumors but not fibroadenomas and normal mammary parenchyma. Both ghrelin and synthetic GHS inhibit cell proliferation in different human breast (estrogen-dependent and estrogen-independent) carcinoma cell lines. The same effect is shared by nonacylated ghrelin, indicating that this is a non-GHS-R type 1a-mediated action.

Neuroendocrine carcinoid tumors and even adenocarcinomas of the lung express specific GHS-binding sites. Synthetic peptidyl GHS, but not ghrelin, inhibit human CALU-1 lung cancer cell line proliferation.

On the other hand, the prostate cancer cell line PC-3 shows increased cell proliferation *in vitro* after exposure to ghrelin, suggesting that autocrine-paracrine pathways involving ghrelin might be capable of stimulating cell proliferation (at least of this cell line).

### CLINICAL PERSPECTIVES

GHS have been considered to represent a potential diagnostic and therapeutic tool based on their strong and reproducible GH-releasing effect even after oral administration.

Particularly when combined with GHRH, GHS represent one of the most potent and reliable tests to evaluate the pituitary GH-releasable pool for the diagnosis of GHD, provided that appropriate cutoff limits are assumed.

On the other hand, the promise of GHS as growth-promoting factor in children with GHD and as an anti-aging drug has not been fulfilled thus far.

Due to evidence that ghrelin possesses other relevant endocrine and nonendocrine actions, the possibility that ghrelin and/or GHS analogues, acting as either agonists or antagonists on different GH-unrelated activities, might have clinical impact in internal medicine, metabolism, gastroenterology, immunology, oncology, and cardiology has been suggested and is receiving increasing attention.

### CONCLUSIONS

Ghrelin, a 28-residue acylated peptide predominantly produced by the stomach, displays strong GH-releasing action but also shows other central and peripheral endocrine and nonendocrine actions via the activation of the GHS-R. Indeed, ghrelin isolation helped elucidate GH-IGF-I physiology but also

opened new promising perspectives in other medical fields, such as internal medicine, gastroenterology, immunology, oncology, and cardiology. Researchers are working on the possibility that ghrelin and/or GHS analogues may have potential diagnostic or therapeutic application.

### See Also the Following Articles

Ghrelin • Growth Hormone (GH)

### Further Reading

- Arvat, E., Maccario, M., Di Vito, L., Broglio, F., Benso, A., Gottero, C., Papotti, M., Muccioli, G., Dieguez, C., Casanueva, F. F., Deghenghi, R., Camanni, F., and Ghigo, E. (2001). Endocrine activities of ghrelin, a natural growth hormone secretagogue (GHS), in humans: Comparison and interactions with hexarelin, a nonnatural peptidyl GHS, and GH-releasing hormone. *J. Clin. Endocrinol. Metab.* **86**, 1169–1174.
- Bowers, C. Y. (1998). Growth hormone-releasing peptide (GHRP). *Cell. Mol. Life Sci.* **54**, 1316–1329.
- Kojima, M., Hosoda, H., Date, Y., Nakazato, M., Matsuo, H., and Kangawa, K. (1999). Ghrelin is a growth-hormone-releasing acylated peptide from stomach. *Nature* **402**, 656–660.
- Kojima, M., Hosoda, H., and Kangawa, K. (2001). Purification and distribution of ghrelin: The natural endogenous ligand for the growth hormone secretagogue receptor. *Horm. Res.* **56**, 93–97.
- Muccioli, G., Tschop, M., Papotti, M., Deghenghi, R., Heiman, M., and Ghigo, E. (2002). Neuroendocrine and peripheral activities of ghrelin: Implications in metabolism and obesity. *Eur. J. Pharmacol.* **440**, 235–254.
- Smith, R. G., Van der Ploeg, L. H., Howard, A. D., Feighner, S. D., Cheng, K., Hickey, G. J., Wyvratt, M. J., Jr., Fisher, M. H., Nargund, R. P., and Patchett, A. A. (1997). Peptidomimetic regulation of growth hormone secretion. *Endocr. Rev.* **18**, 621–645.

The neuroendocrine theory of aging evolved from observations that animals exhibit predictable and progressive impairments in a number of physiological processes with time, including, but not limited to, decreases in reproductive and immune function, decreases in muscle mass and function, accumulation of adipose tissue, and decreases in glucose utilization and cognitive function. The role of the neuroendocrine system in these processes was related to the observations that many of the hormones regulated by the neuroendocrine system had an important trophic and integrative role in maintaining tissue function and that withdrawal of hormonal support mimicked some of the phenotypes observed in aging animals and humans.

The neuroendocrine system includes the hypothalamus and associated brain structures as well as the pituitary gland. This system includes neurotransmitters and neuropeptides within the brain that regulate hypothalamic-releasing and-inhibiting hormones secreted into hypophysial portal blood that reaches the pituitary gland. The release of these hormones influences the secretion of anterior pituitary hormones into the bloodstream and subsequently regulates tissue function. The posterior pituitary is also an important part of the neuroendocrine system; however, in contrast to the anterior pituitary gland, it is composed of long axons from specific hypothalamic nuclei. The hypothalamus and pituitary gland have the capacity to detect neural activity and/or humoral secretions from target tissues and adjust activity to maintain an optimal internal environment or "milieu" for tissue function. It is well established that the neuroendocrine system has critical roles in regulating tissue growth and metabolism through the release of growth hormone and thyroid-stimulating hormone; reproductive function through the release of luteinizing hormone (LH), follicle-stimulating hormone (FSH), and prolactin; and plasma electrolytes and responses to stress through secretion of vasopressin and adrenocorticotropin (ACTH), respectively. In addition, the hypothalamus also has an important role in the integration of parasympathetic and sympathetic nervous system activity and can thereby affect a wide variety of functions, including heart rate, blood pressure, vascular responses, and glucose metabolism. The hypothalamus also regulates biological rhythms. Recently, fat metabolism and food intake were shown to be regulated through the hypothalamus by its response to leptin and the synthesis of neuropeptide Y and other orexigenic peptides. Unfortunately, the categorization of hormones and their primary function noted previously is an overly simplistic view of the neuroendocrine system since critical interactions

occur between hormones that contribute to the regulation of cellular function. Because many of the early events of aging include alterations in systems regulated by the neuroendocrine axis, it was proposed (and subsequent studies supported the conclusion) that age-dependent alterations in the neuroendocrine system result in a progressive series of events that are manifest as biological aging.

Although the etiology of the age-related changes in the neuroendocrine system is unknown, it has been proposed that cellular and molecular events in specific subpopulations of neurons within the hypothalamus and brain and/or supporting structures are a contributing factor in the dysregulation of this system. The cause of the specific perturbations may be related to genetic errors or increased free radicals that lead to progressive aberrations in tissue function. As a result, the neuroendocrine theory of aging is unique compared to other theories of aging in that alterations in this system are not considered the primary causative factor in biological aging but rather are important mediators of aging initiated by cellular changes in specific subpopulations of neurons or systems that closely interact with hypothalamic neurons. In this article, the major alterations within the neuroendocrine axis are discussed, with emphasis on the regulation of growth hormone since it exhibits some of the more important physiological changes that occur with age.

## GROWTH HORMONE AND INSULIN-LIKE GROWTH FACTOR-I

Since the early 20th century, it was known that a substance present in blood promotes growth, but it was only after the isolation of pure bovine growth hormone from the pituitary gland in 1945 that the biological effects of the hormone became evident. It was subsequently shown that growth hormone stimulates cellular amino acid uptake and DNA, RNA, and protein synthesis, and that it has a role in cellular division and hypertrophy.

Plasma levels of growth hormone reveal a pattern of discrete pulses that increase after the onset of sleep. Although the precise function of this ultradian pattern remains unknown, the pulsatile release of growth hormone has been confirmed in every species examined and is closely related to the biological actions of the hormone. Growth hormone in humans is characterized by relatively low-amplitude pulses throughout the day and a large pulse after the onset of sleep. In rats, growth hormone is characterized by an ultradian rhythm with high-amplitude secretory pulses every

3–3.5 h; between pulses, growth hormone decreases to almost undetectable levels. This secretory pattern is regulated by two hormones released from the hypothalamus: growth hormone-releasing hormone (GHRH), which increases growth hormone release, and somatostatin, which inhibits growth hormone release. The results of several studies suggest that both hormones are secreted in a phasic manner, with GHRH contributing to high-amplitude growth hormone pulses and somatostatin to trough levels. A dynamic interrelationship between these hypothalamic hormones is necessary for pulsatile secretion. A number of other factors either within the hypothalamus or circulating in plasma contribute to the regulation of growth hormone release either by acting directly on the pituitary gland or by regulating hypothalamic somatostatin and GHRH secretion. Both growth hormone and insulin-like growth factor-1 (IGF-1) inhibit growth hormone secretion in a typical feedback relationship: These hormones increase somatostatin and decrease GHRH release from the hypothalamus.

Recently, there has been increased interest in small peptides that stimulate the release of growth hormone. One of these peptides, Ghrelin, is produced in the stomach and stimulates growth hormone release. Although it does not interact with the GHRH receptor, it has been shown to act synergistically with GHRH to regulate both growth hormone release and possibly food intake. This peptide is particularly important since oral administration of peptide and non-peptidyl analogs of Ghrelin increase growth hormone secretion.

Although growth hormone is one of the most potent anabolic factors in the body, most of the anabolic actions of the hormone are regulated through the secretion of another hormone, IGF-1. IGF-1 binds with high affinity to the type 1 IGF receptor found in tissues throughout the body. The IGF-1 receptor shares 50% amino sequence homology with the insulin receptor, and competitive binding and affinity cross-linking studies have demonstrated that IGF-1 binds to the insulin receptor with 100 times lower affinity than the IGF-1 receptor. IGF-1 is synthesized mainly in the liver under the regulation of growth hormone but is also synthesized in smaller quantities in almost all tissues. Although regulation of local IGF-1 production or paracrine activity is poorly understood, this represents an important source of the hormone that also appears to be regulated by growth hormone.

IGF-1 secreted from liver circulates in the blood either free or bound to specific binding proteins that prolong the half-life of the peptide. At least six binding

proteins have been identified and constitute an intricate transport system for the IGFs that regulate their availability to specific tissues. It is clear that IGF binding proteins are important regulators of IGF-1 activity and may have actions independent of IGF-1.

The actions of growth hormone are not entirely mediated through IGF-1, and in muscle and adipose tissue growth hormone can have direct effects. Specifically, growth hormone decreases insulin sensitivity by direct effects on skeletal muscle and increases free fatty acids (and decreases fat mass) by increasing the activity of hormone-sensitive lipase in adipose tissue. These findings have led to the concept that the primary anabolic actions of growth hormone are mediated through IGF-1 (through either endocrine or paracrine regulation) and that the effects of growth hormone on fuel utilization are mediated by direct actions on muscle and adipocytes.

Alterations in the neuroendocrine axis and specifically in the regulation of growth hormone have an important role in the physiological and biochemical changes normally associated with aging. Early studies in humans indicated that the amplitude of growth hormone pulses and the increase in the concentrations of growth hormone were blunted after insulin-induced hypoglycemia. Subsequent studies demonstrated a prominent decrease in the amplitude of growth hormone pulses in aged animals, although there were no changes in either the basal concentrations or the ultradian rhythm of growth hormone. The decline in growth hormone results in a progressive reduction in plasma concentrations of IGF-1 in both animals and humans. Subsequent studies in various strains of rats and mice, nonhuman primates, and humans have consistently confirmed the decline in growth hormone and IGF-1 concentrations with age and suggest that these are a robust marker of biological aging in mammalian species.

Traditionally, endocrinologists investigated the impact of hormone deficiency by replacement therapy. Initial investigations demonstrated that purified growth hormone preparations reversed the age-related decline in protein synthetic capacity. Although these studies did not address the question of whether tissue response to growth hormone diminished with age, they clearly indicated that cellular protein synthesis decreases with age in an environment of reduced growth hormone and IGF-1 levels and that growth hormone and/or IGF-1 replacement have the capacity to increase protein synthesis in aged animals. Thus, the age-related decline in tissue function results, at least in part, from deficits within the neuroendocrine system.

Subsequent studies demonstrated that growth hormone or IGF-1 administration could partially reverse the decline in immune function; increase the expression of aortic elastin, lean body mass, skin thickness, and vertebral bone density; and improve cognitive function in aged animals and humans. Although the beneficial effects of growth hormone administration have been repeatedly confirmed, deleterious side effects of growth hormone replacement severely limit its usefulness as a treatment to modulate physiological changes with age. For example, growth hormone increases cartilage growth and contributes to carpal tunnel syndrome, increases glucose levels contributing to diabetes, and, in several cases, has been found to increase pathology. In fact, elevated IGF-1 levels are a risk factor for breast, gastrointestinal, prostate, and lung cancer. Nevertheless, it is evident that decreases in growth hormone and IGF-1 have clinical significance and are a contributing factor in the genesis of tissue dysfunction in aged animals and humans, although replacement therapy in the elderly population is currently unwarranted.

### ACTH, CORTISOL, AND ADRENAL FUNCTION

One of the controversies regarding neuroendocrine aging is whether the regulation of ACTH and cortisol is altered in aged animals and humans and the significance of these changes for the development of the aged phenotype. ACTH released from the pituitary gland is regulated by a synergistic action between corticotropin-releasing hormone and vasopressin on the pituitary. ACTH released into the blood subsequently stimulates activity of an enzyme within the adrenal gland that results in the conversion of cholesterol to pregnenolone, initiating the biosynthetic pathway for the synthesis of cortisol (in humans and nonhuman primates) or corticosterone (in rodents). These hormones have a diurnal rhythm, with highest levels usually observed at the end of the dark phase of the light–dark cycle. Cortisol and corticosterone exert a major role in response to stress and the regulation of immune function but are also important factors in fat metabolism and glucose regulation. Increased blood levels result in an increase in free fatty acids through their actions on hormone-sensitive lipase in adipocytes and glucose levels by stimulating gluconeogenesis and glycogenolysis and decreasing tissue sensitivity to insulin. Furthermore, elevated levels of cortisol or corticosterone result in protein degradation in skeletal muscle that causes a muscle-wasting condition similar to that found in aged animals and humans.

The similarities between the physiological changes that accompany excess cortisol secretion and the aging phenotype have led to the hypothesis that cortisol secretion increases with age (the glucocorticoid hypothesis of aging). However, the results from analyses of basal corticosterone and cortisol levels have been controversial, with investigators reporting either no change or increased basal levels of the hormones. Since ACTH and cortisol levels can be dramatically influenced by environmental variables, it is probable that many of these differences can be reconciled by differences in time of blood sampling and/or variations in animal housing conditions that produce minor stresses that differentially affect young and old animals. Similarly, in humans, there appear to be few differences in the diurnal levels of ACTH or cortisol. Perhaps the most important differences between these hormones occur when animals and humans are assessed under conditions of mild stress. In response to stressors, similar acute increases in hormone levels are noted with age but the levels of the hormones remain elevated for longer periods of time in older subjects, leading to the conclusion that relatively small but important increases in cortisol and corticosterone occur. The consequences of the potential age-related increases in corticosterone or cortisol for the development of tissue impairments associated with age are unknown. Certainly, high levels of these hormones (independent of age) have been associated with atrophy of brain structures linked to learning and memory (e.g., the hippocampus), loss of muscle and bone mass, and diminished insulin sensitivity, whereas a reduction in the levels of these hormones has been reported to delay or prevent age-related changes in the hippocampus. Whether blood levels of glucocorticoids or alterations in brain response to glucocorticoids are an important part of the normal cognitive decline with age remains to be determined.

Interestingly, there are several reports of a reduced ability of ACTH to stimulate adrenal hormone synthesis and release with age, consistent with the reduction of other adrenal hormones indirectly regulated by ACTH. Aldosterone (involved in sodium regulation) decreases modestly with age, whereas there is a profound decrease in the weak androgen, dehydroepiandrosterone (DHEA). Age-related changes in these latter hormones appear to be the result, at least in part, of a reduced response to ACTH (in the case of aldosterone) and a profound loss of specific cells in the adrenal gland that produce DHEA. The significance of alterations in these hormones is beyond the scope of this article and is detailed elsewhere in this encyclopedia.

## LH/FSH/PROLACTIN AND REPRODUCTIVE FUNCTION

Although it is well recognized that the primary cause of reproductive decline in women is an age-related loss of follicles from the ovary, alterations in hypothalamic and pituitary regulation of reproductive function play an important role in reproductive decline in both genders. Similar changes within neuroendocrine structures responsible for regulating reproductive function have been noted in nonhuman primates and rodents. In both males and females, reproductive function is regulated by the secretion of gonadotropin-releasing hormone (GnRH) from the hypothalamus, which stimulates the release of the gonadotropins, FSH and LH, from the anterior pituitary gland. These hormones are secreted in a pulsatile manner to regulate gonadal function. In the male, FSH regulates spermatogenesis and LH stimulates testosterone production within specific cells of the testis. Sex steroids and other gonadal-derived hormones inhibit the secretion of GnRH from the hypothalamus and LH and FSH release from the pituitary gland. In the female, FSH regulates folliculogenesis and estrogen secretion, whereas LH stimulates luteinization of the mature follicle as well as estrogen and progesterone secretion. The cyclical release of hormones in the female requires a complex interaction between steroid secretion from the ovary, the detection of alterations in hormone levels by the hypothalamus and pituitary, the integration of physiological responses that lead to a midcycle surge of gonadotropins, and the preparation of estrogen-responsive tissues (e.g., oviduct and uterus) for fertilization and implantation.

Because of the complex interrelationships between hormone secretion and physiological responses that are necessary to maintain the female menstrual cycle, it is not surprising that alterations in the regulation of this system are some of the most consistent age-related events noted. Prior to the decline in ovarian follicles that is responsible for the onset of menopause in humans, there are decreases in the release of GnRH from the hypothalamus and impaired sensitivity of the pituitary to GnRH stimulation. These changes undoubtedly contribute to irregular menstrual cycles several years before the onset of menopause and may be a contributing factor in the acceleration of follicular loss in the premenopausal years. The resulting decline in estrogen levels leads to atrophy of estrogen-sensitive tissues and appears to be a risk factor for a number of other diseases (e.g., atherosclerosis and osteoporosis).

In rodents, decreased secretion of GnRH from hypothalamic neurons, rather than loss of follicles, is the primary factor leading to a decline in gonadotropins and the loss of reproductive cycles. Rodents typically proceed from regular cycles to irregular cycles, a state of constant estrus, repetitive pseudopregnancy, followed by an anestrus state later in life. Administration of GnRH and several other factors that act through the hypothalamus reinitiate estrous cycles in aged animals. Although the manifestations of reproductive aging are clearly different in human and rodent models, the neuroendocrine alterations (e.g., GnRH release and pituitary response) appear to share common mechanisms. Recent studies suggest that the absence of GnRH release with age may result from lack of plasticity within the hypothalamus and/or interactions between neurons and glia.

In the male, decreased GnRH secretion also contributes to a decline in gonadotropins and subsequently androgen levels, with a corresponding loss of skeletal muscle mass and reproductive function. Although this decrease appears to be consistent in animal models, the decline in humans exhibits greater variability, with many older men demonstrating relatively minor decreases in total or free androgen levels throughout life. Nevertheless, testosterone replacement in androgen-deficient individuals has been shown to reverse the age-related decline in muscle mass and other aspects of reproductive function.

Prolactin secretion from the anterior pituitary is tonically inhibited by dopamine release from tuberoinfundibular neurons within the hypothalamus. This hormone has a role in the maintenance of luteal function during the menstrual cycle of the female and is responsible for milk production in the postpartum female. With increasing age, the production of dopamine is decreased, resulting in hypertrophy and hyperphasia of prolactin-secreting cells within the pituitary. In rodent models, prolactin levels generally increase in plasma, which may contribute to the genesis of mammary tumors. In many strains, prolactin-secreting pituitary tumors develop that may suppress reproductive function (through actions on GnRH release) and/or interfere with the secretion of other anterior pituitary hormones by mechanical displacement. Although the specific etiology of the decrease in dopamine release with age is poorly understood, data suggest that high levels or prolonged estrogen treatment may damage tuberoinfundibular neurons. In humans, prolactin microadenomas are commonly observed in the elderly at autopsy, but the increase in prolactin levels is not sufficient in many cases to produce significant pathology.

## THYROID-STIMULATING HORMONE, TRIIODOTHYRONINE, THYROXINE, AND THYROID FUNCTION

Metabolism is regulated by plasma levels of triiodothyronine ( $T_3$ ) and thyroxine ( $T_4$ ).  $T_3$  has greater biological activity than  $T_4$ , and tissues express enzymes that convert  $T_4$  to  $T_3$  (e.g., deiodinase). Levels of these hormones regulate the basal metabolic activity and function of numerous tissues.  $T_3$  and  $T_4$  are controlled by feedback regulation of thyroid-stimulating hormone released from the anterior pituitary gland. Apart from specific thyroid diseases, there appear to be subtle changes in the levels of thyroid hormones with age. There appears to be a reduction in deiodinase activity, which may lead to a reduction in levels of  $T_3$  within target organs, but the significance of these changes for the development of physiological changes in tissue function has not been determined.

## POSTERIOR PITUITARY HORMONES

The posterior pituitary hormones, oxytocin and vasopressin (also known as antidiuretic hormone), regulate myometrial and myoepithelial contractions in the female and water balance in both genders, respectively. In response to decreased blood volume/pressure or increased plasma osmolality, vasopressin constricts arterioles and increases water transport in the kidney. Both oxytocin and vasopressin have a diurnal rhythm that diminishes with age, but the significance of this decrease is ill defined. In response to water deprivation, studies suggest that there are no major impairments in either rodent models or humans with age. However, the results of other studies suggest an attenuated response to oropharyngeal factors that regulate vasopressin and/or alterations in kidney response to vasopressin with age. The significance of these changes for the regulation of water balance in the elderly population remains to be determined.

## CONCLUSION

Alterations in the neuroendocrine system contribute to physiological manifestations of aging. Decreases in growth hormone and IGF-1, for example, are important factors in the accumulation of body fat mass, the reduction in cellular protein synthesis, and impairments in some aspects of immune function with age. Similarly, alterations in the neuroendocrine regulation of reproductive function are responsible for cessation of cyclicity in female rodents, whereas in humans the

decline in reproductive function is related primarily to a loss of follicles that is exacerbated by neuroendocrine impairments. In the male, neuroendocrine alterations are a contributing factor to the decline in androgens with age. Although the specific etiology of the age-related neuroendocrine dysfunctions is unknown, increases in oxidative damage in specific hypothalamic nuclei, impairments in neurotransmitter levels and turnover, and diminished plasticity within hypothalamic nuclei are some of the potential mechanisms.

An alternative explanation for the alterations within the neuroendocrine system is that an active regulation of hypothalamic function exists throughout the life span. Decreases in growth hormone and IGF-1, for example, are initiated immediately after puberty and decline progressively with age. In fact, age-related events within this system may be consistent with a model of antagonistic pleiotropy, a recently described theory on the evolution of aging that suggests that the expression of particular genes is beneficial early in life but becomes detrimental as the organism ages. Certainly, the beneficial actions of growth hormone and IGF-1 on muscle mass early in life would be expected to increase general as well as reproductive fitness, whereas continued high levels may increase pathology and limit life span. Thus, the age-related decrease in growth hormone and IGF-1 may be a physiologically regulated, adaptive process rather than the result of damage to specific hypothalamic nuclei. The potential significance of such changes for this and other systems regulated by the neuroendocrine axis awaits further investigations.

## See Also the Following Articles

Aging and Longevity of Human Populations • Aging and the Male Reproductive System • Aging, Animal Models for • Aging, Immunology and • Aging: Muscle • Autonomic Nervous System, Aging and • Body Weight, Body Composition, and Aging • Functional Genomics of Aging • Oxidative Stress and Aging • Stress, Aging, and Central Nervous System Interactions

## Further Reading

- Carter, C. S., Ramsey, M. M., Ingram, R. L., Cashion, A. B., Cefalu, W. T., Wang, Z. Q., and Sonntag, W. E. (2002). Models of growth hormone and IGF-1 deficiency: Application to studies of aging processes and life span determination. *J. Gerontol. Biol. Sci.* 57, B177–B188.
- Griffin, J. E., and Ojeda, S. R. (1996). *Textbook of Endocrine Physiology*, 3rd ed. Oxford Univ. Press, New York.

- Landfield, P. W. (1990). The role of glucocorticoids in brain aging and Alzheimer's disease: An integrative physiological hypothesis. Nathan Shock Memorial Lecture 1990. *Exp. Gerontol.* **29**, 3–11.
- Leitol, H., Behrends, J., and Brabant, G. (2002). The thyroid axis and aging. *Novartis Foundation Symp.* **242**, 193–201.
- Sapolsky, R. M. (1999). Glucocorticoids, stress, and their adverse neurological effects: Relevance to aging. *Exp. Gerontol.* **34**, 721–732.
- Wise, P. M. (2002). Neuroendocrine modulation and repercussions of female reproductive aging. *Recent Prog. Horm. Res.* **57**, 235–256.

**Table I** Diagnostic Criteria for NF1

---

Two or more criteria are necessary for a diagnosis of NF1:

- Six or more café-au-lait spots >5 mm in greatest diameter in prepubertal individuals and >15 mm in greatest diameter in postpubertal individuals
- Two or more neurofibromas of any type or one or more plexiform neurofibromas
- Freckling in the axillary or inguinal region
- Two or more Lisch nodules (iris hamartomas)
- Optic pathway glioma
- A distinctive osseous lesion, such as sphenoid dysplasia or thinning of long bone cortex, with or without bowing and pseudoarthrosis
- A first-degree relative (parent, sibling, or offspring) with NF1 according to the preceding criteria

---

of Schwann cells, perineural fibroblasts, endothelial cells, and mast cells.

### Dermal Neurofibromas

Dermal neurofibromas are cutaneous or subcutaneous tumors that originate from terminal nerve branches in the skin (Fig. 2). Cutaneous neurofibromas are characteristically reddish-bluish, soft, and nearly gelatinous to touch. Subcutaneous neurofibromas are discrete spherical or ovoid shaped and often firm to touch; they are more often tender or painful. Both types generally appear prior to puberty and develop mainly on the trunk. During adulthood, these benign tumors may become a major cosmetic burden. A large study on pregnancy outcome in NF1 patients reported growth of new neurofibromas in 60% of cases and enlargement of existing neurofibromas in 52%.

**Table II** Diagnostic Criteria for NF2

---

Individuals with the following clinical features definitely have NF2:

- Bilateral (VS) *or*
- Family history of NF2 (first-degree relative) *plus*
- Unilateral VS < 30 years *or*
- Any two of the following: meningioma, glioma, schwannoma, juvenile posterior subcapsular lenticular opacities/juvenile cortical cataract

Individuals with the following clinical features possibly have NF2:

- Unilateral VS < 30 years *plus* at least one of the following: meningioma, glioma, schwannoma, juvenile posterior subcapsular lenticular opacities/juvenile cortical cataract
- Multiple meningiomas (two or more) *plus* unilateral VS < 30 years *or* one of the following: glioma, schwannoma, juvenile posterior subcapsular lenticular opacities/juvenile cortical cataract

---

Note. VS, vestibular schwannomas.



**Figure 1** Café-au-lait spots and skin fold freckling in the axilla.

### Plexiform Neurofibromas

Plexiform neurofibromas originate from major peripheral nerves and, as a consequence, may cause severe complications (Fig. 3). They are present in approximately one-third of the NF1 population and are congenital, although they may present later due to growth. The overlying skin may be abnormal, with signs of hypertrophy, hyperpigmentation, or hypertrichosis. Two types of plexiform neurofibromas have been described. Diffuse plexiform neurofibromas are soft subcutaneous swellings with ill-defined margins. Nodular plexiform neurofibromas are ovoid or spherically shaped, feel firm, and are well circumscribed. Extension into adjacent tissues must always be suspected because particularly diffuse plexiform neurofibromas grow with numerous finger-like fronds.

### Lisch Nodules

Lisch nodules are small pigmented benign hamartomas of the iris, presenting at slit lamp presentation. They are seen in 92% of NF1 patients over 6 years of age. As a consequence, these iris hamartomas are of great diagnostic importance for NF1.

### Optic Pathway Glioma

Optic pathway glioma (OPG), or pilocytic astrocytoma of the optic pathway, is an important complication of NF1 during childhood. In general, OPG associated with NF1 is less progressive than OPG without NF1. Symptomatic OPG seldom appears after 6 years of age. Listerick and colleagues reported OPG in 33 of 176 (19%) NF1-affected children after routine screening with magnetic resonance imaging (MRI). However, only half of these children developed glioma-associated signs or symptoms, and

Table III Prevalence Rates of Diagnostic Criteria, Symptoms, and Complications (Percentages)

	Huson <i>et al.</i> , 1989 ( <i>n</i> = 39)	Cnossen <i>et al.</i> , 1998 ( <i>n</i> = 150)	DeBella <i>et al.</i> , 2000 ( <i>n</i> = 1893)
Diagnostic criteria			
Café-au-lait spots	97–100	96.7	99
Freckling	70	85.3	90
Neurofibromas			48
Dermal	0–40	40.0	—
Plexiform	43.6	26.6	—
Lisch nodules	79–85	52.0	>70
Distinctive osseous lesions	—	7.8	14
Symptoms and complications			
Mental retardation	2.9	17.2	—
Optic pathway glioma	5.1	11.3	4.0
Epilepsy	5.1	0.7	—
Severe plexiform neurofibroma	12.8	18.0	—
Bowing	—	2.7	—
Pseudoarthrosis	7.8	2.0	—
Hemihypertrophy	—	2.7	—
Severe scoliosis	2.6	2.0	—
Aqueduct stenosis	2.1	0.7	—
Atlantoaxial dislocation	—	0.7	—
<i>Endocrinological disorder</i>	—	4.7	—
Central precocious puberty	—	2.0	—
Growth hormone deficiency	—	2.0	—
Diencephalic syndrome	—	0.7	—
<i>Malignancy</i>	1.5	4.0 <sup>a</sup>	—
Malignant nerve sheath tumor	—	2.0	—
Malignant brain tumor	—	2.7	—

Note. —, not mentioned.

<sup>a</sup>One patient developed two malignancies.

progression of tumor growth and deteriorating vision were observed in only 3 children (9%).

### Osseous Lesions

Osseous lesions specific for NF1 are sphenoid wing dysplasia and congenital bowing or thinning of long cortical bones with or without pseudoarthrosis. Sphenoid wing dysplasia is rare in NF1 patients but occasionally may present with pulsatile exophthalmia. Congenital bowing is generally located in the tibia or fibula. It is reported in 3% of NF1 patients. However, NF1 has been implicated in more than half of cases of congenital pseudoarthrosis of the tibial bone.

### Minor Disease Features

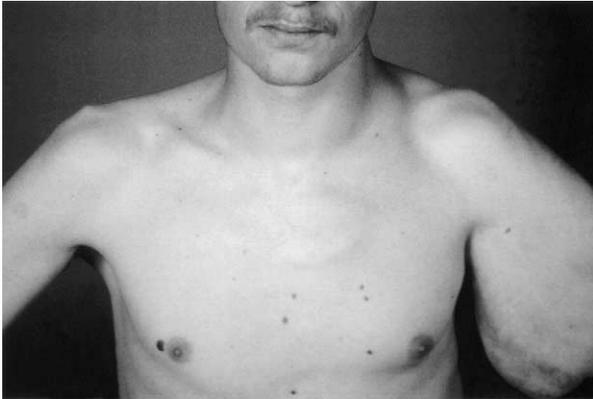
#### **Physical Characteristics**

In addition to the diagnostic signs, physical characteristics such as macrocephaly (53%), hypertelorism

(63%), thorax abnormalities (e.g., pectus excavatum, pectus carinatum) (38%), and short stature (25%) are observed more frequently than in the general



Figure 2 Dermal neurofibromas.



**Figure 3** Plexiform neurofibroma located on the upper left arm.

population. The “facies neurofibromatosis,” as described by Lin, encompasses hypertelorism, a broad nasal bridge or root, pigmentary abnormalities (“dirty skin”), and facial asymmetries. Furthermore, a variety of mild to severe learning difficulties, including motoric speech and behavioral problems such as attention deficit hyperactivity disorder, are common.

#### **Hyperintense Regions on T2-Weighted MRIs**

An additional minor disease feature is the “increased intensity” lesions on T2-weighted MRIs or “unidentified bright objects” that are typically seen in the basal ganglia, thalamus, cerebellum, and brainstem regions of the brain in up to 79% of NF1 patients. Some discussion has evolved to include these typical lesions as a diagnostic criterion. They present as well-circumscribed, hyperintense foci, nonenhancing after administration of contrast medium, without mass effect, and resolving with increasing age (average age of presentation: 7 years). Histopathologically, the hyperintense regions on T2-weighted MRIs correspond to areas of vacuolar or spongiotic change. Correlations have been found between these lesions and learning difficulties. However, studies are not decisive. These lesions should be carefully differentiated from brainstem tumors that cause focal or diffuse brainstem enlargement, present with mass effect, and cause gadolinium enhancement.

## **COMPLICATIONS AND MANAGEMENT**

Complications in NF1 patients may occur in every organ or tract and may cause severe morbidity or even mortality. It does not seem possible to differentiate

between patients at high risk for complications and those at low risk for complications as risk factors for development of complications have not been identified and incidence rates of complications do not differ significantly between groups of children with complications at presentation and those without complications at presentation. Therefore, regular screening by an experienced team (every 1–2 years), even in groups without complications, seems justified.

### **Plexiform Neurofibromas**

Depending on their localization and symptomatology, plexiform neurofibromas may cause considerable morbidity such as functional disturbances, neurological deficits, and cosmetic disfigurement. Moreover, malignant degeneration into a malignant peripheral nerve sheath tumor (neurofibrosarcoma or malignant schwannoma) occurs in 5% of plexiform neurofibromas. Alarming signals include growth and tenderness. In those cases where surgery is indicated, complications as bleeding and recurrence of the tumor are well-known problems. In general, excessive bleeding is caused by vascular abnormalities and the presence of mast cells within plexiform neurofibromas. Total excision of a plexiform neurofibroma is usually impossible due to finger-like fronds that insinuate themselves into adjacent tissues.

### **Optic Pathway Gliomas and Other Ophthalmological Abnormalities**

The NF1 OPG task force recommends annual ophthalmological examinations by an experienced ophthalmologist until 6 years of age, supplemented by visual field evaluations when possible and, most important, visual-evoked potentials (VEPs) given the fact that they monitor the functioning of the optic pathway (Table IV). Evidence of dysfunction of the optic nerve should be followed by MRI of the brain, with contrast enhancement and special attention to the orbits (Table V). Management of children with OPG depends on the localization of the tumor; chiasmatic tumors have been reported to progress more frequently and may be accompanied by hormonal disturbances. If clinical progression is observed, treatment may be indicated and may include chemotherapy or (less commonly) radiation therapy or surgery.

Other ophthalmological problems, such as congenital ptosis, cornea abnormalities, strabismus, and myopia, do not seem specific to NF1.

**Table IV Ophthalmological Screening Protocol According to the NF OPG Task Force**


---

All children under 6 years of age: once yearly complete examination  
 All newly diagnosed NF1 patients: complete examination  
 Follow-up schedule of children over 6 years of age:  
   Ages 8, 13, 20 years: short examination  
   Ages 10, 16, 25 years: complete examination  
 Complete ophthalmological examinations should be performed whenever signs or symptoms suggest an ophthalmological abnormality

---

*Note.* Complete examination = visual acuity, color vision, visual fields (when possible) and visual-evoked potentials, eye movements, papillary light responses, refractive status, funduscopy, (indirect and direct), and slit lamp examination. Short examination = visual acuity, color vision, and slit lamp examination.

### Orthopedic Problems

In addition to the earlier described osseous lesions, scoliosis and hemihypertrophy are associated with NF1 and occur in 2.0 and 2.5% of patients, respectively. Congenital bowing must be monitored by an orthopedic surgeon with experience in NF1 as prevention of bone fractures and pseudoarthrosis is essential. Atlantoaxial dislocation in patients with abnormalities of the cervical spine is a severe complication in NF1 patients undergoing situations associated with traction and hyperextension of the spine such as intubation during general anesthesia. Bidirectional cervical vertebral column X rays, and supplementary flexion–extension imaging in case of abnormalities, is recommended.

### Neurological Symptoms

Mental retardation has been reported in 3 to 15% of NF1 patients. Many of these studies are biased in that they consist of a high frequency of severely affected patients. A slightly increased risk of epilepsy, which is

**Table V Ophthalmological Follow-up Protocol for Children with OPG According to the NF1 OPG Task Force**

Time interval following diagnosis	Complete ophthalmological examination	MRI
First year	Every 3 months	Every 6 months
Second year	Every 6 months	Every 6 months
Thereafter	Yearly	Yearly

*Note.* If there is no evidence of progression 2 years after diagnosis, the interval may be increased gradually.

only partly accounted for by macroscopic abnormalities, has been reported. Also, it has been suggested that headache of various types (e.g., tension headaches, migraines) may be more common in NF1 patients than in the general population. Although high prevalences of psychiatric illness have been reported, these have not been confirmed by others.

### Endocrinological Abnormalities, Short Stature, and Hypertension

Precocious puberty, delayed puberty, and growth hormone (GH) deficiency are more frequent in children with NF1 than in the general population. Therefore, deviations from the growth curve and premature manifestations of puberty should be recognized as possible complications of NF1. Although central precocious puberty (CPP) in NF1 patients may be associated with OPG, several studies have reported children with NF1 and CPP or GH deficiency without OPG. Possible mechanisms may be slow-growing hamartomas, undetectable at the neuroimaging level, or abnormalities at the cellular level given that the protein encoded by the NF1 gene, neurofibromin, plays a regulating role in signal transduction pathways. No adverse side effects of treatment with luteinizing hormone-releasing hormone (LHRH) analogues or recombinant human GH have been reported, although continuous monitoring is recommended.

Short stature is observed in 25% of NF1 patients and is often without endocrinological basis. Hypertension in NF1 patients may be associated with either renal artery stenosis or pheochromocytoma. In children with NF1, pheochromocytoma is quite rare. However, 5 to 25% of adult patients with pheochromocytoma are reported to have NF1.

### Malignancies

A descriptive analysis by Friedman and Birch of nearly 1800 NF1 patients observed malignancies (excluding OPG) in approximately 7.0% of probands and 4.5% of affected relatives. In addition to malignant peripheral nerve sheath tumors that may arise in preexisting plexiform neurofibromas, astrocytomas, juvenile chronic myeloid leukemia, rhabdomyosarcoma, adenocarcinoma of the ampulla Vater, and duodenal somatostatinoma have been reported to be associated with NF1. Moreover, secondary tumors develop in 21% of NF1 patients with primary malignancies in comparison with 4% of the general population. Brainstem tumors have been described as being less

aggressive in NF1 patients than in those not affected by NF1.

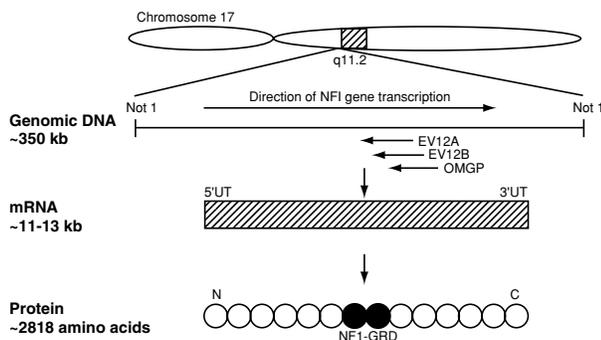
**MOLECULAR GENETICS**

**Gene, Protein, and Mutation Analysis**

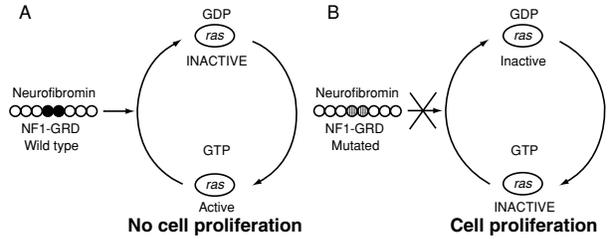
In 1990, the NF1 gene was mapped to chromosome 17q11.2. The gene spans a region of about 350 kb of genomic DNA, contains approximately 60 exons, and encodes an 11- to 13-kb mRNA that is markedly expressed in neurons, oligodendrocytes, and nonmyelinating Schwann cells. Three genes are embedded within intron 27: EV12A, EV12B, and OMGP. Interestingly, these genes are transcribed in the opposite direction of the NF1 gene (Fig. 4).

The protein encoded by the NF1 gene, neurofibromin, consists of 2818 amino acids with a predicted molecular mass of 327 kDa. Various studies report that neurofibromin is involved in microtubule-mediated signal transduction pathways. In addition, a certain region of the predicted protein product shows homology to the GTPase-activating protein (GAP) family in yeast and mammals. The GAP-related domain (NF1-GRD) of neurofibromin is the only known functional domain of the NF1 gene and plays a role in the regulation of cellular proliferation and differentiation through hydrolysis of active RAS-GTP into inactive RAS-GDP (Fig. 5).

Mutations in the NF1 gene are spread throughout and are of different types and sizes. Most unrelated patients have their own unique mutation. Genotype-phenotype relationships have not been found with the exception of correlations between large deletions, encompassing the entire NF1 gene, and a severe phenotype with mental retardation and possibly early presentation of dermal neurofibromas. Modifying



**Figure 4** The NF1 gene location and size, with embedded genes. Illustration shows the NF1 gene transcript and protein with only known functional domain, NF1-GRD.



**Figure 5** (A) Hydrolysis of active RAS-GTP into inactive RAS-GDP by neurofibromin. Cell proliferation is regulated. (B) Loss of function of neurofibromin, leading to high levels of active RAS-GTP. Cell proliferation is no longer regulated.

genes may play a role in the extreme clinical variability of the disease. Candidate genes have not been identified.

**Mutation Rate and Mosaicism**

The high mutation rate of  $1 \times 10^{-4}$  per gamete per generation of the NF1 gene is impressive and cannot be explained by the large size of the gene alone. Various biological mechanisms may be responsible, including the presence of many NF1 pseudogenes in the genome, a high frequency of methylatable CpG residues within the NF1 gene that are prone to mutations, and particular sequence patterns predisposing toward insertions and/or deletions.

Genetic mosaicism may occur more often in NF1 than in other genetic diseases due to the high mutation rate. Genetic mosaicism is a condition in which genetically distinct cell populations are observed within one individual. Mutations during early embryonic development, before the determination of the germ line, will cause gonosomal mosaicism (affecting both somatic tissues and the germ line). Mutations occurring later may affect either the germ cells alone (germ line mosaicism) or the somatic cells alone (somatic mosaicism). This phenomenon has important implications for mutation analysis and genetic counseling in NF1 because risks for affected offspring are higher when germ line cells are affected.

**Tumor Suppressor Gene**

Characteristic of an autosomal dominant disorder (mosaics are not taken into account) is that all body cells have one wild-type (normal) allele and one mutated allele of the disease gene. If the gene involved is a tumor suppressor gene, tumor formation takes place when a mutation or “second hit” occurs in the wild-type allele. This leads to complete lack of functional

protein and loss of “tumor suppression.” Such is the case in NF1; absence of neurofibromin leads to defective regulation of proliferation and differentiation of the cell, causing cellular proliferation or tumor growth. A number of symptoms and complications in NF1 (e.g., neurofibromas, malignancies) seem to follow this second hit model by Knudson. This phenomenon has been confirmed at the molecular genetic level. However, other symptoms in NF1 patients, such as learning difficulties, cannot be explained by this model and are more likely due to haploinsufficiency of the gene.

## GENETIC COUNSELING

Consultation of a clinical genetics center is highly recommended as information on the inheritance of NF1, its natural history, and the possibilities and limitations of molecular genetic analysis and prenatal diagnosis is essential for patients and family members. Prenatal diagnosis may be offered to NF1 families in which a gene mutation has been detected or the tested intragenic polymorphic markers have proved to be informative (the mutated allele can be followed within an NF1-affected family). Because the probability of genetic mosaics within the NF1 population is high, prenatal diagnosis should be considered and offered to the parents of a *de novo* NF1 patient with an identified mutation. In practice, parents at risk for NF1 in their offspring rarely request prenatal testing. Presumably, the extreme clinical variability of the disorder and the low mortality of the disease, in combination with the fact that mainly familial NF1 patients are eligible for

prenatal testing, leads to a low uptake of prenatal testing.

## See Also the Following Articles

Adrenal Tumors, Molecular Pathogenesis • Hypertension, Endocrine

## Further Reading

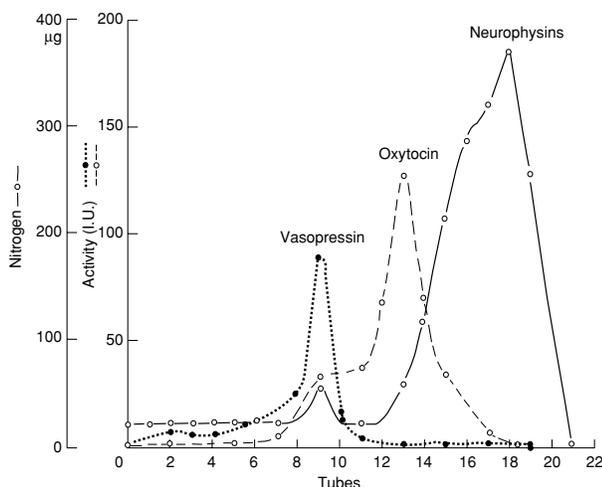
- Cnossen, M. H., De Goede-Bolder, A., van den Broek, K. M., *et al.* (1998). A prospective 10 year follow-up study of patients with neurofibromatosis type 1. *Arch. Dis. Child.* **78**, 408–412.
- Cnossen, M. H., Stam, E. N., Cooman, C. M. G., *et al.* (1997). Endocrinological disorders and optic pathway glioma in children with neurofibromatosis type 1. *Pediatrics* **100**, 667–670.
- DeBella, K., Szudek, J., and Friedman, J. M. (2000). Use of the National Institutes of Health criteria for diagnosis of neurofibromatosis type 1 in children. *Pediatrics* **105**, 608–614.
- Friedman, J. M., and Birch, P. H. (1997). Type 1 neurofibromatosis: A descriptive analysis of the disorder in 1,728 patients. *Am. J. Med. Genet.* **70**, 138–143.
- Huson, S. M., Compston, D. A. S., and Harper, P. S. (1989). A genetic study of von Recklinghausen neurofibromatosis in South East Wales: II. Guidelines for genetic counselling. *J. Med. Genet.* **26**, 712–721.
- Listernick, R., Louis, D. N., Packer, R. J., *et al.* (1997). Optic pathway gliomas (OPG) in children with neurofibromatosis type 1 (NF1): Consensus statement from the NF1 OPG task force. *Ann. Neurol.* **41**, 143–149.
- Messiaen, L. M., Callens, T., Mortier, G., *et al.* (2000). Exhaustive mutation analysis of the NF1 gene allows identification of 95% of mutations and reveals a high frequency of unusual splicing defects. *Hum. Mutat.* **15**, 541–555.
- Sorensen, S. A., Mulvihill, J. J., and Nielsen, A. (1986). A long term follow-up of von Recklinghausen neurofibromatosis: Survival and malignant neoplasms. *N. Engl. J. Med.* **314**, 1010–1015.

Table I Approximate Number of Individual Proteins Involved in the Building Up of the Neuroendocrine Osmoregulatory Reflex

<b>A. Afferent Neural Branch</b>			
1. Sensory Neurons			
Baroreceptors	DEG / MEC	4	
High pressure	Carotid tissue, aortic arc, and jugulary artery	3	
Low pressure	Atrial	3	
Osmoreceptors	Brain receptors	5	
Subtotal			15
2. Interneurons			
Brain nuclei/Hypothelamic neurons		20	20
3. Secretory Neurons			
Signaling	Cation channels	5	
Transduction	Kinases	5	
Transcription	Transcription factors Nucleopores, gene activators, and RNA polymerases	20	
Translation ER	Protein-conducting channels	20	
	Quality control	10	
	Golgi shuttle vesicles	10	
	Golgi processing enzymes	5	
Secretory granules	Kinesin motors	5	
	Actin microtubules	5	
	Targeting proteins	5	
	Fusion proteins	5	
			95
<b>Total</b>			<b>130</b>
<b>B. Efferent Endocrine Branch</b>			
1. Exocytosis and Blood Transport			
Neurohormone secretion		5	
Neurohormone carriers		5	
Subtotal			10
2. Hydroosmotic Target Cells			
Receptors and G proteins		10	
AC, PKA, AKAP, and phospholipase C		10	
Nucleoporins, gene activation, transcription factors, and polymerases		20	
Translation, translocation ER, Golgi shuttle vesicles, quality control, proteasomes, and processing enzymes		25	
Vesicle targeting, fusion proteins, kinesin motors, and microtubules, actins		20	
Subtotal			85
3. Effectors in Apical Membrane			
Aquaporins		5	
Urea transporters (UT1A, UT1B, UT2)		5	
Sodium channel ENaC		5	
Potassium channels ROM-K1		5	
Chloride channels CFTR		5	
Subtotal			25
<b>Total</b>			<b>120</b>
<b>Grand Total</b>			<b>250</b>

activities. Dissociation could also be performed by acetic acid dialysis, trichloroacetic acid precipitation, and countercurrent distribution (Fig. 1). From the cathodic fraction obtained by electro dialysis, two

peptides migrating like pure oxytocin and pure vasopressin and having their respective amino acid compositions and activities were isolated by paper zone electrophoresis. On the other hand, the inactive



**Figure 1** Dissociation of the “mother molecule,” a noncovalent complex, into vasopressin, oxytocin, and neurophysins by counter-current distribution.

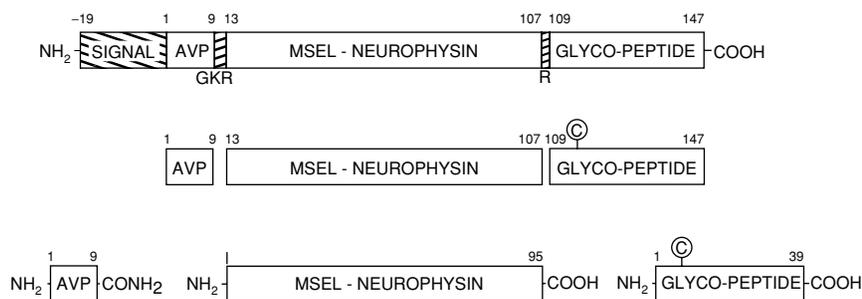
protein, for which we suggested the name neurophysin, was later shown to include two highly homologous components termed MSEL-neurophysin (M, methionine-2; S, serine-3; E, glutamic acid-6; L, leucine-7) and VLDV-neurophysin (V, valine-2; L, leucine-3; D, aspartic acid-6; V, valine-7), according to the residues in positions 2, 3, 6, and 7.

The complex was purified from posterior pituitary glands of ox, sheep, pig, horse, rat, guinea pig, and whale and was revealed to be a four-component assembly containing oxytocin, vasopressin, MSEL-neurophysin, and VLDV-neurophysin. Because of the rough stoichiometry between vasopressin and MSEL-neurophysin, on the one hand, and oxytocin and VLDV-neurophysin, on the other, the hypothesis that each tandem hormone-neurophysin derived from a common precursor was suggested.

### The Four-Component Neurohormone-Neurophysin Complex

The first direct data on the common neurohormone-neurophysin precursors were obtained by Sachs and collaborators, who injected [ $^{35}$ S]cysteine into the third ventricle of dogs and removed the hypothalamus 1.5 h later. One-half of the tissue, examined right away, contained proteins but practically no vasopressin, whereas the other half, incubated for 4.5 h in a medium containing puromycin and unlabeled cysteine, yielded a significant amount of a labeled vasopressin. Sachs was able to identify a neurophysin in the dog hypothalamus and to show that biosynthesis and secretion of vasopressin and of this neurophysin are parallel events. Later, Brownstein and colleagues showed that [ $^{35}$ S]cysteine injected into the supraoptic nucleus of the rat is incorporated into 20,000-Da proteins (pI 5.4 and 6.1) that are in time converted into 12,000-Da-labeled proteins reacting with rat neurophysin antiserum. The vasopressin precursor was found to be glycosylated, whereas the oxytocin precursor was not. On the other hand, a glycosylated 39-residue peptide found in the human neural lobe in proportions approximately equivalent to those of neurophysins was suspected of being a constituent of the vasopressin precursor. Finally, in 1982–1983, the amino acid sequences of preprovasopressin and prooxytocin were deduced through cDNA technology by Richter and co-workers. Aside from a signal peptide, provasopressin is organized into three domains and is comprised of vasopressin linked by a processing sequence Gly-Lys-Arg to MSEL-neurophysin, itself linked by an arginine residue to a glycopeptide or copeptin (Fig. 2). Prooxytocin has two domains and is comprised of oxytocin linked by the tripeptide Gly-Lys-Arg to VLDV-neurophysin.

Thus, in approximately a century, a neurohypophysial multiactive compound first was supposed to be a



**Figure 2** Domain organization and processing of bovine preprovasopressin. Signal, signal peptide; AVP, arginine vasopressin; C, carbohydrate. GKR and R are the enzymatic processing sites.

single protein, the “mother molecule; then was dissociated into four components, two active peptides (oxytocin and vasopressin) and two inactive protein carriers (neurophysins); and finally was shown to originate from two neurohormone–neurophysin tandem precursors. Progressive advances in techniques led to elucidation of the biological status of oxytocin and vasopressin. However, it remains to be explained why these hormones are stored in the glands of most mammals in approximately equal molar amounts despite their very distinct physiological functions. The finding that the human vasopressin gene is closely linked to the oxytocin gene on the same chromosome, the two genes being transcribed on opposite DNA strands, might suggest some relationship in their expressions.

### The Two Precursor Genes

The genes of mammalian neurohypophysial hormones, namely those of oxytocin and vasopressin of man, rat, ox, and mice, have been identified by Sausley, Richter, Schutz, and Gainer, respectively, along with their collaborators. They are built in the same way, namely, with the same three-exon organization, with the first exon encoding the signal peptide, the hormonal nonapeptide, the processing tripeptide sequence, and the first 9 residues of neurophysin; the second exon encoding the central part of neurophysin (residues 10–76); and the third exon encoding the C-terminal part of neurophysin (residues 77–93/95) and, for vasopressin genes, an arginine processing site linking to a 37/39-residue glycopeptide or copeptin. The gene product, preprovasopressin, is a four-domain protein (Fig. 2), and this organization is preserved in the nonmammalian tetrapod preprovasotocin. On the other hand, preprooxytocin and its nonmammalian tetrapod counterpart, prepromesotocin, are three-domain proteins, with copeptin extension being absent.

The promoters and the regulatory elements of these genes, as well as their expressions in the central nervous system and the peripheral tissues, have been investigated intensively. Because of a striking similarity in the nucleotide sequences of the second exons of oxytocin and vasopressin genes in all mammalian species investigated, it is assumed that recent gene conversion events occurred. In contrast, the third exons are clearly different. The three exons, and therefore the corresponding encoded peptide sequences, have likely undergone distinct evolutionary fates (composite evolution). Furthermore, mutations in the regions coding for hormones in the first exons should have been limited by the need of the mature peptides to fit specific receptors, whereas the constraints on

neurophysins, if any, seem to be determined by particular interactions, either with endoplasmic reticulum (ER) chaperones or with specific trans-Golgi sequences.

Neurohypophysial hormone genes are also known in lower vertebrates. In nonmammalian tetrapods, mesotocin and vasotocin replace oxytocin and vasopressin, respectively. The chicken vasotocin gene has been identified by Ivell's team; its organization is comparable to that of mammalian vasopressin genes. In the bony fishes, isotocin and vasotocin correspond to mammalian oxytocin and vasopressin. Isotocin and vasotocin genes of a tetraploid catostomid fish, the white sucker *Catostomus commersoni*, and of the tetraodontoid Japanese pufferfish, *Fugu rubripes*, have been characterized by Richter's and Venkatesh's groups, respectively. Whereas in the first species the two vasotocin genes appear to be built essentially like the mammalian neurohypophysial genes (i.e., with a three-exon organization), the two isotocin genes are revealed to be intronless. It is still premature to integrate this unusual feature and to deduce that evolution operated by insertion of introns. The Fugu is remarkable because its genome has been subjected to compaction, with a large part of the usual intergene sequence being deleted and very few repetitive sequences being present. With 400 Mb, the Fugu genome is approximately 7.5 times smaller than the human genome, but it has a gene repertoire similar to that of the human. The two genomes apparently are syntenic. Whereas in mammals the vasopressin and oxytocin genes are closely linked in a tail-to-tail orientation, Fugu vasotocin and isotocin genes are linked head-to-tail and are separated by five genes. A possible peculiar evolution of fish neurohypophysial hormone genes, derived from an ancestral vasotocin gene through an inverted duplication and differentiation of the two copies into isotocin and vasotocin genes and then a fragment of DNA containing the vasotocin gene, has undergone inversion in the *Fugu* lineage. Interestingly, Venkatesh's experiments with transgenic rats reveal functional conservation of regulatory controls between the Fugu isotocin and rat oxytocin genes. Finally, in the vasotocin gene of the Pacific hagfish *Eptatretus stouti*, belonging to the most primitive vertebrate class, Cyclostomes, not only are the two usual introns found at positions similar to those in the mammalian vasopressin and oxytocin genes, but also a third one is detected interrupting the 5'-untranslated region.

In summary, the general organization of neurohypophysial precursor genes is remarkably preserved during evolution, with some punctual secondary

variations in a few species. The discovery of “neurohypophysial-like” precursors in invertebrates displaying the four-domain structure suggests that this type of protein antedates the vertebrate–invertebrate divergence (some 700 million years ago) and so is very ancient. The physiological functions that these molecules could fulfill in the primitive organisms are still unknown, but they should have been involved in the early evolutionary shaping of the neurohypophysial-like peptide conformations.

## Structures and Conformations of Neurohypophysial Peptides

### **Precursor Self-Folding and Domain Organization**

The precursor polypeptide chain, emerging from the ribosome, is translocated into the rough ER through the signal recognition particle (SRP) and the SRP receptor. Whereas the signal prepeptide is removed by a signalase, the translocated polypeptide folds in the lumen to take its native conformation. The right folding is subjected to quality control in the ER and may be helped by a chaperone. This folding allows the 16 cysteine residues of the precursor to pair in eight precise disulfide bridges. In both vasopressin and oxytocin precursors, there is one disulfide in the hormonal moiety and seven disulfides in the neurophysin moiety. It can be stated that the hormonal nonapeptide domain folds first and that the sequence connecting the hormone to the neurophysin moiety adopts a conformation allowing the accessibility for the specific processing endopeptidase. The crucial processing site is a Gly–Lys–Arg sequence (Fig. 2). The 93/95-residue neurophysin domain is likely folded in a second step. In the vasopressin or vasotocin precursor, there is a third additional domain, a 39-residue glycopeptide or copeptin, devoid of disulfide bridges and cut off, in the case of provasopressin only, during the processing at a monobasic site, an arginine residue (Fig. 2). Prooxytocin has only two domains, oxytocin and VLDV–neurophysin, but the maturation of the hormone moiety is virtually identical to that of vasopressin. However the fish proisotocin owns an additional C-terminal domain similar to the one found in provasopressin or provasotocin.

### **Precursor Processing and Mature Conformations**

#### *The Processing Enzyme Cascade*

Most of hormonal peptides have a C-terminal amidated residue, and this feature—essential for biological activities—is the result of a four-enzyme

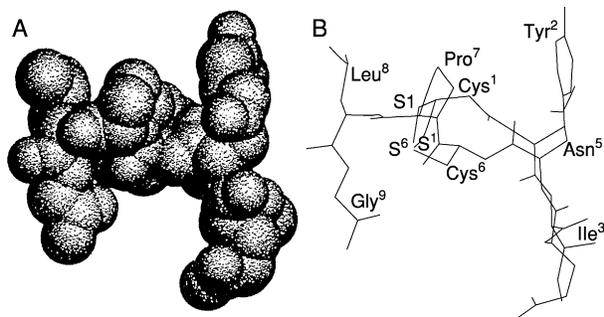
processing machinery very likely operating in the last compartment of the secretory pathway, the secretory granules. Sequentially, the four enzymes are (1) a dibasic endopeptidase, Ca<sup>2+</sup> dependent, often termed PC1/PC3 and PC2 (prohormone convertases); (2) a carboxypeptidase B-like enzyme termed carboxypeptidase E or H, which removes the two basic residues; (3) a peptidylglycine mono-oxygenase giving a peptidylhydroxyglycine; and (4) an alpha-amidating lyase that splits the hydroxyglycine into the amide group and glyoxylic acid. The last two enzymes share a common precursor and may operate like a two-domain single protein (PAME), as shown by Eipper and Mains. The four enzymes are likely granule membrane bound and work as machinery. They have pH optima between 5 and 6, and we have identified them in the neurohypophysial secretory granules of the ox and rat. The pH in these granules is 5.5 to 5.8. With enzymes purified from secretory granules, we have reconstituted the processing *in vitro* using vasopressinyl–Gly–Lys–Arg converted into vasopressin, identified by chromatography with the synthetic peptide and its specific rat pressor activity. It can be assumed that all neurohypophysial amidated nonapeptides are produced from protein precursors through this multienzyme mechanism.

In the case of mammalian provasopressin, a second cleavage (specific monobasic endopeptidase) occurs in the precursor, separating the neurophysin domain from the copeptin domain. Release of mature mediators in circulating blood involves regulated fusion mechanisms of secretory vesicles with the plasma membrane and subsequent exocytosis.

#### *The Conformation-Shaping Mechanisms*

The processing machinery not only tailors the proper size of the peptide but also secures the right conformation necessary for recognizing and binding to the specific receptor. After being detached from the precursor by the processing endopeptidase, the intermediate dodecapeptide, and particularly its terminal tail, self-folds in a way that is different from the one it exhibits within the prohormone. Its conformation becomes flexible. Removal of the last three residues and building of the C-terminal carboxamide group of the hormone through the alpha-amidating lyase are crucial for shaping the active conformation able to bind the specific receptor (Fig. 3). Alpha-deamidation entails complete loss of the biological activity.

In anuran Amphibia, an adaptive differential processing of provasotocin involving a down-regulation of either carboxypeptidase E, leading to vasotocinyl–Gly–Lys–Arg (hydrin 1) in *Xenopus laevis*, or the



**Figure 3** Three-dimensional structure of deamino-oxytocin: (A) the space filling model and (B) the equivalent stick model. Reprinted with permission from Wood, S. P., *et al.* (1986). Crystal structure analysis of deamino-oxytocin conformational flexibility and receptor binding. *Science* 232, 633-636. Copyright 1986, American Association for the Advancement of Science.

alpha-amidating enzymes, giving vasotocinyl-Gly (hydrin 2) in frogs and toads, has been discovered in our laboratory. These intermediates have conformations different from that of mature vasotocin, not only because of the hindrance due to the extension

but also because the C-terminal negative carboxyl induces a different folding. They display hydro-osmotic activities on frog urinary bladder and skin but not on the kidney, in contrast to vasotocin, and may act through distinct receptors, playing an essential role in the rehydration of the animal.

**Neurohypophysial Hormones**

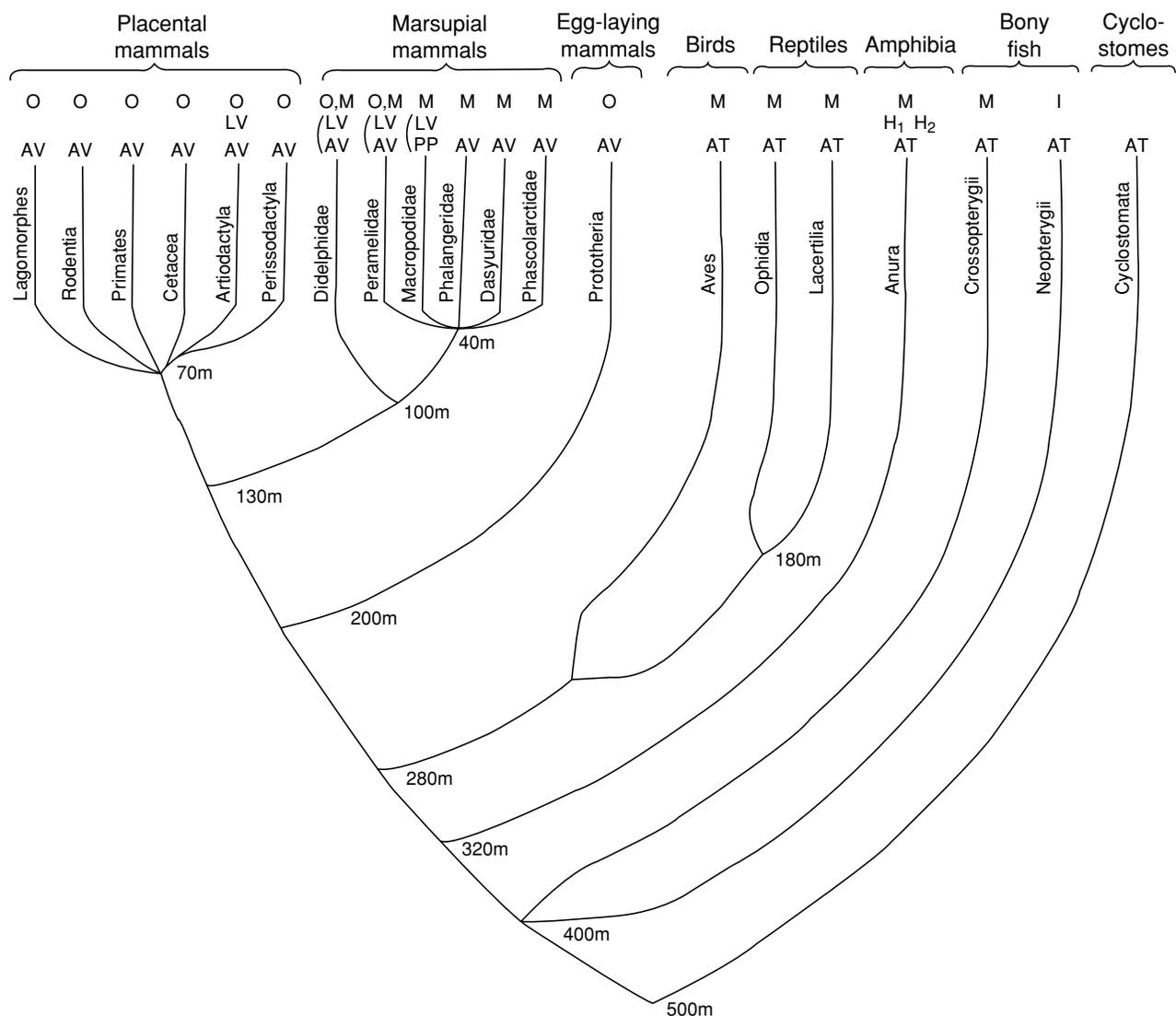
*The Vertebrate Neurohypophysial Hormones*

Since the chemical characterization in 1953 of bovine oxytocin and arginine vasopressin by Du Vigneaud, Tuppy, and our laboratory, we have isolated 10 distinct neurohypophysial peptides from marsupials and nonmammalian vertebrates (Table II). About 80 species belonging to the seven vertebrate classes have been investigated. There is a great evolutionary stability in the structures given that species within the same class of vertebrates, with the exception of cartilaginous fishes, almost always possess the same hormones. Furthermore, two peptides, one oxytocin-like and the other vasopressin-like, are usually found in

**Table II** Structures of Vertebrate Neurohypophysial Peptides

		Origin
<b>Oxytocin-like Peptides</b>		
	1 2 3 4 5 6 7 8 9	
Oxytocin	Cys-Tyr-Ile-Gln-Asn-Cys-Pro-Leu-Gly(NH <sub>2</sub> )	Placentals Some marsupials Ratfish
Mesotocin	----- Ile ---	Nonmammalian tetrapods Marsupials Lungfishes
Seritocin	----- Ser ----- Ile ---	<i>B. regularis</i>
Isotocin	----- Ser ----- Ile ---	Bony fishes
Glunitocin	----- Ser ----- Gln ---	Rays
Valitocin	----- Val ---	} Spiny dogfish
Aspargtocin	----- Asn -----	
Asvatocin	----- Asn ----- Val ---	} Spotted dogfish
Phasvatocin	--- Phe --- Asn ----- Val ---	
<b>Vasopressin-like Peptides</b>		
	1 2 3 4 5 6 7 8 9	
Vasopressin	Cys-Tyr-Phe-Gln-Asn-Cys-Pro-Arg-Gly(NH <sub>2</sub> )	Mammals
Lysipressin	----- Lys ---	Pig Macropodids Didelphids Peramelids
Phenypressin	--- Phe -----	Macropodids
Vasotocin	----- Ile -----	Nonmammalian vertebrates

Note. Residues identical to those of oxytocin or vasopressin are indicated by dashes.



**Figure 4** The two evolutionary lineages of vertebrate neurohypophysial hormones. The peptides identified in cartilaginous fishes are not included because these fishes are not located in the evolutionary pathway from bony fishes to mammals and were subjected to a separate evolution. O, oxytocin; AV, arginine vasopressin; LV, lysine vasopressin; M, mesotocin; PP, phenylpressin; AT, arginine vasotocin; I, isotocin; H<sub>1</sub>, hydrin 1; H<sub>2</sub>, hydrin 2. The line connecting LV and AV or connecting LV and PP means that both peptides exist in the same species. The numbers give the time in million years (m) since the divergence.

each species, so that two paralogue lineages can be traced from Cyclostomes to mammals (Fig. 4). The vasopressin-like hormones are characterized by a basic residue, arginine or lysine, in position 8, a key position for receptor recognition, whereas oxytocin-like peptides almost always have a neutral aliphatic hydrophobic residue in this position.

The preservation of the hexacyclonapeptide pattern with invariant residues in positions 1, 6, 7, and 9 and C-terminal amidated glycine strongly suggests, on the one hand, that the general conformation of

the molecule (Fig. 3) has been largely conserved in the course of evolution and, on the other hand, that the two lineages likely originated from a primordial duplication of a common ancestral gene. This latter assumption is reinforced by the fact that the most primitive vertebrates, the Cyclostomes, have a single neurohypophysial peptide, namely vasotocin. Table II gives the names, structures, and origins of the 13 vertebrate neurohypophysial peptides chemically identified. Whereas mesotocin ([Ile<sup>8</sup>]-oxytocin) has been chemically identified in two species of

African lungfish (*Protopterus annectens* and *Protopterus dolloi*) and in two individuals of the Australian lungfish (*Neoceratodus forsteri*), Hyodo and colleagues have deduced the occurrence of a variant, Phe2–mesotocin, from a cDNA of an Australian lungfish.

Most species have the two types of neurohypophysial hormones, but secondary duplications of either or both can be observed in some groups. In marsupial mammals, duplication of the vasopressin-like peptide is found in Australian macropodids (lysiipressin and phenyppressin), in Australian peramelids (lysiipressin and vasopressin), and in American didelphids (lysiipressin and vasopressin). In the same way, duplication of the oxytocin-like peptide, namely oxytocin and mesotocin, is observed in Australian peramelids and in American didelphids. Occurrence of additional neurohypophysial hormone duplications in both American and Australian marsupials can be explained by common ancestors living at the time of the association of South America and Australia in Gondwanaland and their later separation through the continental drift. Curiously, the nonmammalian vasotocin has been replaced completely by the vasopressin-like hormone in Metatherians, whereas the nonmammalian mesotocin is still present. In the cartilaginous fishes, which form a heterogeneous class, duplication of the oxytocin-like peptide has also been detected in sharks, namely valitocin and aspartocin in the spiny dogfish and asvatocin and phasvatocin in the spotted dogfish (Table II). It is likely that secondary duplications of the genes have led to duplications of the protein precursors and finally to duplications of the processed hormones. Apparently, these secondary duplications were random and restricted to limited groups. It will be of great interest, when the genomes are sequenced, to check whether “silent” duplications with unexpressed genes exist in the other classes of vertebrates.

When substitutions in the oxytocin/vasopressin superfamily are examined (Table II), it is clear that whereas some positions are invariant (e.g., positions 1, 6, 7, and 9), others are particularly affected (e.g., positions 4 and 8). Substitutions in position 4 are always polar, and those in position 8 are virtually always aliphatic hydrophobic in the oxytocin family and always basic in the vasopressin family. Position 3 is occupied by an aliphatic hydrophobe (isoleucine) in the oxytocin family and by an aromatic hydrophobe (phenylalanine) in the vasopressin family except in vasotocin. Position 2 is usually occupied by tyrosine, but occasionally phenylalanine replaces tyrosine (phasvatocin and phenyppressin).

The biological significance of the substitutions is not clearly understood except for the replacement of

vasotocin by vasopressin in mammals (replacement of Ile3 by Phe3) that leads to a peptide that acts only on kidney and vascular receptors and no longer on uterus and mammary gland receptors, as does nonmammalian vasotocin. On the other hand, replacement of mesotocin by oxytocin in placental mammals (replacement of Ile8 by Leu8) could be neutral given that the two peptides have roughly equal uterotonic and milk-ejecting activities in rat and rabbit, respectively, and given that mesotocin was conserved in marsupial mammals with the same function as oxytocin in placentals.

The finding of oxytocin in the ratfish, a holocephalian fish (cartilaginous fishes), is very puzzling. Cartilaginous fishes display a great diversity of oxytocin-like peptides because rays, sharks, and chimera have different peptides. It has been suggested that this apparent freedom for mutations might be related to the fact that these fishes regulate their osmoregulation with urea rather than with salts. Simultaneously, the amount of vasotocin stored in neurohypophysis is dramatically reduced when compared with that in other nonmammalian vertebrates.

**The Invertebrate Neurohypophysial Hormone-like Peptides**  
“Neurohypophysial hormone-like” peptides have been discovered in various tissues of some invertebrates: locupressin in the brain of *Locusta migratoria* (insects); Lys–conopressin and Arg–conopressin in the venom of two marine *Conus* species, the freshwater snail *Lymnaea stagnalis* and the seawater snail *Aplysia kurodai* (gastropod molluscs); cephalotocin in the “neurosecretory system of the vena cava” of *Octopus vulgaris* (cephalopod molluscs); and annetocin in the body of the earthworm *Eisenia foetida* (Table III). When present, there is always a single neurohypophysial hormone-like peptide in invertebrate species, in contrast to two neurohypophysial hormone-like peptides in virtually all vertebrates. The possible physiological functions of these peptides are poorly known, except for locupressin, the dimer of which would be a diuretic hormone.

Whatever their functions, the occurrence of neurohypophysial hormone-like peptides in invertebrates indicates that the ancestral gene antedates the single preprovasotocin gene of Cyclostomes and was present before the invertebrate–vertebrate divergence at least some 700 million years ago.

### The Neurophysins

#### Mammalian Neurophysins

Neurophysins are small proteins initially found to be stoichiometrically associated with neurohypophysial hormones in noncovalent complexes. Mammalian

Table III Structures of Invertebrate Neurohypophysial Hormone-like Peptides

		Origin
Vertebrates		
	1 2 3 4 5 6 7 8 9	
Vasopressin	Cys-Tyr-Phe-Gln-Asn-Cys-Pro-Arg-Gly(NH <sub>2</sub> )	Mammals
Lysipressin	-----Lys---	Pigs
		Some marsupials
Vasotocin	-----Ile-----	Nonmammalian-vertebrates
Oxytocin	-----Ile-----Leu---	Placentals
Insects		
Locupressin	---Leu-Ile-Thr-----	<i>L. migratoria</i>
Molluscs		
Arg-conopressin	---Ile-Ile-Arg-----	<i>C. geographus</i>
Lys-conopressin	---Phe-Ile-Arg-----Lys---	<i>C. striatus</i>
		<i>L. stagnalis</i>
		<i>A. kurodai</i>
Cephalotocin	-----Arg-----Ile---	<i>O. vulgaris</i>
Annelids		
Annetocin	---Phe-Val-Arg-----Thr---	<i>E. foetida</i>

Note. Residues identical to those of vasopressin are indicated by dashes.

neurophysins are 93/95 residues long and contain seven disulfide bridges, so that a reticulated conformation can be suspected. When neurophysin amino acid sequences from several mammalian species are compared (Table IV), it is clear that three segments have different variabilities. The N-terminal part (residues 1–9), which permits the two types to be distinguished, is rather constant in each family; the central part (residues 10–76) is strongly preserved within the family and between the two families; and the C-terminal part (residues 77–93/95) is subject to many variations. The near identity between the central parts of MSEL- and VLDV-neurophysins from man, ox, and rat has been explained by gene conversion events, with the oxytocin and vasopressin genes being closely linked on the same chromosome and transcribed on opposite DNA strands.

In the case of provasopressin, the precursor has a third domain, a 39-residue C-terminal glycopeptide extension termed copeptin, which is cut off from MSEL-neurophysin at the level of an arginine residue (Fig. 2). This glycopeptide has no disulfide bridge and may be proteolytically truncated during the extraction. It is also encoded by the third exon of the provasopressin gene, as is the C-terminal part (residues 77–93/95) of neurophysin, and displays the same high evolutionary variability. In mammals, copeptin is usually detached from MSEL-neurophysin and found separated in the secretory granules. However, in the guinea pig, about 20 % of the initial domains remain

linked, so that an intermediate having 132 residues can be isolated along with separate neurophysin and copeptin. The separation does not occur in nonmammalian tetrapods, so that MSEL-neurophysins reach 129 to 132 residues (“big” neurophysin) (Table IV).

#### Avian and Amphibian Neurophysins

In nonmammalian tetrapods, oxytocin and vasopressin are replaced by mesotocin and vasotocin, respectively. These hormones are processed from protein precursors similar to those of mammalian hormones, and the two types of neurophysins (VLDV type [mesotocin associated] and MSEL type [vasotocin associated]) can be recognized. The numbers of amino acid substitutions increase when avian, amphibian, and fish neurophysins are compared with their mammalian counterparts (Table IV).

In all neurophysins, the seven disulfide bridges occupy the same location, so that the general three-dimensional structure, as determined by X ray for bovine MSEL-neurophysin (Fig. 5), is very likely preserved. However, for the avian and amphibian MSEL-neurophysins, the presence of the additional extension copeptin might modify to some extent the conformation of the C-terminal neurophysin domain.

#### Fish Neurophysins

In bony fishes, isotocin and vasotocin replace the mammalian oxytocin and vasopressin, respectively. Bony fish neurophysins have not been chemically

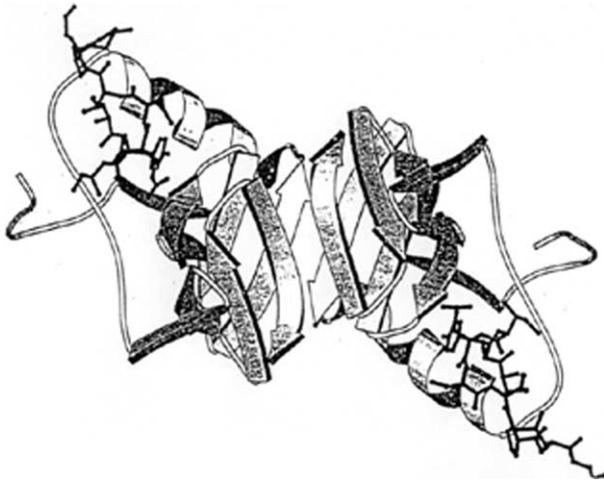
Table IV Amino Acid Sequences of MSEL-Neurophysins from Some Species

	1								10									20				25				
Bovine	Ala.	Met.	Ser.	Asp.	Leu.	Glu.	Leu.	Arg.	Gln.	Cys.	Leu.	Pro.	Cys.	Gly.	Pro.	Gly.	Gly.	Lys.	Gly.	Arg.	Cys.	Phe.	Gly.	Pro.	Ser	
Ovine	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
Porcine	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
Equine	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
Whale	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
Rat	---	Thr	---	---	Met	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
Guineapig	---	Leu	---	---	Thr	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
Human	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
Goose	---	Leu.	Ala	---	Thr.	Ala	---	---	---	---	---	---	---	---	---	---	Asn.	Arg	---	---	---	---	---	---	Gly	
Ostrich	---	Leu.	Ala	---	Ala.	Ala	---	---	---	Met	---	---	---	---	---	---	Asp.	Arg	---	Asn	---	---	---	---	---	
Chicken	---	Leu.	Gly	---	Thr.	Ala	---	---	---	---	---	---	---	---	---	---	Asn.	Arg	---	---	---	---	---	---	Gly	
Frog	Ser.	Tyr.	Pro	---	Thr	---	Val	---	---	---	Ile	---	---	---	---	---	Asn.	Arg	---	Asn	---	---	---	---	Asn	
Toad	Ser.	Tyr.	Pro	---	Thr.	Ala.	Val	---	---	---	Ile	---	---	---	---	---	Asn.	Arg	---	Asn	---	---	---	---	Asn	
	26				30				33	33'	34					40									50	
Bovine	Ile.	Cys.	Cys.	Gly.	Asp.	Glu.	Leu.	Gly	[ ]	Cys.	Phe.	Val.	Gly.	Thr.	Ala.	Glu.	Ala.	Leu.	Arg.	Cys.	Gln.	Glu.	Glu.	Asn.	Tyr.	Leu
Ovine	---	---	---	---	---	---	---	---	[ ]	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	Ile	---
Porcine	---	---	---	---	---	---	---	---	[ ]	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Equine	---	---	---	---	---	---	---	---	[ ]	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Whale	---	---	---	---	---	---	---	---	[ ]	---	---	Met	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Rat	---	---	---	Ala	---	---	---	---	[ ]	---	---	Leu	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Guineapig	---	---	---	Ala	---	Ala	---	---	[ ]	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Human	---	---	---	Ala	---	---	---	---	[ ]	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Goose	---	---	---	---	Val	---	---	---	Gly	---	Tyr	---	---	---	---	---	Thr.	Arg	---	---	Ala	---	---	Asp	---	
Ostrich	---	---	---	---	Ala	---	---	---	[ ]	---	Tyr	---	---	---	---	---	Thr	---	---	---	Ala	---	---	---	---	---
Chicken	---	---	---	---	Ala	---	---	---	[ ]	---	Tyr-Leu	---	---	---	---	---	Thr.	Arg	---	---	Ala	---	---	Asp	---	Met
Frog	---	---	---	---	Glu.	Asp	---	---	[ ]	---	Tyr-Ile	---	---	Pro	---	Thr	---	---	---	---	Val	---	---	---	---	
Toad	---	---	---	---	Glu.	Asp	---	---	[ ]	---	Tyr	---	---	Pro	---	Thr	---	---	---	---	Val	---	---	Thr	---	
	51									60																
Bovine	Pro.	Ser.	Pro.	Cys.	Gln.	Ser.	Gly.	Gln.	Lys.	Pro.	Cys.	Gly.	Ser.	Gly.	Gly.	Arg.	Cys.	Ala.	Ala.	Ala.	Gly.	Ile.	Cys.	Cys.	Asn	
Ovine	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Porcine	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Equine	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Whale	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Rat	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	Ser
Guineapig	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	Asn	---	Val	---	---	---
Human	---	---	---	---	---	---	---	---	---	Ala	---	---	---	---	---	---	---	---	---	---	Phe	---	Val	---	---	---
Goose	---	---	---	---	---	Pro	---	Gly.	Gln	---	---	---	---	---	---	---	---	---	---	---	Asp	---	Val	---	---	Ser
Ostrich	---	---	---	---	Arg.	Ala	---	Gly.	Gln	---	---	---	Ala	---	---	---	---	---	---	---	Pro	---	Ile	---	---	Ser
Chicken	---	---	---	---	---	Ala	---	Gly.	Gln	---	---	---	---	Asp	---	---	---	---	---	---	Asn	---	Val	---	---	Ser
Frog	---	---	---	---	Glu.	Ala	---	Gly	---	---	---	---	Ala	---	---	---	---	---	---	---	Pro	---	Val	---	---	---
Toad	---	---	---	---	Glu.	Ala	---	Gly	---	---	---	---	Ser	---	---	---	---	---	---	---	Pro	---	Val	---	---	Ser

continues

Table IV *continued*

	Neurophysin														Copeptin											
	76	80			Ile 90				93	95		96	100													
Bovine	Asp.	Glu.	Ser.	Cys.	Val.	Thr.	Glu.	Pro.	Glu.	Cys.	Arg.	Glu.	Gly.	Val.	Gly.	Phe.	Pro.	Arg.	Arg.	Val.	Arg.	Ala.	Asn.	Asp.	Arg	
Ovine	---	---	---	---	---	---	---	---	---	---	---	---	lle	---	---	---	---	---	---	---	---	---	---	Ser	---	---
Porcine	---	---	---	---	---	---	---	---	---	---	---	---	Ala.	Ser	---	Leu	---	---	---	Ala	---	---	Ser	---	---	
Equine	---	---	---	---	---	---	---	---	---	---	---	---	Ala	---	Leu	---	---	---	---	Ala	---	---	---	---	---	
Whale	---	---	---	---	---	---	---	---	---	---	---	---	Ala-Ser	---	---	---	---	---	---	Ala	---	---	---	---	---	
Rat	---	---	---	---	---	Ala	---	---	---	---	---	---	Phe.	Phe [			]	---	Leu.	Thr	---	---	---	Arg-Glu-Gln		
Guinea pig	---	---	---	---	---	Ile	---	---	---	---	---	---	Glu.	Phe.	His [		]	---	Pro	---	---	---	---	Gly	---	---
Human	---	---	---	---	---	---	---	---	---	---	---	---	Phe.	His [		]	---	---	Ala	---	---	---	Ser	---	---	
Goose	Ala.	Asp.	Thr	---	Ala.	Ala.	Asp.	Ala.	Ser	---	Leu	---	Glu.	Gly.	Ser.	Glu [		]	---	Ala.	Glu.	Glu.	Ala.	Ala.	Glu	
Ostrich	---	---	Thr	---	Ser.	Leu	---	---	Ala	---	Leu	---	Glu.	Ala	---	Glu [		]	---	Gly.	Gly.	Glu.	Pro.	Ala.	Gln	
Chicken	Ala.	Asp.	Thr	---	Ala.	Met.	Asp.	Ala.	Val	---	Leu	---	Glu.	Gly.	Ser.	Glu [		]	Gln.	Ala.	Glu.	Glu.	Ala.	Ala.	Glu.	
Frog	---	Gln	---	---	Thr.	Met.	Asp.	Ser.	Ser	---	Leu	Asp	Glu.	Asp.	Ser.	Glu [		]	---	Gln	---	Val.	Ser.	Pro.	Asp	
Toad	---	Asp.	Thr	---	---	Val.	Asp.	Ser.	Ser.	---	Leu.	Asp.	Glu.	Asp.	Ser.	Glu [		]	---	Arg	---	Val.	Thr.	Pro.	Glu	
	101				110						120		125													
Bovine	Ser.	Asn.	Ala.	Thr.	Leu.	Leu.	Asp.	Gly.	Pro.	Ser.	Gly.	Ala.	Leu.	Leu.	Leu.	Arg.	Leu.	Val.	Gln.	Leu.	Ala.	Gly.	Ala.	Pro.	Glu	
Ovine	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	Ala	---	---	---
Porcine	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Rat	---	---	---	---	Gln	---	---	---	Ala.	Arg.	Glu	---	---	---	---	---	---	---	---	---	---	---	---	Thr.	Gln	---
Guinea pig	---	---	Val	---	Gln	---	---	---	Ala	---	---	---	---	---	---	---	---	Met	---	---	---	---	---	---	---	---
Human	---	---	---	---	Asn	---	---	---	Ala	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Goose	Lys	---	Leu	---	Val	---	---	Ser.	Ala	---	Asp	---	---	---	---	---	---	Met	---	---	---	---	---	Arg.	Gln.	Gln
Ostrich	Lys	---	Leu	---	Gly	---	---	Ala.	Ser.	Ala	---	Asp.	Phe	---	---	Lys	---	Met.	His	---	---	---	Ala.	Asn.	Arg.	Gln
Chicken	Lys	---	Leu	---	Val	---	---	Ala.	Ala	---	Asp	---	---	---	---	---	---	Met.	His	---	---	---	Asn.	Arg.	Gln.	Gln
Frog	Gln	---	Met	---	Gln.	Met.	Asn	---	Ser.	Ala.	Ser.	Asp	---	---	---	---	---	Met.	His.	Met	---	---	Asn.	Arg.	Gln.	Gln
Toad	Gln	---	Met	---	Gln.	Met	---	---	Ser.	Ala.	Ser.	Asp	---	---	---	---	---	Met.	His.	Met	---	---	Asn.	Arg.	Gln.	Gln
	126				130	131	132	135																		
Bovine	Pro.	Ala.	Glu.	Pro.	Ala.	Gln.	Pro.	Gly.	Val.	Tyr																
Ovine	---	---	---	---	---	---	---	---	---	---																
Porcine	---	---	---	---	---	---	---	---	---	---																
Rat	Ser.	Val.	Asp.	Ser	---	Lys	---	Arg	---	---																
Guinea pig	---	Gln [		]	---	---	Lys	---	Arg	---	---															
Human	---	Phe	---	---	---	---	---	Asp.	Ala	---																
Goose	Gly.	Arg.	Gln	---	Glu [		]	Leu.	Leu																	
Ostrich	Gln.	Gln.	Gly.	Gly.	Lys.	Gly	---	Leu.	Leu																	
Chicken	Gln.	Gly [		]	Lys	---	---	---	Leu																	
Frog	Gln.	Gln.	Thr.	Lys.	His.	Tyr																				
Toad	Gln [		]	Ser.	Lys.	His	---	Phe.	Tyr																	



**Figure 5** Three-dimensional structure of the neurophysin (ribbon representation)-oxytocin (stick model) complex dimer. Reprinted with permission from Rose, J. P., *et al.* (1996). Crystal structure of the neurohypophysin-oxytocin complexes. *Nature Structural Biology* 3, 163-169.

characterized extensively. Partial sequence information is available for pollack (*Pollachius virens*) MSEL- and VLDV-neurophysins. Amino acid sequences of the two types of neurophysins have been deduced from cDNA or genes for the white sucker *C. commersoni* by Richter's group and the chum salmon *Oncorhynchus keta* by Hyodo and colleagues. Catostomids and salmonids are tetraploid fishes, so that two MSEL-neurophysins (vasotocin-associated neurophysins) and two VLDV-neurophysins (isotocin-associated neurophysins) have been detected. The two types of neurophysins have also been sequenced in the Japanese puffer fish *Fugu rubripes* by Venkatesh and Brenner. The numbers of substitutions increase when compared with the bovine counterparts. The new features are a shortening of the C-terminal copeptin extension in provasotocins and, conversely, the occurrence of a copeptin extension in proisotocins, so that the two precursors have approximately the same numbers of residues. In Cyclostomes, a single neurohypophysial peptide, vasotocin, has been chemically identified in the lamprey *Petromyzon marinus*; the provasotocin sequence has been deduced from the provasotocin gene of the Pacific hagfish *Epastretus stouti*. The length of the neurophysin-copeptin domain (127 residues) is nearly similar to that of white sucker neurophysins (121/123 residues), assuming that copeptin extensions are not cut off. Sequence comparison with other vertebrates shows that the hagfish vasotocin-associated neurophysin contains 14 conserved cysteine residues at identical positions and

suggests a similar polypeptide chain folding with seven disulfide bridges.

### **Composite Evolution and the Two Phylogenetic Lineages**

Because neurohypophysial hormone-neurophysin precursors are usually encoded by three-exon genes, with the first coding for the hormone moiety, the tripeptide processing signal, and the first 9 residues of neurophysin; the second coding for the central part (residues 10-76); and the third coding for the C-terminal part (residues 77-93/95 or 77-135), three autonomous evolutions can be considered (Fig. 6).

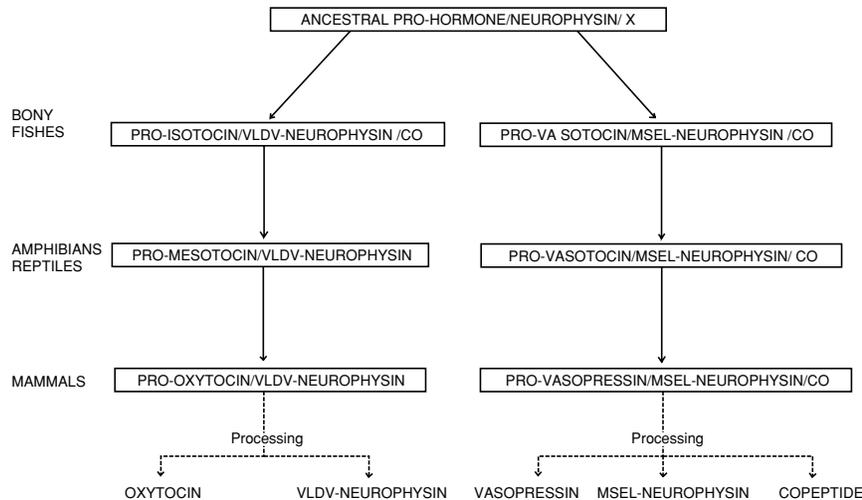
For the first exon, it is plausible that the hormone moiety, subjected to coevolution with its receptor, has its specific structural constraints in each lineage. Whereas the processing signal is perfectly preserved with a strict conservation of the Gly-Lys-Arg sequence, with all hormones being nonapeptides, amino acid substitutions in the hormone moiety may be regarded either as selective (Ile3/Phe3 for the change of vasotocin into vasopressin) or as neutral (Ile8/Leu8 for the change of mesotocin into oxytocin). For the first 9 residues of neurophysins, the relative preservation of the four tracer residues in each lineage (MSEL or VLDV) may be explained by the proper folding of the processing region. However, variations in these nine positions are more numerous than those in the nine positions of the hormone moiety (Table IV).

The central part of neurophysins encoded by the second exon appears to be strongly invariant in mammals and, furthermore, nearly identical in the two types of a given species due to gene conversion. It is difficult to explain the particular evolutionary stability within each lineage, but it may be a requisite for proper folding.

The C-terminal part of the precursors encoded by the third exon is very variable, either for the last part of the neurophysin domain (residues 77-93/95) or for the copeptin domain (95/97-133/135). Moreover, copeptin, present in fish proisotocins, disappears in avian/amphibian promesotocins and mammalian prooxytocins (Fig. 6).

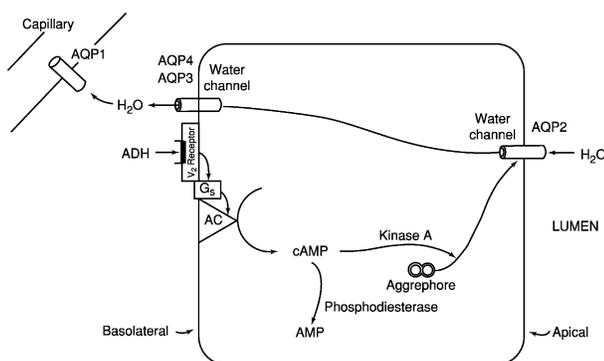
### **MOLECULAR MACHINERIES WITHIN THE NEUROHYPOPHYSEAL HORMONE TARGET CELLS: SYSTEMIC FUNCTIONS**

Neurohypophysial hormone target cells contain G protein-coupled receptors that are associated with



**Figure 6** The two evolutionary lineages of vertebrate neurohypophysial hormone precursors. The three domains (hormone, neurophysin, and copeptin [CO]) are likely present in the ancestral prohormone. After gene duplication, they remain at the level of the two bony fish precursors, but copeptin disappears in the prooxytocin lineage from the tetrapods onward. In the provasopressin-like lineage, neurophysin and copeptin remain linked in amphibians and birds, whereas the complete processing into three fragments appears at the level of mammals.

either  $G_{\alpha s}$  proteins and the adenylate cyclase–protein kinase A cascade or  $G_{\alpha q}$  proteins and the phospholipase C–protein kinase C cascade (cf. Fig. 7). The physiological specificities of the hormones depend essentially on the nature of the effectors. A number of biological or pharmacological functions have been ascribed to the vasopressin-like hormones (Table V) as well as to the oxytocin-like peptides. Diversification of functions has been generated in the course of evolution by recruitment of novel effectors rather than by changes in the structures of hormones or receptors.



**Figure 7** A hydro-osmotic principal cell of the nephron-collecting duct involved in water reabsorption through the vasopressin (ADH) action. Water uptake occurs by the vasopressin-regulated aquaporin 2 (AQP2), water output occurs by aquaporins 3 and 4 (AQP3 and AQP4), and return to blood occurs by the capillary aquaporin 1 (AQP1).  $G_s$ , heterotrimeric  $G_s$  protein; AC, adenylate cyclase; Aggrephore, cytoplasmic vesicles containing AQP2 targeted to apical membrane through PKA activation.

## Vasopressin/Vasotocin Receptors

On the basis of pharmacological and functional studies, Mitchell in 1979 proposed the occurrence of two types of vasopressin receptors. Activation of vasopressin receptors in the liver triggers an increase in inositol triphosphate and diacylglycerol, leading to activation of protein kinase C. Mitchell suggested the term V1 for this type of receptor. Renal vasopressin receptors involved in water reabsorption activate adenylate cyclase and are called V2. The vasopressin receptor present in anterior pituitary corticotropes has a pharmacological profile different from that of the hepatic/vascular receptors, so these receptors were subdivided into V1a and V1b subtypes, respectively. Mammalian vasopressin receptors were sequenced by Brownstein, Brinbaumer, and their colleagues in 1992 through molecular biology.

**Table V** Pleiotropic Effects of Vasopressin

Antidiuresis	Blood volume
Antinatriuresis	Blood osmolality
Urea reabsorption	Renal regulation
Vasoconstriction	Vascular tone
Corticotrope action	Adrenocorticotrophic hormone secretion
Glycogenolysis	Hepatic metabolism
Platelet aggregation	Coagulation
Coagulation factors	Factors release
Neurotransmission	Synaptic action
Mitogenesis	Cell proliferation

### **V1a Vasopressin Receptors**

The hepatic/vascular V1a vasopressin receptors have been identified through complementary DNAs from the rat and human. They comprise 394 and 418 residues, respectively, with 72% sequence identity between them, and they possess seven putative transmembrane segments. Measurements of vasopressin-induced inositol trisphosphate production and calcium mobilization in COS-7 cells and Chinese hamster ovary cells where the cloned receptors have been expressed indicate that these receptors are coupled to phospholipase C via a pertussis toxin-insensitive pathway. A coupling with phospholipase D, leading to the breakdown of phosphatidylcholine into phosphatidic acid, diacylglycerol, and choline, has also been observed in hepatocytes and smooth muscle cells. These receptors mediate different physiological effects in vascular smooth cells (vasoconstriction) and in hepatic cells (glycogenolysis), despite identical second messengers, because of the differentiated molecular environment and different interaction cascades. V1a receptor transcripts are also present in the larger renal blood vessels, including the interlobular arteries and branches. V1a receptors have been detected by *in situ* hybridization histochemistry in many brain areas, suggesting a neurotransmitter function for vasopressin.

### **V1b Vasopressin Receptors**

The corticotrope V1b vasopressin receptors of the rat and human have been cloned. These receptors (424 residues for the human receptor) display a high percentage sequence identity with their V1a and V2 vasopressin receptor counterparts (53 and 43%, respectively, for rat receptors and 45 and 39% for human receptors). They also show a high degree of sequence identity with oxytocin receptors (49% for rat and 45% for human), so that a superfamily of oxytocin/vasopressin receptors can be defined.

corticotropin-releasing factor (CRF) and vasopressin are found colocalized in parvocellular cells of the paraventricular nucleus. A cooperation between CRF action through its specific receptor and vasopressin action through the V1b receptor seems involved in corticotropin secretion. Vasopressin, in contrast to CRF, does not increase pro-opiomelanocortin mRNA transcription but induces corticotropin release (secretagogue activity), perhaps by a separate secretory pathway.

### **V2 Vasopressin Receptors**

The renal V2 vasopressin rat and human receptors have also been cloned and characterized. They comprise 370 and 371 residues, respectively, and display

significant sequence identity either with or between their V1a vasopressin receptor counterparts (36% for human receptors). In the transmembrane domains, the rat V2 vasopressin receptor has approximately 60% identity with the rat V1a vasopressin receptor. The transmembrane hydrophobic helices are particularly preserved in the course of evolution.

### **Vasotocin Receptors**

In all nonmammalian vertebrates, vasopressin is replaced by vasotocin ([Ile3]vasopressin). Whereas vasotocin is the antidiuretic hormone of nonmammalian tetrapods, its function in fishes is not yet clearly clarified. A vasotocin receptor from a bony fish, the white sucker *C. commersoni*, has been cloned by Richter. It comprises 435 residues and has the seven stretches of hydrophobic residues that are assumed to span the membrane, as found for mammalian vasopressin receptors. This receptor is present in various teleost tissues such as pituitary, liver, gills, swim bladder, and lateral line. The pharmacological properties of the vasotocin receptor expressed in *X. laevis* oocytes seem to indicate an ancestral nonapeptide receptor; vasotocin, vasopressin, oxytocin, mesotocin, and aspartocin activated the receptor, but isotocin ([Ser4,Ile8]oxytocin), the second neurohypophysial hormone of the white sucker, did not. This receptor belongs to the V1-type vasotocin receptor.

## **Oxytocin-like Hormone Receptors**

### **Oxytocin Receptor**

Identification of the human oxytocin receptor has been performed through complementary DNA isolated by expression cloning. Messenger RNAs have been isolated from breast, ovary, uterine endometrium, and myometrium. The mRNA level in the myometrium is very high at term of labor, with a dramatic increase in the number of oxytocin receptors. Both oxytocin and vasopressin elicit electrophysiological responses on the oxytocin receptor expressed in *X. laevis* oocytes, but a vasopressin V1 receptor agonist ([Phe2,Ile3,Orn8]vasopressin) has no effect at high concentration ( $10^{-5}$  M). A specific oxytocin antagonist ([d(CH2)5,Tyr(OMe)2-Orn8]-vasotocin) markedly decreases the membrane current at  $10^{-6}$  M concentrations.

### **Mesotocin Receptor**

Mesotocin ([Ile3]-oxytocin) is the oxytocin-like hormone of nonmammalian tetrapods (lungfishes, amphibians, reptiles, and birds) and in most marsupials,

particularly macropodids (kangaroos and wallabies). The cloning and functional characterization of the cDNA for the mesotocin receptor from the urinary bladder of the toad *Bufo marinus* has been carried out by Farhenholz. The cloned cDNA encodes a polypeptide of 389 amino acids and shows great similarities to the mammalian oxytocin receptor (70 % sequence identity) and to the teleost fish isotocin receptor (66 % sequence identity). When expressed in COS cells, it exhibits the following relative order of ligand affinity: mesotocin > vasotocin = oxytocin > vasopressin > hydrin 1, isotocin, hydrin 2. This receptor is coupled to the inositol phosphate/calcium pathway. It is found not only in bladder but also in kidney, muscle, and brain of the toad.

#### **Isotocin Receptor**

Teleost fish possess a specific oxytocin-like hormone, isotocin ([Ser4,Ile8]-oxytocin) that can be regarded as the evolutionary precursor of nonmammalian tetrapod mesotocin. The isotocin receptor of the white sucker *C. commersoni* has been cloned by Richter. The isotocin receptor can also be activated by vasotocin, mesotocin, oxytocin, and arginine vasopressin, although these have lower potencies than does isotocin.

The general conclusion can be drawn that there is a high evolutionary conservation in the structures of both the hormones and the receptors.

### **Evolution of Oxytocin/Vasopressin-like Peptide Receptors**

#### **An Ancestral Invertebrate Lys-Conopressin Receptor**

In invertebrates, cooccurrence of vasopressin- and oxytocin-related peptides has never been demonstrated in the same species. Lys-conopressin, a hormone first characterized in *Conus geographicus* and then found in other molluscs, controls reproductive functions that are analogous to oxytocin functions in vertebrate reproductions despite its being structurally related to vasopressin (Table III). Geraerts and colleagues have identified in *L. stagnalis* a G protein-coupled receptor (LSCPR) mediating both vasopressin- and oxytocin-like functions in this freshwater snail. The receptor is a classical seven-helix transmembrane protein consisting of 462 amino acids. LSCPR may represent an ancestral receptor to both vertebrate vasopressin and oxytocin receptors.

#### **Evolution of Vertebrate Neurohypophysial Hormone Receptors**

Because the most primitive vertebrates, the Cyclostomes, have a single neurohypophysial peptide,

vasotocin, it can be assumed that the first vertebrate receptor was a vasotocin receptor and that successive gene duplications have allowed diversification of receptors in the more evolved vertebrate classes. We can suppose that two main lineages of receptors, namely oxytocin- and vasopressin-like hormone receptors, correspond to the two main neurohypophysial hormone lineages. A phylogenetic tree depicting a possible evolution of the receptors of the neurohypophysial hormone family derived from a single ancestral gene is proposed.

### **CONCLUSIONS**

The described neurohypophysial endocrine regulatory cascade outlines the general features of any intercellular communication mediated by a peptide messenger. The precursor biosynthesis and the mediator storage in the hypothalamic secretory neurons are regulated by both systemic effects and neural signals, for instance, for vasopressin by signals originating from osmoreceptors and baroreceptors.

The physiological function can now be analyzed as a cascade of protein-protein interactions occurring partly in soluble medium and partly in membrane-bound systems (Table I). From a mechanical viewpoint, the key factor is the elasticity of proteins (i.e., the reversible change of conformation) through covalent or noncovalent actions. The succession of specific recognitions is ordered by the conformational shift at each level. However, dynamic events, such as the translocation of the precursor into the rough ER, its transport, and its processing along the secretory pathway; the exocytotic secretion of the mediator in blood; the hormone fishing by a target cell membrane receptor; the transduction signaling with release of a second messenger; and the intracellular cascade leading to effector mobilization, cannot be explained only by static-specific interactions. These events involve energy consumption such as adenosine triphosphate (ATP) or guanosine triphosphate (GTP) hydrolysis. On the other hand, a regulated network of cascades, with each one triggered by a specific hormone, must ensure the fine-tuning of the physiological function.

The array of surface membrane receptors make, in part, the identity of the cells. Emission and treatment of the hormonal message are dependent on the differentiated intracellular machineries of the two cell partners: the messenger secretory cell and the target cell.

The duplication propensity of genomic DNA determines the occurrence of multigene families and,

therefore, multiprotein families. As a result, several sequence-related messengers, receptors, and effectors exist in evolved organisms, allowing for diversification of the regulatory cascades through combinatorial mechanisms. The two mammalian neurohypophysial hormones, oxytocin and vasopressin, very likely derived from an ancestral duplication in early vertebrates. Each peptide can stimulate a variety of sequence-related receptors, themselves originating from gene duplications. For example, vasopressin in mammals has at least three types of receptors—V1a, V1b, and V2—displaying high-identity percentages among them and between them and oxytocin receptors. Each receptor, in turn, activates a set of effectors: water channels, amiloride-sensitive Na<sup>+</sup> channels, inward rectifier potassium channel, and urea transporters for the V2-type and smooth muscle myosin regulatory light chain; glycogenolytic enzymes; and corticotrope secretory vesicles for the V1 type. So, diversity of physiological functions was achieved by evolution, with a limited number of molecules interacting in cascades ordered through precise conformation recognitions.

Adaptation may involve a switch from one cascade to another, and a peptide messenger, whose initial function became obsolete in the frame of organismal evolution, can recruit a new receptor and (indirectly) new effectors. The deciphering of the biological roles of invertebrate neurohypophysial peptides should shed light on the evolutionary molecular flexibility of neuroendocrine systems.

### Acknowledgments

The author is deeply indebted to researchers and technicians of his laboratory who worked enthusiastically over the past decades in obtaining some of the results described in this article. He is grateful

to many colleagues from Africa, Asia, Europa, North and South America, and Oceania who, by providing pituitary glands from approximately 80 vertebrate species, permitted him to partly decipher the evolutionary history of neurohypophysial hormones. The research received financial support from the Ministère de l'Enseignement Supérieur et de la Recherche, the CNRS, and the INSERM.

### See Also the Following Articles

G Protein-Coupled Receptors • Hypothalamic Regulation of Appetite and Obesity • Opioid/Orphanin Gene Family, Evolution of • Peptide Hormones, Regulation and Gene Expression

### Further Reading

- Acher, R., and Chauvet, J. (1995). The neurohypophysial endocrine regulatory cascade: Precursors, mediators, receptors and effectors. *Front. Neuroendocrin.* **16**, 237–289.
- Bentley, P. J. (2002). “Endocrines and Osmoregulation: A Comparative Account in Vertebrates,” 2nd ed., *Zoophysiology*, Vol. 39. Springer, Berlin.
- Caporale, L. H. (ed.) (1999). “Molecular Strategies in Biological Evolution,” *Annals of the New York Academy of Sciences*, vol. 870. New York Academy of Sciences, New York.
- Cold Spring Harbor Symposia on Quantitative Biology (1995). In “Protein Kinesis: The Dynamics of Protein Trafficking and Stability,” Vol. 60. Cold Spring Harbor Laboratory Press, New York.
- Greger, R., Rosenthal, W. (eds.) (2000) “Water Transport across Biological Membranes,” *Physiology, Biochemistry, and Pharmacology*, Vol. 141. Springer, Berlin.
- Ivell, R., Russel, J. A. (eds.) (1995). “Oxytocin, Cellular, and Molecular Approaches in Medicine and Research,” *Advances in Experimental Medicine and Biology*, Vol. 395. Plenum Press, New York.
- Richter, D. (ed.) (1999). “Regulatory Peptides and Their Cognate Receptors.” Springer, Berlin.
- Saito, T., Kurokawa, K., and Yoshida, S. (eds.) (1995). “Neurohypophysial: Recent Progress of Vasopressin and Oxytocin research.” Elsevier, Amsterdam.

Substance P		RPKPQQFFGLM.NH <sub>2</sub>
Eledoisin		<EPSKDAFIGLM.NH <sub>2</sub>
Physalaemin		<EADPNKFYGLM.NH <sub>2</sub>
Kassinin		DVPKSDQFVGLM.NH <sub>2</sub>
		HKTDSFVGLM.NH <sub>2</sub>
Neuropeptide K	DADSSIEKQVALLKALYGHGQISHKRHKTD	DFVGLM.NH <sub>2</sub>
Neuropeptide $\gamma$	DA	GHGQISHKRHKTD
Neurokinin B		DMHDFVGLM.NH <sub>2</sub>

**Figure 1** A comparison of the primary structures of substance P and the neurokinins that have been identified in human tissues with selected nonmammalian tachykinins. The peptides are aligned to emphasize the regions of sequence identity. The symbol “.NH<sub>2</sub>” indicates that the peptides contain a C-terminally,  $\alpha$ -amidated amino acid residue. A, alanine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; Y, tyrosine; <E, pyrroglutamyl.

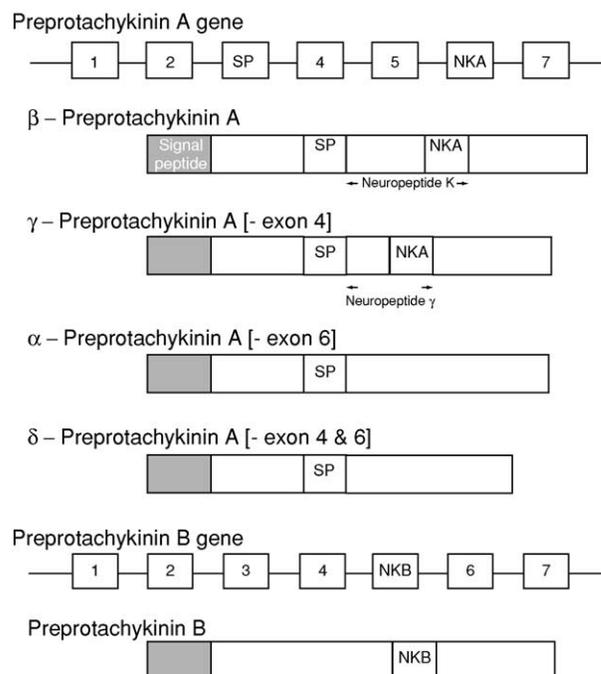
in the guinea pig ileum was much more sensitive to NKB than to other tachykinins.

## BIOSYNTHESIS OF NEUROKININS

The synthesis of the tachykinins in mammals is directed by two separate genes termed preprotachykinin A and preprotachykinin B, each of which includes seven exons. Sequence analysis of cloned cDNAs from various human tissues have identified mRNAs directing the synthesis of four biosynthetic precursors of SP ( $\alpha$ -,  $\beta$ -,  $\gamma$ -, and  $\delta$ -preprotachykinin A) that arise from the preprotachykinin A gene by an alternative RNA splicing mechanism. As shown in Fig. 2, the mRNA encoding  $\beta$ -preprotachykinin A is derived from transcription of all seven exons of the gene, so that  $\beta$ -preprotachykinin A contains SP, NKA, and its 36-amino acid residue, NH<sub>2</sub>-terminally extended form, NPK. The mRNA encoding  $\gamma$ -preprotachykinin A lacks exon 4, so that  $\gamma$ -preprotachykinin A contains the sequence of SP, NKA, and its 21-amino acid residue, NH<sub>2</sub>-terminally extended form, NP $\gamma$ . The mRNA encoding  $\alpha$ -preprotachykinin A lacks exon 6, which precisely specifies the NKA region, and the mRNA encoding  $\delta$ -preprotachykinin lacks exons 4 and 6, so that these biosynthetic precursors contain the sequence of SP only. The organization of the preprotachykinin B gene is similar to that of the preprotachykinin A gene, suggesting evolution from a common ancestral gene, but preprotachykinin B contains the sequence of NKB only.

## TACHYKININ RECEPTORS

The biological effects of the tachykinins are mediated through interaction with three discrete and fully



**Figure 2** A simplified schematic representation of the organization of the preprotachykinin A and preprotachykinin B genes and the biosynthetic precursors of the mammalian tachykinins.

characterized receptors. The original nomenclature used to describe these receptors (SP-P, SP-E, and SP-N) has now been replaced by NK<sub>1</sub>, NK<sub>2</sub>, and NK<sub>3</sub>. The receptors are defined pharmacologically in terms of the binding affinities of their endogenous ligands, so that SP may be regarded as the preferred agonist of the NK<sub>1</sub> receptor (rank order of potency SP > NKA > NKB), NKA is the preferred agonist of the NK<sub>2</sub> receptor (NKA > NKB > SP), and NKB is the preferred agonist of the NK<sub>3</sub> receptor (NKB > NKA > SP). The N-terminally extended forms of NKA, neuropeptide K, and neuropeptide  $\gamma$  bind with highest affinity to the NK<sub>2</sub> receptor.

Molecular cloning studies have confirmed unequivocally the existence of three classes of tachykinin receptor in mammalian tissues. Nucleotide sequence analysis demonstrates that the receptors are homologous (high sequence identity within and between species) and belong to the extensive family of rhodopsin-type (seven transmembrane domains), G protein-coupled receptors. Binding of ligand results in activation of phospholipase C with a subsequent rise in inositol 1,4,5-trisphosphate and diacylglycerol concentrations, although in some systems activation of adenylate cyclase, leading to an elevation in intracellular concentrations of cyclic AMP (cAMP), has also been reported.

## DISTRIBUTION OF THE NEUROKININS AND THEIR RECEPTORS

### Neurokinins

Because SP and NKA are biosynthetically related, the tissue distribution of the two peptides within the central and peripheral nervous systems overlaps closely and colocalization within primary C-afferent sensory neurons has been reported. However, the distribution of the mRNAs encoding the different preprotachykinin A precursors is both tissue and species dependent. In the rat, the relative abundance  $\gamma > \beta > \alpha$  is common to all tissues, whereas in the ox,  $\alpha$ -preprotachykinin A predominates in the brain and  $\beta$ -preprotachykinin A predominates in the intestine. All four mRNA isoforms ( $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$ ) have been identified in human mononuclear phagocytes and lymphocytes. NKB has been detected by immunohistochemistry in several regions of the brain, with the highest concentrations in the hypothalamus, substantia nigra, and olfactory tubercle as well as in the dorsal horn of the spinal cord. *In situ* hybridization immunohistochemistry reveals a similar, but not identical, distribution pattern of preprotachykinin B mRNA compared with preprotachykinin A mRNA in most regions of the brain. NKB is present in extracts of the intestines of those mammalian species examined, but the concentrations are low ( $< 1$  pmol/g tissue). Selective expression of preprotachykinin B mRNA in human neuroblastomas has been reported.

### Neurokinin Receptors

Few if any mammalian tissues contain a single homogeneous population of tachykinin receptors, and all natural tachykinins act as full agonists at the three receptor subtypes, and so pharmacological data must be interpreted in this light. The tissue distribution of the NK<sub>1</sub> receptor was described earlier. NK<sub>2</sub> receptors have a wide distribution, particularly on the smooth muscle of peripheral tissues, and rat gastric fundus and uterus, endothelium-denuded rabbit pulmonary artery, and hamster trachea and bladder have been used as a source of NK<sub>2</sub> receptors in pharmacological studies. In contrast, NK<sub>3</sub> receptors are found primarily in tissues of the central nervous system, with abundant expression in rat cerebral cortex, cerebellum, and hypothalamus. *In situ* hybridization studies have shown that the levels of NK<sub>3</sub> receptor mRNA in the gastrointestinal tract are much lower than the levels of NK<sub>1</sub> and NK<sub>2</sub> receptor mRNAs. Expression of NK<sub>3</sub> receptor mRNA and preprotachykinin

B mRNA in the uterus of pregnant rats has suggested a role for NKB in the regulation of uterine functions.

### PHYSIOLOGICAL AND PATHOPHYSIOLOGICAL ROLES

In general, the diverse biological actions displayed by SP are mimicked by NKA with a greater or lesser potency depending on the relative levels of NK<sub>1</sub> and NK<sub>2</sub> receptors in the target tissue. For example, NKA is approximately 10-fold more potent than SP in producing contraction of the rat stomach but is more than 10-fold less potent in stimulating secretion of rat salivary secretion. NKA is also more potent than SP in stimulating the motility of the urinary bladder and in activating the micturition reflex. Exposure of the airways to mechanical or chemical irritants such as cigarette smoke produces reflex responses such as bronchoconstriction, vasodilatation, and increased vascular permeability, mucous secretion, and mucociliary activity. There is strong evidence that tachykinins released from capsaicin-sensitive afferent fibers plays an important role in mediating these responses. NKA is appreciably more potent than SP in humans with respect to bronchoconstrictor action, whereas SP is more potent in vascular and secretory regulation. Inappropriate release of tachykinins may be involved in the pathogenesis of allergic reactions in the respiratory system and in the hyperresponsiveness of the airways in asthma.

The metabolic clearance rate of NKA, like SP, is extremely high, whereas the half-lives of the N-terminally extended forms (NPK and NP $\gamma$ ) in the circulation are appreciably longer. Consequently, intra-arterial infusions of NPK and NP $\gamma$  produce hypotensive and bronchoconstrictor responses that are of enhanced magnitude and duration compared with NKA. Similarly, NPK is appreciably more potent than NKA in stimulating adrenal corticosterone release in rats. Elevated concentrations of NPK have been measured in the circulation of patients with metastatic carcinoid tumors. The powerful vasodilator and bronchoconstrictor actions of this peptide are consistent with a role in the etiology of the flushing and bronchoconstriction observed in patients with carcinoid syndrome. Data have demonstrated that the human placenta is an abundant source of NKB, and enhanced release of placental NKB into the maternal circulation during the third trimester of pregnancy has been found in women suffering from preeclampsia.

## See Also the Following Articles

Kininases • Neurotransmitters, Overview • Substance P

## Further Reading

- Burcher, E., Mussap, C. J., and Stephenson, J. A. (1994). Autoradiographic localization of receptors in peripheral tissues. In "The Tachykinin Receptors" (S. H. Buck, ed.), pp. 125–163. Humana Press, Totowa, NJ.
- Gao, Z., and Peet, N. P. (1999). Recent advances in neurokinin receptor antagonists. *Curr. Med. Chem.* **6**, 375–388.
- Holzer, P., and Holzer-Petsche, U. (1997a). Tachykinins in the gut: I. Expression, release, and motor function. *Pharmacol. Therapeut.* **73**, 173–217.
- Holzer, P., and Holzer-Petsche, U. (1997b). Tachykinins in the gut: II. Roles in neural excitation, secretion, and inflammation. *Pharmacol. Therapeut.* **73**, 219–263.
- Krause, J. E., Blount, P., and Sachais, B. S. (1994). Molecular biology of receptors: Structures, expression, and regulatory mechanisms. In "The Tachykinin Receptors" (S. H. Buck, ed.), pp. 165–218. Humana Press, Totowa, NJ.
- Raffa, R. B. (1998). Possible role(s) for neurokinins in CNS development and neurodegenerative or other disorders. *Neurosci. Behav. Rev.* **22**, 789–813.

## PATHOGENESIS OF DIABETIC AUTONOMIC NEUROPATHY

Hypotheses concerning the multiple etiologies of diabetic autonomic neuropathy include a metabolic insult to nerve fibers, neurovascular insufficiency, auto-immune damage, and neurohormonal growth factor deficiency. Several different factors have been implicated in this pathogenic process. Hyperglycemic activation of the polyol pathway leading to the accumulation of sorbitol and potential changes in the NAD:NADH ratio may cause direct neuronal damage and/or decreased nerve blood flow. Activation of protein kinase C induces vasoconstriction and reduces neuronal blood flow. Increased oxidative stress, with increased free radical production, causes vascular endothelium damage and reduces nitric oxide bioavailability. Alternatively, excess nitric oxide production may result in the formation of peroxynitrite and damage to the endothelium and neurons, a process referred to as nitrosative stress. In a subpopulation of individuals with neuropathy, immune system mechanisms may also be involved. Reduction in neurotrophic growth factors, deficiency of essential fatty acids, and formation of advanced glycosylation end-products also result in reduced endoneurial blood flow and nerve hypoxia with altered nerve function. The result of this multifactorial process may be the activation of poly(ADP) ribosylation depletion of ATP, resulting in cell necrosis and the activation of genes involved in neuronal damage.

## CLINICAL MANIFESTATIONS OF DIABETIC AUTONOMIC NEUROPATHY

The metabolic disorders of diabetes lead to diffuse and widespread damage of nerves and small vessels. Clinical manifestations of autonomic dysfunction frequently occur concurrently but in inconsistent patterns (Table I). Diabetic autonomic neuropathy is typically assessed by focusing on symptoms or dysfunction attributable to a specific organ system. Cardiac autonomic neuropathy is the most prominent focus because of the life-threatening consequences of this complication and the availability of direct tests of cardiovascular autonomic function. However, neuropathies involving other organ systems should also be considered in the optimal care of patients with diabetes.

Table I Clinical Manifestations of Autonomic Neuropathy

<b>Cardiovascular</b>
Resting tachycardia
Exercise intolerance
Orthostatic hypotension
Silent myocardial ischemia
<b>Gastrointestinal</b>
Esophageal dysmotility
Gastroparesis diabeticorum
Constipation
Diarrhea
Fecal incontinence
<b>Genitourinary</b>
Neurogenic bladder
Erectile dysfunction
Retrograde ejaculation
Female sexual dysfunction (loss of vaginal lubrication)
<b>Metabolic</b>
Hypoglycemia unawareness
Hypoglycemia-associated autonomic failure
<b>Sudomotor</b>
Anhidrosis
Heat intolerance
Gustatory sweating
Dry skin
<b>Pupillary</b>
Argyll-Robertson pupil
Decreased diameter of dark-adapted pupil

## Cardiac Autonomic Neuropathy

Cardiac autonomic neuropathy results from damage to the autonomic nerve fibers that innervate the heart and blood vessels and results in abnormalities in heart-rate control and vascular dynamics. The 5-year mortality rate from this serious complication is five times higher for individuals with cardiac autonomic neuropathy than for individuals without cardiovascular autonomic involvement. Autonomic dysfunction can impair exercise tolerance. The severity of cardiac autonomic neuropathy has been shown to correlate inversely with an increase in heart rate at any time during exercise and with the maximal increase in heart rate. Cardiac autonomic neuropathy also leads to a reduced cardiac ejection fraction, systolic dysfunction, and decreased diastolic filling. There is a consistent association between cardiac autonomic neuropathy and the presence of silent myocardial ischemia. The

perception of angina is severely impaired in diabetic patients, allowing these individuals to exercise longer after the onset of myocardial ischemia with dire consequences. Chest pain in any location in a patient with diabetes should be considered to be of myocardial origin until proven otherwise; however, of equal importance, unexplained fatigue, confusion, tiredness, edema, hemoptysis, nausea and vomiting, diaphoresis, arrhythmias, cough, or dyspnea should alert the clinician to the possibility of silent myocardial infarction. Mortality rates after a myocardial infarction are also higher for diabetic patients than for nondiabetic patients. This may be due to regional differences in autonomic nerve damage within the heart, increasing the tendency for the development of ventricular arrhythmia and cardiovascular events after infarction.

Orthostatic hypotension, another sign of autonomic neuropathy, is a fall in systolic blood pressure of greater than 30 mm Hg on standing. It is aggravated by peripheral vascular pooling of blood such as occurs with hot baths or drugs used to treat hypertension. Patients with orthostatic hypotension typically present with lightheadedness and presyncopal symptoms. Symptoms such as dizziness, weakness, fatigue, visual blurring, and neck pain also may be due to orthostatic hypotension.

### Gastrointestinal Autonomic Neuropathy

Gastrointestinal symptoms are relatively common among patients with diabetes and often reflect diabetic gastrointestinal autonomic neuropathy. Esophageal dysfunction results at least in part from vagal neuropathy; symptoms include heartburn and dysphagia for solids. Gastroparesis in diabetes is usually clinically silent and may be the most important cause of "brittle diabetes." Severe diabetic gastroparesis is one of the most debilitating of all diabetic gastrointestinal complications. Via the use of radioisotopic techniques that quantify gastric emptying, it appears that ~50% of patients with longstanding diabetes have delayed gastric emptying. Major clinical features of this disorder are early satiety, anorexia, nausea, vomiting, epigastric discomfort, and bloating. Episodes of nausea or vomiting may last for days to months or occur in cycles. Even with mild symptoms, gastroparesis interferes with nutrient delivery to the small bowel and therefore disrupts the relationship between glucose absorption and exogenous insulin administration. These changes may result in wide swings in glucose levels and unexpected episodes of postprandial hypoglycemia and apparent brittle diabetes. Gastroparesis should therefore always be

suspected in patients with erratic glucose control. Diarrhea is evident in 20% of diabetic patients, particularly those with known diabetic autonomic neuropathy. Diarrhea is typically intermittent, but bowel movements may occur 20 or more times per day with urgency, and the stools are often watery. Individuals with constipation may have less than three bowel movements per week and these may alternate with diarrhea. Fecal incontinence due to poor sphincter tone is common in individuals with diabetes and may be associated with severe paroxysmal diarrhea or constitute an independent disorder of anorectal dysfunction.

### Genitourinary Autonomic Neuropathy

#### **Neurogenic Bladder**

The earliest bladder dysfunction is a sensory abnormality that results in impaired bladder sensation, an elevated threshold for initiating the micturition reflex, and an asymptomatic increase in bladder capacity with urinary retention. Overflow incontinence occurs because of denervation of the external and internal sphincters. Individuals with bladder dysfunction are predisposed to the development of urinary tract infections, including pyelonephritis, which may accelerate or exacerbate renal failure.

#### **Erectile Dysfunction**

Erectile dysfunction is the most common form of organic sexual dysfunction in males with diabetes, with an incidence estimated to be between 35 and 75%. It is defined as the consistent inability to attain and maintain an erection adequate for sexual intercourse. Erectile dysfunction may be the presenting symptom of diabetes and more than 50% of men with diabetes notice the onset of erectile dysfunction within 10 years of onset of the diabetes. Erectile dysfunction is a marker for the presence and development of generalized vascular disease and for premature demise from a myocardial infarct. Penile failure may portend an upcoming and possibly preventable, cardiovascular event. Erectile dysfunction should alert physicians to perform cardiovascular evaluations for these patients. Retrograde ejaculation into the bladder also occurs in diabetic males.

### Hypoglycemic Unawareness

The counterregulatory hormone responses and awareness of hypoglycemia are reduced in patients with diabetes mellitus. Unawareness of hypoglycemia and unresponsiveness to it are serious problems that

hamper the patient's ability to manage his or her diabetes. In most diabetic patients, catecholamine release, triggered by low glucose levels, produces noticeable symptoms, such as tremulousness and cold sweat, that alert the patients to eat and take other measures to prevent coma. Diabetic autonomic neuropathy impairs catecholamine release and prevents the warning signs of hypoglycemia, leaving the patient unaware of it. The related problem of hypoglycemic unresponsiveness occurs when impaired autonomic responses derange glucose counter-regulation during fasting or periods of increased insulin activity. In healthy people and in patients with early-stage diabetes, these autonomic responses result in the release of glucagon and epinephrine for short-term glucose counter-regulation and in the release of growth hormone and cortisol for long-term regulation. Failure in glucose counter-regulation can be confirmed by the absence of glucagon and epinephrine response to hypoglycemia induced by a controlled dose of insulin. The glucagon response becomes impaired after 1 to 5 years of type 1 diabetes. After 15 to 30 years, the glucagon response is almost undetectable, and it is absent in patients with autonomic neuropathy. The difficulty is in distinguishing the loss of counter-regulation due to autonomic neuropathy from that due to the dampening effects of intensive glycemic control and recurrent hypoglycemia, called "hypoglycemia-associated autonomic failure."

### Neurovascular Disturbances

Skin blood flow is important in maintaining nutrition, maintaining regional and whole body temperature, and healing traumatized skin. The apical (glabrous) skin is present in the palmar surface of the hand, the plantar surface of the foot, and the face. It contains a large number of arteriovenous anastomoses or shunts and functions in thermoregulation. In contrast, non-apical (hairy) skin is present over most of the body surface. There are relatively few arteriovenous shunt vessels and blood flow is primarily nutritive in function. Microvascular skin flow is under the control of the autonomic nervous system and is regulated by both the central and peripheral components. In diabetes, the rhythmic contraction of arterioles and small arteries is disordered. Microvascular blood flow can be accurately measured noninvasively using laser Doppler flowmetry. Defective blood flow in the small capillary circulation is found with decreased responsiveness to mental arithmetic, cold pressor, handgrip, and heating. The defect is associated with a reduction in the amplitude of vasomotion and

resembles premature aging. There are differences in the glabrous and hairy skin circulations. In hairy skin, a functional defect is found before the development of neuropathy; it occurs in family members and it may precede the development of diabetes. The clinical impact is dry skin, loss of sweating, and the development of fissures and cracks that are portals of entry for microorganisms leading to infectious ulcers and ultimately gangrene. The effect of autonomic neuropathy on the risk of developing a foot ulcer is independent of other measures of sensory neuropathy. Autonomic neuropathy may also lead to hyperperfusion of bone with increased osteoclastic activity, resulting in reduced bone density. Thus, a hot foot should alert the physician to impending Charcot's neuroarthropathy.

## DIAGNOSTIC TESTS OF DIABETIC AUTONOMIC NEUROPATHY

### Assessing Cardiovascular Autonomic Function

Tests of cardiovascular reflexes are sensitive, reproducible, simple, and noninvasive and they allow extensive evaluation of diabetic cardiovascular autonomic neuropathy. These include measurements of the resting heart rate, beat-to-beat heart-rate variation, blood pressure response to the Valsalva maneuver, changes in heart rate and systolic blood pressure in response to sustained exercise, and the QT interval. An increased resting heart rate and loss of heart-rate variation in response to deep breathing are primary indicators of parasympathetic dysfunction. Tests for sympathetic dysfunction include measurements of changes in heart rate and blood pressure in response to standing, exercise, and handgrip. Reduced 24h heart-rate variability (HRV), a newer test, is believed to be more sensitive than standard reflex tests and can detect cardiac autonomic neuropathy earlier. A 24h recording of HRV can reveal abnormal circadian rhythms regulated by sympathovagal activity. In vagal dysfunction, the high-frequency (HF) component of HRV is reduced; in sympathetic dysfunction, the low-frequency (LF) and very low-frequency components are reduced. Furthermore, in advanced cardiac autonomic neuropathy, all three components are reduced, as is the LF:HF ratio, which represents sympathovagal balance.

HRV should be assessed during deep breathing, the Valsalva maneuver, and standing. It may be used to diagnose the cause, for example, of gastroparesis or erectile dysfunction, to give an incentive for intensive

glycemic control, to quantitate the risk of a macrovascular event, and to monitor the response to therapy. Abnormalities in two or more of these tests suggest a diagnosis of autonomic neuropathy. Some centers recommend that patients with type 1 diabetes be tested 5 years after diagnosis and yearly thereafter and that patients with type 2 diabetes be tested at diagnosis and yearly thereafter. The sympathetic innervation of the heart can also be visualized and quantified by single-photon emission computed tomography with metaiodobenzylguanidine in a research setting.

### Assessing Gastrointestinal Autonomic Function

The finding of retained food in the stomach after an 8 to 12 h fast in the absence of obstruction is diagnostic of gastroparesis. Basic diagnostic tests include upper gastrointestinal (GI) endoscopy or barium series to rule out structural or mucosal abnormalities of the GI tract. Gastric emptying can be visualized by scintigraphic imaging after the patient consumes radio-nuclide-labeled food, but the scintigraphic results do not always correlate with the severity of the symptoms. The blood glucose should be normal at the time of testing because hyperglycemia decreases gastric motility. Gastroduodenal manometry may be helpful in patients with symptoms but apparently normal emptying because it can help identify pylorospasm or incoordinate gastric and duodenal motility.

Before attributing constipation to diabetic autonomic neuropathy, the clinician should rule out other causes of constipation such as hypothyroidism, side effects of drugs such as amitriptyline or calcium channel blockers, and colonic carcinoma. All patients should have a careful digital rectal examination and women should have a bimanual pelvic examination. Three stool specimens should be tested for occult blood. Anorectal manometry may be used to assess the rectal anal inhibitory reflex, which can distinguish rectosigmoid dysfunction and outlet-obstructive symptoms from colonic hypomotility.

### Assessing Genitourinary Autonomic Function

A thorough workup for erectile dysfunction in men should include a medical and sexual history, physical and psychological evaluations, blood tests for testosterone, prolactin, and thyroid hormones, a test for nocturnal erections, tests to assess penile, pelvic, and spinal nerve functions, and tests to assess penile blood supply and blood pressure. Physical examination must

include an evaluation of the autonomic nervous system, vascular supply, and hypothalamic–pituitary–gonadal axis. Autonomic neuropathy that causes erectile dysfunction is usually accompanied by loss of the HRV, ankle jerk reflex, and absent or reduced vibration sense over the large toes. To determine the integrity of sacral parasympathetic divisions, the physician should assess perianal sensation, sphincter tone, and the bulbocavernous reflex. Stenosis of the internal pudental artery is another potential cause of impotence. A penile/brachial index of less than 0.7 indicates diminished blood supply. A venous leak manifests as unresponsiveness to vasodilators and must be evaluated by penile Doppler sonography. In simple terms, unresponsiveness to intracavernosal injection of a direct vasodilator means arterial or venous insufficiency and requires inflatable devices, prostheses, or vascular reconstruction. A response to the vasodilator indicates autonomic insufficiency or a psychogenic cause that can be further defined using the HRV.

## TREATMENT OF AUTONOMIC NEUROPATHY

Intensive glycemic control is critical to preventing the onset of diabetic autonomic neuropathy and slowing its progression. The Diabetes Control and Complications Trial (DCCT) provided extensive clinical evidence that good metabolic control reduces diabetic complications. Specifically with regard to cardiovascular autonomic function, the DCCT showed that intensive glycemic control prevented the development of abnormal HRV and slowed the deterioration of autonomic dysfunction over time in individuals with type 1 diabetes. The Steno Memorial Study emphasized a multifactorial approach including blood pressure, lipid, and glucose control and the use of vitamins and antioxidants and the researchers were able to reduce the likelihood of development of autonomic neuropathy by 68%. Early identification of cardiac autonomic neuropathy permits the timely initiation of therapy with the antioxidant  $\alpha$ -lipoic acid, which appears to slow or reverse the progression of neuropathies in some studies. The use of cardioselective (e.g., atenolol) or lipophilic (e.g., propranolol) beta-blockers may also modulate the effects of autonomic dysfunction. By opposing the sympathetic stimulus, they may restore the parasympathetic–sympathetic balance. Studies using graded exercise, angiotensin-converting enzyme inhibitors, and aldosterone antagonists as a means to improve HRV have resulted in suggestive results in small populations in

research studies. Cardiovascular evaluation should include macrovascular risk factors and their elimination.

Orthostasis is best treated by avoiding precipitating factors such as hot baths and drugs that accentuate the fall in blood pressure. Patients should stand slowly and useful drugs include clonidine, midodrine, and somatostatin. The use of fludrosterone and the liberalization of salt intake usually result in weight gain edema and inadequate control of the orthostatic symptoms.

Neurovascular dysfunction requires the use of emollient creams, scrupulous attention to foot care, vasodilators, and bisphosphonates for impending Charcot's neuroarthropathy.

Treatment of gastroparesis should stress improvement of glycemic control and correction of other metabolic abnormalities. It also includes dietary modification, gastric suction, and prokinetic agents such as metoclopramide, domperidone, or erythromycin. In some severe cases, jejunostomy may be needed to provide for nutrient intake and to allow the stomach to rest until such time that it recovers its function.

The severe and intermittent nature of diabetic diarrhea makes treatment difficult. Because afferent denervation may contribute to the problem, a bowel program that includes restriction of soluble fiber and regular effort to move the bowels is indicated. In addition, trials of a gluten-free diet and restriction of lactose, cholestyramine, clonidine, somatostatin analogue, pancreatic enzyme supplements, and antibiotics, such as metronidazole, may be indicated. Treatment of constipation should begin with an emphasis on good bowel habits, including regular exercise and maintenance of adequate hydration and fiber consumption. Sorbitol and lactulose may be helpful.

A grossly overdistended bladder should be drained by catheter to improve contractility and the patient should be instructed to void by the clock rather than waiting for the sensation of bladder distension. Cholinergic agents or clean intermittent self-catheterization may also be used to facilitate emptying.

Treatment of erectile dysfunction may include withdrawal from offending medications coupled with psychological counseling, medical treatment, or surgery. Medical treatment may include the 5'-phosphodiesterase inhibitors, such as sildenafil, tadalafil, or vardenafil. A lower dosage is needed for individuals with renal failure or liver dysfunction. It may, however, take several doses to achieve an effect in patients

with diabetes. These drugs should not be taken by individuals with unstable ischemic heart disease or those using nitroglycerin or other nitrate-containing medications. Alternative treatments include suction devices with or without a constriction ring, injections of vasodilators into the corpus cavernosum, and ultimately prostheses.

## Prospects for the Future

With improved understanding of the pathogenesis of autonomic neuropathy, agents that have the potential for treating the underlying biologic defect rather than purely symptomatic therapy are being investigated. The situation is much less gloomy than the 1994 *Lancet* editorial that stated, "all we can do is make the diagnosis and commiserate with the patient."

## See Also the Following Articles

Autonomic Nervous System, Aging and • Cardiovascular Disease in Diabetes • Diabetes, Type 1 • Diabetes, Type 2 • Diabetic Nerve Disease, Neuropathy • Eye Disease in Diabetes • Foot Disease in Diabetes • Kidney Disease in Diabetes

## Further Reading

- DCCT Research Group. (1998). The effect of intensive diabetes therapy on measures of autonomic nervous system function in the Diabetes Control and Complications Trial (DCCT). *Diabetologia* **41**, 416–423.
- Maser, R. E., Mitchell, B. D., Vinik, A. I., and Freeman, R. (2003). The association between cardiovascular autonomic neuropathy and mortality in individuals with diabetes: A meta-analysis. *Diabetes Care* **26**, 1895–1901.
- Vinik, A., Erbas, T., and Stansberry, K. (1999). Gastrointestinal, genitourinary and neurovascular disturbances in diabetes. *Diabetes Rev.* **7**, 358–378.
- Vinik, A., and Richardson, D. (1998). Erectile dysfunction in diabetes. *Diabetes Rev.* **6**, 16–33.
- Vinik, A. I., and Erbas, T. (2001). Recognizing and treating diabetic autonomic neuropathy. *Cleveland Clin. J. Med.* **68**, 928–944.
- Vinik, A. I., Erbas, T., Pifer, M. A., Feldman, L., Stevens, M. J., and Russell, J. W. (2003). Diabetic autonomic neuropathy. In "Ellenberg & Rifkin's Diabetes Mellitus" (Porte, Sherwin, and Baron, eds.), 6th ed., pp. 789–804. McGraw Hill, New York.
- Vinik, A. I., Erbas, T., Tae, S., Stansberry, K., Scanelli, J. A., and Pittenger, G. L. (2001). Dermal neurovascular dysfunction in type 2 diabetes. *Diabetes Care* **24**, 1468–1475.
- Vinik, A. I., Maser, R. E., Mitchell, B. D., and Freeman, R. (2003). Diabetic autonomic neuropathy. *Diabetes Care* **26**, 1553–1579.
- Ziegler, D. (1999). Cardiovascular autonomic neuropathy: Clinical manifestations and measurement. *Diabetes Rev.* **7**, 300–315.

receptors in general elicit a similar signaling response as other receptors coupling to  $G_{i/o}$  proteins.

## NPY AND METABOLIC FUNCTION

NPY is one of the most powerful stimulators of food intake. Endogenously, this appears to involve neurons from the arcuate nucleus that project to the paraventricular nucleus and act on a combination of  $Y_1$  and  $Y_5$  receptors. Increases in body weight have been demonstrated in a variety of animal models upon central NPY administration. Since food intake, body weight homeostasis, and thermogenesis are closely linked, it is noteworthy that central NPY administration reduces brown adipose tissue thermogenesis. Thus, it has been proposed that NPY may play a role in obesity. This is further supported by the finding that alterations in the hypothalamic NPY system precede the disturbances in feeding in genetically obese animals, such as *ob/ob* and *db/db* mice or the *fa/fa* Zucker rat, that are characterized by hyperphagia and hyperinsulinemia.

Central administration of NPY raises levels of circulating insulin independent of the availability of food, whereas peripheral NPY administration was reported to lower blood glucose without accompanying alterations of insulin levels. Alterations of circulating insulin, on the other hand, can markedly affect hypothalamic expression of NPY, as observed in animal models of genetic and acquired diabetes. The pathophysiological importance of these findings remains to be determined.

## NPY AND PITUITARY FUNCTION

NPY can lower circulating levels of growth hormone, possibly via the release of somatostatin, which can be stimulated in the hypothalamus by low NPY doses. Correspondingly, pathophysiological conditions associated with elevated NPY tone in the hypothalamus, such as obesity, are associated with reduced growth hormone levels.

NPY can stimulate the release of corticotrophin-releasing factor to enhance the secretion of ACTH from the anterior pituitary and hence plasma levels of corticosterone. This appears to be part of a regulatory feedback loop since glucocorticoids can regulate NPY gene expression.

NPY can also facilitate luteinizing hormone (LH) secretion, an effect that is dependent on the presence of sex steroids. Although acute facilitatory effects on LH secretion are well documented, the chronic

effect of NPY is less well understood due to possible interference by concomitant alterations of food intake.

## NONENDOCRINE NPY EFFECTS

In both the central and the peripheral nervous systems, NPY can inhibit transmitter release via prejunctional receptors, usually belonging to the  $Y_2$  subtype. This can occur as autoinhibition (i.e., in neurons releasing NPY, such as sympathetic neurons) or as heteroinhibition (i.e., in neurons not releasing NPY, such as parasympathetic neurons). For a number of central NPY actions, it is not yet known whether they occur directly (i.e., via postjunctional receptors) or indirectly (i.e., via prejunctional receptors).

In addition to its effects on metabolism and neuroendocrine regulation, NPY can also cause other central nervous system (CNS) effects such as anxiolysis, which apparently occurs via  $Y_1$  receptors. An anti-seizure activity of NPY has also been demonstrated. A greater susceptibility to seizures in NPY knockout mice suggests the endogenous importance of this effect, which seems to occur via  $Y_5$  receptors. Various types of stress-related behaviors as well as the regulation of voluntary alcohol intake are also prominent CNS-based NPY effects.

The most prominent peripheral nonendocrine effect of NPY is vasoconstriction. This involves a direct effect on vascular smooth muscle cells and also a potentiation of the effects of other vasoconstricting agents; in most cases, both parts of the enhancement of vascular smooth muscle tone occur via  $Y_1$  receptors. Although NPY-induced vasoconstriction is also present in the renal circulation, NPY has repeatedly been shown to induce diuresis and natriuresis. However, the exact localization of the receptor, which modulates tubular function and resembles a  $Y_5$  receptor, is not known. NPY can also modulate the contractility of nonvascular smooth muscle cells, including those of the gut.

While all of these peripheral effects occur acutely, extended NPY exposure can also affect cell growth by causing, for example, cardiomyocyte hypertrophy or vascular smooth muscle cell proliferation and angiogenesis. Data from studies of knockout mice indicate that NPY may also be a neuroproliferative factor.

## See Also the Following Articles

G Protein-Coupled Receptors • GI Hormone Development (Families and Phylogeny) • GI Hormones Outside the Gut:

Central and Peripheral Nervous System • Neuropeptide Y, Evolution of • Pancreatic Polypeptide (PP) • Peptide YY (PYY)

### Further Reading

- Cerda-Reverter, J. M., and Larhammar, D. (2000). Neuropeptide Y family of peptides: Structure, anatomical expression, function, and molecular evolution. *Biochem. Cell. Biol.* **78**, 371–392.
- Heilig, M., and Thorsell, A. (2002). Brain neuropeptide Y (NPY) in stress and alcohol dependence. *Rev. Neurosci.* **13**, 85–94.
- Krzsiak, R., Obuchowicz, E., and Herman, Z. S. (1999). Interactions between the neuropeptide Y system and the hypothalamic–pituitary–adrenal axis. *Eur. J. Endocrinol.* **140**, 130–136.
- Malmström, R. E. (2002). Pharmacology of neuropeptide Y receptor antagonists. Focus on cardiovascular functions. *Eur. J. Pharmacol.* **447**, 11–30.
- Michel, M. C., Beck-Sickinger, A. G., Cox, H., Doods, H. N., Herzog, H., Larhammar, D., Quirion, R., Schwartz, T. W., and Westfall, T. C. (1998). XVI International Union of Pharmacology recommendations for the nomenclature of neuropeptide Y, peptide YY and pancreatic polypeptide receptors. *Pharmacol. Rev.* **50**, 143–150.
- Parker, E., van Heek, M., and Stamford, A. (2002). Neuropeptide Y receptors as targets for anti-obesity drug development: Perspective and current status. *Eur. J. Pharmacol.* **440**, 173–187.
- Thorsell, A., and Heilig, M. (2002). Diverse functions of neuropeptide Y revealed using genetically modified animals. *Neuropeptides* **36**, 182–193.

	1	10	20	30	36																																				
NPY human	Y	P	S	K	P	D	N	G	E	D	A	P	A	E	D	M	A	R	Y	Y	S	A	L	R	H	Y	I	N	L	I	T	R	Q	R	Y	-amide					
NPY rat	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....		
NPY pig	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....		
NPY sheep	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....		
NPY chicken	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....		
NPY alligator	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....		
NPY marsh frog	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....		
NPY river lamprey	F	N	.....	S	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....		
PYY human	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....		
PYY marsh frog	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....		
PP human	A	L	E	V	Y	.....	D	N	T	P	.....	Q	.....	Q	A	A	D	.....	R	.....	M	L	.....	Q	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
PP chicken	G	.....	Q	T	Y	.....	D	.....	V	.....	L	I	.....	F	N	D	.....	Q	.....	L	V	.....	H	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	

**Figure 1** Alignment of NPY family peptide sequences. Human NPY is shown as master sequence (top); in the other sequences, only differences are shown. Alligator is *Alligator mississippiensis*, river lamprey is *Lampetra fluviatilis*, and marsh frog is *Rana ridibunda*.

## GENE STRUCTURE AND MECHANISMS OF GENE DUPLICATION

The genes for NPY, PYY, and PP all have three introns in equivalent positions. This confirms evolution from a common ancestral gene by duplication followed by sequence divergence. The genes for NPY and PYY are located on different human chromosomes close to the HoxA and HoxB gene clusters, respectively. Many other gene families have members on both of these chromosomes, suggesting that the entire chromosome was duplicated, after which the NPY and PYY genes accumulated sequence differences and became differentially expressed. This duplication probably took place very early in vertebrate evolution (Fig. 2). The PP gene is located only 11 kb from the PYY gene and in the same orientation, suggesting that it arose by a local tandem duplication. Because PP is known only in tetrapods, this gene duplication probably occurred in the common ancestor of tetrapods (Fig. 2). Fish PY and the second lamprey PYY-like peptide probably arose by independent gene duplications.

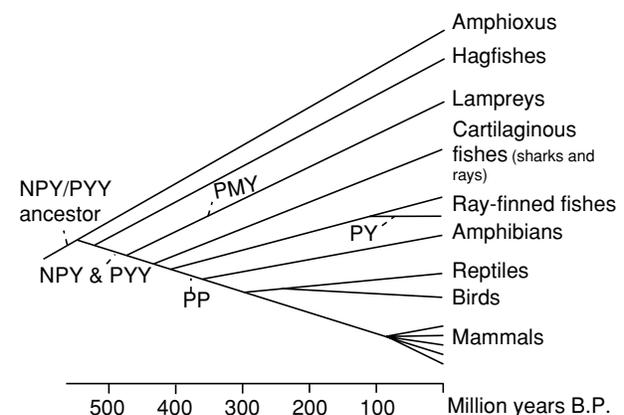
## EVOLUTION OF FUNCTIONS

The functions of NPY have primarily been studied in mammals, where NPY influences a very large number of physiological parameters. The effects include stimulation of food intake, vasoconstriction, release of pituitary hormones, influence on circadian rhythms, and modulation of neuronal activity in the hippocampus. Beyond mammals, the functions of NPY are less well known. The most studied effect is that on food intake, which is stimulated in virtually all vertebrates

investigated, including chicken, garter snake, and goldfish.

## EVOLUTION OF NPY RECEPTORS

The NPY family peptides have at least five distinct receptor subtypes in mammals, all of which belong to the superfamily of G protein-coupled receptors. The subtypes Y1, Y2, and Y5 probably arose before the origin of vertebrates, and these receptors display only approximately 30% overall amino acid sequence identity to each other. The genes for these three subtypes are located in close proximity to each other on chromosome 4 in the human genome, and the genes for Y1 and Y5 even seem to overlap in their regulatory



**Figure 2** Probable gene duplication events for the NPY family of peptides shown in a schematic tree for vertebrates, hagfishes, and amphioxus.

regions. This gene cluster was probably duplicated as part of a chromosome duplication during early vertebrate evolution, in analogy with the NPY–PYY gene duplication mentioned previously, forming the two Y1-like genes: Y4, which is the primary receptor for PP, and y6, which is a pseudogene in human and several other mammals (hence indicated with lower-case “y”) but appears to be functional in mouse, rabbit, chicken, and a shark. A Y2 duplicate exists in teleost fishes and frogs but has probably been lost in mammals. Little is yet known about the functional evolution of the receptor subtypes, but both Y1 and Y5 seem to stimulate food intake in response to NPY in the species that have been investigated. Receptor subtypes Y2 and Y4 inhibit food intake in response to PYY(3–36) and pancreatic polypeptide, respectively.

### See Also the Following Articles

GI Hormone Development (Families and Phylogeny) • G Protein-Coupled Receptors • Natriuretic Peptide System, Evolution of • Neuropeptide Y • Pancreatic Polypeptide (PP) • Peptide YY (PYY)

### Further Reading

- Cerdá-Reverter, J. M., and Larhammar, D. (2000). Neuropeptide Y family of peptides: Structure, anatomical expression, function, and molecular evolution. *Biochem. Cell Biol.* **78**, 371–392.
- de Jong-Brink, M., ter Maat, A., and Tensen, C. P. (2001). NPY in invertebrates: Molecular answers to altered functions during evolution. *Peptides* **22**, 309–315.
- Kalra, S. P., Dube, M. G., Pu, S., Xu, B., Horvath, T. L., and Kalra, P. S. (1999). Interacting appetite-regulating pathways in the hypothalamic regulation of body weight. *Endocrine Rev.* **20**, 68–100.
- Larhammar, D. (1996). Evolution of neuropeptide Y, peptide YY, and pancreatic polypeptide. *Regul. Pept.* **62**, 1–11.
- Larhammar, D., Wraith, A., Berglund, M. M., Holmberg, S. K. S., and Lundell, I. (2001). Origins of the multiple NPY-family receptors in mammals. *Peptides* **22**, 295–307.
- Michel, M. C., Beck-Sickinger, A., Cox, H., Doods, H. N., Herzog, H., Larhammar, D., Quirion, R., Schwartz, T., and Westfall, T. (1998). XVI. International Union of Pharmacology recommendations for the nomenclature of neuropeptide Y, peptide YY and pancreatic polypeptide receptors. *Pharmacol. Rev.* **50**, 143–150.
- Wang, Y., Nielsen, P. F., Youson, J. H., Potter, I. C., Lance, V. A., and Conlon, J. M. (1999). Molecular evolution of peptide tyrosine-tyrosine. *Regul. Pept.* **79**, 103–108.
- Wraith, A., Törnsten, A., Chardon, P., Harbitz, I., Chowdhary, B. P., Andersson, L., Lundin, L.-G., and Larhammar, D. (2000). Evolution of the neuropeptide Y receptor family: Gene and chromosome duplications deduced from the cloning of the five receptor subtype genes in pig. *Genome Res.* **10**, 302–310.

Table I The Neurotensin Family

Bovine/Human/Canine/Porcine/Rat NT	pGlu	Leu	Tyr	Glu	Asn	Lys	Pro	Arg	Arg	Pro	Tyr	Ile	Leu
Guinea pig NT	pGlu	Leu	Tyr	Glu	Asn	Lys	Ser	Arg	Arg	Pro	Tyr	Ile	Leu
Chicken NT	pGlu	Leu	His	Val	Asn	Lys	Ala	Arg	Arg	Pro	Tyr	Ile	Leu
Possum NT	pGlu	Leu	His	Val	Asn	Lys	Ala	Arg	Arg	Val	Tyr	Ile	Leu
Amphibian xenopsin						pGlu	Gly	Lys	Arg	Pro	Trp	Ile	Leu
Turkey/Canine xenopsin					Phe	His	Pro	Lys	Arg	Pro	Trp	Ile	Leu
Porcine neuromedin N								Lys	Ile	Pro	Tyr	Ile	Leu
Chicken Lys Asp NT6 (LANT6)								Lys	Asn	Pro	Tyr	Ile	Leu

regulatory peptides, such as cholecystokinin, to influence gastrointestinal functions. Neurotensin is also secreted from neurosecretory cells in the hypothalamus and nerve terminals in the median eminence, with elevated levels seen in hypophyseal-portal blood of the rat. In contrast, the neurotensin originating from the gonadotropes and thyrotropes of the anterior pituitary probably functions as paracrine/autocrine agents.

## RECEPTORS

Following the first demonstration of neurotensin-binding sites in rat brain synaptic terminals in 1977, there have been many biochemical and autoradiographic studies on the nature and distribution of neurotensin-binding sites. These have culminated with the cloning of three receptor subtypes, all of which recognize the C-terminal, biologically active part of neurotensin.

### Neurotensin Type 1 (NTR1)

The cDNA clone was obtained from the rat brain cDNA library using an electrophysiological assay in *Xenopus* oocytes as the end point. The cDNA encodes a 424-amino acid protein (similar in size to that reported earlier by chemical isolation) and, as suggested by pharmacological studies, is a member of the G protein-coupled receptor family. The human receptor has been cloned from the colonic adenocarcinoma cell line HT29. The receptor has a high affinity (0.2 nM), is insensitive to levocabastine, and is expressed in the intestine and brain. Signal transduction mechanisms are primarily intracellular Ca and inositol 1,4,5-trisphosphate.

### Neurotensin Type 2 (NTR2)

This is a low-affinity (5–7 nM) receptor that shares about 60% homology with NTR1 and is sensitive to levocabastine. Although it has the characteristics of a

G protein-coupled receptor, the transduction mechanisms have not been completely characterized and appear to be system and species specific. Furthermore, neurotensin antagonists have paradoxical agonist functions in some species. The receptor is expressed mostly in the brain, but in a distribution pattern different from NTR1, so that there is rarely coexpression.

### Neurotensin Type 3 (NTR3)

A third neurotensin-binding site was isolated by affinity cross-linking of brain extracts and following cloning shown to have 100% homology with the previously cloned human gp95/sortilin receptor. This protein has a large luminal domain, a single transmembrane domain, and a short cytoplasmic tail. The signal transduction mechanism and biological role remain to be determined.

## BIOLOGICAL EFFECTS

Neurotensin has a variety of biological effects, with more than 30 different *in vitro* and *in vivo* effects known. Peripheral effects include vasodilation, cyanosis, increased histamine release, stimulation of the endocrine and exocrine pancreas, effects on gastrointestinal tract smooth muscle activity and motility, stimulation of intestinal secretion, and inhibition of blood flow to adipose tissue. Centrally, neurotensin has hypothermic and antinociceptive effects, and it modulates brain dopamine systems, luteinizing hormone (LH), and prolactin release.

### Gastrointestinal Effects

Neurotensin secretion is stimulated by a fat-containing meal and has been postulated as the mediator of many of the effects of fat ingestion. These include inhibition of gastric acid secretion and motility, stimulation of pancreatic and intestinal secretions, decrease in adipose tissue blood flow, and increase in small intestinal blood flow. Neurotensin potentiates

the actions of secretin, cholecystokinin, and the vagus on pancreatic secretion and acts in concert with other enterogastrones, such as secretin, to inhibit gastric acid secretion.

### Central Effects

The neuroendocrine effects of neurotensin are extremely complex and dependent on the route of administration, the sex steroid status, and the species. Furthermore, there are outstanding issues of separating the physiological effects from the pharmacological effects and the mode of action (neurocrine, autocrine, paracrine, or endocrine). In this regard, the development of specific receptors and the use of anti-neurotensin antiserum have been most useful. SR48692 blocks NTR1-mediated effects, such as potassium-evoked dopamine release and turning behavior but not analgesia or hypothermia. These latter effects appear to be mediated by the NTR2 and are blocked by R142948A, an inhibitor of both NT1 and NT2. However, recent studies using NTR1 knockout mice suggest that this receptor does contribute to the analgesic and hypothermic responses.

As detailed in a recent review by Binder and colleagues, there are intimate and multiple anatomical and functional relationships between neurotensin and dopamine throughout the central nervous system, often with reciprocal modulation of their activities. Neurotensin activates hypothalamic dopaminergic and noradrenergic activity in the basal state but not in the fasted state. Central administration increases plasma adrenocorticotrophic hormone (ACTH) and corticosterone by enhancing the release of corticotrophic releasing factor (CRF) into the hypophyseal portal system. Studies with neurotensin antagonists implicate neurotensin in the stress-initiated activation of the hypothalamic pituitary adrenal axis. In ovariectomized animals (but not in male ones), injection of neurotensin in the preoptic area stimulates LH secretion, whereas injection of NT into the cerebral ventricular system inhibits LH secretion. However, no effect is seen in humans following intravenous administration.

As for gonadotropin release, the route of administration and the experimental conditions are critical in determining the effects of neurotensin on prolactin release. Intracerebroventricular administration decreases circulating prolactin, whereas intravenous administration increases prolactin concentrations. These effects appear to be specific in that they can be reversed by the administration of anti-neurotensin serum. Interestingly, estrogen increases neurotensin mRNA in specific regions of the hypothalamus, further

implicating hypothalamic neurotensin in the estrogen-dependent prolactin secretion. In contrast, estrogen decreases anterior pituitary levels of neurotensin. Despite some conflicting data, studies comparing the effects of neurotensin and neurotensin antiserum administration are consistent with a stimulatory effect of neurotensin on thyroid-stimulating hormone (TSH) secretion via an autocrine or paracrine mode of action.

In summary, neurotensin regulates anterior pituitary function both by its release from the nerve terminal from the median eminence and by autocrine/paracrine effects from neurotensin synthesized within the anterior pituitary. The expression of hypothalamic and pituitary neurotensin is under the regulation of gonadal and steroid hormones.

### PATHOLOGY

Pancreatic tumors producing neurotensin have been reported, but these tumors often produce excess amounts of other peptides, particularly vasoactive intestinal peptide. Because vasoactive intestinal peptide and neurotensin both stimulate intestinal secretion, it is difficult to assess the role, if any, of neurotensin in gastrointestinal symptoms of pancreatic tumors. No clinical syndromes that could be attributed to neurotensin were evident in patients with either lung or hepatic fibrolamellar tumors containing neurotensin as the only detected bioactive peptide. However, the presence of neurotensin in some types of tumors may be of some relevance given that, in experimental studies, neurotensin promotes gastric and colonic carcinogenesis and stimulates the growth of small cell lung cancer cell lines. Neurotensin antagonists inhibited basal and neurotensin stimulated small cell lung cancer growth. Neurotensin is absent from normal prostate but is synthesized and secreted from primary prostate cancers. The proliferative effect of neurotensin may be of theoretical benefit in patients who have atrophy of the bowel associated with long-term parenteral feeding. However, this has not been tested in the human. Neurotensin has been implicated in the pathogenesis of schizophrenia, and a subset of schizophrenic patients have low cerebrospinal fluid neurotensin concentrations that are restored by effective antipsychotic drug treatment.

### See Also the Following Articles

CCK (Cholecystokinin) • GI Hormone Development (Families and Phylogeny) • GI Hormones as Growth Factors • GI Hormones Outside the Gut: Central and Peripheral Nervous System • Substance P

## Further Reading

- Binder, E. B., Kinkead, B., Owens, M. J., and Nemeroff, C. B. (2001). Neurotensin and dopamine interactions. *Pharmacol. Rev.* **53**, 453–486.
- Carraway, R. E., Mitra, S. P., and Spaulding, G. (1992). Posttranslational processing of the neurotensin/neuromedin-N precursor. *Ann. N.Y. Acad. Sci.* **668**, 1–16.
- Ferris, C. (1989). Neurotensin. In “Handbook of Physiology,” section 6: “The Gastrointestinal System” (S. G. Schultz, ed.), volume 2: “Neural and Endocrine Biology” (G. M. Makhlof and B. B. Rauner, eds.), pp. 559–586. American Physiological Society, Bethesda, MD.
- Moody, T. W., Chiles, J., Casibang, M., Moody, E., Chan, D., and Davis, T. P. (2001). SR48692 is a neurotensin receptor antagonist which inhibits the growth of small cell lung cancer cells. *Peptides* **22**, 109–115.
- Pettibone, D. J., Hess, J. F., Hey, P. J., Jacobson, M. A., Leviten, M., Lis, E. V., Mallorga, P. J., Pascarella, D. M., Snyder, M. A., Williams, J. B., and Zeng, Z. (2002). The effects of deleting the mouse neurotensin receptor NTR1 on central and peripheral responses to neurotensin. *J. Pharmacol. Exp. Therapeut.* **300**, 305–313.
- Rostene, W. H., and Alexander, M. J. (1997). Neurotensin and neuroendocrine regulation. *Front. Neuroendocrinol.* **18**, 115–173.
- Shulkes, A. (1994). Neurotensin. In “Gut Peptides: Biochemistry and Physiology” (J. H. Walsh and G. J. Dockray, eds.), pp. 371–388. Raven, New York.
- Tyler-McMahon, B. M., Boules, M., and Richelson, E. (2000). Neurotensin: Peptide for the next millennium. *Reg. Pept.* **93**, 125–136.
- Vincent, J. P., Mazella, J., and Kitabgi, P. (1999). Neurotensin and neurotensin receptors. *Trends Pharmacol. Sci.* **20**, 302–309.

## GABA

GABA was recognized as a unique chemical component of brain in 1950, long before its functions were understood. Its neurotransmitter properties were first validated for the inhibitory nerves of crustacean (lobster) limb muscles, in which GABA content was shown to account for the inhibitory effects of extracts of these nerve fibers, and release of GABA with increased firing of the nerve fibers correlated with the inhibitory effects on the muscle.

In the mammalian CNS, GABA is considered the major mediator for local interneurons and for presynaptic inhibition within the spinal cord. Presumptive GABAergic inhibitory synapses have been demonstrated most clearly between cerebellar Purkinje neurons and their targets in Deiter's nucleus; between small interneurons and the major output cells of the cerebellar cortex, olfactory bulb, cuneate nucleus, hippocampus, and lateral septal nucleus; and between the vestibular nucleus and the trochlear motoneurons. GABA also mediates inhibition within the cerebral cortex and between the caudate nucleus and the substantia nigra. GABAergic neurons and nerve terminals have been localized with immunocytochemical methods that visualize glutamic acid decarboxylase, the enzyme that catalyzes the synthesis of GABA from glutamic acid. GABA-containing neurons have frequently been found to coexpress one or more neuropeptides. The most useful selective antagonists for demonstration of likely GABAergic synapses are bicuculline and picrotoxin.

GABA receptors exist in two main molecular forms, each with a functional representation. The more prominent GABA receptor subtype, the GABA<sub>A</sub> receptor, is a ligand-gated Cl<sup>-</sup> ion channel that opens in the presence of GABA. As a typical ion channel receptor, akin to the classical cholinergic nicotinic receptor, the GABA<sub>A</sub> receptor is composed of four or five subunits, each of which is composed of proteins that exhibit four transmembrane domains and cluster to form the ion channel for Cl<sup>-</sup>. Although there are four main classes of subunits, there are multiple variations, providing for an enormous number of possibly similar but distinct classes of GABA receptors. Many neuroactive drugs, such as benzodiazepines, ethanol, and barbiturates, act on GABA<sub>A</sub> receptors. The GABA<sub>B</sub> receptor is a G protein-coupled receptor able to activate second-messenger signal transduction pathways, and like other receptors of this class, it exhibits a structure composed of seven transmembrane domains. Many of the features described for the GABA<sub>A</sub> receptor family also apply to the glycine receptor, which appears to be

the major inhibitory neurotransmitter in the brainstem and spinal cord.

## Glutamate and Aspartate

Glutamate and aspartate are both abundant in brain, and both are powerful excitants for neurons in every brain region of the CNS, where they are considered the principal fast ("classical") excitatory transmitters. Glutamate receptors, like GABA receptors, exist in two forms—as ligand-gated ion channel receptors with multiple possible combinations of a few isoforms of subunits or as G protein-coupled receptors. The ligand-gated glutamate ion channel receptors are further classified according to the agonists that selectively activate them, including  $\alpha$ -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA), kainate, and *N*-methyl-D aspartate (NMDA) receptors, each of which has selective agonists and antagonists. The diversity of gene expression and, consequently, of the protein structure for glutamate receptors also arises by alternative splicing and RNA editing.

AMPA and kainate receptors mediate fast depolarization at most glutamatergic synapses in the brain and spinal cord. NMDA receptors are also involved in normal synaptic transmission, but activation of NMDA receptors is more closely associated with the induction of various forms of synaptic plasticity. AMPA or kainate receptors and NMDA receptors coexist at many glutamatergic synapses. Activation of NMDA receptors is obligatory for the induction of the form of plasticity known as long-term potentiation. Activation of NMDA receptors requires both the binding of glutamate and the simultaneous depolarization of the postsynaptic membrane to displace a Mg<sup>2+</sup> normally bound within the channel. High concentrations of glutamate are lethal to neurons because of elevation of intracellular Ca<sup>2+</sup> levels and induction of apoptosis. Studies suggest that local depletion of Na<sup>+</sup> and K<sup>+</sup>, as well as small increases in extracellular Zn<sup>2+</sup>, will activate both necrotic and proapoptotic cascades of neuronal death. Disordered glutamatergic transmission has been postulated as a mechanism of chronic neurodegenerative diseases and schizophrenia.

## AMINE TRANSMITTERS

In the following sections, all the neurotransmitters that are chemically amines are discussed, including the monoamines dopamine, norepinephrine, and epinephrine (which together comprise the catecholamines);

5-hydroxytryptamine (or serotonin); the diamine histamine; and the quaternary amine acetylcholine. However, all of these have been studied for far longer than their chemical similarities have been apparent.

## Acetylcholine

Acetylcholine (ACh) became the first established chemical neurotransmitter when it was demonstrated to be the substance released by the vagus nerve to slow the heart rate as the terminal step in the parasympathetic control of heart rate. ACh was later identified as the transmitter at both neuromuscular and parasympathetic neuroeffector junctions, as well as at the major synapse within autonomic ganglia. Eventually, ACh received recognition as a central neurotransmitter, based on its irregular distribution within the CNS and the behavioral effects exerted by cholinergic drugs after central administration. The first established cholinergic central synapse was the recurrent excitatory synapse from Renshaw interneurons to spinal motoneurons.

As for the amino acid receptors, the classical pharmacological forms of the cholinergic receptors were also shown by molecular cloning and sequence determinations to be represented by ligand-gated ion channels (nicotinic receptors) and by G protein-coupled forms (muscarinic receptors). As is the case for the amino acids, nicotinic receptors have multiple combinations of four transmembrane domain subunits, whereas the muscarinic receptors, like all known G protein-coupled receptor forms, exist as a single seven-transmembrane domain. In most regions of the CNS, the effects of ACh, assessed either by iontophoresis or by radioligand receptor-displacement assays, appear to be generated by interaction with a mixture of nicotinic and muscarinic receptors. Several sets of presumptive cholinergic pathways have been described in addition to that of the nicotinic spinal Renshaw cell to motoneuron. Eight major clusters of ACh neurons and their pathways have been characterized, with four separate groups of cell bodies located in the basal forebrain, two groups in the upper pons provide the major cholinergic innervation of thalamus and striatum, and there are two groups for the mid-brain and brainstem. The intense cholinergic projections to neocortex and hippocampal formation are dependent on the trophic growth factors provided to them by retrograde axonal transport from their target neurons, and their loss has been the sole target of treatments for Alzheimer's disease through inhibition of the central catabolic enzyme, acetylcholinesterase.

## Catecholamines

The brain contains separate neuronal systems that use three different catecholamines—dopamine, norepinephrine, and epinephrine. These monoamines are derived from the common precursors tyrosine and L-dihydroxyphenylalanine (L-DOPA), and they form a three-step synthetic metabolic pathway. Tyrosine is converted to L-DOPA by tyrosine hydroxylase, L-DOPA is converted to dopamine by DOPA decarboxylase, dopamine is converted to norepinephrine by dopamine  $\beta$ -hydroxylase, and norepinephrine is converted to epinephrine by the synthetic enzyme phenylethanolamine *N*-methyl transferase. When these enzymatic steps were determined, antibodies raised against these proteins were found to be useful cellular markers and defined three catecholamine-containing neuronal systems: Tyrosine hydroxylase defined dopamine- and norepinephrine-synthesizing neurons, antibodies to dopamine  $\beta$ -hydroxylase revealed exclusively the norepinephrine-synthesizing neurons, and antibodies to phenylethanolamine *N*-methyl transferase defined the epinephrine-synthesizing neurons. Each system is anatomically distinct and serves separate but similar functional roles within their fields of innervation.

### Dopamine

Originally, dopamine was considered to be only a precursor for norepinephrine; however, assays of distinct regions of the CNS eventually revealed that the distributions of dopamine and norepinephrine are markedly different. In fact, more than half the CNS content of catecholamine is dopamine, and extremely large amounts are found in the basal ganglia (especially the caudate nucleus), the nucleus accumbens, the olfactory tubercle, the central nucleus of the amygdala, the median eminence, and restricted fields of the frontal cortex. The anatomical connections of the dopamine-containing neurons are categorized into three major morphological classes: (i) ultrashort neurons within the amacrine cells of the retina and periglomerular cells of the olfactory bulb; (ii) intermediate-length neurons within the tuberobasal ventral hypothalamus, critical for many endocrine regulatory processes within the median eminence and intermediate lobe of the pituitary, that connect the dorsal and posterior hypothalamus with the lateral septal nuclei and other hypothalamic nuclei; and (iii) long projections between the major dopamine-containing nuclei in the substantia nigra (named for the black pigment that these neurons exhibit in the postmortem human and monkey brain and that is lost

when these neurons die in Parkinson's disease) and the ventral tegmentum. These dopamine neurons project to precise targets in the striatum, in the cerebral cortex, and in other major limbic structures except the hippocampus. All dopamine receptors are represented by two functional classes of G protein-coupled receptors—those that activate adenylate cyclase (D1 and D5) and those that inhibit the activation of adenylate cyclase (D2–D4). The D2 receptors couple to multiple effector systems, including the inhibition of adenylate cyclase activity, suppression of  $\text{Ca}^{2+}$  currents, and activation of  $\text{K}^{+}$  currents. D2 dopamine receptors have been implicated in the pathophysiology of schizophrenia and Parkinson's disease.

### **Norepinephrine**

The noradrenergic innervation of the forebrain, including the hypothalamus, cortex, hippocampus, and olfactory bulb, arises from the major cluster of noradrenergic neurons in the pontine nucleus called the locus ceruleus because of the blue pigment they contain in the brains of primates. There are relatively large amounts of norepinephrine within the hypothalamus and in certain zones of the limbic system, such as the central nucleus of the amygdala and the dentate gyrus of the hippocampus. However, norepinephrine is also present in significant but lesser amounts in most brain regions. Detailed mapping studies indicate that most noradrenergic neurons arise either in the locus ceruleus of the pons or in neurons of the lateral tegmental portion of the reticular formation. From these neurons, multiple branched axons innervate specific target cells in a large number of cortical, subcortical, and spinal fields. The pharmacological properties of such synapses are complex, with evidence of mediation by both  $\alpha$ - and  $\beta$ -adrenergic receptors (both G protein-coupled receptors; the  $\beta$ -adrenergic receptors activate adenylate cyclase, whereas  $\alpha$ -adrenergic receptors inhibit cyclic AMP synthesis, and each form has three or more subtypes). For example, stimulation of the locus ceruleus depresses the spontaneous activity of target neurons in the cerebellum mediated by a slowly developing hyperpolarization and a decrease in membrane conductance. However, activation of the locus ceruleus also enhances the higher firing rates produced by stimulation of excitatory inputs to these convergent target neurons to a lesser degree. The afferent circuits innervating the locus ceruleus neurons include medullary cholinergic neurons, opioid peptide neurons, raphe (5-HT) neurons, and corticotropin-releasing hormone neurons from the hypothalamus. Tricyclic antidepressant drugs act on noradrenergic

neurons, augmenting levels in the extracellular space by inhibiting the norepinephrine transporter.

### **Epinephrine**

Epinephrine-containing neurons are found in the medullary reticular formation and make restricted connections to a few pontine and diencephalic nuclei, such as the paraventricular nucleus of the dorsal midline thalamus. Their physiological properties have not been established.

### **5-Hydroxytryptamine**

The early history of this neurotransmitter brought together two sets of previously unconnected work on factors extracted from serum (serotonin) and gut (enteramine) identified chemically as 5-hydroxytryptamine (5-HT), which is functionally detected in large amounts in brain. Through various cytochemical approaches, 5-HT-containing neurons have been characterized in nine pontine and medullary nuclei lying in or adjacent to the midline (raphe). The more rostral raphe nuclei innervate forebrain regions, whereas the caudal raphe nuclei project within the brainstem and spinal cord with some overlap. The median raphe nucleus makes a major contribution to the innervation of limbic cortical structures, and the dorsal raphe nucleus makes a similar contribution to the cerebral cortex and the neostriatum.

Molecular biological approaches have led to the identification of more than 12 distinct mammalian 5-HT receptor subtypes, all of which are G protein-coupled receptors linked to a variety of signal transduction cascades except for the 5HT<sub>3</sub> receptor, which is a ligand-gated ion channel receptor. These subtypes exhibit characteristic ligand-binding profiles, couple to different intracellular signaling systems, exhibit subtype-specific distributions within the CNS, and mediate distinct behavioral effects of 5-HT. The 5-HT<sub>1A</sub> receptors are abundantly expressed on 5-HT neurons of the dorsal raphe nucleus, where they are thought to be involved in temperature regulation. They are also found in regions of the CNS associated with mood and anxiety, such as the hippocampus and amygdala. 5-HT<sub>1D</sub> receptors are potently activated by the drug sumatriptan, which is prescribed for acute management of migraine headaches.

The 5-HT<sub>2</sub> receptor class has three recognized subtypes: 5-HT<sub>2A</sub>, 5-HT<sub>2B</sub>, and 5-HT<sub>2C</sub>. 5-HT<sub>2A</sub> receptors are enriched in forebrain regions such as neocortex and olfactory tubercle as well as in several brainstem nuclei, including most cranial motoneurons. The 5-HT<sub>2C</sub> receptor, which is very similar in

sequence and pharmacology to the 5-HT<sub>2A</sub> receptor, is expressed abundantly in the choroid plexus, where it regulates cerebrospinal fluid production. The hallucinogen LSD interacts through the 5-HT<sub>2</sub> receptors. Serotonin-containing pathways have been implicated in sleep disturbances, appetitive problems, pain, and depression. The class of drugs termed serotonin selective reuptake inhibitors is very effective in some patients with major depression.

## Histamine

Although not recognized as a neurotransmitter for much of the recent history of such signaling molecules, histamine and antihistaminic drugs were long known to possess potent effects on animal behavior. The subsequent biochemical detection of histamine synthesis by neurons, as well as direct cytochemical localization of these neurons, established the existence of a histaminergic system in the CNS. Most histaminergic neurons are located in the ventral posterior hypothalamus and give rise to long ascending and descending tracts to the entire CNS. Histaminergic neurons function in the regulation of arousal, body temperature, and vascular dynamics.

## PEPTIDES

During the 1980s, numerous novel peptides were discovered in the CNS, each capable of regulating one or another aspect of neural function. Furthermore, certain peptides previously thought to exist only in gut or endocrine glands were also found in the CNS. Most CNS peptides are thought to act mainly in concert with coexisting transmitters, both amines and amino acids. Each neuropeptide system has significant implications for endocrinology and endocrine diseases. Three general approaches to the continuously growing families of neurons containing one or more neuropeptides are discussed here.

### Organization by Peptide Families

Peptides can be grouped into families based on the possession of significant homology in amino acid sequences, with each having its own genetic representation through evolutionary gene duplication and subsequent divergence of functions. Such relationships are well illustrated by the tachykinin or the vasotocin (vasopressin/oxytocin) families, in which species differences can be correlated with modest

variations in peptide structure. The multiple members of the tachykinin/substance P family coexist within mammalian neurons in the brain and in the intestines and may account for the apparent existence of subsets of receptors for each of these peptides. The mammalian representatives of the vasotocin family show two concurrent products, vasopressin and oxytocin, each having evolved to perform separate functions once executed in amphibia by a single vasotocin-related peptides and receptor. A different divergent evolutionary pathway is illustrated by the endorphin and by the glucagon–secretin peptide families. In the endorphin superfamily, three major genetically distinct systems of endorphin peptides (proopiomelanocortin, proenkephalin, and prodynorphin) exist in independent neuronal circuits. These natural opioid peptide families have emerged from independent but homologous genes. The peptides all share some actions at receptors once classed generally as “opioid” but that are undergoing progressive refinement with molecular resolution of their structures, functions, and cellular expression patterns. In the glucagon–secretin family, multiple and somewhat homologous peptides are found simultaneously in different cells of the same organism but in separate organ systems: glucagon and vasoactive intestinal polypeptide (VIP) in pancreatic islets; secretin in duodenal mucosa; VIP and related peptides in enteric, autonomic, and central neurons; and growth hormone-releasing factor in central neurons only. The general metabolic effects produced by this family lead to increased blood glucose.

### Organization by Anatomic Pattern

Some peptide systems follow consistent anatomical organizations. Thus, the hypothalamic peptides oxytocin, vasopressin, proopiomelanocortin, gonadotropin-releasing hormone, and growth hormone-releasing hormone all tend to be synthesized by single large clusters of neurons that give off multibranching axons to several distant targets. Others, such as systems that contain somatostatin, cholecystokinin, and enkephalin, can have many forms, with patterns varying from moderately long, hierarchical connections to short-axon, local-circuit neurons that are widely distributed throughout the brain.

### Organization by Function

Since almost all peptides were initially identified on the basis of bioassays, their names reflect these biologically

assayed functions (e.g., thyrotropin-releasing hormone and vasoactive intestinal polypeptide). These names will become trivial if more ubiquitous distributions and additional functions are discovered. Some general integrative role might be hypothesized for widely separated neurons (and other cells) that make the same peptide. However, a more parsimonious view is that each peptide has unique messenger roles at the cellular level and that these are used repeatedly in functionally similar pathways within large systems that differ in their overall functions. Cloning of the major members of the opioid-peptide receptors revealed unexpected and unexplained conservation of sequences with receptors for somatostatin, angiotensin, and other peptides.

### Comparison with Other Transmitters

Peptides differ in several important respects from the monoamine and amino acid transmitters considered earlier. Synthesis of a peptide is performed in the rough endoplasmic reticulum, where mRNA for the propeptide can be translated into an amino acid sequence. The propeptide is cleaved (proteolytically processed) to the form that is secreted as the secretory vesicles are transported from the perinuclear cytoplasm to the nerve terminals. Another major difference between neuropeptides and other neurotransmitters is that no active reuptake mechanisms for peptides have been described. Thus, peptidergic nerve terminals are dependent for their signaling on distant sites of synthesis. Much progress has been made in the development of drugs that act as antagonists for neuropeptides and that are orally active and not peptides, thereby advancing roles for neuropeptides in human diseases such as stress and depression.

### OTHER REGULATORY SIGNALS

In addition to the previously discussed major families of neurotransmitters, other endogenous substances have also gained attention as interneuronal signaling molecules that, if valid, will expand the definition of this process. For example, the purines adenosine monophosphate, adenosine triphosphate, and free adenosine have helped define a purinergic signaling system with two large families of purinergic receptors. ATP-regulated responses have been linked pharmacologically to a variety of supracellular functions, including anxiety, stroke, and epilepsy.

Although all the classical neurotransmitters described previously were detected on the basis of the ability to extract them as factors from their storage sites in brain, gut, or other organ systems, studies have revealed potent classes of signaling molecules that are apparently synthesized on demand and act at significant distances from their sites of synthesis and release through diffusion. Arachidonic acid, normally stored within the cell membrane as a glycerol ester, can be liberated when phospholipases are activated by a variety of neurotransmitter receptors and then converted to highly reactive signals by one of three enzymatic pathways: cyclooxygenases (leading to prostaglandins and thromboxanes), lipoxygenases (leading to the leukotrienes and other transient catabolites of eicosatetraenoic acid), and cytochrome P450 (which is inducible but expressed at low levels in brain). These arachidonic acid metabolites have been implicated as diffusible modulators in the CNS, particularly for long-term potentiation and other forms of plasticity, and include the anandamides, which are endogenous lipid signals identified on the basis of their interactions with the cannabinoid receptors.

Last, nitric oxide was initially recognized as an important regulator of vascular and inflammatory mechanisms and was later recognized to play analogous signaling roles in the brain after determination of the various forms of the enzyme nitric oxide synthase (NOS), through which NO is made and released. At least four isoforms of this biosynthetic enzyme have been identified in the brain: a constitutive form present in some neurons, capillary endothelial cells, and macrophages, as well as inducible forms of the enzyme. The availability of potent activators of NOS has led to reports of the presumptive involvement of NO in a host of phenomena in the brain, including long-term potentiation, guanylyl cyclase activation, neurotransmitter release and reuptake, and enhancement of glutamate (NMDA)-mediated neurotoxicity.

### See Also the Following Articles

Catecholamines • GABA (Gamma-Aminobutyric Acid) • G Protein-Coupled Receptors • Neurokinins • Normetanephrine and Metanephrine • Substance P • Thyrotropin-Releasing Hormone (TRH)

### Further Reading

Bloom, F. E. (1996). Neurotransmission and the central nervous system. *In* "Goodman and Gilman's The Pharmacological

- Basis of Therapeutics" (J. G. Hardman, L. E. Limbird, P. B. Molinoff, R. W. Ruddon, and A. Goodman Gilman, eds.), 9th ed., pp. 267–294. McGraw-Hill, New York.
- Cooper, J. R., Bloom, F. E., and Roth, R. H. (2003). "The Biochemical Basis of Neuropharmacology," 8th ed. Oxford Univ. Press, New York.
- Herrling, P. (1997). "Excitatory Amino Acids. Clinical Results with Antagonists." Academic Press, San Diego.
- Hökfelt, T., Broberger, C., Xu, Z.-Q. D., Sergeev, V., Ubink, R., and Diez, M. (2000). Neuropeptides—An overview. *Neuropharmacology* **39**, 1337–1356.
- Nestler, E. J., Hyman, S. E., and Malenka, R. C. (2001). "Molecular Neuropharmacology: A Foundation for Clinical Neuroscience." McGraw-Hill, New York.
- Schwartz, J.-C., Arrang, J.-M., Garbarg, M., and Traiffort, E. (1994). Histamine. *In* "Psychopharmacology: The Fourth Generation of Progress" (F. E. Bloom and D. J. Kupfer, eds.), pp. 397–406. Raven Press, New York.
- Squire, L. R., Bloom, F. E., McConnell, S. K., Roberts, J. L., Spitzer, N. C., and Zigmond, M. J. (eds.) (2002). "Fundamental Neuroscience," 2nd ed. Academic Press, San Diego.

## DIAGNOSIS OF AMBIGUOUS GENITALIA IN THE NEWBORN

When a genital defect is discovered during the neonatal period, complementary clinical, hormonal, genetic, molecular, and radiographic investigations are needed first to diagnose the intersex state and then to determine the etiology and orient the therapeutic approach.

### Clinical Investigations

Thorough physical examination is key to diagnosis. Methodical general inspection will establish whether the genital defect is a sign of a malformation syndrome. Careful genital inspection will define the extent of the ambiguity by determining the presence, number, size, symmetry, and position of gonads. In some cases, the sexual ambiguity is evident. For example, the genital tubercle may show development that is midway between that of a penis and a clitoris (however, the length and diameter should be evaluated); the genital folds may be completely fused, with a bifid scrotum; the penis may appear abnormally bent and buried or sunken inward; or the urethral orifice may open toward the inner side of the penis, indicating hypospadias. Other signs include posterior fusion of the labia majora and a single perineal orifice at the base of the genital tubercle, between the genital folds, signaling the existence of a urogenital sinus.

However, the genital malformation often is less obvious. When the malformation is not easy to determine, the clinician's attention should be drawn to the following: an association of cryptorchidism and hypospadias, bilateral cryptorchidism, true clitoral hypertrophy of the gland and not of the hood, and an oblong mass in the inguinal position in a newborn with a female phenotype. Precise measurement of the penis should be made. The mean stretched penile length in the normal full-term newborn male is 3.5 cm ( $\pm 0.5$ ).

Once the examination has been completed, the clinician should be ready to classify the degree of ambiguity of the external genitalia (the genital tubercle and folds and the urogenital sinus). Prader's five stages may be used, but Quigley's six stages are used more often because they offer the advantage of greater descriptive detail:

- Grade 1:* individuals with normal male external genitalia such as infertile males with azoospermia and hormonal features of androgen resistance and those with reduced virilization at puberty (so-called minimal androgen resistance)
- Grade 2:* individuals who have a univocally male phenotype but who have mildly defective fetal masculinization, manifested by defects such as isolated hypospadias and/or micropenis
- Grade 3:* individuals with a predominantly male phenotype but with more severely defective masculinization *in utero*, as evidenced by perineal hypospadias, small penis, and cryptorchidism and/or bifid scrotum
- Grade 4:* individuals with an ambiguous phenotype and severely limited masculinization, as evidenced by a phallic structure that is intermediate between a clitoris and a penis, generally accompanied by a urogenital sinus with perineal orifice and labioscrotal folds
- Grade 5:* individuals with an essentially female phenotype (i.e., minimal fetal androgen action), including separate urethral and vaginal orifices, with minimal androgenization, as evidenced by mild clitoromegaly or a small degree of posterior labial fusion
- Grade 6:* individuals with a normal female genital phenotype (i.e., no fetal androgen action) who develop androgen-dependent pubic and/or axillary hair at puberty

Careful palpation to locate gonads below the genital folds or in the inguinal region provides the first element for diagnostic orientation. If the gonads are absent, a diagnosis of female pseudohermaphroditism seems to be advisable. If a gonad or gonads are palpated, a diagnosis of male pseudohermaphroditism is indicated. However, one should bear in mind that cases presenting with very similar clinical expression may be quite different diagnostically (e.g., a masculinized female newborn with congenital adrenal hyperplasia [CAH] and an undervirilized male newborn).

A careful family history should be obtained from the parents. Detailed information on ambiguous genitalia in other siblings or family members, history of neonatal death, and consanguinity all should be sought. Cases of amenorrhea or infertility in the family will be important elements, and careful attention should be paid to clues of maternal ingestion of drugs or exposure to chemical environmental disruptors during pregnancy. Questions about "salt losing" in the family should also be raised.

A differential diagnosis should be formulated by the end of the clinical examination (Fig. 1). A certain number of complementary investigations will then need to be carried out, although perhaps not all need

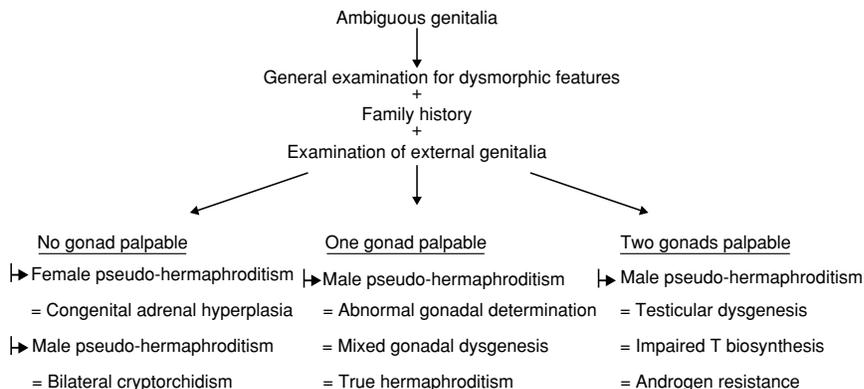


Figure 1 Diagnosis of ambiguous genitalia.

be done in a systematic fashion. However, testing for salt losing is mandatory.

### Cytogenetic and Molecular Investigations

Buccal smears reveal the presence of Barr bodies (chromatin positive) equal to the number of X-1. In practice, this screening test is easy and rapid but insufficiently reliable; thus, it is being increasingly replaced by molecular studies of the X or Y chromosome. Karyotyping is systematic, but results are often not available until several days later. Because of the urgent need for sex assignment—it is intolerable for parents to wait several weeks to know whether their baby is a boy or a girl—many hospitals are now also performing polymerase chain reaction (PCR) analysis of the SRY gene on the Y chromosome given that the results are available within 1 day.

### Hormonal Investigations

Hormonal investigations should be based on the clinical and cytogenetic orientation. Although most are generally easy to perform, they can be quite complex at times. Certain hormones should be measured on an urgent basis—immediately. Other measurements should be made as quickly as possible, and still others can be made within a few days.

Substantial elevation in 17-hydroxyprogesterone (17-OHP), as well as in plasma testosterone (which is of lower amplitude), will confirm the diagnosis of CAH, which in most cases is due to a deficiency in 21-hydroxylase.

Basal plasma testosterone levels evaluate the presence of functional Leydig cells. Testicular stimulation

with human chorionic gonadotropin (hCG) (1000 U/day for 3 days or 1500 U/every 2 days for 2 weeks) is required to determine the functional value of testicular tissue. An insufficient response (< 3 ng/ml) suggests a final diagnosis of gonadal dysgenesis. This same test is also needed to show evidence of an inborn error of testosterone biosynthesis by an augmentation in the precursors (e.g., 17-OHP, dehydroepiandrosterone,  $\Delta^4$  androstenedione), which contrasts with the absence of variation in the plasma testosterone.

In all cases of undermasculinization of the external genitalia that are associated with an often elevated secretion of testosterone, peripheral androgen receptivity must be investigated. Depending on the group, this is accomplished either by evaluation of the clinical response of the genital tubercle to exogenous testosterone or by measurement of the concentration of receptor sites in the external genitalia.

There is no consensus regarding the dose, method of administration, timing, or duration of therapeutic trials in the newborn with micropenis and ambiguous genitalia. To test for a clinical response to testosterone, an intramuscular injection of either 25 or 100 mg every 4 weeks for 3 months should bring about an augmentation in length, diameter, and the cavernous bodies of the genital tubercle. An augmentation in phallic length less than 35 mm is insufficient. Failure to respond implies end organ resistance to the action of androgens and introduces uncertainty concerning the appropriate sex of rearing.

For the evaluation of androgen receptor (AR) concentration, the number of receptors and their affinity for testosterone are measured on cultured genital skin fibroblasts. Concentrations less than 300 fmol/mg of DNA suggest partial androgen insensitivity.

## Imaging Investigations

Exploration of the genitourinary axis is carried out principally by ultrasound and genitography. In certain cases, ultrasound confirms the presence of a uterus and ovaries. Genitography can accurately detect the level of implantation of the vaginal cavity on the urethra, an essential consideration in the choice of therapeutic strategy. One of three cases can be found: the vagina opens the length of the vertical urethra, forming a long urogenital sinus; the vaginal opening is in the subvesicle position; or the vagina is located at the junction of the horizontal and vertical portions of the urethra.

## ETIOLOGY OF AMBIGUOUS GENITALIA IN THE NEWBORN

In general, the medical team has a final diagnosis and has determined the etiology of the intersex state by the end of 3 weeks.

### The XX Newborn

In the XX newborn, gonads are not palpable, and this classically leads to the diagnosis of female pseudohermaphroditism. Excessive androgen production can be traced back to the mother (excessive maternal androgen production or maternal exposure to androgens or progestins), the placenta (placental aromatase deficiency), or the fetal adrenal gland (CAH) (Fig. 2).

#### Excessive Maternal Androgens

Any maternal source of elevated androgens can induce virilization in the female fetus. Maternal ingestion of androgens, progestins, or other drugs is a notable cause. For example, several oral progestational compounds given to prevent spontaneous abortion in the past, such as 19-nor testosterone, have been

implicated. Exogenous steroids administered during pregnancy may cause posterior fusion of the labia, clitoridal enlargement, or even increased degrees of androgenization. Other drugs that are also used in pregnancy, such as danazol, have been associated with abnormalities of sexual differentiation. Ovarian tumors include luteoma of pregnancy, arrhenoblastoma, hilar cell tumor, masculinizing ovarian stromal cell tumor, and Krukenberg tumor. Untreated maternal virilizing CAH is another potential cause, although androgen-secreting adrenal tumor in the mother is rare. In both cases, investigation of abnormal androgen production by the mother must be performed immediately after delivery. A discrepancy between marked virilization in the mother and a minimal androgen effect in female offspring indicates placental aromatase activity, which converts androgens to estrogens, or androgen metabolism, which becomes less active.

#### Excessive Placental Androgens

Female pseudohermaphroditism due to placental aromatase deficiency, which has been reported by several investigators, illustrates the critical role of placental aromatization in protecting the fetus from excess androgen exposure. A defect in the placental conversion of androgens to estrogens causes virilization in female offspring. In the absence of aromatase, androgens cannot be converted to estrone, estradiol, or estriol, and large quantities of androstenedione and testosterone are transferred to the maternal and fetal circulation, resulting in masculinization of the urogenital sinus and external genitalia of the female fetus. The mother may undergo virilization during gestation as well. Cloning and sequencing of the CYP19 aromatase gene has provided new opportunities for identifying mutations. Detection of mutations in this gene has confirmed the fundamental role of aromatase for the fetal-placental unit and for sexual differentiation of the female fetus.

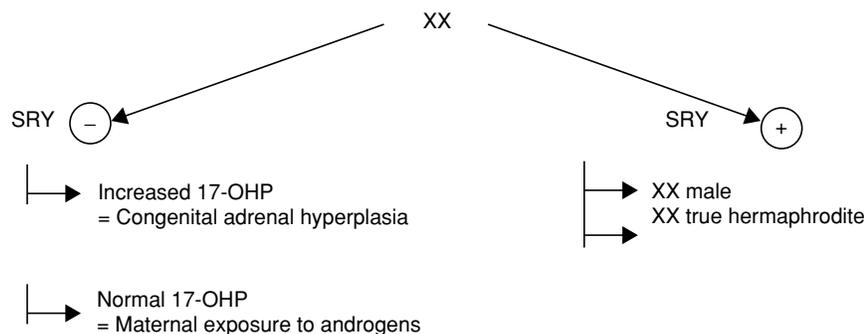


Figure 2 The XX newborn.

Management of excessive androgens in the female fetus is somewhat difficult and will differ, for example, if there is a family history of female pseudohermaphroditism or if androgen excess is suspected from ultrasonography.

### **Excessive Fetal Androgens**

CAH is the most frequent cause of androgen excess and ambiguous genitalia in the female newborn, and the various forms of CAH are due to defects in the biosynthesis of cortisol, with the subsequent excessive adrenocorticotropic hormone (ACTH) production leading to an accumulation of adrenal androgens and steroid precursors. Adrenal androgens undergo peripheral conversion to testosterone and dihydrotestosterone, and steroid precursors produce specific findings depending on the exact enzyme deficiency. The enzymatic defects causing female virilization involve  $3\beta$ -hydroxysteroid dehydrogenase ( $3\beta$ -HSD)/ $\Delta^5$ - $\Delta^4$  isomerase, P450C21 hydroxylase (21-OH), and P450C11 hydroxylase (11-OH). The association of ambiguous genitalia and salt loss at birth signals an enzymatic block of the adrenal glands. The differential diagnosis is made by the marked accumulation of the steroid above the enzymatic block: 17-OHP for 21-OH deficiency, 11-deoxycortisol for 11-OH deficiency, and pregnenolone and other  $\Delta^5$  precursors for  $3\beta$ -HSD deficiency. The molecular genetics of these defects has been studied extensively. More than 90% of CAH is due to 21-hydroxylase deficiency, and during the newborn period salt losing is present in 75% of patients. Glucocorticoid treatment should be started as soon as possible and adjusted on the basis of length, weight, and steroid levels. Mineralocorticoid should be added if salt losing is documented. 11 $\beta$ -OH deficiency is associated with hypertension.  $3\beta$ -hydroxylase dehydrogenase is exceptional.

Prenatal diagnosis of these defects is possible in high-risk families, and prenatal treatment can be offered to the mother of a 21-OH-deficient girl for future pregnancies. In preterm infants, differential diagnosis with the exceptional clitoridal hypertrophy associated with a high concentration of adrenal androgen should be considered.

### **True Hermaphroditism**

The true hermaphrodite has both testicular and ovarian tissues present in either the same or opposite gonads. Both the external genitalia and the internal duct structures display gradations between male and female. The initial manifestation is ambiguous genitalia in 90% of the cases. More rarely, isolated clitoromegaly or penile hypospadias is seen. Two-thirds of

true hermaphrodites are raised as males. Among those raised as females, two-thirds will have clitoromegaly. Virtually all patients have a urogenital sinus, and in most cases a uterus is present. The most common peripheral karyotype is 46,XX, but mosaicisms are observed (XX/XY). The SRY gene is present in 10 to 30% of patients, suggesting that true hermaphroditism is a heterogeneous condition in terms of its genetic background.

The most critical factor in managing true hermaphroditism is gender assignment. A decision of male sex of rearing should be based on both the findings at laparotomy and the potential for adequate penile length. Another important consideration is that true hermaphrodites have the potential for fertility.

### **The XY Newborn**

In the XY newborn, gonads are usually palpable and the diagnosis will be male pseudohermaphroditism (Fig. 3).

### **Normal Testosterone Increases**

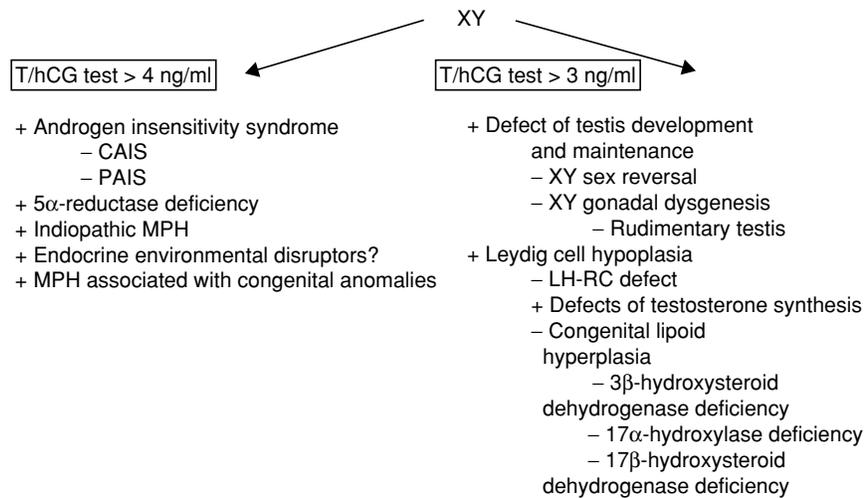
If testosterone rises normally after hCG stimulation, a defect in androgen sensitivity or  $5\alpha$ -reductase deficiency ( $5\alpha$ -RD) is indicated. Androgen resistance is the most common cause of ambiguous genitalia in male pseudohermaphroditism.

### **Complete Androgen Insensitivity**

Complete androgen insensitivity is characterized by an unambiguous female phenotype with a blind vagina pouch and no uterus. Underdevelopment of the clitoris and labia minora may also be observed. The development of an inguinal hernia signals the possibility of complete androgen insensitivity during infancy, whereas this diagnosis is evoked by primary amenorrhea during puberty. However, pubertal breast development is normal or augmented in the majority of cases, in contrast to absent or scanty axillary and pubic hair. Patients with complete androgen insensitivity develop female habitus. The major treatment decision for complete androgen insensitivity syndrome primarily concerns the optional timing of gonadectomy. Our group performs gonadectomy before puberty and prescribes estrogens during puberty.

### **Partial Androgen Insensitivity Syndrome**

Partial androgen insensitivity syndrome (PAIS) covers a wide spectrum of clinical phenotypes, from patients with a predominantly female phenotype (i.e., mild clitoromegaly) to an undervirilized male phenotype. In



**Figure 3** The XY newborn. MPH, male pseudohermaphroditism.

addition, the Wolffian duct may develop to a variable extent. Simple hypospadias or micropenis in children, or undervirilization and gynecomastia in adolescent boys, should also come to medical attention.

Newborns with PAIS have increased luteinizing hormone (LH) and testosterone secretion, whereas estrogen production is also higher (serum sex hormone-binding globulin [SHBG] concentrations are intermediate between those of normal male and normal female). The SHBG response to the increase in serum testosterone induced by an hCG stimulation test has been used as an aid in the differential diagnosis between PAIS and other forms of male pseudohermaphroditism.

Androgen binding in genital skin fibroblasts has revealed that defects are heterogeneous in that they range among reduced capacity, reduced affinity, thermolability, increased ligand dissociation rate, and altered ligand specificity. However, no research group has been able to report any consistent correlation between the concentration of AR and the degree of undervirilization.

Since the AR gene was cloned, the tools of molecular biology have made it possible to identify mutations within the gene from patients with different phenotypes of PAIS. Screening procedures with sequencing of the AR gene allow identification of subtle changes responsible for missense or nonsense mutations. Measurements of AR mRNA have been useful in identifying mutations that cause PAIS by altering the state levels or the size of the mRNA. Transfection of constructs expressing the mutant AR in mammalian cells is the main approach for demonstrating the

causative role of the mutation in the development of the androgen insensitivity.

The management of patients with partial androgen insensitivity syndrome must be individualized depending on the degree of genital ambiguity, the growth response of the penis to supraphysiological doses of testosterone, and the type of AR mutation. Although certain AR defects may be amenable to androgen therapy, multiple reconstructions of external genitalia and azoospermia are good arguments to prefer a female sex of rearing.

#### *5α-Reductase Deficiency*

Patients with 5α-RD are characterized at birth by an undervirilized phenotype. Affected newborns exhibit ambiguous genitalia with a hypospadiac phallus resembling a clitoris, a bifid scrotum that is labial-like, and a urogenital sinus opening on the perineum. It is worth noting that the testes have been found in the inguinal canal, labia majora, or scrotum.

The clinical presentation can actually range from a nearly normal female phenotype to a clear-cut male phenotype with isolated hypospadias, but in all cases Wolffian ducts have been differentiated normally into vas deferens, epididymis, and seminal vesicles.

The main characteristic of patients with 5α-RD is the virilization of the external genitalia that occurs at puberty along with the acquisition of male genetic identity in these patients usually raised as females.

Diagnosis should be based on physical examination, pedigree analysis, analysis of basal and post-hCG stimulation plasma T and DHT levels, 5β/5α urinary steroid metabolite ratio, measurement of

5 $\alpha$ -RD activity in cultured genital skin fibroblasts, and analysis of the 5 $\alpha$ -R2 gene.

The characteristic endocrine features of 5 $\alpha$ -RD are as follows: normal male to high levels of plasma T and low levels of plasma DHT, an elevated ratio of the concentration of plasma T to DHT during adulthood and after stimulation with hCG during childhood, and elevated ratios of urinary 5 $\beta$ - to 5 $\alpha$ -metabolites of androgen and C21 steroids. The biological diagnosis of 5 $\alpha$ -RD is supported mainly by an increased plasma T/DHT ratio.

From a biochemical point of view, the decrease in 5 $\alpha$ -reductase activity in the intact genital skin fibroblasts supports the diagnosis of 5 $\alpha$ -RD, but enzymatic activity is sometimes in the normal range. The decreased activity in sonicated cell extracts at acidic pH provides strong evidence that the mutation resulted in a loss of type 2 enzyme activity.

Isolation and sequencing of the cDNA encoding 5 $\alpha$ -reductase type 2 provides the molecular tools required for definition of the gene abnormalities responsible for 5 $\alpha$ -RD.

The management of 5 $\alpha$ -RD is primarily dependent on the phenotypic findings and gender at the time of diagnosis. Given the severe defect of the external genitalia, most newborns are raised as female. Gonadectomy should be performed early to prevent masculinization along with vaginoplasty and clitoral reduction. If the diagnosis is made during puberty, one can consider raising the child as male.

#### **When Testosterone Does Not Increase**

If testosterone does not rise after the therapeutic trial, either testicular dysgenesis (which is an anomaly in testicular determination) or a disturbance in testicular biosynthesis may be responsible. Genitography is useful for differential diagnosis. When a vagina and uterine cavity are found, the diagnosis of testicular dysgenesis is likely. When a blind vagina without a uterine cavity is found, the diagnosis of disturbance in testicular biosynthesis is made.

#### **Abnormal Testicular Determination**

- *Dysgenetic testis.* Newborns with dysgenetic testis present with bilateral dysgenetic testes, persistent Müllerian structures, cryptorchidism, and inadequate virilization. This disorder shows wide clinical heterogeneity. Because the uterus is present, the sex of rearing should be female and gonadectomy is recommended, as for mixed gonadal dysgenesis. These patients should be screened routinely for tumor.

- *Defect of testis maintenance.* Bilateral vanishing testis (or embryonic testicular regression) is characterized

by an XY karyotype and absent or rudimentary testes. The syndrome entails the presence of testes that vanish during embryogenesis, although the etiology is unclear. The regression of the testes *in utero* may be due to genetic mutation, a teratogen factor, or bilateral torsion. Clinically, the syndrome encompasses a spectrum of phenotypes, ranging in severity from genital ambiguity to a male phenotype with an empty scrotum. Patients with a defect in testis maintenance will be managed according to their position in the clinical spectrum of the disorder. In true agonadism, external genitalia are ambiguous and Müllerian derivatives are absent or rudimentary due to complete or partial anti-Müllerian hormone secretion without secretion of testosterone. Patients with rudimentary testis have a male phenotype with micropenis and small atrophic testis with pre-Sertoli and Leydig cells. Some patients present with perineal hypospadias and persistent Müllerian duct structures. Congenital anorchia is characterized by the complete absence of testicular tissue at birth but with normal male sexual differentiation without Müllerian structures. Patients with ambiguous genitalia require meticulous assessment to determine the optimal sex of rearing. In the absence of palpable gonads, measurement of basal- or hCG-stimulated testosterone secretion above the female range is informative, but laparotomy or celioscopy followed by histological analysis of the gonads is absolutely essential. In general, the sex of rearing should be male with testosterone replacement therapy, which will lead to normal puberty and sexual function. Phenotypic males require long-term androgen replacement, beginning at the time of expected puberty.

#### **Leydig Cell Hypoplasia**

Male pseudohermaphrodites with Leydig cell hypoplasia have impaired Leydig cell differentiation and testosterone production. The phenotype is usually female, although it may be ambiguous. In all cases, Müllerian structures are absent. Inhibiting mutations in the LH receptor gene have been reported in these patients.

#### **Defect in Testosterone Synthesis**

Decreased androgen production caused by an alteration in the enzymes involved in the testosterone biosynthesis pathway is another cause of male pseudohermaphroditism. On the other hand, increased fetal production of androgens results in female pseudohermaphroditism.

- *Congenital lipoid adrenal hyperplasia.* Congenital lipoid adrenal hyperplasia (CLAH) is a rare disease

characterized by a defect in the synthesis of the three classes of steroid hormones, resulting in severe salt wasting and a female phenotype. Few mutations in humans have been found in the gene encoding for P450scc, which is the principal candidate gene. Recently, the gene responsible for CLAH was cloned and validated by nonsense mutation. This gene encodes for the StAR (steroidogenic acute regulatory) protein, which is thought to be responsible for the transport of cholesterol to the inner membrane of mitochondria and, thereby, to the P450scc enzyme complex.

- *17 $\alpha$ -hydroxylase deficiency.* Defects in P450c17 lead to male pseudohermaphroditism with various degrees of ambiguous genitalia. This is frequently a severe form and is most often diagnosed at puberty, with a female phenotype associated with hypertension. Cytochrome P450c17 catalyzes the transformation of progesterone and pregnenolone into 17-OH progesterone and 17-OH pregnenolone (17 $\alpha$ -hydroxylase activity), respectively, and then into dehydroepiandrosterone and  $\Delta^4$ -androstenedione (17-20 lyase activity). The gene encoding for this enzymatic complex, CYP17, is located on chromosome 10q24-q25. Several different mutations have been reported in the CYP17 gene leading to either a complete or partial form of the disease.

- *3 $\beta$ -hydroxysteroid dehydrogenase deficiency.* Defects in the function of 3 $\beta$ -HSD result in 46,XY individuals with male pseudohermaphroditism who sometimes show salt wasting in the classic form. Nearly 15 mutations of the type II 3 $\beta$ -HSD enzyme have been reported. However, no mutation of type I has been found, and this explains how virilization can occur in 46,XX individuals by peripheral nonsteroidogenic conversion of elevated testosterone precursors.

- *17 $\beta$ -hydroxysteroid dehydrogenase deficiency.* Type 3 17 $\beta$ -HSD deficiency is a rare autosomal recessive cause of male pseudohermaphroditism. The typical patient is a 46,XY male born with female external genitalia and testes located in the inguinal canals or labia majora. This disorder is particularly puzzling. The deficiency in testosterone synthesis and the defect in virilization, both of which are usually more complete during embryogenesis than in later life, contrast with the well-differentiated Wolffian duct structures, suggesting that androgen acts *in utero* by an alternate mechanism in these tissues. Substantial virilization is seen at puberty in association with elevated levels of androstenedione but low to normal plasma levels of testosterone. As in the case of 5 $\alpha$ -RD, pubertal virilization may result from extraglandular testosterone formation due to peripheral conversion of

increased testicular androstenedione by unaffected 17 $\beta$ -HSD isoenzymes. Missense and nonsense mutations, splice junction abnormalities, and a small deletion resulting in a frame shift have been described, among other gene alterations. Expression of mutant enzymes after site-directed mutagenesis showed that the missense mutations caused nearly complete loss of enzymatic activity.

Although affected newborns are generally considered to be female, the choice of sex of rearing will be greatly influenced by family values and cultural background. When female sex of rearing is maintained, we strongly urge that orchidectomy be carried out during infancy or childhood. When diagnosis is not made before puberty, a gender change to male at that time is acceptable.

Male pseudohermaphroditism may, in some cases, be part of a multiple malformation syndrome. In other cases, etiology cannot be determined, and the so-called idiopathic pseudohermaphroditism should raise the suspicion of environmental contamination by pesticides during gestation.

#### **The 46,XY/XO Newborn**

For the 46,XY/XO newborn, the diagnosis is mixed gonadal dysgenesis. The most common karyotype in mixed gonadal dysgenesis is 45,XO/46,XY, but other mosaics have been reported with structurally abnormal or normal Y chromosome. The characteristics include a unilateral testis that is often intra-abdominal, a contralateral streak gonad, and persistent Müllerian duct structures. Because affected patients are at great risk for gonadal tumor, the gonads should be removed so that the patients can be reared as females. However, when the abnormality of the external genitals is minor, parents may perceive their children as boys. Male sex of rearing will then impose lifelong surveillance of the gonads. In any case, mixed gonadal dysgenesis is associated with varying degrees of inadequate masculinization, and such males would be infertile. It should be remembered that the distinction between mixed gonadal dysgenesis and Turner syndrome with Y material is unclear.

### **SEX ASSIGNMENT OF THE INTERSEX CHILD AT BIRTH**

When a child is born with ambiguous genitalia, the medical team must mobilize for a neonatal emergency. In addition to the urgent need to rule out life-threatening conditions such as salt wasting, determination of the most appropriate sex for rearing must be made as rapidly as possible. It cannot be emphasized enough

that the parents desperately need to know whether their baby is a girl or a boy. The unambiguous designation of a baby's sex is a key step in the birth process, and for the parents of an intersex neonate, the period of waiting to learn their baby's sex is often agonizing. Birth is the meeting of parents' dreams and a real infant, alive and present in the world. When a child is born, one of the first announcements is whether the mother has delivered a boy or a girl. We can easily imagine the profound distress and helplessness of parents who are confronted with the intersex status of their newborn.

The clinical examination provides an assessment of the degree of undervirilization and the presence of gonads. Biological assessments are mandatory for plasma 17-OHP and the SRY gene.

Once the investigations have been concluded, the medical team should be ready to undertake its greatest responsibility: the assignment of sex for rearing. This decision must be guided by three parameters: the anatomical condition and functional abilities of the genitalia, the etiology of the genital malformation, and family considerations (e.g., cultural factors, religious convictions). The choice of declared sex must be—and this bears repeating—the result of full discussion among all protagonists.

In cases of female pseudohermaphroditism, the newborn with ambiguous genitalia should always be declared to be of the female sex. With normal ovaries and uterus, the female pseudohermaphrodite is potentially capable of bearing children.

In cases of male pseudohermaphroditism, great care should be taken in the declaration of the male sex. Major considerations will be the potential for reconstructive surgery, the probability of pubertal virilization, and the “programmed” response of the external genitalia to exogenous and endogenous testosterone. On the other hand, the presence of testicular tissue is not an essential factor in this decision.

In cases of gonadal dysgenesis, sex assignment will be based on several criteria. A defect in 5 $\alpha$ -reductase is an indication for male sex for rearing because pubertal virilization will lead to penile development (although it will be subnormal), normal pubic hair development, and the acquisition of male sexual identity. In contrast, for inborn errors of testosterone biosynthesis, female orientation is advisable if effective male reconstructive surgery appears to be highly unlikely. When a vagina and uterus are present, female orientation is also preferable if it appears that vaginoplasty can be easily accomplished.

In cases of androgen resistance, a female orientation is unquestionably correct for complete resistance, and

it is preferable in cases of partial resistance confirmed by therapeutic testing.

In true hermaphroditisms, which are very rare, female sex assignment is to be preferred because ovarian function may be preserved.

Castration is indicated for male pseudohermaphrodites reared as girls. When they are reared as boys, careful gonadal follow-up throughout life is crucial. Because the cumulative risk of gonadal tumor in the intersex child is present even before puberty, we advocate castration in the XY intersex infant with testicular dysgenesis.

## CONCLUSION

A methodical clinical examination and hormonal, radiographic, molecular, and genetic investigations are the bases for determining the diagnosis, etiology, and optimal management of ambiguous genitalia in the newborn. In conversations with family members, terms such as “ambiguous genitalia” and “pseudohermaphrodite” should be avoided and more neutral terms such as “genital malformation” should be substituted. The family should be informed of the various difficulties and therapeutic options, but excessive detail is unnecessary. Once a decision for the assignment of sex is made, it should be definitive for the parents. The baby will hereafter be referred to by name and as either “he” or “she” by all staff to help the parents fix their child's gender firmly in their minds.

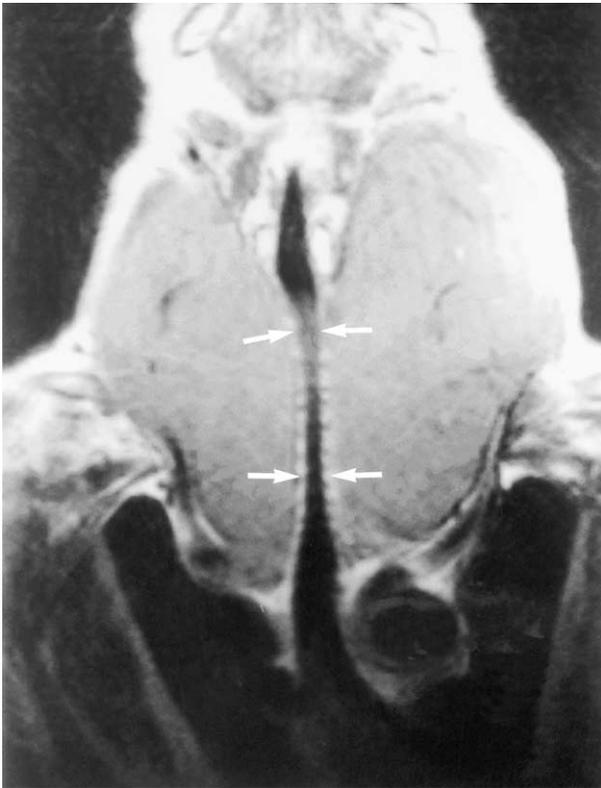
Because of the far-ranging consequences, the assignment of sex for rearing should never be the decision of a single physician. The entire multidisciplinary team should remain involved in every step of the diagnostic procedure, the choice of sex for rearing, and the treatment strategy. The team also has an important role in guiding the family and ensuring family agreement with and support for the decision, which is the sole means of guaranteeing that the child will be raised with an unambiguous sexual identity.

## See Also the Following Articles

Androgen Insensitivity Syndrome • Androgens, Gender and Brain Differentiation • Congenital Lipoid Adrenal Hyperplasia • Gender Assignment and Psychosocial Management • Genes and Gene Defects Affecting Gonadal Development and Sex Determination • 3 $\beta$ -Hydroxysteroid Dehydrogenase Deficiency • 17 $\alpha$ -Hydroxylase/17,20-Lyase Deficiency • Sexual Maturation, Female • Sexual Maturation, Male • Undescended Testes

## Further Reading

- Dewing, P., Bernard, P., and Vilain, E. (2002). Disorders of gonadal development. *Sem. Reprod. Med.* **20**, 189–197.
- Hughes, I. A. (2002). Congenital adrenal hyperplasia: 21-hydroxylase deficiency in the newborn and during infancy. *Sem. Reprod. Med.* **20**, 229–241.
- Migeon, C. J., Berkovitz, G. D., and Brown, T. R. (1994). Sexual differentiation and ambiguity. In “The Diagnosis and Treatment of Endocrine Disorders in Childhood and Adolescence” (M. S. Kappy, R. M. Blizzard, and C. J. Migeon, eds.), pp. 573–681. Charles C Thomas, Springfield, IL.
- Miller, W. (2002). Disorders of androgen biosynthesis. *Sem. Reprod. Med.* **20**, 205–215.
- New, M. I. (1995). Steroid 21-hydroxylase deficiency (congenital adrenal hyperplasia). *Am. J. Med.* **98**, S2–S8.
- Peter, M. (2002). Congenital adrenal hyperplasia: 11 $\beta$ -hydroxylase deficiency. *Sem. Reprod. Med.* **20**, 249–254.
- Quigley, C. A., Debellis, A., Marschke, K. B., Elawady, M. K., Wilson, E. M., and French, F. S. (1995). Androgen receptor defects: Historical, clinical, and molecular perspectives. *Endocrine Rev.* **16**, 271–321.
- Simard, J., Moisan, A. M., and Morel, Y. (2002). Congenital adrenal hyperplasia due to 3 $\beta$ -hydroxysteroid dehydrogenase/ $\Delta^5$ - $\Delta^4$  isomerase deficiency. *Sem. Reprod. Med.* **20**, 255–276.
- Sultan, C., Lumbroso, S., Paris, F., Jeandel, C., Terouanne, B., Belon, C., Andran, F., Poujol, N., Georget, V., Gobinet, J., Jalaguier, S., Auzou, G., and Nicolas, J. C. (2002). Disorders of androgen action. *Sem. Reprod. Med.* **20**, 217–227.
- Sultan, C., Lumbroso, S., Poujol, N., and Lobaccaro, J. (1997). Intrauterine virilization of female fetuses. In “Androgen Excess Disorders in Women” (R. Azziz, J. E. Nestler, and D. Dewailly, eds.), pp. 593–599. Lippincott–Raven, Philadelphia.
- Sultan, C., and Savage, M. O. (1998). Intersex states. In “Clinical Endocrinology” (A. B. Grossman, ed.), pp. 795–809. Blackwell Science, London.
- Themmen, A., and Verhoef-Post, M. (2002). LH receptor defects. *Sem. Reprod. Med.* **20**, 199–204.
- Toppari, J. (2002). Environmental endocrine disruptors and disorders of sexual differentiation. *Sem. Reprod. Med.* **20**, 305–311.
- White, P. C., Curnow, K. M., and Pascoe, L. (1994). Disorders of steroid 11 $\beta$ -hydroxylase isozymes. *Endocrine Rev.* **15**, 421–438.
- Wilson, J. D., Griffin, J. E., and Russell, D. W. (1993). Steroid 5 $\alpha$ -reductase 2 deficiency. *Endocrine Rev.* **14**, 577–593.



**Figure 1** Coronal T1-weighted MRI of a goiter with intrathoracic growth and bilateral tracheal compression (arrows).

of one or both of the *nervi recurrentes*, and it will result in hoarseness and dyspnea. Paralysis of the *nervus phrenicus* and Horner's syndrome are rare complications of goiters.

### Laboratory Tests

The level of thyroid-stimulating hormone (TSH) in blood should be measured for all patients with goiter to either determine or exclude (subclinical) hyperthyroidism or hypothyroidism. If the TSH concentration is low, it is necessary to measure the level of free thyroxine (T<sub>4</sub>) to distinguish between subclinical or manifest hyperthyroidism. If the concentration of TSH is low and the concentration of free T<sub>4</sub> is normal, triiodothyronine (T<sub>3</sub>) toxicosis must be excluded by measuring the serum level of T<sub>3</sub>. If the concentration of TSH is high and the concentration of free T<sub>4</sub> is low, chronic autoimmune thyroiditis should be considered as a possible cause of the thyroid enlargement. This diagnosis can be confirmed by measuring the level of antibodies against microsomes (thyroperoxidase) in serum and/or by fine-needle aspiration biopsy.

### Imaging Diagnostics and Lung Function Tests

The presence of tracheal compression and intrathoracic growth of the goiter should be considered in all patients with goiter, especially in the case of a large goiter. Taking X-rays of the thorax and trachea is a simple but not very sensitive screening method for trachea compression. Lung function tests can provide additional indications for tracheal compression, especially tests with so-called flow-volume loops. Computed tomography (CT) and magnetic resonance imaging (MRI) are relatively expensive but very sensitive methods to demonstrate tracheal compression and intrathoracic growth of a goiter. Iodine-containing contrast agents should not be used for CT because of the risk of inducing hyperthyroidism. If it is absolutely necessary to use iodine-containing contrast agents in order to show venous obstruction with phlebography or CT, pretreatment with a thyrostatic drug or perchlorate is needed. Ultrasonography of the thyroid gives detailed information about the number and size of the nodules within the goiter, but this investigation is not indicated routinely. X-rays of the esophagus can show esophageal compression by a goiter, but these are rarely necessary. Thyroid scintigraphy is also not indicated routinely. Thyroid scintigraphy and the measurement of radioiodine uptake in the thyroid are useful when radioiodine treatment is considered to reduce the goiter volume. If radioiodine uptake in the thyroid is low or if large parts of the thyroid hardly take up any radioiodine, treatment with radioiodine is not very effective.

### Fine-Needle Aspiration Biopsy

In patients with nontoxic goiter, clinically significant thyroid carcinoma is rare, given the high prevalence of nontoxic goiter and the very low incidence of thyroid carcinoma in the population. For this reason, fine-needle aspiration biopsy is not indicated routinely. Biopsy is needed, however, if nodules are fast growing or if they have a significantly firmer consistency than other nodules.

### THERAPY

Nontoxic goiters usually grow slowly. The presence of a diffuse or multinodular goiter is not an indication for treatment. The most important indications for treatment are compression of the trachea and esophagus and obstruction of the large veins. Treatment should also be considered in the case of progressive

growth of either the whole goiter or individual nodules, particularly when the goiter shows significant intrathoracic growth. Indeed, it is not possible to palpate the intrathoracic part of the goiter or take a biopsy from it. In addition, bleeding within a nodule or a cyst can result in acute and life-threatening compression of the trachea. Sometimes, treatment is indicated for cosmetic problems.

The main therapeutic options are thyroidectomy, administration of radioiodine, and treatment with thyroxine. Iodine supplementation can be tried for patients with a small diffuse goiter caused by iodine deficiency. Iodine supplementation is not advised for patients with multinodular goiter, even if it is caused by iodine deficiency. The administration of iodine to these patients is rarely effective and can even induce hyperthyroidism.

### Thyroidectomy

Surgical treatment of goiter results in fast decompression of vital structures and provides tissue material for pathological examination. In most cases, bilateral subtotal thyroidectomy is performed, including the removal of all macroscopically abnormal tissue. Almost all nontoxic goiters can be removed through an incision in the neck, even goiters that have grown substantially into the thoracic cavity. The most common complication of thyroid surgery is tracheal obstruction caused by bleeding, damage to the nervus recurrens, or tracheomalacia. Other postsurgical complications are hypoparathyroidism, change in voice caused by damage to the nervus laryngeus superior, and hypothyroidism. The incidence of postsurgical hypothyroidism is dependent on the extensiveness of the operation. The highest morbidity is found in patients with very large goiters and those who have had secondary surgery. The mortality after thyroid surgery on patients with nontoxic goiters is less than 1%.

### Radioactive Iodine

During the past decade, it has been conclusively shown that administration of radioactive iodine (radioiodine) is an effective treatment that results in reduction of the thyroid volume in more than 90% of patients with nontoxic goiter. Usually, a single dose of approximately 100  $\mu\text{Ci}$  (3.7 MBq) radioactive iodine ( $^{131}\text{I}$ ) per gram of thyroid tissue is administered. Twelve to 18 months after treatment with radioiodine, the thyroid volume in patients with a diffuse goiter is reduced by 50–60%. Radioiodine treatment of patients with a multinodular goiter results in a reduction of thyroid volume of approximately 40% after

1 year and 50% after 3–5 years. There is a positive correlation between the reduction of the goiter volume and the administered dose of radioiodine per gram of thyroid tissue and a negative correlation between the reduction of the goiter volume and the goiter volume before treatment. In addition to the reduction in thyroid volume, the treatment also results in a reduction of compression symptoms for a majority of patients.

Early side effects, such as radiation thyroiditis or radiation esophagitis, are generally mild and temporary. Exacerbation of compression symptoms following treatment with radioiodine is rare. Therefore, the administration of glucocorticoids is not routinely indicated. The most important late complication, which appears months after radioiodine treatment in 5% of patients, is the development of autoimmune (Graves') hyperthyroidism, which is probably induced by thyroid antigens that are released as a result of radiation. The prevalence of hypothyroidism after treatment is 20–50% after 5 years. Recurrent goiter growth after 3–5 years occurs in approximately 10% of patients. A second treatment with radioiodine can be effective in these cases.

Treatment with high doses of radioactive iodine is potentially carcinogenic. Risk assessments show that the risk of developing a malignancy that is caused by treatment with a high dosage of radioiodine in patients 65 years of age or older is approximately 0.5%, which is equal to the surgical mortality after subtotal thyroidectomy.

Treatment of nontoxic goiter with radioiodine is mainly given to older patients, particularly those who have an increased surgical risk and those who refuse surgery. However, radioiodine is also a good alternative to surgery for younger patients, provided that the dosage of radioiodine is relatively low (only patients with a small goiter and sufficient radioiodine uptake).

### Thyroxine

TSH is the most important growth stimulator of normal thyroid tissue. Treatment of patients with nontoxic goiter is based on the hypothesis that the growth of a goiter is also TSH dependent and that suppression of TSH secretion will result in the reduction of the goiter volume or at least prevent further growth of the goiter.

Treatment with thyroxine is an effective method to reduce the thyroid volume in patients with diffuse goiter. Only one placebo-controlled, randomized trial of patients with multinodular goiters has been published. In this trial, thyroid volume was reduced in

58% of patients (by an average of 25%). Only patients with relatively small goiters were treated. The effectiveness of thyroxine treatment in patients with larger multinodular goiters is probably considerably lower. Thyroxine treatment of patients with a multinodular goiter and a low serum TSH level is not advisable because this treatment can cause overt hyperthyroidism, and a reduction of the goiter volume is not to be expected in the case of low TSH levels.

Long-term thyroxine treatment with doses that reduce the serum TSH concentration to subnormal level reduces bone density, particularly in women after menopause. It also causes left ventricular hypertrophy and possibly cardiac dysfunction and arrhythmia.

### Choice of Treatment

The pros and cons of the three treatments—thyroidectomy, radioiodine, and thyroxine—should be weighed carefully before advice is given about the treatment of a patient with nontoxic goiter. Thyroidectomy is the standard treatment, especially for young and otherwise healthy patients and when prompt decompression of vital structures is required. Radioiodine treatment is an attractive alternative to surgery for older patients and for patients with cardiopulmonary diseases or recurrent goiters. There are few indications for thyroxine treatment of patients with nontoxic goiters. Thyroxine treatment can be tried in young patients with small, diffuse goiter and a normal TSH level. In this case, the goal is to obtain a TSH concentration between 0.1 and 0.5 mU/liter.

### Acknowledgments

This article is a translation of an article originally published in Dutch (*Ned. Tijdschr. Geneesk.* **144**, 1623–1627, 2000); it appears with permission.

### See Also the Following Articles

Goitrogens, Environmental • Iodine Deficiency • Iodine, Radioactive • Thyroid Hormone Action • Thyroidectomy • Toxic Multinodular Goiter • TSH Function and Secretion

### Further Reading

- Hegedüs, L., Hansen, B. M., Knudsen, N., and Hansen, J. M. (1988). Reduction of size of thyroid with radioactive iodine in multinodular non-toxic goitre. *Br. Med. J.* **297**, 661–662.
- Hermus, A. R., and Huysmans, D. A. (1998). Treatment of benign nodular thyroid disease. *N. Engl. J. Med.* **338**, 1438–1447.
- Huysmans, D. A. K. C., Hermus, A. R. M. M., Corstens, F. H. M., Barentsz, J. O., and Kloppenborg, P. W. C. (1994). Large, compressive goiters treated with radioiodine. *Ann. Intern. Med.* **121**, 757–762.
- Kaplan, E. L., Shukla, M., Hara, H., and Ito, K. (1995). Surgery of the thyroid. In “Endocrinology” (L. J. DeGroot, M. Besser, and H. G. Burger, *et al.*, eds.), 3rd ed., Vol. 1, pp. 900–914. Saunders, Philadelphia.
- Nygaard, B., Hegedüs, L., Gervil, M., Hjalgrim, H., Søe-Jensen, P., and Hansen, J. M. (1993). Radioiodine treatment of multinodular non-toxic goitre. *Br. Med. J.* **307**, 828–832.
- Nygaard, B., Knudsen, J. H., Hegedüs, L., Scient, A. V., and Hansen, J. E. M. (1997). Thyrotropin receptor antibodies and Graves’ disease, a side-effect of <sup>131</sup>I treatment in patients with nontoxic goiter. *J. Clin. Endocrinol. Metab.* **82**, 2926–2930.
- Toft, A. D. (1994). Thyroxine therapy. *N. Engl. J. Med.* **331**, 174–180.
- Vanderpump, M. P. J., Tunbridge, W. M. G., French, J. M., *et al.* (1995). The incidence of thyroid disorders in the community: A twenty-year follow-up of the Whickham Survey. *Clin. Endocrinol.* **43**, 55–68.
- Wesche, M. F., Tiel-van Buul, M. M., Lips, P., Smits, N. J., and Wiersinga, W. M. (2001). A randomized trial comparing levothyroxine with radioactive iodine in the treatment of sporadic nontoxic goiter. *J. Clin. Endocrinol. Metab.* **86**, 998–1005.
- Wiersinga, W. M., Drexhage, H. A., Smits, N. J., and Touber, J. L. (1990). Comparison of placebo with L-thyroxine alone or with carbimazole for treatment of sporadic non-toxic goitre. *Lancet* **336**, 193–197.

test will become available. Until that time, Noonan syndrome will remain a clinical diagnosis.

## Diagnosis

The diagnosis of Noonan syndrome is based on clinical findings. Characteristic facies are helpful in diagnosis but change with age. In the newborn, Noonan syndrome is difficult to diagnosis by facial appearance alone. The forehead is often sloping and the ears may be thick, broad, and posteriorly angulated. Newborns often have redundant nuchal skin and experience significant weight loss due to the edema present at birth. In infancy, from neonatal to approximately 2 years of age, the head often appears relatively large. The malar eminencies are flat and the eyes are more prominent and round. The nasal bridge is depressed. The neck appears short but is no longer webbed. At 2 years of age, the body appears more stocky and the chest deformity appears more prominent. As childhood progresses, the facial appearance becomes more coarse and assumes a more triangular shape as the chin lengthens. The eyes become less prominent and ptosis becomes more apparent. The neck appears longer and the low hair line and webbing may become more obvious. The teenager and young adult develop triangular facial features that become much sharper. The nose has a pinched root and a thin, high bridge. The older adult often has prominent nasal labial folds and a high anterior hair line, and the skin often appears transparent and wrinkle (Fig. 1).

Although the majority of patients with Noonan syndrome have an unremarkable prenatal history, in one-third of patients the pregnancy is complicated by polyhydramnios. Since fetal ultrasound has become

more widely available, a number of fetuses later diagnosed with Noonan syndrome have been reported to show increased nuchal fluid and hydrops as early as 14 weeks of gestation. Follow-up studies frequently revealed resolution of the nuchal fluid. Excessive weight loss common in the first week of life suggests fetal edema may be quite common.

Short stature is noted in more than 80% of patients with Noonan syndrome. The cause of the short stature is unknown. After an apparently normal birth history, many patients develop feeding problems in early infancy. Approximately 35% have mild to moderate feeding problems, mainly vomiting, whereas nearly 25% develop severe feeding problems, often requiring tube feedings. Studies have indicated an increased incidence of gastroesophageal reflux and poor gut motility. In general, the feeding problems resolve later in infancy.

Children with Noonan syndrome often present to the endocrinologist with short stature, delayed puberty, and undescended testes in males. Weight and length are usually normal at birth. Height drops off within the first few months. In general, there is a 2-year delay between bone age and chronological age. This results in a delay in the onset of puberty. Continued growth may occur until the early twenties. Height and weight are usually below the third percentile, with a mean final adult height of 162.5 cm in males and 151 cm in females. In approximately half of males, one or both testes fail to descend. Delayed puberty is frequent in females. Females seem to possess normal fertility, whereas males, as expected due to undescended testes, appear to have decreased fertility.

Developmental delay is frequent. Motor delay may be partially attributed to muscular hypotonia. In a study by van der Burgt *et al.*, individual IQ scores varied between 48 and 130. The overall experience suggests that approximately one-third of children will have some degree of mental retardation or learning disability. Graduation from college and achievement of Ph.D. degrees have been reported. Conductive hearing loss is relatively common, and very frequent visual problems may contribute to learning disability.

Almost every system of the body may be affected in Noonan syndrome. Eye findings are very frequent and include apparent hypertelorism, ptosis, refracted errors, strabismus, and amblyopia. An occasional patient will have a coloboma. More than 90% of patients with Noonan syndrome have a chest deformity, such as a pectus carinatum or pectus excavatum. Scoliosis occurs in 10–15%, as does talipes equinovarus. Muscular hypotonia is common and often improves



Figure 1 Mother with Noonan syndrome and her affected infant.

with time. The chest is often shield-like with widely spaced nipples. Unexplained peripheral neuropathy has been seen. Poor coordination is reported by some parents. Spina bifida occulta, tethered cord, and Arnold–Chiari malformation have all been reported.

Although curly hair is often a feature, some patients have both sparse hair and eyebrows. Nevi and freckles are common. A number of patients have findings characteristic of both neurofibromatosis and Noonan syndrome. Children have a tendency to form extensive keloids following surgical procedures. Hepatosplenomegaly, usually unexplained, is present in approximately 25%. Bleeding problems and easy bruising have been frequently noted. A variety of abnormal clotting factors have been reported, including low levels of factor 11 and factor 8, thrombocytopenia, and platelet function defects. Low levels of a wide variety of clotting factors with no specific patterns have also been reported. There may be an increased risk of leukemia in patients with Noonan syndrome, and there have been several reported cases of neuroblastoma. Lymphatic abnormalities occur in less than 20% of patients in Noonan syndrome but may cause serious problems. Intestinal lymphagectasia leading to protein-losing enteropathy has been reported as well as pulmonary lymphagectasia. Spontaneous chylothorax has been reported, and persistent pleural effusions following cardiac surgery is a known risk in patients with Noonan syndrome.

More than 80% of patients with Noonan syndrome have some kind of cardiovascular abnormality. A dysplastic, often stenotic pulmonary valve is the most characteristic lesion of Noonan syndrome, but virtually every type of cardiac defect has been described. Atrial septal defect, branch pulmonary artery stenoses, ventricular septal defect, and tetralogy of Fallot are among those that are frequently reported. Valvar aortic stenosis, subaortic stenosis, patent ductus arteriosus, and coarctation of the aorta have also been reported. In addition, hypertrophic cardiomyopathy, both obstructive and nonobstructive, occurs in 20–30% of patients. This myocardial hypertrophy may be noted at birth or may develop later in infancy or later childhood. Unlike the nonsyndromic, familial hypertrophic cardiomyopathy, patients with Noonan syndrome often have involvement of both the right and the left ventricle. Microscopic examination, however, reveals similar findings in both forms—mainly muscle disarray and thick-walled coronary arteries. In addition to dysplasia of the pulmonary valves, all valves may be dysplastic and mitral valve prolapse is relatively common, occurring with other cardiac

defects or as an isolated finding. An unusual electrocardiogram with an indeterminate left axis deviation and a dominant S wave over the entire precordium is frequent but not clearly related to any specific cardiac malformation. The cause of the electrocardiographic finding is unknown, but it supports the diagnosis of Noonan syndrome.

## DIFFERENTIAL DIAGNOSIS

Differential diagnosis should include Turner syndrome, particularly for a female who has a left-sided cardiac lesion. Cardiofaciocutaneous syndrome and Costello syndrome may be very difficult to diagnose in early infancy, although with time differences become more apparent. Patients with both Watson syndrome and Leopard syndrome may be indistinguishable from those with Noonan syndrome until a specific diagnostic test becomes available. There are children who show evidence of both neurofibromatosis and Noonan syndrome. Careful history to eliminate possible alcohol abuse or other tetragons in the mother as well as chromosome studies are helpful in distinguishing Noonan syndrome from some of the other conditions that must also be considered in the differential diagnosis.

## MANAGEMENT

A child with Noonan syndrome will have special needs. Fortunately, many children are mildly affected and may essentially be treated as “normal.” Because of the high incidence of abnormal eye and cardiac findings, these children should have careful eye and cardiac evaluations. In addition, hearing should be tested. Careful developmental assessment should be carried out before the child begins school so that any potential learning disabilities can be recognized early before the child develops problems in school.

Short stature is of concern to both patients and families. Some studies have been carried out to evaluate the efficacy of growth hormone therapy. No consistent abnormalities in the secretory dynamics of growth hormone have been reported in these children. There has been no consistent relationship between growth hormone studies and the response to growth hormone treatment. Studies by Romano *et al.* and Noordam *et al.* demonstrated a positive effect of growth hormone treatment on linear growth. Romano *et al.*'s study revealed sustained growth for

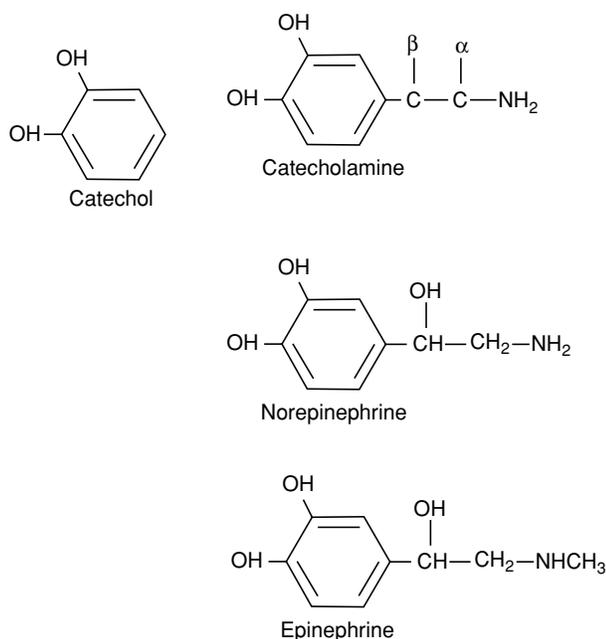
4 years with continued growth hormone therapy. Unfortunately, both studies demonstrated an advance in bone age relative to height age. Clearly, children treated with growth hormone therapy reach final growth earlier, but it is not clear whether their final height is significantly increased. Unfortunately, there are very limited data on final height of patients with Noonan syndrome who have been followed serially. Fortunately, studies indicate that the use of growth hormone treatment for Noonan syndrome appears to be relatively safe. There is no evidence that growth hormone adversely affects the cardiovascular system in children with hypertrophic cardiomyopathy.

### See Also the Following Articles

Short Stature and Chromosomal Abnormalities • Turner Syndrome

### Further Reading

- Allanson, J. E., Hall, J. G., Hughes, H. E., Preus, M., and Witt, R. D. (1985). Noonan syndrome: The changing phenotype. *Am. J. Med. Genet.* **21**, 507–514.
- Jamieson, C. R., *et al.* (1994). Mapping a gene for Noonan syndrome to the long arm of chromosome 12. *Nature Genet.* **8**, 357–360.
- Noonan, J. A. (1968). Hypertelorism with Turner phenotype. *Am. J. Dis. Child.* **116**, 373–380.
- Noonan, J. A. (1994). Noonan syndrome: An update and review for the primary pediatrician. *Clin. Pediatr.* **8**, 548–555.
- Noordam, C., van der Burgt, I., Sengers, R. C. A., Delemans-van de Waal, H. A., and Otten, B. J. (2001). Growth hormone treatment in children with Noonan's syndrome: 4 year results of a partly controlled trial. *Activa Paediatr.* **90**, 889–894.
- Romano, A. A., *et al.* (1996). Growth hormone treatment in Noonan syndrome: The National Cooperative Growth Study experience. *J. Pediatr.* **128**, S18–S21.
- Tartaglia, M., *et al.* (2001). Mutations in PTPN11, encoding the protein tyrosine phosphatase SHP-2, cause Noonan syndrome. *Nature Genet.* **29**, 465–468.



**Figure 1** Structures of catechol, a generic catecholamine (an ethanolamine with  $\alpha$ - and  $\beta$ -carbons, as indicated, linked to a catechol moiety), norepinephrine, and epinephrine (*N*-methyl-norepinephrine).

monoamine oxidases, catechol-*o*-methyltransferases) and plasma membrane uptake carriers, to lower extracellular concentrations and terminate the action of norepinephrine.

### ADRENERGIC RECEPTORS: G PROTEIN-COUPLED RECEPTORS THAT LINK VIA G PROTEINS TO EFFECTOR MOLECULES

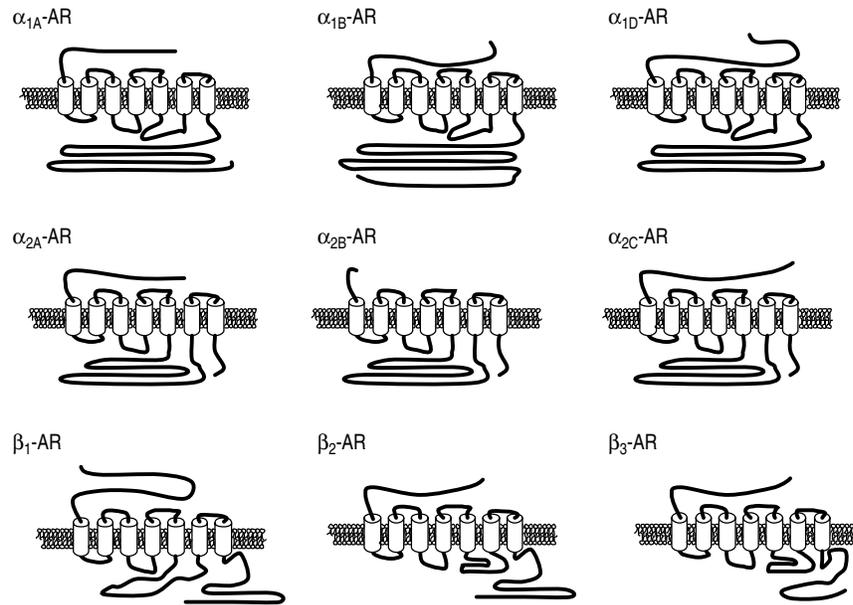
Plasma membrane receptors that recognize and respond to norepinephrine are of three major classes:  $\alpha_1$ -,  $\alpha_2$ -, and  $\beta$ -adrenergic receptors (adrenoceptors), each of which has three subtypes:  $\alpha_{1A,B,D}$ ;  $\alpha_{2A,B,C}$ ; and  $\beta_{1,2,3}$  (Figs. 2 and 3). In humans, the receptors are encoded by separate genes with distinct chromosomal locations (Table I). All of the adrenergic receptors are members of the G protein-coupled receptor (GPCR) superfamily and couple to one or more heterotrimeric ( $\alpha$ ,  $\beta$ ,  $\gamma$ ) GTP-binding (G) proteins (see Table II). Each of the three major classes of adrenergic receptors preferentially couples to G proteins of a particular type:  $\alpha_1$ -adrenergic receptors to  $G_{q/11}$ ,  $\alpha_2$ -adrenergic receptors to  $G_i$ , and  $\beta$ -adrenergic receptors to  $G_s$ . When the G proteins are activated by norepinephrine or another agonist acting at a particular adrenergic

receptor, each of the G proteins preferentially couples to and regulates specific effector enzymes and channels:  $G_q$  couples to phospholipase  $C_\beta$ , which hydrolyzes phosphatidylinositol 4,5-bis-phosphate to inositol 1,4,5-triphosphate and diacylglycerol, which regulate, respectively, increases in cytosolic  $[Ca^{2+}]$  via release from intracellular endoplasmic reticulum stores and activation of protein kinase C;  $G_i$  regulates the inhibition of adenylyl cyclase activity (and the decrease in intracellular cyclic AMP levels), activation of inwardly rectifying potassium channels, and inhibition of voltage-dependent calcium channels; and  $G_s$  primarily regulates the stimulation of adenylyl cyclase activity and also the increase in activity of voltage-dependent calcium channels in some tissues. All the adrenergic receptors regulate the phosphorylation state of intracellular proteins through the action of signaling kinases: protein kinase C for  $\alpha_1$ -adrenergic receptors and protein kinase A for  $\alpha_2$ - and  $\beta$ -adrenergic receptors. Other kinases may be more distally linked in these signaling cascades, including mitogen-activated protein (MAP) kinase/extracellular signal-regulated kinase, calcium-calmodulin kinases, and protein kinase B. In addition, cyclic AMP may regulate cellular functions through additional mechanisms, such as ion channels and guanine nucleotide exchange factors, in particular, Epac 1 and 2, for the monomeric G protein, Rap.

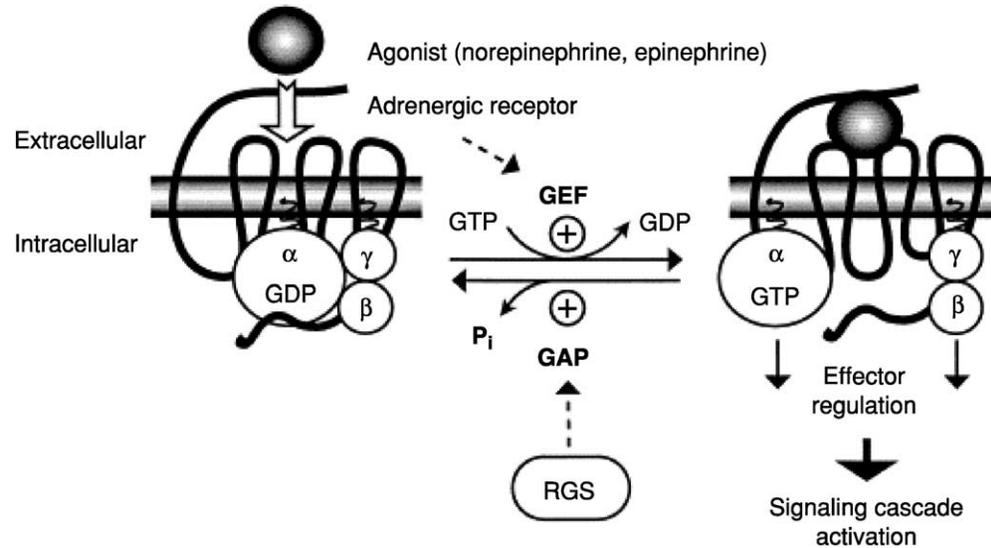
The effects of adrenergic receptor activation have most typically been viewed in the context of their acute regulatory effects, e.g., regulation of vasoreactivity, insulin release, lipolysis, gluconeogenesis, and glycogenolysis. These mechanisms are mostly related to the immediate effects of adrenergic receptor-stimulated kinases to modulate protein function by phosphorylation. However, more chronic effects of adrenergic receptor stimulation via the regulation of transcriptional mechanisms (including MAP kinase- and phosphoinositide 3-kinase-dependent pathways) have been increasingly appreciated.

### ADRENERGIC RECEPTORS: STRUCTURE AND MECHANISMS OF ACTIVATION AND DESENSITIZATION

As is typical of GPCRs, all the adrenergic receptors have a similar general structure (Fig. 3): extracellular amino termini, seven  $\alpha$ -helical domains that cross the lipid bilayer of the plasma membrane, resulting in three intracellular and extracellular loops and in intracellular carboxyl termini. Some of the adrenergic receptors are palmitoylated, with this fatty acid being



**Figure 2** Schematic structures of the nine adrenergic receptor (AR) subtypes. Each of the receptor subtypes has an extracellular amino terminus, seven  $\alpha$ -helical domains that cross the plasma membrane, and an intracellular amino-terminal “tail” domain. Note the similarities among members of certain types of receptors, for example,  $\alpha_1$ -ARs have relatively long carboxy-terminal tails and  $\alpha_2$ -ARs have long intracellular third loops.



**Figure 3** Model of adrenergic receptor activation. Physiologic agonists (norepinephrine, epinephrine) bind to adrenergic receptors, which stimulate signaling through their action as guanine nucleotide exchange factors (GEF) that facilitate GDP release, the subsequent binding of GTP to  $G_\alpha$  subunits, and the release of the  $G_{\beta\gamma}$  dimer from heterotrimeric  $G_{\alpha,\beta,\gamma}$  proteins. Both the  $G_\alpha$  and  $G_{\beta\gamma}$  subunits are able to regulate effectors and modulate the activity of signaling cascades, in large part via phosphorylation mechanisms. The activation of the G protein and the regulation of the effector are reversible, in part via intrinsic GTPase activity of the  $G_\alpha$  subunits and, in addition, via the action of RGS (regulators of G protein signaling) proteins, which are GTPase-activating proteins (GAP).  $G_\alpha$  and  $G_{\beta\gamma}$  subunits interact with the plasma membrane via lipid modifications, which are indicated by squiggly lines. Although the agonists are shown sitting on the outside of the receptor, the binding sites for norepinephrine and epinephrine are thought to be located within the lipid bilayer. Modified from Sondak, J., and Siderovski, D. (2001). *Biochem. Pharmacol.* 61, 1329–1337, with permission.

**Table I Human Adrenergic Receptors: Chromosomal Locations**

Receptor	Chromosomal location
$\alpha_{1A}$ -AR	8p21-p11.2
$\alpha_{1B}$ -AR	5p23-q32
$\alpha_{1D}$ -AR	20p13
$\alpha_{2A}$ -AR	10q24-q26
$\alpha_{2B}$ -AR	2p13-q13
$\alpha_{2C}$ -AR	4p16
$\beta_1$ -AR	10q24-q26
$\beta_2$ -AR	5q31-q32
$\beta_3$ -AR	8p12-p11.2

added to a cysteine residue in the proximal end of the carboxy-terminal regions, thereby potentially creating a fourth intracellular loop. Adrenergic receptors are N-linked glycoproteins, i.e., with asparagine-linked carbohydrates found in the amino-terminal regions. G protein-interacting domains are predominantly in the third intracellular loop and proximal portions of the carboxy-terminal regions. The third intracellular loop and carboxy-terminal “tails” also possess residues that can be phosphorylated by various protein kinases.

Norepinephrine and other agonists of adrenergic receptors appear to bind to regions of the receptors within the lipid bilayer, apparently as a consequence of folding of the receptors so as to create a polar core region within the hydrophobic transmembrane  $\alpha$ -helices that facilitates binding of the charged amines. Certain amino acid residues within the hydrophobic domains of adrenergic receptors have been implicated as sites that preferentially bind and complex with the hydroxyl groups of the catechol moiety and with amines of the side chain (e.g., Ser203, Ser204, and Ser207 bind the hydroxyl groups of the catechol moiety and Asp113 binds amines of the side chain,

in the  $\beta_2$ -adrenergic receptor). Data indicate that at least some of the adrenergic receptors may exist as preformed (constitutive) dimers within the membrane, perhaps preferentially in certain membrane microdomains, which are enriched for cholesterol and sphingolipids. These lipid rafts can also contain the protein caveolin (caveolae) and appear to be sites in which one finds not only the receptors, but also portions of the cognate G proteins and effector enzymes and channels.

The precise mechanism whereby occupancy of adrenergic receptors by norepinephrine or other agonists leads to receptor activation is poorly understood but most evidence points to agonist-promoted conformational changes in the receptor proteins. Of key importance (Fig. 4) is the ability of agonist-occupied receptors to activate G proteins by facilitating the exchange of cytosolic GTP for  $G_{\alpha}$ -bound GDP. In this action, the agonist-occupied receptors function as guanine nucleotide exchange factors (see Figs. 3 and 4). Subsequent to the exchange of GTP for GDP, the G protein heterotrimer dissociates into  $G_{\alpha}$  and  $G_{\beta\gamma}$  (the  $\beta$ - and  $\gamma$ -subunits of  $G_{\beta\gamma}$  are tightly bound and function as a dimer) subunits, both of which are able to regulate the activity of effector molecules. The dissociated  $G_{\alpha}$ GTP and  $G_{\beta\gamma}$  subunits represent the active state of G proteins; the lifetime of G protein activation is determined by intrinsic GTPase activity of  $G_{\alpha}$  and by accessory RGS (regulators of G protein signaling) proteins, which serve as GTPase-activating proteins (see Fig. 3). The hydrolysis of GTP and generation of GDP decrease the activity of  $G_{\alpha}$  and facilitate the re-formation of the inactive  $G_{\alpha\beta\gamma}$  heterotrimer.

Deactivation of signaling in response to norepinephrine or other agonists that activate adrenergic receptors occurs not only via the reversible activation of G proteins, but also by several mechanisms that “turn off” (desensitize) signaling (Fig. 4). These include

**Table II Adrenergic Receptors: Subtypes and Their Signaling Pathways**

	Receptors		
	$\alpha_{1A,B,D}$	$\alpha_{2A,B,C}$	$\beta_{1,2,3}$
<b>G proteins</b>	$G_q$	$G_i$	$G_s$
<b>Effectors</b>	Phospholipase $C_{\beta}$	Adenylyl cyclase $Ca^{2+}$ Channels $K^+$ Channels	Adenylyl cyclase $Ca^{2+}$ Channels
<b>Second messengers</b>	Inositol 1,4,5-trisphosphate Diacylglycerol	Cyclic AMP $Ca^{2+}$ $K^+$	Cyclic AMP $Ca^{2+}$
<b>Protein kinases</b>	Protein kinase C	Protein kinase A	Protein kinase A

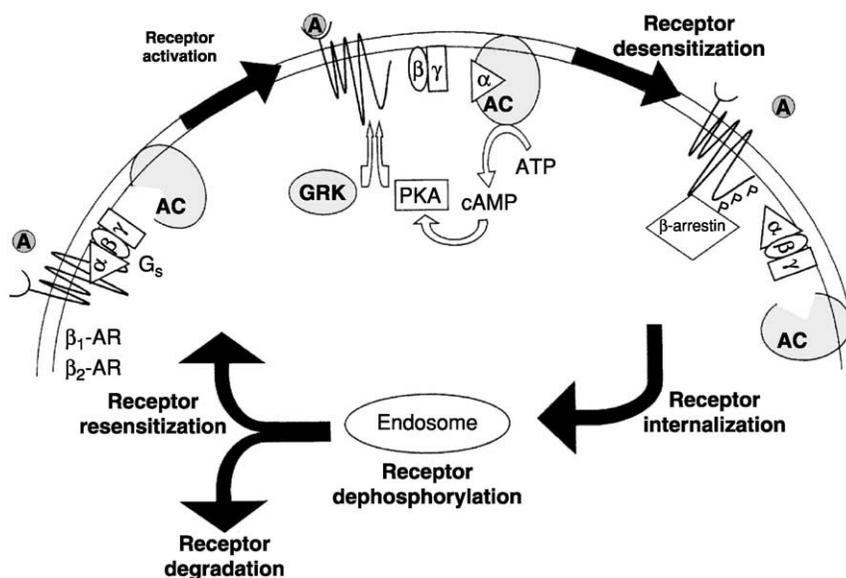
agonist-promoted activation of G protein receptor kinases or second-messenger kinases that phosphorylate receptors; binding of accessory proteins, e.g.,  $\beta$ -arrestins, to phosphorylated receptors; biochemical uncoupling and physical sequestration of receptors from G proteins; and internalization of receptors (which is facilitated by  $\beta$ -arrestin interaction with clathrin-coated pits); and degradation or removal of agonist-promoted second messengers and of receptors, the latter occurring at least in part by lysosomal enzymes, and perhaps by proteasomes. In some cases, internalized receptors recycle to the plasma membrane and are resensitized, events that are especially detectable after an agonist is withdrawn. In the continued presence of an agonist, the expression of surface receptors thus represents the balance between internalization and resensitization/externalization. Some results indicate that adrenergic receptor internalization may also facilitate activation of certain intracellular events and signaling pathways.

Beyond the importance of desensitization in the modulation of adrenergic receptor-activated responses, this process has important pharmacological implications. It is notable that by and large, the use of

adrenergic agonists is primarily confined to either acute or intermittent use—in part related to the impact of desensitization on their actions. In contrast, the effects of adrenergic antagonists are much more stable over time and consequently are more broadly utilized therapeutically. Moreover, antagonists can blunt the tonic desensitization and degradation of receptors that occur physiologically. This can precipitate withdrawal syndromes if administration of the antagonist is abruptly stopped and tissue levels fall, prior to the return to normal levels of the “supersensitized” receptors.

## ADRENERGIC RECEPTOR EXPRESSION IN TISSUES

In addition to norepinephrine, a key physiologic agonist of adrenergic receptors is the methylated derivative, epinephrine (adrenaline) (Fig. 1), which is particularly prominent in the adrenal medulla and other chromaffin cells, as a consequence of the action of phenylethanolamine *N*-methyltransferase. Relative to norepinephrine, epinephrine shows preferential interaction, i.e., binds with higher affinity to certain



**Figure 4** Schematic diagram of adrenergic receptor activation and desensitization using  $\beta_1$ -adrenergic receptors and  $\beta_2$ -adrenergic receptors ( $\beta_1$ -AR and  $\beta_2$ -AR) as examples. Occupancy of the  $\beta_1$ -AR or  $\beta_2$ -AR by norepinephrine, epinephrine, or other agonist (A) promotes the activation of  $G_s$  (see Fig. 3) and stimulation of adenylyl cyclase (AC) to enhance the synthesis of cyclic AMP (cAMP) from ATP. cAMP can activate protein kinase A (PKA), which promotes the phosphorylation of cellular substrates, including the  $\beta$ -adrenergic receptors. Agonist-occupied receptors are also phosphorylated by G protein receptor kinases (GRK). The phosphorylated receptors (which are probably phosphorylated primarily by GRK-mediated phosphorylation) subsequently interact with  $\beta$ -arrestin, thereby uncoupling the receptors from activation of the  $G_s$  protein. Subsequently, the receptors are internalized via a clathrin-mediated endocytic pathway. In the endosome, receptors can be dephosphorylated and recycled back to the cell surface (resensitization) or degraded by lysosomes or perhaps proteasomes. Modified, with permission, from Post, S. R., Hammond, H. K., and Insel, P. A. (1999). *Annu. Rev. Pharmacol. Toxicol.* 39, 343–360, © 1999 by Annual Reviews, [www.annualreviews.org](http://www.annualreviews.org).

classes and subtypes of adrenergic receptors (e.g.,  $\beta_2$ -adrenergic receptors). Some data indicate that  $\alpha_1$ - and  $\beta_1$ -adrenergic receptors, which show relatively high affinity for binding norepinephrine, are located at postsynaptic neuroeffector junctions in the sympathetic nervous system, whereas other receptors, such as  $\beta_2$ -adrenergic receptors, which recognize norepinephrine with lower affinity than epinephrine, are located at extrasynaptic sites.  $\alpha_{1A}$  receptors appear to be particularly enriched in the heart and vascular smooth muscle in humans.  $\alpha_{2A}$  and  $\alpha_{2C}$  receptors appear to play a key role in presynaptic feedback inhibition of norepinephrine release from sympathetic nerve endings and at central sites, whereas  $\alpha_{2B}$  receptors are located at postsynaptic sites. Although previously it had been suggested that certain tissues exclusively contained single subtypes of adrenergic receptors (e.g., heart and fat cells,  $\beta_1$  versus lung,  $\beta_2$ ), it has become apparent that most major organs, tissues, and cells express multiple types and subtypes of adrenergic receptors. Thus, one can identify  $\beta_1$ -,  $\beta_2$ -, and  $\beta_3$ -adrenergic receptors as well as  $\alpha_{1A}$ - and  $\alpha_{1B}$ -adrenergic receptors in the heart; adipocytes express multiple subtypes of  $\alpha_1$ -,  $\alpha_2$ -, and  $\beta$ -adrenergic receptors, including  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$  in brown fat.

## ADRENERGIC RECEPTORS: ROLE IN CLINICAL MEDICINE

Norepinephrine receptors are important clinically for three reasons: (1) their key role in physiology (e.g., cardiovascular, metabolic, renal); (2) the large number of drugs that target these receptors as agonists or antagonists; and (3) the contribution of these receptors in disease, including endocrine disorders. Information about physiologic actions and pharmacological derivatives is available in other reference texts. In terms of endocrine disorders, a few key disorders should be briefly mentioned: disorders of the endocrine pancreas, hyperthyroidism/hypothyroidism, and pheochromocytoma.

Endocrine diseases of the pancreas, in particular,  $\beta$ -cell disorders that are associated with hypoglycemia, lead to activation of the sympathetic nervous system and increases in circulating norepinephrine and epinephrine, the latter primarily as a result of adrenal medullary stimulation. Such counterregulatory responses can occur with hypoglycemia of any etiology, although excess administration of insulin or sulfonylurea drugs is probably the most common cause. Activation of receptors by neuronally released and circulating catecholamines plays a key role

in several signs and symptoms that characterize hypoglycemia, including pallor, tremor (a  $\beta_2$ -adrenergic receptor response), tachycardia, and actions of adrenergic receptors in the liver, muscle, and fat to increase blood glucose concentration.

Hyperthyroidism is associated with a number of signs and symptoms that suggest a "hyperadrenergic state," e.g., tremor, tachycardia, arrhythmias, and systolic hypertension. Data in experimental animals and patients have indicated that hyperthyroidism is associated with increased  $\beta$ -adrenergic and decreased  $\alpha_1$ -adrenergic receptor activity, at least in part as a consequence of altered receptor expression in several thyroid hormone-responsive tissues and cell types. The molecular basis of these changes may relate to the ability of thyroid hormone to act as a transcriptional regulator of adrenergic receptor mRNA, but post-receptor events also appear to be altered.  $\beta$ -Adrenergic antagonists thus provide rational therapy to treat the signs and symptoms of hyperthyroidism that result from the enhanced  $\beta$ -adrenergic receptor activity.

Conversely, some of the clinical manifestations of hypothyroidism, such as bradycardia, suggest a "hypoadrenergic state." Consistent with this idea are data that indicate opposite changes to those described above for hyperthyroidism, i.e., decreased  $\beta$ -adrenergic receptor activities, increased  $\alpha_1$ -adrenergic receptor activities, and altered activity of postreceptor components.

Pheochromocytomas, chromaffin cell tumors that are most commonly found in the adrenal medulla, produce high circulating levels of catecholamines, stimulation of adrenergic receptors, and predictable effects of such stimulation in target cells. In addition to the direct activation of these receptors, the elevation of norepinephrine (and epinephrine when present) can produce desensitization and down-regulation of adrenergic receptors. These feedback responses help buffer hypertension and other effects of chronic or episodic elevations in catecholamines, but also may contribute to the hypotension that can occur, especially in settings of receptor blockade or postoperatively with removal of the catecholamine-producing tumors.

Data have suggested another potentially important aspect of adrenergic receptors in endocrine and other disorders: the role of genetic variants of these receptors. Several classes of adrenergic receptors, especially  $\alpha_2$  receptor subtypes, as well as  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$  receptors are polymorphic, with coding sequence variants that change amino acids involved in receptor activation and desensitization or change amino acids at noncoding regions that can influence receptor expression. Of particular importance in terms of endocrine disorders are data related to the possible contribution of  $\alpha_2$ - and

$\beta_3$ -adrenergic receptor variants in obesity. It is likely that substantial further information will be forthcoming regarding genetic variants of adrenergic receptors or their role as disease-modifying genes in endocrine and other disorders—both genetic and acquired.

### See Also the Following Articles

Catecholamines • Hypertension, Neurogenic • Norepinephrine Transporter

### Further Reading

Insel, P. A. (1996). Adrenergic receptors—Evolving concepts and clinical implications. *N. Engl. J. Med.* **334**, 580–585.

Kirstein, S. L., and Insel, P. A. (2003). Autonomic nervous system pharmacogenomics—A progress report. *Pharmacol. Rev.*, in press.

Kohout, T. A., and Lefkowitz, R. J. (2003). Regulation of G protein-coupled receptor kinases and arrestins during receptor desensitization. *Mol. Pharmacol.* **63**, 9–18.

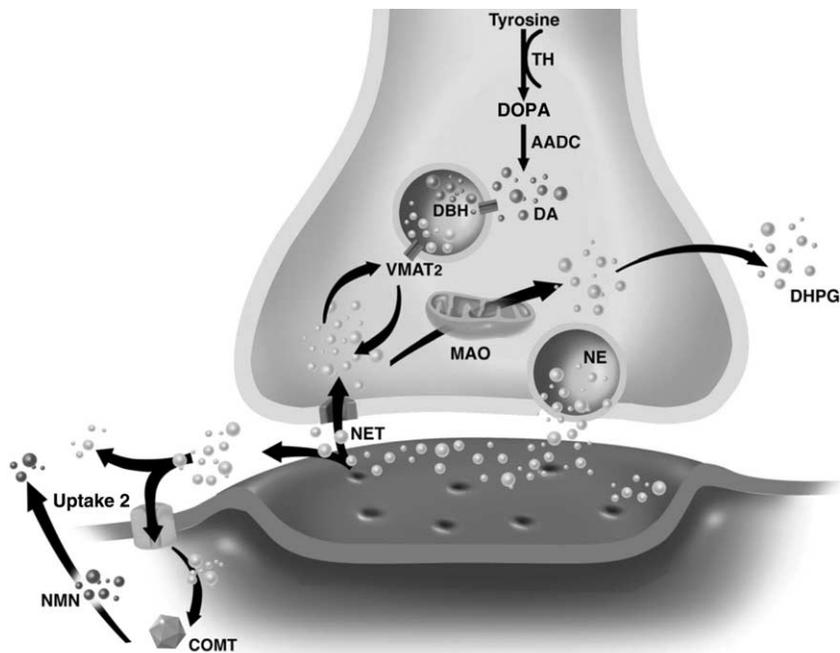
Neer, E. J. (1995). Heterotrimeric G proteins: Organizers of transmembrane signals. *Cell* **80**, 249–257.

Neves, S. R., Ram, P. T., and Iyengar, R. (2002). G protein pathways. *Science* **296**, 1636–1639.

Piasecki, M. T., and Perez, D. M. (2001).  $\alpha_1$ -Adrenergic receptors: New insights and directions. *J. Pharmacol. Exp. Ther.* **298**, 403–410.

Pierce, K. L., Premont, R. T., and Lefkowitz, R. J. (2002). Seven-transmembrane receptors. *Nat. Rev. Mol. Cell. Biol.* **3**, 639–650.

Xiang, Y., and Kobilka, B. K. (2003). Myocyte adrenoceptor signaling pathways. *Science* **300**, 1530–1532.



**Figure 1** The noradrenergic neuron. The initial and rate-limiting step in the production of norepinephrine (NE) is the hydroxylation of tyrosine by tyrosine hydroxylase (TH) to L-dihydroxyphenylalanine (DOPA). DOPA is subsequently decarboxylated by L-aromatic amino acid decarboxylase (AADC) to dopamine (DA). Dopamine is mobilized into vesicles through vesicular monoamine transporter 2 (VMAT2), where it is subsequently hydroxylated by dopamine β-hydroxylase (DBH) to form NE. The majority of released NE is removed from the synaptic cleft through reuptake into the presynaptic neuron by the norepinephrine transporter (NET). The remaining NE diffuses away from the synapse and may be taken into other cells by extraneuronal monoamine transporter. Catechol-*O*-methyltransferase (COMT) is an extraneuronal enzyme that catalyzes the methylation of NE to normetanephrine (NMN). NE taken into the presynaptic neuron is either repackaged into vesicles or oxidized by monoamine oxidase (MAO) to dihydroxyphenylglycol (DHPG).

increased aggression in humans homozygous for the low-activity COMT allele, but more study is required on this topic.

### Monoamine Oxidase

Monoamine oxidase A (MAO-A) and MAO-B are mitochondrial isoenzymes that are widely expressed among tissues; MAO-A predominates in neural tissue and catalyzes the oxidative deamination of DA to DOPAC and NE to DHPG. MAO-B is selectively expressed in platelets and has been localized to a variety of brainstem structures. MAO-B may have an important modulatory role in limbic output. Dopamine is a substrate for both subtypes. The genes encoding the two isoenzymes are distinct and in adjacent locations on the human X chromosome (Xp11.23). MAO-A-deficient animals showed increased aggressive behavior, which can be corrected by administration of serotonin (5HT) inhibitor. Developmental studies showed an uptake of 5HT by noradrenergic neurons and the animals exhibited cytoarchitectural

abnormalities in the somatosensory cortex. MAO-B-deficient mice showed a reduction in mean arterial pressure and heart rate. Elevated plasma levels of phenylethylamine, a dietary amine, have been reported after administration into MAO-B-deficient mice. Thus, humans deficient in MAO-B may exhibit an increased sensitivity to dietary amines and it brings into focus the association between MAO and migraine headache.

### TRANSPORTERS

#### Vesicular Monoamine Transporter

NE taken up into the presynaptic neuron by norepinephrine transporter (NET) (Uptake 1) is translocated into storage vesicles or deaminated by monoamine oxidase. Vesicular monoamine transporter 2 (VMAT2) mobilizes monoamines from the neuronal cytoplasm into vesicles, where they are repackaged for release at the synapse. Reserpine effectively blocks VMAT-2 and also shuts down the conversion of DA to NE. This

prevents the recycling of NE and results in a gradual and prolonged depletion of NE from the nerve terminal and has been used therapeutically in the treatment of hypertension. VMAT2 null mutation is lethal in mice 2 weeks after birth. The absence of biogenic amine storage in the central nervous system (CNS) of neonatal VMAT2-deficient mice confirms that VMAT2 is the only functional biogenic amine vesicular transporter in the mammalian CNS and that vesicular transporter activity is required for storage. VMAT2 protein levels were reduced by 50% in mice with a single VMAT2 gene and their neural tissue contains approximately 50% of the normal amount of norepinephrine, dopamine, and serotonin. After 10 months, 28% of the mice with a single VMAT2 gene experienced sudden death. Body temperature and heart rate were normal but QT and QTc intervals were lengthened. The failure to completely compensate for the changes in NE signaling may lead to arrhythmias and consequently sudden death. VMAT2 is a potential pharmacological target in the treatment of neuropsychiatric disease in humans.

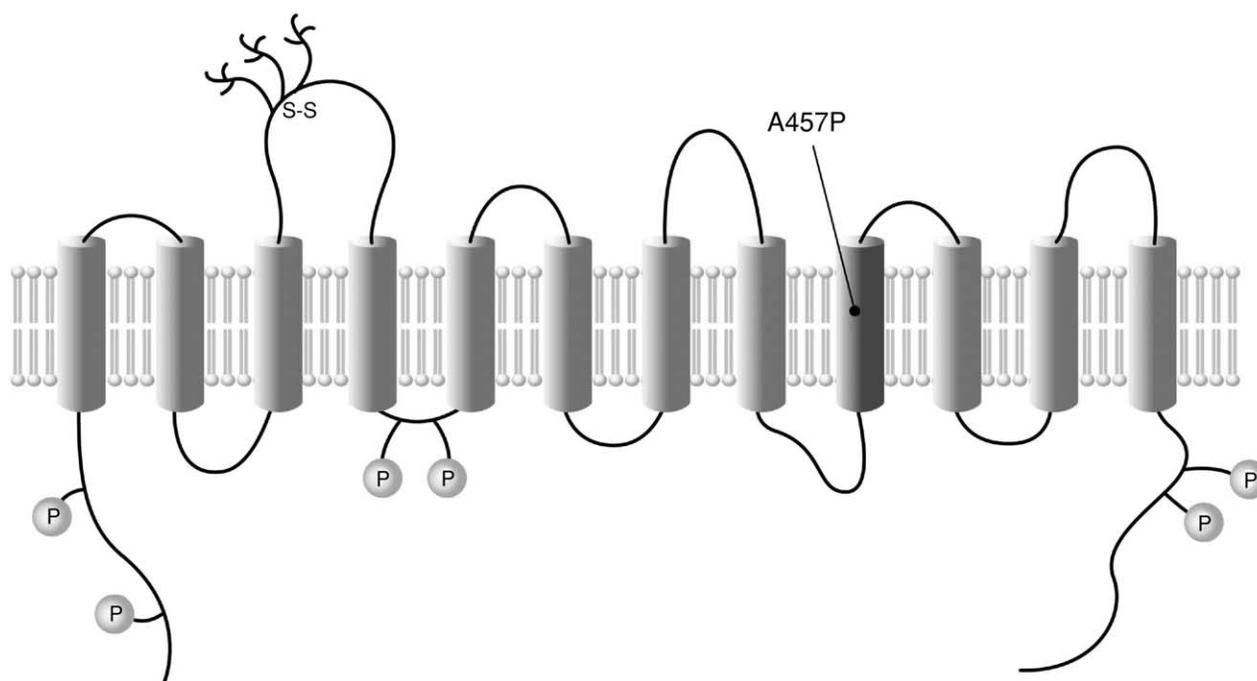
### Norepinephrine Transporter

Neuronal uptake (Uptake 1) by NET accounts for inactivation of 70–90% of released NE from the synaptic cleft in the central and peripheral nervous systems. In

humans, a single gene located on chromosome 16q12.2 encodes a 617-amino-acid protein with 12 putative transmembrane domains (TMDs), and intracellular amino and carboxyl termini (Fig. 2). NET protein undergoes posttranslational N-glycosylation on a large extracellular loop between TMD3 and TMD4, before the protein is subsequently trafficked to the plasma membrane. This modification enhances NET protein stability and catalytic function. NET is expressed presynaptically on noradrenergic axons and varicosities. It is a target of many pharmacological agents including tricyclic antidepressants (TCAs).

TCAs are potent inhibitors of NE uptake and elevate circulating catecholamines. The heart is particularly dependent on reuptake of NE for its inactivation and an increase in heart rate with standing after systemic administration of desipramine, sibutramine, and reboxetine has been reported. Desipramine, which has greater activity on NE reuptake than imipramine, decreases sympathetic nerve traffic, whole body NE spillover, and renal and forearm NE spillover, whereas imipramine was associated with orthostatic hypotension. These effects are similar to those of NET deficiency.

Genetic variability in the human NET gene could contribute to altered reuptake of NE and susceptibility to diseases involving impairment of function at NE synapses, including autonomic nervous system and

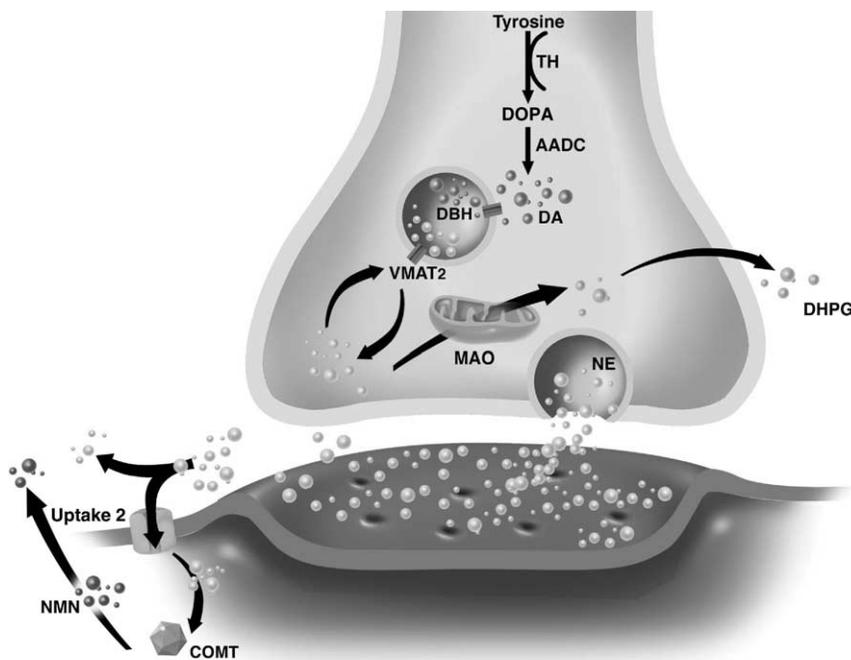


**Figure 2** The structure of NET encompassing 12 membrane domains is shown. The arginine to proline mutation at amino acid 457 is shown in the ninth transmembrane domain.

psychiatric disorders. Single-nucleotide polymorphisms are the most common type of genetic variability. Among the human NET variants identified thus far, a functional missense mutation in exon 9 in which guanine was replaced by cytosine at position 237 (G237C) has been described. This polymorphism results in a change from alanine to proline (A457P) within the highly conserved region of TMD9 of the NET protein. Proline residues in a transmembrane-spanning  $\alpha$ -helix have the ability to kink the  $\alpha$ -helix by preventing the formation of backbone hydrogen bonds, which results in disrupted surface expression and function. Transient transfection of A457P into heterologous expression systems revealed a protein with a near complete loss of transporter activity and greatly diminished cell surface expression of the mature, glycosylated form of the transporter. Furthermore, A457P exerts a dominant-negative effect on wild-type NET uptake activity and suggests that individuals heterozygous for A457P may be affected to a greater extent than predicted for harboring one mutant allele. The allele segregates with postural tachycardia, elevated plasma NE, decreased NE clearance, diminished conversion to the intraneuronal metabolite DHPG, and a blunted response to tyramine (Fig. 3).

Molecular studies showed that NET is regulated via transporter surface expression and intrinsic catalytic activity. The regulatory mechanisms involve transporter phosphorylation and regulated associations of accessory proteins. NET has been reported to complex with the catalytic subunit of protein phosphatase 2A (PP2Ac), syntaxin 1A, a presynaptic soluble N-ethylmaleimide-sensitive factor attachment protein receptor (SNARE) protein, and the scaffolding protein PICK1. NET internalization is mediated by protein kinase C through disruption of Syntaxin and PP2Ac association. NET catalytic activity is regulated by insulin- and mitogen-activated protein kinase-linked pathways through PP2Ac. It is quite possible that trafficking and catalytic function are coregulated through kinase-dependent mechanisms.

The management of NET dysfunction requires pharmacological and nonpharmacological interventions, which must be tailored to the individual patient. The pharmacological intervention is directed at attenuating tachycardia in response to standing. Clonidine, an  $\alpha_2$ -adrenoceptor partial agonist, could reduce NE release by stimulating  $\alpha_2$ -autoreceptors and reduce sympathetic outflow. It can also reduce heart rate responses to standing by binding to central  $\alpha_2$ -adrenoceptors involved in brainstem cardiovascular regulation. Beta-blockers can also



**Figure 3** NET deficiency. The norepinephrine transporter is responsible for clearance of a majority of the released NE from the synaptic cleft. Loss of NET from the neuron results in elevated synaptic levels of NE. Furthermore, the inability to recycle NE leads to reductions in intraneuronal levels of NE and a concomitant reduction in the production of the NE metabolite DHPG.

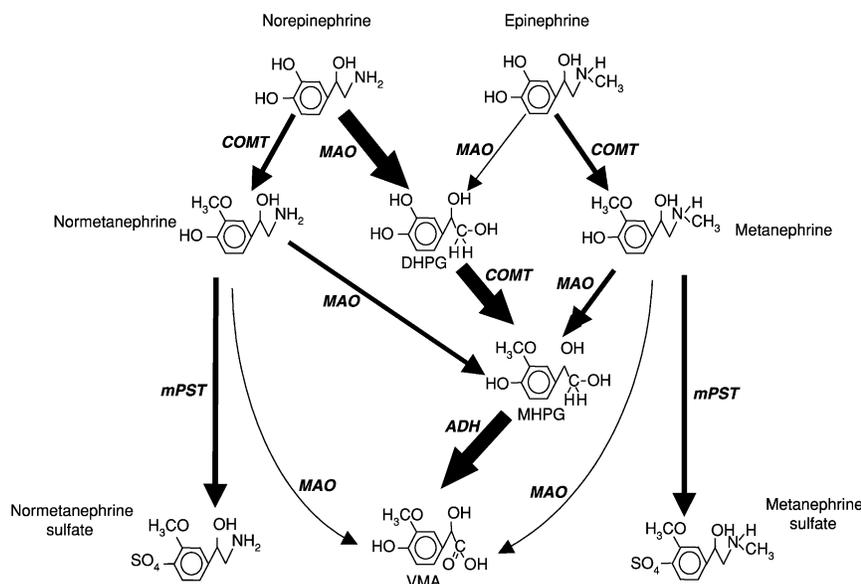
attenuate the cardiovascular symptoms.  $\alpha$ -Methyldopa appears to be the drug of choice in this group of patients.

### See Also the Following Articles

Adrenergic Mechanisms • Adrenergic Receptors • G Protein-Coupled Receptors • Hypertension, Neurogenic • Norepinephrine Receptors

### Further Reading

- Blakely, R. D., De Felice, L. J., and Hartzell, H. C. (1994). Molecular physiology of norepinephrine and serotonin transporters. *J. Exp. Biol.* **196**, 263–281.
- Carson, R. P., and Robertson, D. (2002). Genetic manipulation of noradrenergic neurons. *J. Pharmacol. Exp. Ther.* **301**, 410–417.
- Goldstein, D. S. (1996). Noradrenergic neurotransmission. In “Primer on the Autonomic Nervous System” (D. Robertson, P. A. Low, and R. J. Polinsky, eds.), pp. 91–98. Academic Press, San Diego, CA.
- Shannon, J. R., Flattem, N., Jordan, J., *et al.* (2000). Orthostatic intolerance and tachycardia associated with norepinephrine transporter deficiency. *N. Engl. J. Med.* **342**, 541–549.
- Sung, U., Apparsundaram, S., Galli, A., Kahlig, K. M., *et al.* (2003). A regulated interaction of syntaxin 1A with the antidepressant-sensitive norepinephrine transporter establishes catecholamine clearance capacity. *J. Neurosci.* **23**, 1697–1709.
- Vincent, S., and Robertson, D. (2002). The broader view: Catecholamine abnormalities. *Clin. Auton. Res.* **12**, I44–I49.



**Figure 1** Pathway of metabolism of norepinephrine and epinephrine. *Enzymes:* MAO, monoamine oxidase; COMT, catechol-*O*-methyltransferase; mPST, monoamine preferring sulfotransferase (SULT1A3); ADH, alcohol dehydrogenase. *Metabolites:* MHPG, 3-methoxy-4-hydroxyphenylglycol; DHPG, 3,4-dihydroxyphenylglycol; VMA, vanillylmandelic acid.

this is represented by circulating and urinary levels of NMN.

COMT is not located in neurons but is found in nonneuronal tissues, such as liver, muscles, and kidneys. An additional important location of COMT is the adrenal medulla, in which the enzyme is responsible for the conversion of EPI to MN and that of NOREPI to NMN. Circulating EPI is also subject to uptake-2, but less than 10% of MN is derived from this source. More than 90% of circulating MN and 23–40% of NMN are derived from *O*-methylation of EPI and NOREPI within adrenal medullary chromaffin cells. The importance of the adrenal medulla for production of MN and NMN reflects the fact that COMT is predominantly present in this tissue as the membrane-bound enzyme, an isoform of COMT with particularly high affinity for catecholamines. Therefore, plasma metanephrines are specific markers for extraneuronal and intra-adrenal metabolism of catecholamines, whereas plasma DHPG is a specific marker for neuronal metabolism of catecholamines. From a quantitative standpoint, the neuronal pathway of metabolism greatly predominates over extraneuronal and intra-adrenal pathways for metabolism of NOREPI, whereas the extraneuronal and intra-adrenal pathways are more important for metabolism of EPI.

## KINETICS AND METABOLISM OF METANEPHRINES

Metanephrines can be measured in plasma and urine by high-pressure liquid chromatography (HPLC) with electrochemical detection (ED). In plasma, the levels of free plus conjugated metanephrines are approximately 20-fold higher than the levels of the free (unconjugated) metanephrines. In human urine, most metanephrines are in the conjugated form. Conjugation of metanephrines by the specific sulfotransferase isoenzyme, SULT1A3, represents an important route of NMN and MN metabolism (Fig. 1). This enzyme is located mainly in extraneuronal tissue of the mesenteric organs, particularly the wall of the gut. These conjugates are excreted by the kidneys. Other routes of metabolism of free metanephrines involve their conversion by MAO to MHPG. MHPG is further oxidized in the liver by alcohol dehydrogenase to the final end product of catecholamine metabolism, vanillylmandelic acid (VMA). The MHPG that is excreted in the urine is mainly conjugated to MHPG sulfate.

Plasma levels of free metanephrines in healthy subjects under resting baseline conditions average approximately 0.6 nmol/liter for NMN and 0.3 nmol/liter for MN. The plasma half-lives of free (unconjugated)

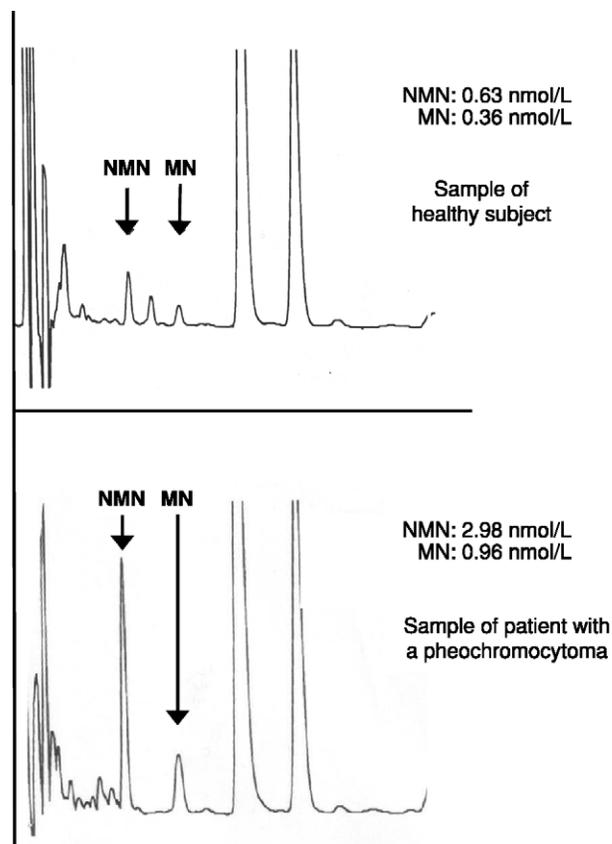
metanephrines are similar to those of catecholamines (approximate 3 or 4 min). The plasma clearances of free NMN and MN are approximately 1.5 liters/min, slightly lower than that of plasma catecholamines (2 liters/min). In contrast, the plasma half-lives of conjugated NMN and MN are much longer (>60 min) and their circulatory clearances are much lower (0.1 liters/min). Since the circulatory clearance of conjugated metanephrines is determined almost exclusively by their extraction by the kidneys, plasma levels of conjugated metanephrines do not always reflect production of free metanephrines. For example, plasma levels of conjugated metanephrines are substantially increased in patients with renal failure, whereas those of free metanephrines are minimally affected.

Plasma levels of metanephrines increase during activation of the sympathetic nervous and/or adrenomedullary hormonal systems. However, increases in plasma levels of NMN and MN above baseline levels are relatively less than those of their parent catecholamines. Even during intense adrenomedullary stimulation induced by hypoglycemia, plasma MN levels increase only 3-fold, in contrast to a 25-fold increase in plasma EPI levels. During less intense stressful conditions, such as mental stress, the responses of catecholamines are proportionally much greater than those of metanephrines. Consequently, metanephrines are inferior to catecholamines for gauging stress responses.

## MEASUREMENT OF METANEPHRINES

Measurements of metanephrines in urine have been available for clinical purposes since the early 1960s. Historically, the term total metanephrines was coined to indicate the combined measurement of both NMN and MN by early spectrophotometric assays that did not allow separate measurement of NMN and MN. The development of HPLC techniques has allowed separate measurement of both amines, hence the term fractionated metanephrines. However, urinary measurements of NMN and MN are usually carried out after subjecting samples to acid hydrolysis or enzymatic deconjugation with sulfatase, which liberates the free from conjugated metanephrines. These measurements therefore reflect levels of both the free and conjugated forms of NMN and MN.

In plasma, metanephrines can be measured as free metanephrines or as free plus conjugated metanephrines by HPLC and ED (Fig. 2). In view of the



**Figure 2** Chromatogram of free metanephrines in a healthy subject (top) and a patient with a pheochromocytoma (bottom) as measured by HPLC with electrochemical detection. NMN, normetanephrine; MN, metanephrine.

much lower plasma concentrations of free metanephrines, the required analytical sensitivity is much higher for free metanephrines. Many factors should be considered as potential sources of falsely high or falsely low test results. Despite the fact that plasma metanephrines are less sensitive than plasma catecholamines to sympathoadrenal excitation, sympathoadrenal excitation must still be considered a potential cause of falsely high test results. Analytical interference with the analgesic drug acetaminophen can be a problem. Several antidepressant drugs may cause pharmacokinetic interference with the measurement of plasma free metanephrines. Tricyclics can elevate NMN levels, whereas MAO inhibitors increase both NMN and MN levels. Conversely, COMT inhibitors and central sympatholytic drugs lower NMN and MN levels. Despite the availability of this high specific assay, caution for drug interference remains warranted in the interpretation of test results.

## APPLICATION OF MEASUREMENT OF METANEPHRINES

In the past 10 years, a new assay for plasma metanephrines has enabled investigators to obtain more detailed and comprehensive insights into the metabolism of catecholamines. This has led to several promising applications. First, measurement of plasma free metanephrines provides a highly sensitive test for diagnosis of pheochromocytoma. The abundant presence of COMT in pheochromocytoma tumor tissue is responsible for a continuous intratumoral conversion of catecholamines to metanephrines. These metanephrines diffuse to the circulation, independently of the release of catecholamines by the tumor. More than 94% of the increased circulating levels of NMN and MN in patients with pheochromocytoma is derived from catecholamine metabolism within tumor cells. Thus, only small increases in NMN and MN levels indicate metabolism of catecholamines after their secretion by the tumor into the circulation. This largely explains why plasma free metanephrines have an approximately 100% sensitivity. Elevations of plasma metanephrines in patients with pheochromocytoma show larger relative increases above normal levels compared to the parent catecholamines. This also contributes to the higher diagnostic sensitivity of measurements of free metanephrines over catecholamines. The continuous production of metanephrines within the tumor tissue also explains why levels of plasma metanephrines are less apt to increase compared to plasma catecholamines during paroxysmal catecholamine release from a pheochromocytoma or during surgical manipulation of the tumor.

Other applications of plasma free metanephrines include a role in the assessment of COMT activity. In conjunction with measurements of plasma DHPG levels, measurements of free NMN can also be used to assess the activity of MAO: An increased NMN/DHPG ratio indicates inhibition of MAO and can also be used to establish MAO-deficiency states. Finally, since plasma free MN is almost completely derived from intramedullary conversion from

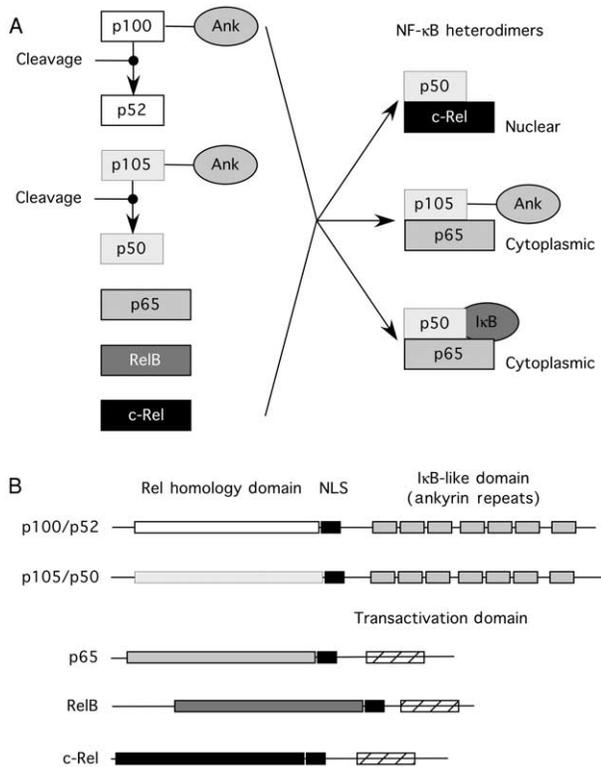
adrenomedullary EPI, plasma MN levels can serve as a marker of the adrenomedullary EPI stores.

## See Also the Following Articles

Catecholamines • Neurotransmitters, Overview

## Further Reading

- Eisenhofer, G., Friberg, P., Pacak, K., Goldstein, D. S., Murphy, D. L., Tsigos, C., Quyyumi, A. A., Brunner, H. G., and Lenders, J. W. M. (1995). Plasma metanephrines: Do they provide useful information about sympatho-adrenal function and catecholamine metabolism? *Clin. Sci.* **88**, 533–542.
- Eisenhofer, G., Rundqvist, B., Åneman, A., Friberg, P., Dakak, N. A., Kopin, I. J., Jacobs, M.-C., and Lenders, J. W. M. (1995). Regional release and removal of catecholamines and extraneuronal metabolism to metanephrines. *J. Clin. Endocrinol. Metab.* **80**, 3009–3017.
- Eisenhofer, G., Åneman, A., Hooper, D., and Friberg, P. (1996). Mesenteric organ production, hepatic metabolism, and renal elimination of norepinephrine and its metabolites in humans. *J. Neurochem.* **66**, 1565–1573.
- Eisenhofer, G., Keiser, H., Friberg, P., Mezey, E., Huynh, T. T., Hiremagalur, B., Ellingson, T., Duddempudi, S., Eijssbouts, A., and Lenders, J. W. M. (1998). Plasma metanephrines are markers of pheochromocytoma produced by catechol-O-methyltransferase within tumors. *J. Clin. Endocrinol. Metab.* **83**, 2175–2185.
- Eisenhofer, G., Lenders, J. W. M., Linehan, W. M., Walther, M. M., Goldstein, D. S., and Keiser, H. R. (1999). Plasma normetanephrine and metanephrine for detecting pheochromocytoma in Von Hippel Lindau disease and multiple endocrine neoplasia type 2. *N. Engl. J. Med.* **340**, 1872–1879.
- Eisenhofer, G., Huynh, T.-T., Hiroi, M., and Pacak, K. (2001). Understanding catecholamine metabolism as a guide to the biochemical diagnosis of pheochromocytoma. *Rev. Endocr. Metab. Disorders* **2**, 297–311.
- Goldstein, D. S., Eisenhofer, G., and Kopin, I. J. (2003). Sources and significance of plasma levels of catechols and their metabolites in humans. *J. Pharmacol. Exp. Ther.* **305**, 800–811.
- Lenders, J. W. M., Eisenhofer, G., Armando, I., Keiser, H. R., Goldstein, D. S., and Kopin, I. J. (1993). Determination of plasma metanephrines by liquid chromatography with electrochemical detection. *Clin. Chem.* **39**, 97–103.
- Lenders, J. W. M., Pacak, K., McClellan, M. W., Linehan, W. M., Mannelli, M., Friberg, P., Keiser, H. R., Goldstein, D. S., and Eisenhofer, G. (2002). Biochemical diagnosis of pheochromocytoma. Which test is best? *J. Am. Med. Assoc.* **287**, 1427–1434.

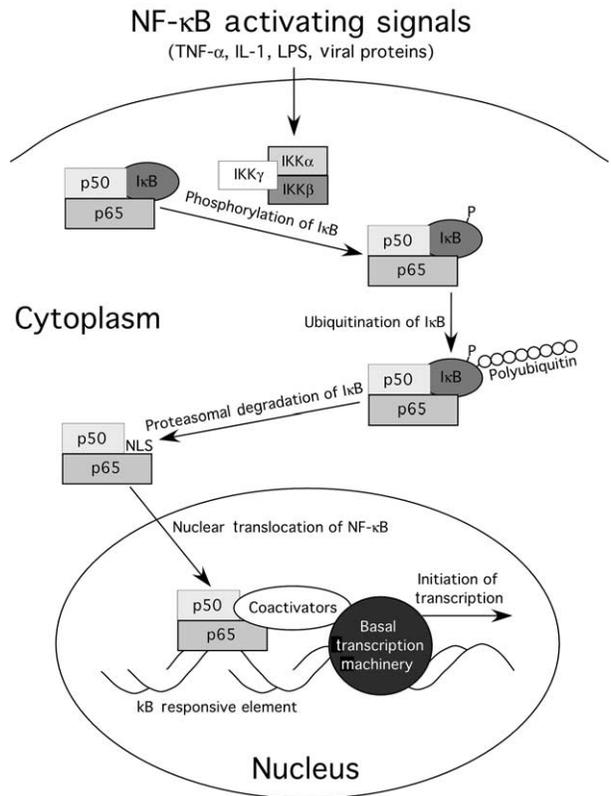


**Figure 1** (A) Nuclear factor- $\kappa$ B family of transcription factors and representative heterodimers between different nuclear factor- $\kappa$ B proteins. (B) Schematic diagram of the primary structure and functional domains of nuclear factor- $\kappa$ B proteins.

NF- $\kappa$ B is considered an inducible transcription factor. In cells with inactive, cytoplasmic NF- $\kappa$ B, the I $\kappa$ B protein is phosphorylated by specific I $\kappa$ B kinases (IKKs) in response to various endogenous and exogenous signals, including viral infection, bacterial components, DNA damage, oxidative stress, and chemotherapeutic agents. The phosphorylated I $\kappa$ B is then marked for degradation by the ubiquitin-26S proteasome pathway. Degradation and removal of I $\kappa$ B from the NF- $\kappa$ B heterodimer exposes a nuclear localization signal, which directs the movement of NF- $\kappa$ B to the nucleus. Once in the nucleus, NF- $\kappa$ B binds  $\kappa$ B responsive elements in the promoter of various genes and stimulates transcription by interacting with other transcription factors, coactivators, and the basal transcription machinery. Full transcriptional activity of NF- $\kappa$ B appears to require its phosphorylation.

### Inhibitory Proteins Mask the Nuclear Localization Signal in NF- $\kappa$ B

The I $\kappa$ Bs are a family of proteins that play a critical role in regulating NF- $\kappa$ B activity by physically



**Figure 2** General mechanism of action for cytoplasmic nuclear factor- $\kappa$ B proteins as described in the text.

interacting with and retaining NF- $\kappa$ B in the cytoplasm. The I $\kappa$ B family includes I $\kappa$ B $\alpha$ , I $\kappa$ B $\beta$ , I $\kappa$ B $\gamma$ , I $\kappa$ B $\epsilon$ , Bcl-3, p105, p100, and the homologous protein Cactus, which is found in *Drosophila*. Interestingly, different I $\kappa$ Bs interact preferentially with different members of the NF- $\kappa$ B family and may have specific patterns of expression in different cell types. For example, I $\kappa$ B $\alpha$  and I $\kappa$ B $\beta$  have a high affinity for the p65 subunit of NF- $\kappa$ B. As noted previously, ankyrin repeats are a common feature of all I $\kappa$ B proteins as well as I $\kappa$ B-like portions of p100 and p105. These ankyrin repeats are involved in the physical interaction between I $\kappa$ B proteins and NF- $\kappa$ B. This interaction essentially masks the nuclear localization signal and prevents interactions with other proteins that would transport NF- $\kappa$ B to the nucleus (Fig. 2). An amino acid sequence containing acidic residues is also characteristic of the C-terminal region of I $\kappa$ B family members. Evidence suggests that this acidic domain may interact with the DNA-binding and nuclear localization domains of NF- $\kappa$ B. Nevertheless, the function of this region of I $\kappa$ B is not as well defined as that of the ankyrin repeats.

### Protein Phosphorylation Leads to Ubiquitination and Degradation of I $\kappa$ B

Diverse signals activate cytoplasmic NF- $\kappa$ B. Nevertheless, most signaling pathways activate a common I $\kappa$ B kinase or IKK (Fig. 2). Although IKK is a multisubunit complex that contains IKK $\alpha$ , IKK $\beta$ , and IKK $\gamma$  proteins in a stoichiometric ratio, the composition of the entire IKK complex has not been completely defined. It is clear that the activated IKK complex phosphorylates conserved serine residues in the amino terminus of I $\kappa$ B proteins. For example, I $\kappa$ B $\alpha$ , which interacts with p65/p50 heterodimers, is phosphorylated by IKK at Ser32 and Ser36. Phosphorylated I $\kappa$ B $\alpha$  is then recognized by specific enzymes that initially attach one ubiquitin peptide at nearby lysine residues within I $\kappa$ B $\alpha$ . Attachment of the first ubiquitin leads to ubiquitin–ubiquitin conjugation by other enzymes to form a ubiquitin polymer (Fig. 2). Polyubiquitin-marked I $\kappa$ B $\alpha$  is then recognized and degraded by the 26S proteasome (Fig. 2). Although most stimuli that activate NF $\kappa$ B make use of this general pathway, hypoxia and ultraviolet irradiation appear to activate NF- $\kappa$ B by distinct mechanisms.

### NF- $\kappa$ B Translocates to the Nucleus, Binds DNA, and Acts as a Transcription Factor

The phosphorylation, ubiquitination, and degradation of I $\kappa$ B releases NF- $\kappa$ B from its tonic state of inhibition and exposes its nuclear localization signal. Nuclear localization signals are generally composed of a single cluster of five to seven basic amino acids, often including lysine (K) and arginine (R). For example, the core nuclear localization signal is KRKR in p65, KRQ(glutamine)K in p50, and KKQK in c-Rel. Specific transporter proteins called karyopherins recognize these nuclear localization signals, act as molecular chaperones, and direct the movement of NF- $\kappa$ B into the nucleus (Fig. 2). Mutations of the DNA sequence coding for these amino acids prevent nuclear import and block NF- $\kappa$ B transcriptional activity, thus demonstrating that these residues are critical for the function of NF- $\kappa$ B. Once in the nucleus, NF- $\kappa$ B binds  $\kappa$ B responsive elements in the promoter of various genes (Fig. 2). Although NF- $\kappa$ B regulates many different genes at the transcriptional level, most are key modulators of inflammatory and immune responses. Among this general class of genes are cytokines and chemokines such as interleukin-1 (IL-1), IL-2, IL-6, IL-8, tumor necrosis factor- $\alpha$

(TNF- $\alpha$ ),  $\beta$ -interferon, and  $\gamma$ -interferon. Other NF- $\kappa$ B-induced genes important for inflammation and immunity are certain cell adhesion molecules, complement factors, antiapoptotic genes, and immune cell receptors.

Upon binding a consensus  $\kappa$ B responsive element (i.e., 5'-GGGRNYYCC-3'), NF- $\kappa$ B stimulates transcription of such genes by interacting with other transcription factors or with proteins called coactivators. Although an increasing number of transcription factors and coactivators can act as a bridge between NF- $\kappa$ B and the basal transcription machinery to enhance transcription, some of these proteins can also repress NF- $\kappa$ B transactivation in a context-specific manner. For instance, CCAAT-enhancer binding protein- $\beta$  can repress NF- $\kappa$ B at the promoter of the angiotensinogen gene or synergize with NF- $\kappa$ B at the promoter of the IL-6 and IL-8 genes. The cyclic AMP-response element binding protein (CBP/p300) acts in concert with NF- $\kappa$ B to stimulate gene transcription. CBP/p300 is particularly important because it serves to mediate mutually antagonistic interactions between NF- $\kappa$ B and the glucocorticoid receptor, which mediates the anti-inflammatory and immunosuppressive effects of glucocorticoids. Interestingly, AP-1 (Fos-Jun) is another transcription factor that displays positive interactions with NF- $\kappa$ B but negative interactions with the glucocorticoid receptor.

### FACTORS AND RECEPTORS THAT ACTIVATE NF- $\kappa$ B

Having outlined the common pathway for NF- $\kappa$ B activation, it is now fruitful to examine some of the diverse signaling cascades that converge on the IKK/I $\kappa$ B/NF- $\kappa$ B system. NF- $\kappa$ B activation can occur in response to general cellular stress or in response to specific signaling molecules. For example, oxidative stress and DNA damage, which are produced by many agents, lead to NF- $\kappa$ B activation. In contrast, more specific receptors transduce endogenous signals from cytokines such as TNF- $\alpha$  and IL-1 into an intracellular cascade that activates NF- $\kappa$ B. A wide range of pathogens, including bacteria and viruses, also activate NF- $\kappa$ B, either via specific membrane-bound receptors or by coopting intracellular components of signaling pathways. There is also significant crosstalk or interaction between NF- $\kappa$ B and other transcription factors such as the steroid receptors. One of the most prominent is the glucocorticoid receptor, which antagonizes the activated form of

NF- $\kappa$ B, thereby blocking inflammatory and immune responses. A brief overview of some of the receptors and the signaling mechanisms involved in NF- $\kappa$ B activation and inhibition is provided in the following sections.

### TNF- $\alpha$ and TNFR1

The TNF family of ligands, including TNF- $\alpha$ , lymphotoxin- $\alpha$  (LT- $\alpha$ ), LT- $\beta$ , and LIGHT, are generally transmembrane proteins that act in a juxtacrine fashion to influence the function of adjacent cells. The sole exception is LT- $\alpha$ , which is normally a secreted factor. Despite being active in its membrane-bound form, TNF- $\alpha$  can also be released by proteolytic cleavage to form an active, soluble ligand. The TNF family of ligands is trimeric and usually composed of one type of protein. Lymphotoxin- $\beta$ , however, forms biologically active heterotrimers with LT- $\alpha$ . Although each ligand has its own membrane-bound receptor, TNF- $\alpha$  and its cognate receptor, TNFR1, are the best characterized pair with regard to the intracellular signaling pathways that activate NF- $\kappa$ B.

The TNF- $\alpha$  and TNFR1 signaling pathway is of particular interest because it can either induce or inhibit programmed cell death (apoptosis). In fact, NF- $\kappa$ B plays a pivotal role in determining whether an apoptotic program is initiated or whether a survival program is activated. TNF- $\alpha$  triggers apoptosis in NF- $\kappa$ B-deficient cells or when NF- $\kappa$ B-induced gene expression is blocked. In contrast, TNF- $\alpha$ -induced activation of NF- $\kappa$ B results in the synthesis of survival factors that block apoptosis. The earliest events in the latter signaling pathway include binding of trimeric TNF- $\alpha$  to at least two TNFR1 molecules and the juxtaposition of the intracellular domains of TNFR1. Association of two or more intracellular domains leads to the recruitment of TNFR1-associated death domain, which in turn recruits receptor-interacting protein 1, TNF receptor-associated factor 2 (TRAF2), and TRAF1. The events downstream from this step are still unclear but appear to involve activation of kinase cascades that ultimately phosphorylate and activate the IKK complex. Although TNF- $\alpha$  undoubtedly activates NF- $\kappa$ B, a great deal of work remains to be done to provide a clear picture of the entire cascade leading up to the common NF- $\kappa$ B activation pathway.

### IL-1 and IL-1R1

Interleukins are a family of 10 cytokines that exert inflammatory effects on diverse tissues. The first interleukins discovered were IL-1 $\alpha$  and IL-1 $\beta$ .

Considering that these two molecules are regulated independently, it is somewhat surprising that they exhibit essentially the same biological effects by interacting with the same membrane-bound receptors. Interleukin-1 receptor 1 (IL-1R1) is a functional receptor that can be activated by both IL-1 $\alpha$  and IL-1 $\beta$ . In contrast, IL-1R2 is nonfunctional and appears to act as a "decoy receptor" that can competitively inhibit IL-1 signaling via IL-1R1. This functional difference is not due to any difference in the extracellular domain of the receptors, which are highly homologous in primary structure and contain three immunoglobulin-like domains. Rather, IL-1R1 and IL-1R2 differ in their intracellular domain: IL-1R1 contains a domain that is truncated in IL-1R2. Indeed, fusion of the intracellular domain from IL-1R1 to the extracellular domain of IL-1R2 makes the latter receptor functionally active and responsive to IL-1 $\alpha$  and IL-1 $\beta$ .

The intracellular domain of IL-1R1 therefore plays a critical role in transducing the extracellular signal into an inflammatory cellular response. Specifically, IL-1 binds to IL-1R1, resulting in a conformational change in the receptor. The ligand-receptor complex then recruits IL-1R1 accessory protein. This ternary complex is required for intracellular signaling events that include recruitment of the adapter and effector protein MyD88. MyD88 activates interleukin receptor associated kinase (IRAK) and TRAF6. In turn, TRAF6 activates a downstream kinase cascade, including TGF- $\beta$ -activated kinase 1 and NF- $\kappa$ B-inducing kinase (NIK). Ultimately, NIK appears to phosphorylate and activate IKK. Although the last step in this cascade has not been definitively established, it is clear that the proinflammatory effects of IL-1 are mediated by activation of the IKK/I $\kappa$ B/NF- $\kappa$ B system.

### Microbial Pathogens and Toll-like Receptors

Toll was initially discovered in *Drosophila* as a protein involved in the development of the dorsal-ventral axis, but it was subsequently found to be a key protein in triggering innate immune responses. Twelve mammalian homologs of Toll, called toll-like receptors (TLRs), have been found and more may well be discovered. Although TLRs have a distinct extracellular domain, they have an intracellular domain very much like that of interleukin receptors. Consequently, TLRs signal through an intracellular mechanism very much like that of IL-1R1. The unique extracellular domain, however, varies among different TLRs and appears to

play a role in recognition of specific groups of pathogens, such as gram-negative bacteria, gram-positive bacteria, mycobacteria, and yeasts. The specificity of particular TLRs for certain components of each class of pathogens may be important in triggering tailored immune responses. For example, TLR4 recognizes components of the bacterial cell wall such as lipopolysaccharide and does so in conjunction with CD14. Lipopolysaccharide is a well-known and widely used activator of NF- $\kappa$ B. Deletion of either TLR4 or CD14 leads to an extremely blunted inflammatory response to lipopolysaccharide. After the external domain binds lipopolysaccharide, the internal domain of TLR4 undergoes a conformational change and physically interacts with MyD88. This protein then engages IRAK, TRAF6, as well as other downstream components typical of the IL-1 signaling cascade. Although TLR4 activates the IKK/I $\kappa$ B/NF- $\kappa$ B system, additional work needs to be done on other TLRs to determine if their effects are also mediated by NF- $\kappa$ B and to explore more rigorously the hypothesis that different TLRs produce tailored responses to particular pathogens.

### Viral Infection and Viral Proteins

A broad range of viruses are capable of activating the IKK/I $\kappa$ B/NF- $\kappa$ B signaling pathway: NF- $\kappa$ B is activated by human immunodeficiency virus (HIV-1), human T cell leukemia virus type I (HTLV-1), hepatitis B virus, hepatitis C virus, influenza virus, and others. To be more accurate, viral products are the proximate factors that lead to NF- $\kappa$ B activation. Nevertheless, these factors do not appear to trigger NF- $\kappa$ B by activating specific pattern-recognition receptors analogous to TLRs, as do bacterial pathogens. Instead, viral proteins interact at various points in different NF- $\kappa$ B signaling pathways. Because of the large number of viruses that activate NF- $\kappa$ B and the diverse mechanisms involved, it is beyond the scope of this article to review more than a few well-characterized and medically important pathways.

HTLV-1 and a protein called Tax that it encodes are a major cause of adult T cell leukemia. The Tax protein is required and sufficient on its own to cause transformation of primary human lymphocytes into cancer cells. Although this protein interacts with a number of transcription factors to alter patterns of gene expression, we are primarily interested in how this protein takes over NF- $\kappa$ B signaling. The mechanism appears to be fairly simple in that Tax directly interacts with the IKK- $\gamma$  subunit of IKK, leading

to persistent activation of IKK- $\alpha$  and IKK- $\beta$ , continuous degradation of I $\kappa$ B, and sustained activation of NF- $\kappa$ B. Given that Tax-mediated activation of NF- $\kappa$ B plays a key role in the expression of downstream genes involved in malignant transformation by HTLV-1, it is important to note that disruption of NF- $\kappa$ B signaling may be involved in a number of other cancers.

Multiple mechanisms activate NF- $\kappa$ B during HIV-1 infection. For example, double-stranded (ds) RNA-activated kinase PKR senses the intermediate retroviral dsRNA, is thereby activated, and phosphorylates NIK and IKK. This pathway of NF- $\kappa$ B activation is presumably common to most, if not all, retroviruses, including HTLV-1. Binding of HIV-1 to CD4 also appears to activate NF- $\kappa$ B by way of two additional pathways that have not been discussed and are not as clearly defined. In one, HIV-1 interacts with CD4 and crosslinks the receptor. Crosslinked CD4 stimulates Ras, Raf, and mitogen-activated protein kinase pathways that appear to phosphorylate IKK. In the other pathway, HIV-1 binding to CD4 results in phosphorylation of phosphatidylinositol 3-kinase, the subsequent activation of Akt serine-threonine kinase, and the ultimate phosphorylation and activation of IKK. Overall, viral activation of NF- $\kappa$ B is prolonged relative to the normal pattern of activation in response to endogenous signals. This occurs because viruses circumvent regulatory mechanisms that would normally inactivate the NF- $\kappa$ B signaling pathway.

### ANTAGONISTIC INTERACTIONS BETWEEN NF- $\kappa$ B AND GLUCOCORTICOIDS

Glucocorticoids are a class of lipophilic, steroidal hormones produced by the adrenal cortex. These hormones are synthesized and secreted in response to physical, physiological, and social stressors, including pain, starvation, and agonistic interactions with other individuals of the same species. Much like the other well-known adrenal hormone adrenaline, glucocorticoids play a central role in the fight or flight response. Glucocorticoids do so by binding to specific glucocorticoid receptors and profoundly influencing the physiology and function of numerous tissues in vertebrates. For instance, glucocorticoids have well-characterized effects on the brain and behavior. In addition, they mobilize energy stores by promoting gluconeogenesis in the liver, degradation of proteins to free amino acids in muscle, and lipolysis in adipose tissues. This class of hormones also has medically

important immunosuppressive and anti-inflammatory effects that are, in part, mediated by antagonistic interactions with the NF- $\kappa$ B signaling pathway. In the following sections, we discuss how glucocorticoids and the glucocorticoid receptor block NF- $\kappa$ B signaling and vice versa.

### General Mechanism of Glucocorticoid Action

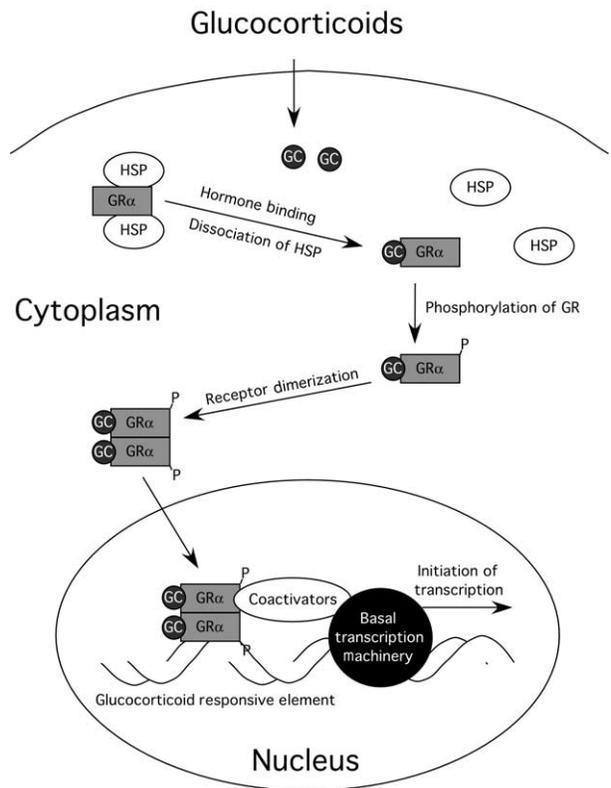
Before we discuss the influence of glucocorticoids on NF- $\kappa$ B signaling, we briefly describe the classical mode of action of these hormones. In general, glucocorticoids have their effect by entering cells and interacting with the intracellular, cytoplasmic glucocorticoid receptor (GR $\alpha$ ). Upon binding ligand with a high affinity, the monomeric GR $\alpha$  undergoes a conformational change, dissociates from regulatory heat shock proteins, and is phosphorylated. The glucocorticoid-GR $\alpha$  complex then forms homodimers and translocates to the nucleus, where it binds to specific DNA sequences called glucocorticoid responsive elements (GREs). In this way, the glucocorticoid-GR $\alpha$  complex usually stimulates gene transcription by interacting with coactivators and the basal transcription machinery (Fig. 3).

### Glucocorticoid Induction of I $\kappa$ B

Glucocorticoids can inhibit NF- $\kappa$ B signaling by inducing the expression of new protein. In particular, the glucocorticoid-GR $\alpha$  complex binds to GREs within the promoter of the I $\kappa$ B gene and stimulates its transcription (Fig. 4A). This messenger RNA is translated into the inhibitory I $\kappa$ B protein, which then sequesters activated, nuclear NF- $\kappa$ B and directs its movement out of the nucleus. Synthesis of new I $\kappa$ B protein is a general mechanism for terminating NF- $\kappa$ B activity and is observed both *in vivo* in whole animals and *in vitro* in cell culture. However, an important caveat is that glucocorticoid-induced expression of I $\kappa$ B has not been observed in all studies and may be cell-type specific. Moreover, this effect is relatively slow and may be more important for resetting the system to its basal homeostatic state rather than inhibiting the initial transcriptional activity of NF- $\kappa$ B.

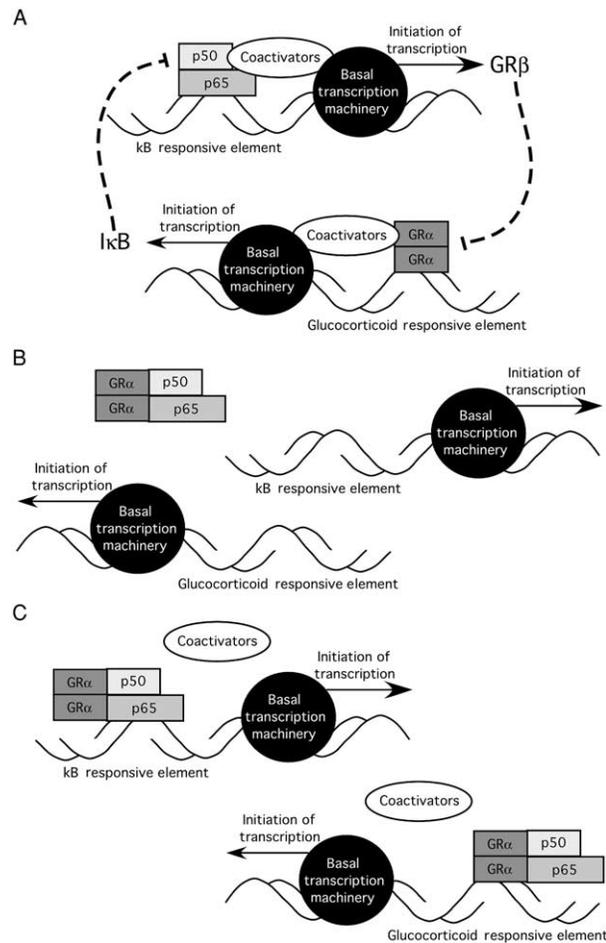
### Direct Interactions between NF- $\kappa$ B and GR $\alpha$

Interestingly, GR $\alpha$  can inhibit the transcriptional activity of NF- $\kappa$ B more directly by physically



**Figure 3** General mechanism of action for glucocorticoid receptors as described in the text.

interacting with the NF- $\kappa$ B protein. Given that this interaction does not require new protein synthesis, it is perhaps the most important mechanism whereby glucocorticoids immediately antagonize NF- $\kappa$ B signaling. There is evidence that such protein-protein interactions induce conformational changes that prevent NF- $\kappa$ B from binding  $\kappa$ B responsive elements (Fig. 4B). In other cases, however, the interaction with GR $\alpha$  does not block DNA binding by NF- $\kappa$ B. In these instances, GR $\alpha$  actually associates with the transactivation domain of the p65 subunit of NF- $\kappa$ B while it sits on  $\kappa$ B responsive elements (Fig. 4C). The glucocorticoid-GR $\alpha$  complex thereby disrupts NF- $\kappa$ B interactions with the basal transcription machinery that are mediated by coactivators. Another hypothesis that has less support is that GR $\alpha$  competes for a limited number of coactivators that are required to form a bridge between NF- $\kappa$ B and the basal transcription machinery. Finally, it is possible that GR $\alpha$  binds to DNA sequences called negative GREs within the promoter of NF- $\kappa$ B responsive genes. Negative GREs of this sort are known to mask binding sites for other transcription factors as well as the basal transcription machinery. Additional studies are required to clarify the



**Figure 4** (A) Induction of inhibitory proteins I $\kappa$ B and GR $\beta$ . (B) GR-NF- $\kappa$ B interactions that block DNA binding. (C) Disrupted interactions with transcriptional machinery.

relative importance of these mechanisms and whether they contribute to cell- or tissue-specific patterns of glucocorticoid antagonism of NF- $\kappa$ B signaling.

### NF- $\kappa$ B Induction of Dominant Negative GR $\beta$

There is reciprocal antagonism of glucocorticoid signaling by activated NF- $\kappa$ B. The mechanisms responsible for this effect are essentially the converse of those previously described, including interference at the level of DNA binding (Fig. 4B) and obstruction of interactions with the basal transcription machinery (Fig. 4C). However, studies indicate that activation of NF- $\kappa$ B by IL-1 or TNF- $\alpha$  may also cause glucocorticoid insensitivity in a manner analogous to the induction of I $\kappa$ B. In particular, IL-1 and TNF- $\alpha$  stimulate the preferential accumulation of a dominant

negative form of the glucocorticoid receptor called GR $\beta$ . This alternatively spliced isoform of GR is only found in humans, is constitutively nuclear, and blocks the transcriptional activity of GR $\alpha$  when it is more abundant than the normal isoform. Although the physiological importance of GR $\beta$  has been debated, it has been reported that levels of GR $\beta$  are elevated in disease states refractory to treatment with glucocorticoids. Thus, stimuli that activate NF- $\kappa$ B may block the anti-inflammatory effects of glucocorticoids by inducing GR $\beta$  and glucocorticoid resistance (Fig. 4A).

### CONCLUSION

NF- $\kappa$ B is a ubiquitously expressed transcription factor that regulates the expression of numerous genes involved in stress and inflammatory responses. Although one form of NF- $\kappa$ B is constitutively nuclear and transcriptionally active in B cells, forms of NF- $\kappa$ B in other cell types are retained in the cytoplasm in association with inhibitory proteins called I $\kappa$ Bs. Inactive cytoplasmic NF- $\kappa$ B, however, is an inducible transcription factor. Activation of NF- $\kappa$ B in these cells depends on the phosphorylation, ubiquitination, and degradation of the I $\kappa$ B protein, which exposes a nuclear localization signal on NF- $\kappa$ B. Subsequent translocation of NF- $\kappa$ B to the nucleus results in binding of this transcription factor to specific DNA sequences called  $\kappa$ B responsive elements within the promoters of inducible genes, including important cytokines and chemokines. A number of receptors transduce endogenous signals such as TNF- $\alpha$  and IL-1 into intracellular signaling cascades that activate NF- $\kappa$ B. A wide range of exogenous stimuli, including viral and bacterial pathogens, DNA-damaging agents, and oxidative stress, will also trigger the core IKK/I $\kappa$ B/NF- $\kappa$ B signaling pathway. In summary, NF- $\kappa$ B is an extremely important transcription factor that has pleiotropic effects and integrates endogenous and exogenous signals. Because it plays such a pivotal role in normal physiology, misregulation of NF- $\kappa$ B can lead to various disease states, including asthma, arthritis, autoimmune disease, and certain cancers.

Glucocorticoids have an important influence on NF- $\kappa$ B activity by opposing its action at a number of levels. Glucocorticoids bind to the GR, which then dissociates from regulatory heat shock proteins. The GR subsequently undergoes phosphorylation and translocates to the nucleus, where it binds to specific DNA sequences called glucocorticoid responsive elements within the promoters of inducible genes such as I $\kappa$ B. Newly synthesized I $\kappa$ B then interacts

with the activated, nuclear form of NF- $\kappa$ B and escorts it out of the nucleus. Importantly, the activated GR can also physically interact with NF- $\kappa$ B and directly block its transcriptional activity. Antagonistic interactions between GR and NF- $\kappa$ B signaling pathways explain an important part of the anti-inflammatory and immunosuppressive effects of glucocorticoids, which are widely prescribed for the treatment of disease states related to NF- $\kappa$ B misregulation.

### See Also the Following Articles

Glucocorticoid Receptor • Glucocorticoids, Overview • Interferons • Interleukin-2 • Tumor Necrosis Factor (TNF)

### Further Reading

- Auphan, N., DiDonato, J. A., Rosette, C., Helmberg, A., and Karin, M. (1995). Immunosuppression by glucocorticoids: Inhibition of NF-kappa B activity through induction of I kappa B synthesis. *Science* **270**, 286–290.
- Hiscott, J., Kwon, H., and Genin, P. (2001). Hostile takeovers: Viral appropriation of the NF- $\kappa$ B pathway. *J. Clin. Invest.* **107**, 143–151.
- Karin, M., and Ben-Neriah, Y. (2000). Phosphorylation meets ubiquitination: The control of NF- $\kappa$ B activity. *Annu. Rev. Immunol.* **18**, 621–663.
- Legrand-Poels, S., Schoonbroodt, S., Matroule, J. Y., and Piette, J. (1998). NF- $\kappa$ B: An important transcription factor in photobiology. *J. Photochem. Photobiol. B Biol.* **45**, 1–8.
- McKay, L. I., and Cidlowski, J. A. (1999). Molecular control of immune/inflammatory responses: Interactions between nuclear factor- $\kappa$ B and steroid receptor-signaling pathways. *Endocr. Rev.* **20**, 435–459.
- Mercurio, F., and Manning, A. M. (1999). NF- $\kappa$ B as a primary regulator of the stress response. *Oncogene* **18**, 6163–6171.
- Takeuchi, O., and Akira, S. (2001). Toll-like receptors: Their physiological role and signal transduction system. *Int. Immunopharmacol.* **1**, 625–635.
- Webster, J. C., Oakley, R. H., Jewell, C. M., and Cidlowski, J. A. (2002). Proinflammatory cytokines regulate human glucocorticoid receptor gene expression and lead to the accumulation of the dominant negative b isoform: A mechanism for the generation of glucocorticoid resistance. *Proc. Natl. Acad. Sci. USA* **98**, 6865–6870.

## CLINICAL FEATURES OF PATIENTS WITH PAGET'S DISEASE

As noted above, the majority of patients with Paget's disease are asymptomatic and bone pain is the most common symptom. Skeletal deformities are usually seen in the skull, face, or lower extremities, with the lumbar spine, femur, and pelvis being the most commonly affected bones. Deformities of bone, such as an increase in head size, bowing of a limb, or curvature of the spine, may occur in advanced cases. Pagetic bone is susceptible to fracture with moderate stress. Bone pain associated with Paget's disease is usually due to degenerative arthritis in joints contiguous to bones affected with Paget's disease. Severe headaches or deafness from nerve entrapment can develop from skull involvement with Paget's disease. Dental problems due to temporomandibular involvement can also occur. Involvement of the spine can result in compression fractures, spinal stenosis, and neurological impairment, as well as degenerative arthritis. Because of the increased vascularity of pagetic lesions, the lesions can feel warm to the touch and elderly patients can develop high-output cardiac failure. In addition, patients with Paget's disease can develop hypercalciuria and hypercalcemia, due to accelerated bone resorption induced by immobilization.

Bone radiographs and isotopic bone scans are abnormal in patients with Paget's disease. Bone scans show marked accumulation of the radiolabeled bisphosphonates in the region where the bone formation is markedly increased and can outline early lesions that are not detectable by radiographic techniques. Bone scans are the most sensitive method of detecting pagetic lesions and can be used to follow the activity of the disease in these patients. In the skull, the patients can develop discrete round or oval lesions in the frontal occipital bones, called "osteoporosis circumscripta." In the long bones, the lesions usually begin in the subchondral region of either epiphysis and advance proximally or distally at a rate of approximately 1 cm per year in untreated patients. Initial lesions appear osteolytic, followed by a chaotic sclerotic appearance, and finally become osteosclerotic. Considerable thickening of the sclerotic bone results in bone deformity. Paget's disease usually does not cross the joint space to affect an adjacent bone and, as noted above, it is extremely rare for new lesions to develop in previously uninvolved bones in patients with Paget's disease over the course of their disease.

## BIOCHEMICAL MARKERS AND HISTOLOGIC ABNORMALITIES

The biochemical markers provide an integrated assessment of the cellular events occurring throughout the skeleton of patients with Paget's disease. The earliest index of bone matrix resorption was measurement of urinary hydroxyproline excretion while ingesting a low-gelatin diet. This index is well correlated with the extent of the disease despite the fact that hydroxyproline is a prominent component of extraskelatal connective tissue as well as skeletal collagen. Newer markers, such as collagen cross-links and associated peptides used, reflect the primary lesion in Paget's disease, the increase in bone resorption. Urinary N-telopeptide, pyridinoline, and deoxypyridinoline have all been reported to be more specific indices of skeletal matrix resorption and are not influenced by dietary gelatin. Serum tartrate-resistant acid phosphatase, presumably released by osteoclasts, appears to be another index of bone resorption in Paget's disease but is not routinely used. The most useful markers for the increased osteoblast activity in Paget's disease are the total alkaline phosphatase and bone-specific alkaline phosphatase activity levels in serum. Patients also showed significantly higher endothelin-1 circulating levels than controls, with a positive correlation with serum alkaline phosphatase, but not with urinary hydroxyproline. However, serum calcium levels are typically normal in Paget's disease and serum osteocalcin levels appear to be a poor index of the progression of the disease.

Histologic examination of bone biopsy revealed abundant structurally abnormal osteoclasts. Osteoclasts are increased in number and size and contain as many as 100 nuclei per multinucleated cell compared to 3 to 5 nuclei for a normal osteoclast. These osteoclasts have characteristic ultrastructural abnormalities including microfilaments, paracrystalline arrays located in the nucleus and sometimes in the cytoplasm, which are absent in nonpagetic bone or bone marrow cells. These inclusions closely resemble nucleocapsids of viruses of the *Paramyxoviridae* family. Osteoblasts are also increased in lesions in patients with Paget's disease and they appear to be morphologically normal. Osteoblasts contain abundant rough endoplasmic reticulum and mitochondria in a well-developed Golgi zone, consistent with the increased bone formation activity that occurs in the active lesions. In advanced lesions in patients with Paget's disease, the marrow is also abnormal. Normal

hematopoietic elements are usually absent and replaced by mononuclear cells intermixed with highly vascular connective tissue. The bone matrix in Paget's disease is highly abnormal in structure due to disordered bone remodeling. The bone matrix consists of erratic patterns of "cement lines" and demonstrates a "mosaic" pattern. The matrix is interspersed with numerous foci of woven bone, reflecting the increased rates of bone deposition.

## TREATMENT OF PAGET'S DISEASE

Bisphosphonates are the most common treatment for patients with Paget's disease. These inorganic phosphate compounds inhibit osteoclast-mediated bone resorption and induce osteoclast apoptosis. These compounds have been shown to inhibit osteoclast formation and osteoclastic bone resorption through their effects on the cholesterol biosynthesis pathway. Etidronate was the original bisphosphonate used in patients with Paget's disease. It is an orally active bisphosphonate that lowers bone turnover, urinary hydroxyproline, and serum alkaline phosphatase levels by approximately 50% in patients with Paget's disease. Since that time, more potent bisphosphonates have been used in patients with Paget's disease. Pamidronate, an intravenous bisphosphonate, induces long-term remissions in patients with Paget's disease at doses of 60 or 90 mg given by single infusions. Additional courses of treatment can be given to patients to achieve normal or near-normal serum alkaline phosphatase levels for long periods of time. Other more potent bisphosphonates, including tiludronate, alendronate, and residronate, have been used in patients with Paget's disease. Residronate is 1000 times more potent than etidronate in its capacity to inhibit bone resorption and given orally over 2 to 3 months can normalize serum alkaline phosphatase levels in 50–70% of patients. All oral bisphosphonates induce mild upper gastrointestinal symptoms and rarely cause iritis. Clodronate treatment has also been shown to reduce serum levels of interleukin-6 (IL-6) soluble receptor in patients with Paget's disease. Zoledronate is the most potent bisphosphonate under investigation to treat Paget's disease. Patients also have been treated with salmon calcitonin or human calcitonin, which induced remissions and lowered bone turnover, urinary hydroxyproline, and serum alkaline phosphatase levels by approximately 50%. However, resistance to calcitonin frequently developed in the majority of patients. Mithramycin, a cytotoxic chemotherapeutic agent, has been used

to treat patients with Paget's disease who are unresponsive to calcitonin or the bisphosphonates. This agent has the potential for severe toxic side effects and is used only in patients who are refractory to other therapies.

## GENETICS OF PAGET'S DISEASE

Familial incidence is common in Paget's disease and 15–40% of patients with the disease have an affected first-degree relative. Therefore, genetic factors play an important role in the pathogenesis of Paget's disease of bone. The disease often is inherited in an autosomal-dominant manner, manifesting genetic heterogeneity and incomplete penetrance. Familial Paget's disease has an equal incidence in males and females. A genetic locus for Paget's disease has been identified on chromosome 18q in a region near the familial expansile osteolysis (FEO) locus in several large families with Paget's disease. FEO is a disease related to Paget's disease but occurs in patients at a much younger age and is a much more severe disease. FEO is an extremely rare disease, affecting only a very limited number of kindreds in the world; it is also mapped to chromosome 18q and is linked to activating mutations in the TNFRSF11A gene, which encodes receptor activator of nuclear factor  $\kappa$ B (NF- $\kappa$ B) (gene symbol RANK). In patients with juvenile Paget's disease, a homozygous deletion of the gene on chromosome 8q24.2 that encodes osteoprotegerin, a member of the superfamily of tumor necrosis factor receptors, has been reported. However, linkage studies, coupled with mutation screening, have excluded involvement of RANK and also osteoprotegerin in the majority of patients with Paget's disease of bone. Studies also indicated that patients with Paget's disease have an increased incidence of osteosarcoma, with approximately 1% of patients with Paget's disease developing osteosarcoma in an affected bone. This incidence of osteosarcoma is 1000 times higher than that in the general population for this age group. Genetic studies have demonstrated linkage in 7 of 7 patients with osteosarcoma to loss of heterozygosity in a region of 18q that is adjacent to or within a locus for Paget's disease on 18q.

A genome-wide search of familial Paget's disease of bone further indicated genetic heterogeneity of the disease, with candidate loci on chromosomes 2q, 10q, and 5q. In the gene encoding sequestosome 1 (SQSTM1/p62), mapped within the critical region on chromosome 5q35–qter, a proline-leucine amino acid change at codon 392 (P392L) was identified in French

Canadian patients with Paget's disease of bone. The frequency of mutation was 16 and 46% for sporadic and familial cases tested, respectively. Further studies also identified different mutations affecting the highly conserved ubiquitin-binding domain of SQSTM1/p62 protein in patients with familial and sporadic Paget's disease. It is well known that SQSTM1/p62 mediates IL-1 and tumor necrosis factor  $\alpha$  cytokine signaling to activate NF- $\kappa$ B. However, the precise role that SQSTM1/p62 may play in the pathogenesis of Paget's disease of bone remains to be elucidated.

## VIRAL ETIOLOGY

Since the early 1970s, a variety of studies have implicated paramyxoviruses in Paget's disease. A viral etiology has been proposed for Paget's disease, with an initial description of nucleocapsid-like structures in the nuclei and cytoplasm of pagetic osteoclasts by electron microscopy. Immunocytochemical studies further confirmed that these nuclear inclusions cross-reacted with antibodies that recognized measles virus or respiratory syncytial virus nucleocapsid antigens. *In situ* hybridization techniques also identified the presence of measles virus mRNA sequences in up to 90% of osteoclasts and other mononuclear cells in pagetic bone specimens. Similarly, canine distemper virus nucleocapsid antigens were also detected in osteoclasts from patients with Paget's disease. These paramyxoviral-like nuclear inclusions are not unique to Paget's disease and were reported in patients with FEO and rarely in patients with osteopetrosis, pycnodysostosis, otosclerosis, or oxalosis. This has raised the possibility that the virus is a nonetiologic agent in a cell altered by a genetic defect.

To further explore the viral etiology, using reverse transcription-polymerase chain reaction (RT-PCR) analysis, the measles virus nucleocapsid transcripts from freshly isolated bone marrow cells from patients with Paget's disease were amplified. These measles virus nucleocapsid transcripts contain mutations clustered at the C-terminal end of the messenger RNA. All these mutations were sense mutations and resulted in amino acid substitutions in the nucleocapsid gene product. The mutations occurred at a rate of 1% in the total measles virus nucleocapsid gene isolated from a patient with Paget's disease. It was further demonstrated that osteoclast precursors, the granulocyte/macrophage colony-forming units (GM-CFU), as well as mature osteoclasts from patients with Paget's disease, expressed measles virus nucleocapsid transcripts. Since GM-CFU circulate and also give rise to

monocytes and granulocytes in the peripheral blood, peripheral blood mononuclear cells from patients with Paget's disease and normals were then examined for expression of measles virus nucleocapsid transcripts. RT-PCR analysis demonstrated that peripheral blood samples from 9 of 10 patients with Paget's disease contained measles virus nucleocapsid transcripts, whereas none of the 10 normals tested expressed measles virus nucleocapsid transcripts. The authors have been unable to find canine distemper virus or respiratory syncytial virus nucleocapsid transcripts in patients. In contrast, canine distemper virus nucleocapsid transcripts were detected in affected bones from 100% of patients tested using *in situ* RT-PCR techniques. Furthermore, it has also been demonstrated that infecting canine bone marrow cells with canine distemper virus results in the development of multinucleated cells that share some of the phenotypic characteristics of pagetic osteoclasts. However, other workers have been unable to detect paramyxoviral nucleocapsid transcripts in samples obtained from patients with Paget's disease.

The presence of measles virus or canine distemper virus transcripts in osteoclasts and osteoclast precursors from patients with Paget's disease does not indicate a pathophysiologic role for these genes in the development of the pagetic lesions. It is possible that these paramyxoviral transcripts and paramyxoviral-like inclusions are simply markers for the disease and have no pathophysiologic significance. Studies found that normal osteoclast precursors (GM-CFU) transduced with retroviral vectors expressing the measles virus nucleocapsid gene formed large osteoclasts more rapidly and with increased numbers of nuclei, were hypersensitive to 1,25-dihydroxyvitamin D<sub>3</sub> [1,25-(OH)<sub>2</sub>D<sub>3</sub>], and had increased bone-resorbing capacity compared to normal osteoclasts. In contrast, normal osteoclast precursors transduced with the measles virus matrix gene did not express an abnormal phenotype. In further studies, we have targeted CD46, human measles virus receptor to cells of the osteoclast lineage in transgenic mice, and demonstrated that measles virus infection of osteoclast precursors from CD46 transgenic mice form osteoclasts, which express a pagetic phenotype *in vitro*. Taken together, these data suggest a potential pathophysiologic role for the paramyxoviral nucleocapsid gene that is expressed in patients with Paget's disease.

However, measles virus infection has a similar incidence worldwide and occurs in very young patients, whereas Paget's disease is a disease of the elderly. These observations suggest that if paramyxoviruses have an etiologic role in Paget's disease, these viral infections must persist for long periods of time. To

further investigate a potential site for the initial infection of osteoclast precursors with Paget's disease, the hypothesis that very early pluripotent hematopoietic stem cells, which can persist for long periods of time in a quiescent phase, may be the initial target for the paramyxoviral infection in patients with Paget's disease was tested. It was found that other hematopoietic lineages from patients with Paget's disease in addition to the osteoclast lineage, including the erythroid and the erythroid precursors (burst-forming unit-erythroid, or BFU-E) and multipotent myeloid precursors (colony-forming unit granulocyte, erythrocyte, megakaryocyte, macrophage, or CFUGEMM), which give rise to megakaryocytes, monocytes, erythroid cells, and granulocytes, also contain paramyxoviral nucleocapsid transcripts. Thus, if the initial site of infection occurs in a small number of primitive pluripotent hematopoietic stem cells that predominantly remain in G<sub>0</sub>, this might explain the chronicity of the infection. Furthermore, there may be a genetic predisposition for chronic paramyxoviral infections of hematopoietic precursors in patients with Paget's disease. However, a cause-and-effect relationship of paramyxoviruses in Paget's disease remains to be proven as no infectious virus has been isolated from pagetic cells and, in addition, it is not clear how the initial lesion occurs in Paget's disease.

## PAGETIC OSTEOCLASTS AND OSTEOCLAST PRECURSORS

Bone marrow culture techniques identified several abnormalities in osteoclast formation and osteoclast precursors from patients with Paget's disease. These studies provided new insights into the potential pathogenesis of Paget's disease. Osteoclast-like multinucleated cells formed more rapidly with increased numbers (10- to 100-fold) and with more nuclei per osteoclast and they expressed high levels of tartrate-resistant acid phosphatase in marrow cultures from patients with Paget's disease compared to normals. In addition, osteoclast formation in pagetic bone marrow cultures was induced at 1,25-(OH)<sub>2</sub>D<sub>3</sub> concentrations that were 10 to 100 times lower than those required in normal marrow cultures. Structural examination of the osteoclast-like cells that formed in these cultures showed that they had many of the features of pagetic osteoclasts but lacked the characteristic nuclear and cytoplasmic inclusions. Immunocytochemical studies confirmed that measles virus and respiratory syncytial virus nucleocapsid antigens were expressed in osteoclasts formed *in vitro* in these cultures.

Osteoclasts from patients with Paget's disease also appear to produce increased levels of IL-6 and express high levels of IL-6 receptors compared to normal osteoclasts. *In situ* hybridization studies have further identified increased levels of IL-6, c-fos proto-oncogene, and Bcl-2 anti-apoptotic gene mRNA expression in pagetic osteoclasts. IL-6 receptor and nuclear factor-IL-6 mRNA levels were also increased in osteoclasts from bone samples from patients with Paget's disease compared to those with osteoarthritis. These data suggest that IL-6, which stimulates human osteoclast formation, may act as an autocrine/paracrine factor to enhance osteoclast formation in patients with Paget's disease and increase the osteoclast precursor pool. IL-6 levels were also shown to increase in bone marrow plasma and peripheral blood of patients with Paget's disease. In addition, the increased levels of IL-6 in the peripheral blood of patients with Paget's disease may in part explain the increased bone remodeling seen in bones not clinically involved with Paget's disease.

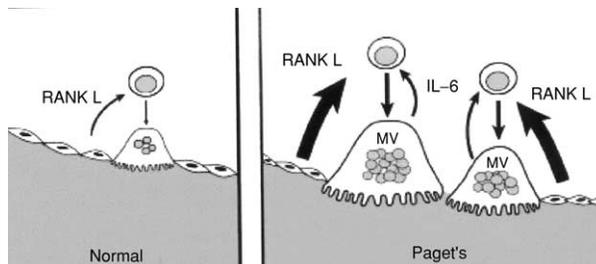
To further investigate the potential abnormalities in osteoclast precursors in patients with Paget's disease, studies were conducted to assess the number of osteoclast precursors in marrow aspirates from involved bones from patients with Paget's disease and compared these to normals. It has been found that the number of early osteoclast precursors (GM-CFU) was increased significantly in marrow aspirates from patients with Paget's disease compared to normals. Interestingly, when the osteoclast precursors were separated from the marrow microenvironmental elements present in the marrow aspirates, similar numbers of osteoclast precursors were detected in these aspirates. These data suggested that the marrow microenvironment enhanced osteoclast precursor growth compared to the normal marrow microenvironment.

To determine the potential role of the marrow microenvironment and enhanced osteoclast formation in patients with Paget's disease, reconstitution experiments were conducted using highly purified populations of osteoclast precursors from patients with Paget's disease or normals and marrow stromal cells from patients with Paget's disease and normals. Coculture of normal osteoclast precursors with marrow stromal cells from patients with Paget's disease resulted in enhanced growth of the osteoclast precursors from normals. Interestingly, when osteoclast precursors from patients with Paget's disease were cocultured with marrow stromal cells from normals, they also showed increased growth. These data suggest that both the marrow microenvironment and the

osteoclast precursors are abnormal in patients with Paget's disease.

These studies also confirmed that the osteoclast precursors were hypersensitive to  $1,25\text{-(OH)}_2\text{D}_3$  compared to normals. The increased sensitivity to  $1,25\text{-(OH)}_2\text{D}_3$  of osteoclast precursors from Paget's patients is mediated through the vitamin D receptor. This was confirmed by up-regulation of 24-hydroxylase mRNA expression in pagetic osteoclast precursors at concentrations of  $1,25\text{-(OH)}_2\text{D}_3$  that are 1 log less than that required for normal osteoclast precursors. The increased sensitivity to  $1,25\text{-(OH)}_2\text{D}_3$  was not due to increased numbers of vitamin D receptors in pagetic osteoclast precursors compared to normals, but appeared to be due to enhanced affinity of the vitamin D receptor in pagetic cells for its ligand compared to normals. The basis for the increased affinity of  $1,25\text{-(OH)}_2\text{D}_3$  for the vitamin D receptor is unknown. Potentially it could be due to enhanced expression of coactivators that interact with the vitamin D receptor and  $1,25\text{-(OH)}_2\text{D}_3$  to induce gene expression or could result from decreased levels of suppression of corepressors that interact with  $1,25\text{-(OH)}_2\text{D}_3$  and the vitamin D receptor.

The osteoclast precursors from patients with Paget's disease also appear to be hyperresponsive to RANK ligand and marrow stromal cells from pagetic lesions have increased RANK ligand expression. RANK ligand is a critical osteoclast differentiation factor that is expressed on marrow stromal and osteoblast cells in response to several osteotropic factors. The increased sensitivity of osteoclast precursors from Paget's patients to RANK ligand appears to be due to interactions of these precursors with IL-6. The addition of neutralizing antibodies to IL-6 decreased the sensitivity to RANK ligand in the osteoclast precursors from patients with Paget's disease to normal levels. Similarly, the addition of IL-6 to cultures of normal osteoclast precursors enhanced the responsiveness of these precursors to RANK ligand to the levels seen with pagetic osteoclast precursors. The enhanced expression of RANK ligand and IL-6 in pagetic lesions could contribute to the abnormal osteoclast development and highly localized nature of Paget's disease (Fig. 1). It has been shown that SHIP mice, inositol 5' phosphatase-deficient mice, are severely osteoporotic with increased numbers of osteoclast precursors and hyperactive osteoclasts. In addition, serum levels of IL-6 are markedly increased in these mice as in Paget's disease. However, the basis for these abnormalities in both osteoclasts and osteoclast precursors from patients with Paget's disease is still unknown.



**Figure 1** Abnormal osteoclast formation in Paget's bone micro-environment. The osteoclast precursors contain measles virus transcripts (MV) and are hyperresponsive to RANK ligand (RANK L). The osteoclasts overexpress IL-6 and the marrow stromal/osteoblast cells overexpress RANK ligand.

## SUMMARY

Paget's disease of bone is a chronic focal skeletal disorder that affects up to 2 to 3% of the population over the age of 60 years. The disease is an autosomal-dominant trait with genetic heterogeneity. Recurrent mutations in the ubiquitin-associated domain of sequestosome 1 have been identified in patients with Paget's disease. Osteoclasts and osteoclast precursors from patients with Paget's disease are abnormal; they appear to be hyperresponsive to  $1,25\text{-(OH)}_2\text{D}_3$  and RANK ligand and contain paramyxoviral transcripts. The basis for these abnormalities and the role that the paramyxoviruses may play in this disease are still unclear.

## See Also the Following Articles

Bisphosphonates • Bone Remodeling, Dynamics of • Bone Structure • Bone Turnover Markers • Collagen Metabolism Disorders • Osteoporosis in Older Men • Osteoporosis in Older Women • Osteoporosis, Overview

## Further Reading

- Hansen, M. F., Nellisery, M. J., and Bhatia, P. (1999). Common mechanisms of osteosarcoma and Paget's disease. *J. Bone Miner. Res.* **14**, 39–44.
- Helfrich, M. H., Hobson, R. P., Grabowski, P. S., Zurbriggen, A., Cosby, S. L., Dickson, G. R., Fraser, W. D., Ooi, C. G., Crisp, A. J., Wallace, R. G., Kahn, S., and Ralston, S. H. (2000). A negative search for a paramyxoviral etiology of Paget's disease of bone: Molecular, immunological and ultrastructural studies in UK patients. *J. Bone Miner. Res.* **15**, 2315–2329.
- Hughes, A. E., Ralston, S. H., Marken, J., Bell, C., MacPherson, H., Wallace, R. G. H., van Hul, W., Whyte, M. P., Nakatsuka, K., Hovy, L., and Anderson, D. M. (2000). Mutations in TNFRSF11A, affecting the signal peptide of RANK, cause familial expansile osteolysis. *Nat. Genet.* **24**, 45–48.

- Laurin, N., Brown, J. P., Morissette, J., and Raymond, V. (2002). Recurrent mutation of the gene encoding sequestosome 1 (SQSTM1/p62) in Paget's disease of bone. *Am. J. Hum. Genet.* **70**, 1582–1588.
- Mee, A. P., and Sharpe, P. T. (1993). Dogs, distemper and Paget's disease. *BioEssays* **15**, 783–789.
- Neale, S. D., Smith, R., Wass, J. A. H., and Athanasou, N. A. (2000). Osteoclast differentiation from circulating mononuclear precursors in Paget's disease is hypersensitive to 1,25-dihydroxyvitamin D3 and RANKL. *Bone* **27**, 409–416.
- Reddy, S. V., Kurihara, N., Mena, C., Landucci, G., Forthal, D. F., Koop, B. A., Windle, J. J., and Roodman, G. D. (2001). Osteoclasts formed by measles virus infected osteoclast precursors from hCD46 transgenic mice express characteristics of pagetic osteoclasts. *Endocrinology* **142**, 2898–2905.
- Singer, F. R. (1999). Update on viral etiology of Paget's disease. *J. Bone Miner. Res.* **14**, 29–33.
- Siris, E. S. (1999). Paget's disease of bone. In "Primer on the Metabolic Bone Diseases and Disorders of Mineral Metabolism" (M. J. Favus, ed.), 4th ed., pp. 415–425. Lippincott, Williams & Wilkins, Philadelphia, PA.
- Whyte, M. P., Obrecht, S. E., Finnegan, P. M., Jones, J. L., Podgornik, M. N., McAlister, W. H., and Mumm, S. (2002). Osteoprotegerin deficiency and juvenile Paget's disease. *N. Engl. J. Med.* **347**, 175–184.

90%. Spiral computed axial tomography (CT), which is up to 98% sensitive, most commonly demonstrates a mass in the head of the pancreas. Pancreatic adenocarcinoma appears as a hypodense mass compared to the surrounding pancreatic parenchyma with bolus injection of intravenous contrast material. Furthermore, the relationship to the mesenteric arterial blood supply and superior mesenteric vein and portal vein confluence is well visualized on CT and hepatic metastasis can also be demonstrated. A mass in the head of the pancreas or a dilated pancreatic and biliary duct without a mass in a jaundiced patient suggests the diagnosis of pancreatic cancer. Endoscopic retrograde cholangiopancreatography (ERCP) is the best test to visualize the pancreatic duct and biliary system and it is warranted in patients with duct dilation without a mass on CT or in whom the diagnosis is otherwise unclear. Sampling the pancreatic duct by brush or biopsy is reported to diagnose pancreatic cancer with a sensitivity as high as 92% and a specificity as high as 96%. The diagnosis may be more difficult in patients with chronic pancreatitis who have ductal abnormalities due to chronic inflammation. In addition to diagnosing pancreatic cancer, biliary ductal obstruction can be relieved with a stent placed during ERCP for patients who are unresectable. In patients who are candidates for surgery, the placement of a biliary stent has been demonstrated to increase postoperative infectious complications and placement is not warranted. Other modalities such as magnetic resonance imaging (MRI), endoscopic ultrasound combined with fine-needle aspiration, and positron emission tomography play a limited role in the diagnosis of pancreatic cancer, but they are promising modalities that may be of future benefit. MRI is useful in patients with an allergy to CT contrast or those with renal dysfunction.

## SURGICAL THERAPY

Surgical resection of the head of the pancreas (pancreaticoduodenectomy) was first successfully performed by the German surgeon Walter Kausch in 1912. Allen O. Whipple reported the first successful cases in the United States in 1935 and he popularized the operation. During the last 30 years of the 20th century, the mortality of pancreaticoduodenectomy for pancreatic cancer decreased from 30–35% to less than 5%, making this procedure the standard of care for patients with a resectable cancer in the head of the gland. A standard pancreaticoduodenectomy includes the removal of the gallbladder, common bile duct, duodenum, head of the pancreas, and the antrum

of the stomach. Reconnecting the stomach, pancreas (mucosal-to-mucosal anastomosis), and bile duct (end-to-side anastomosis) to the jejunum to reestablish intestinal continuity completes the operation. The most common alternative to the standard operation is the pylorus-preserving pancreaticoduodenectomy, which was repopularized by Traverso and Longmire in the late 1970s. There appears to be little difference to removing or preserving the pylorus in terms of blood loss, operative time, morbidity, mortality, recurrence rates, or long-term survival. The morbidity rate following pancreaticoduodenectomy is fairly high (35–40%). The common complications of the pancreaticoduodenectomy include pancreatic fistula, hemorrhage, intra-abdominal abscess, and delayed gastric emptying.

Adenocarcinoma of the body and tail of the pancreas is less likely to be amenable to surgical resection because of a later presentation, resulting in larger tumors and an increased likelihood of metastatic disease. A distal pancreatectomy and splenectomy should be performed in patients who are deemed resectable by CT scan. Survival rates for patients with adenocarcinoma of the body and tail are the same stage-for-stage with patients with cancer of the head of the gland.

In patients who are explored for a possible resection and are found to be unresectable, a choledocho- or cholecystojejunostomy should be performed if the patient has preoperative jaundice and does not have a biliary stent. Additionally, a gastrojejunostomy should be performed at the time of exploration if the patient has preoperative signs and symptoms of a gastric outlet obstruction.

## RADIATION AND CHEMOTHERAPY

Despite some success with complete pancreatic resection, local recurrence rates of 50–86% and liver metastasis of 60–90% have been reported following surgery. This has led to the use of pre- and postoperative radiation and chemotherapy in an attempt to reduce the local failure and regional metastasis that are so common in this disease. Additionally, radiation and chemotherapy are the only treatments for the 80% of patients who present with locally advanced disease or metastasis.

### Resectable Pancreatic Cancer

Adjuvant chemoradiation for pancreatic cancer was originally reported in 1985 by the Gastrointestinal Tumor Study Group (GITSG)-9173, which randomized patients to external beam radiation therapy

(40 Gy) and 5-fluorouracil (5-FU) (500 mg/m<sup>2</sup>) given in bolus fashion during the radiation and once a week for 2 years or no therapy. Median survival was improved from 10.9 to 21 months with the therapy and 2-year overall survival was increased from 18 to 43% in favor of the adjuvant therapy. This study included only 18 patients, who were accrued over 8 years. Similar studies from Johns Hopkins University and the Mayo Clinic also reported increased median and 2-year survival with adjuvant chemoradiation. In all of these studies, the patients with recurrent disease eventually failed in the same pattern as those that did not receive any therapy: local disease, liver metastasis, and carcinomatosis with malignant ascites.

Improvements in radiation and chemotherapy administration have occurred, resulting in the treatment of many patients with external beam radiation and continuous 5-FU-based chemotherapy after resection. Although routinely used for the past 10–15 years, studies such as the EORTC Gastrointestinal Tract Cancer Cooperative Group and the European Study Group for Pancreatic Cancer Trial 1 (ESPAC-1) have reported no benefit with adjuvant chemoradiation in patients with pancreatic cancer. Further randomized, prospective studies are needed to clarify this issue and evaluate other drugs, such as gemcitabine, in the adjuvant setting.

Because adjuvant chemoradiation is delayed in up to 25% of patients following pancreatic resection, neoadjuvant chemoradiation has been examined in several centers. This form of therapy ensures that all patients receive multimodality therapy and patients who have disseminated disease on restaging studies are not subjected to laparotomy. A typical protocol consists of rapid-fractionation external beam radiation (30 Gy in 2 weeks) combined with systemic 5-FU or gemcitabine. The patient undergoes restaging by spiral CT and a resection if this appears to be technically feasible. Overall long-term survival appears to be approximately the same as with traditional adjuvant therapy.

### Unresectable Pancreatic Cancer

Patients with unresectable pancreatic cancer have undergone trials similar to those for patients undergoing resection. The first trial sponsored by the GITSG was reported in the early 1970s. Patients with unresectable locally advanced pancreatic cancer (no distant metastasis) received split-dose radiation (60 Gy) and bolus 5-FU. Median survival increased from 22 to 39 weeks and this study is the original basis for offering chemoradiation in patients with unresectable disease.

Several other GITSG trials have examined other agents such as doxorubicin, streptozocin, or mitomycin C in combinations with 5-FU. None of these have been reported to be superior to 5-FU alone and most result in a significant increase in hematologic toxicity. Gemcitabine, a deoxycytidine analogue capable of inhibiting DNA replication and repair, has been shown to provide a survival advantage over 5-FU in patients with locally advanced disease and has become the drug of choice in these patients.

### PALLIATIVE THERAPY

Tumor-associated pain due to invasion of the celiac and mesenteric nerve plexuses is almost universal in patients with pancreatic adenocarcinoma. Narcotic analgesia is effective in only approximately 50% of patients. Neurolytic celiac plexus block with 50–100% alcohol can be performed percutaneously or intraoperatively. This results in immediate pain relief in 90% of patients and 23–70% of patients experience long-term pain relief. Video-assisted thoracoscopic bilateral splanchnectomy relieves pain in over 95% of patients, with 50% experiencing long-term relief. This technique should be considered in patients who fail narcotic analgesia and chemical nerve blocks.

There are several options for the palliation of obstructive jaundice. Endoscopic retrograde cholangiopancreatography and placement of a biliary stent is the first line of therapy and has a success rate of over 95%. If this fails, a percutaneous transhepatic internal–external biliary catheter can be placed radiographically and this catheter can subsequently be internalized. Biliary stents are complicated by obstruction, recurrent jaundice, and cholangitis, but they can be exchanged, thus avoiding an operation. If the endoscopic and percutaneous techniques fail, a surgical cholecysto- or choledochojunostomy should be performed. Surgery is successful 98% of the time; however, the morbidity in the unresectable pancreatic cancer patient approaches 30% and the mortality can be as high as 10%.

Gastric outlet obstruction (GOO) occurs in 3 to 19% of patients with unresectable or metastatic pancreatic cancer. In patients undergoing a staging laparotomy who are deemed unresectable or have evidence of metastasis, a gastrojejunostomy should be performed if the patient had preoperative or intraoperative evidence of a GOO. The decision to bypass a patient with unresectable or metastatic disease who presents with a GOO is more difficult. The development of GOO has been demonstrated to occur as a

near terminal event in most series. Moreover, the mortality of a gastrojejunostomy in these patients ranges from 19 to 80%. Endoscopic stent placement and radiographic stent placement in patients with GOO due to pancreatic cancer have been successful and these techniques should be attempted prior to surgical bypass.

### See Also the Following Articles

GI Hormones and Endocrine Pancreas: Expressional Regulation • GI Hormones and Endocrine Pancreas: Growth • Pancreatic Islet Cell Tumors • Pancreatic Polypeptide (PP) • Prostate Cancer • Thyroid Carcinoma

### Further Reading

- Bell, R. H. (1996). Neoplasms of the exocrine pancreas. In "Digestive Tract Surgery: A Text and Atlas" (R. H. Bell, Jr., L. F. Rikkers, and M. W. Mulholland, eds.), pp. 849–878. Lippincott-Raven, Philadelphia, PA.
- Brand, R. (2001). The diagnosis of pancreatic cancer. *Cancer J.* 7, 287–297.
- Spanknebel, K., and Conlon, K. C. P. (2001). Advances in the surgical management of pancreatic cancer. Pancreaticoduodenectomy: Resection of the head of the pancreas, duodenum, common bile duct, gallbladder, and antrum of the stomach. *Cancer J.* 7, 312–323.
- Trede, M., Richter, A., and Wendl, K. (2001). Personal observations, opinions, and approaches to cancer of the pancreas and periampullary area. *Surg. Clin. N. Am.* 81, 595–609.



## Pancreatic Endocrine Tumors (PET)

see Gastrointestinal Neuroendocrine Tumor Syndromes

Table 1 Main Endocrine Syndromes Caused by Pancreatic Islet Cell Tumors

Type	Incidence	Male:female ratio	Clinical features	Histology	Comment
Insulinoma	1–4 per million	40:60 for benign; 50:50 for malignant	Episodic fasting and/or reactive hypoglycemia with inappropriately raised plasma insulin, C-peptide, and proinsulin levels	90% are solitary benign tumors composed mainly of insulin-containing B cells; 10% are metastatic	1–5% of insulinomas are MEN1; insulinoma is the third commonest tumor in families with MEN1
Zollinger-Ellison syndrome	0.5–1 per million	60:40	Intractable abdominal pain, diarrhea, gastrointestinal bleeding, epigastric tenderness; high gastric acidity with inappropriate and nonsuppressible gastrinemia	60% of tumors are metastatic, 20% are benign, and 20% are diffuse; tumor is usually <2 cm in diameter	33% of gastrinomas occur in families with MEN1
Verner Morrison syndrome	0.2–0.4 per million	25:75	Profuse watery diarrhea, hypokalemia, low gastric acidity, flushing of skin, hyperglycemia, and greatly elevated plasma VIP levels	Tumor is usually >8 cm in diameter and composed predominantly of VIP-containing cells	Not associated with MEN1
Glucagonoma	0.1–0.2 per million	20:80	Characteristic skin rash, mild diabetes, anemia, depression and often an enlarged liver; plasma glucagon grossly elevated	Over 60% have metastasized by the time of diagnosis; cells predominantly glucagon-containing	Not associated with MEN1
Somatostatinoma	<0.1 per million	1:2	Abdominal pain, steatorrhea, cholelithiasis, hyperglycemia (or hypoglycemia); high plasma somatostatin	Generally large (>8 cm) metastasizing tumors containing many different cell types	Often associated with other “endocrine” syndromes due to simultaneous hormone secretion

Note. MEN1, multiple endocrine neoplasia type 1; VIP, vasoactive intestinal polypeptide.

## TREATMENT

Surgical ablation is the treatment of choice, especially—though not necessarily only—when the tumor is confined to the pancreas. When surgery is impracticable or unsafe, palliative treatment with specific inhibitors of hormone secretion, e.g., diazoxide and octreotide, or of hormone action, e.g., omeprazole, is indicated. Radiotherapy and nonspecific anticancer chemotherapy are rarely of value but metastases—because of their usually slow growth and small number in the liver—may be destroyed with benefit by cryotherapy or arterial embolization.

## PROGNOSIS

Prognosis following removal of a benign islet cell tumor is excellent with no reduction in life expectancy. For a patient with metastases, prognosis depends largely on the ability to control the endocrine syndrome by pharmacological means. It

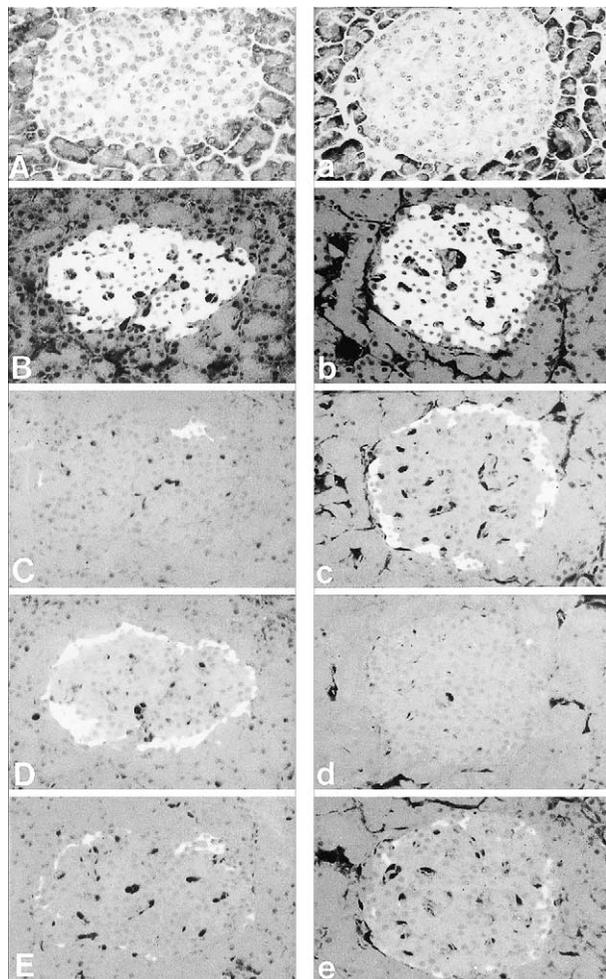
must, however, always be guarded, as tumor growth is inexorable.

## See Also the Following Articles

GI Hormones and Endocrine Pancreas: Expressional Regulation • GI Hormones and Endocrine Pancreas: Growth • GI Hormones in Cancer • Pancreatic Cancer • Pancreatic Polypeptide (PP)

## Further Reading

- Kloppel, G., and Heitz, P. U. (1988). Pancreatic endocrine tumors. *Patbol. Res. Pract.* **183**, 155–168.
- Langley, K. (1994). The neuroendocrine concept today. *Ann. N. Y. Acad. Sci.* **733**, 1–17.
- O’Shea, D. O., and Bloom, S. R. (1996). Gastrointestinal tumours. *Balliere Clin. Gastroenterol.* **10**, 555–766.
- Reed, N. S. (1999). Management of neuroendocrine tumours. *Clin. Oncol.* **11**, 295–302.
- Sheaves, R., Jenkins, P. J., and Wass, J. A. H. (1997). “Clinical Endocrine Oncology.” Blackwell Science, Oxford, UK.
- Thakker, R. V. (1998). Multiple endocrine neoplasia—Syndromes of the twentieth century. *J. Clin. Endocrinol. Metab.* **83**, 2617–2620.



**Figure 1** Histologic anatomy of the islet. Serial sections of a representative islet found in the head or ventral (A–E) and tail or dorsal (a–e) portions of the pancreas. (A, a) Cells stained with hematoxylin and eosin. (B, b) B cells immunohistochemically stained with anti-insulin antiserum. (C, c) A cells stained with anti-glucagon antiserum. (D, d) Pancreatic polypeptide cells stained with anti-pancreatic polypeptide antiserum. (E, e) D cells stained with anti-somatostatin antiserum. Reprinted from Orci, L. (1982). Macro- and micro-domains in the endocrine pancreas. *Diabetes* 31, 538, with permission.

PP is a member of the “PP family” of regulatory peptide hormones. Other structurally related peptides include peptide YY (PYY) and neuropeptide Y (NPY). As their names suggest, the peptide members share considerable amino acid sequence homology, but are found in widely disparate locations, including the pancreas (PP), the distal gut (PYY), and the central nervous system (NPY). The circulating levels of PP have been found to vary with a number of normal and pathologic conditions in human (Table I). The

**Table I** Alterations in PP Levels

Elevated levels in:	Normal aging
	Obesity (accompanied by glucose intolerance)
	Early type 1 diabetes
	Neuroendocrine (islet cell) tumors
Deficient levels in:	Chronic pancreatitis
	(Proximal or total) pancreatic resection
	Diabetic autonomic neuropathy
	Vagal denervation

presence of G protein-coupled receptors specific for PP and its related peptides has been confirmed and the population of hepatic PP receptors has been shown to vary inversely with the circulating levels of PP.

## PHARMACOKINETICS AND METABOLISM

The half-life of exogenously administered PP in the peripheral venous system is approximately 5 to 7 min. The metabolic clearance rate and volume of distribution do not change significantly when blood concentrations of the peptide are increased. Sustained elevations of PP after a meal cannot be accounted for by a low metabolic clearance rate. Therefore, the peptide must be continuously released in the postprandial period to maintain the high level.

PP levels are elevated in patients with chronic renal failure. A close correlation exists between the glomerular filtration rate and serum PP levels, which suggests that the kidney may play a dominant role in the clearance of PP from the circulation.

## REGULATION OF SECRETION

PP is released into the blood following a meal. The response to food is generally biphasic, with plasma concentrations rising four- to sixfold above basal levels within 5 min after food ingestion. The initial release lasts 30 to 60 min and is followed by a secondary response lasting more than 5 h. Protein is the most potent enteral stimulator of PP release, closely followed by fat, whereas glucose has a lesser effect.

Vagal mechanisms are not only the most powerful stimulators of PP secretion, but also are required for nearly all other stimuli of the PP cell. Central vagal activity can be increased either by sham-feeding, a rather specific but mild stimulation, or by hypoglycemia, a less specific but stronger stimulus. PP

secretion is stimulated by “chew-and-spit” sham-feeding. The response is eliminated by vagotomy and can be abolished by atropine, even at very small doses.

Hypoglycemia, whether induced by insulin or not, strongly stimulates PP secretion. Hypoglycemia also elicits parasympathetic and sympathetic adrenal activation, as well as a series of secondary endocrine responses. The PP response, however, is mediated through efferent vagal activity, as indicated by its abolition after vagotomy and inhibition by atropine. The adrenal medulla does not appear to play any major role in PP secretion, as the response is unchanged after adrenalectomy.

The magnitude of the PP response is proportional to the degree of hypoglycemia. PP secretion is surprisingly sensitive to a small decrease in blood glucose. It is even stimulated by rebound hypoglycemia occurring late after an oral glucose tolerance test, despite a nadir in blood glucose of only 55 mg/dl.

A surprisingly long-lasting response is seen when PP secretion is stimulated by activation of endogenous vagal activity. After sham-feeding, PP concentrations remain elevated for hours, even though the stimulus may last for only 15 min. It is quite likely that the prolonged elevation of PP concentrations in plasma in these circumstances is due to continued secretion. Cessation of electrical vagal stimulation is followed by a rapid decline of plasma PP concentrations with a half-life of 5.5 min, similar to the half-life of exogenously administered PP. Thus, hypoglycemia or sham-feeding causes prolonged activation of the “vagal center” responsible for PP secretion. This phenomenon may be an important aspect of the prolonged PP secretion detected after a meal.

The isolated perfused human pancreas model has been used to investigate the neural regulatory mechanism of the endocrine pancreas. In this model, human pancreata are procured from heart-beating organ donors. Following organ procurement, the pancreas is perfused via the splenic artery and the effluent is collected from the splenic vein. Electrical stimulation of the neural plexus along the splenic artery can be performed in the presence or absence of selective neural blockers and hormonal responses in the venous effluent can be examined. In this model, cholinergic stimulation with combined perfusion of phentolamine and propranolol ( $\alpha$ - and  $\beta$ -adrenergic antagonists) during splanchnic nerve stimulation causes a marked stimulation of PP release, consistent with the vagal dependency of *in vivo* PP release. Activation of  $\alpha$ -adrenergic fibers causes strong suppression of PP secretion and activation of  $\beta$ -adrenergic fibers causes mild stimulation.

PP release is also mediated by enteric hormonal agents such as cholecystokinin (CCK), secretin, bombesin, and gastric inhibitory polypeptide. Intravenous infusion of cerulein, an analogue of CCK, stimulates an increase in plasma PP levels in humans and infusions of 20% pure CCK in humans achieve similar results. The most potent stimulus for PP release on a molar basis in dogs is pure CCK, followed by octapeptide of CCK and synthetic human gastrin. Phenylalanine and tryptophan, instilled into the duodenum, cause a significant increase in PP levels. Duodenal acidification also releases PP, but the response is significantly less than that obtained by infusion of the two amino acids. The mechanism of PP release by phenylalanine and tryptophan appears likely to involve the intermediate release of CCK, which in turn stimulates the release of PP. Oleate and liver extract, both strong releasers of CCK, also stimulate the release of significant amounts of PP when infused into the intestine of dogs. These findings suggest that endogenously released CCK plays an important role in the prolonged or so-called intestinal phase of PP release (Fig. 2).

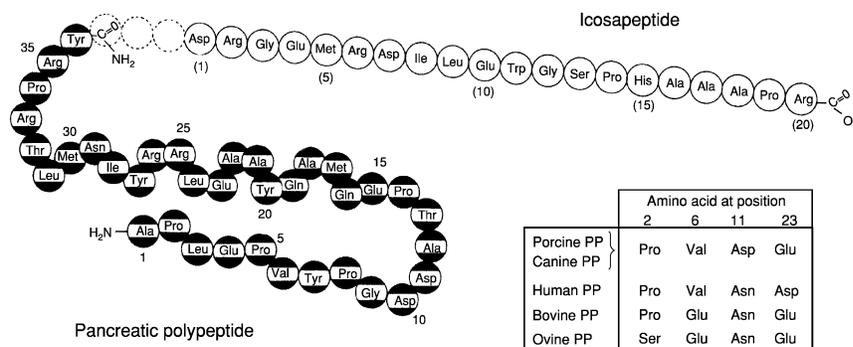
### Gastro-entero-hepatic Actions of PP

Basal pancreatic secretion of fluid and protein is inhibited by physiologic doses of PP in dogs and in humans. At physiologic doses, PP inhibits secretin- and cerulein-stimulated pancreatic protein and bicarbonate secretion in dogs. Secretin- and CCK-stimulated pancreatic secretion of trypsin was markedly inhibited and bicarbonate output was less markedly, but significantly, inhibited by PP. In humans, PP inhibited secretin-stimulated pancreatic secretion in volume, enzyme activity, and total concentration. However, PP appeared to have no effect on pancreatic fluid bicarbonate concentration in humans.

PP infusion does not affect gastric acid or pepsin secretion or the plasma levels of insulin, glucagon, secretin, gastrin, glucose, or lipids. Additionally, there are no signs or symptoms of cardiovascular or gastrointestinal alterations in healthy persons. In dogs, pharmacologic doses of PP had no effect on steady-state bile flow but did cause choledochal resistance and relaxation of the gallbladder. Basal and stimulated output of bilirubin is inhibited significantly by PP infusion at physiologic levels.

### Satiety, Obesity, and Hyperphagia

A number of studies suggest that PP may play a role in feeding and obesity and, therefore, in glucose



**Figure 2** Pancreatic polypeptide is a 36-amino-acid peptide that is secreted by the F cells of the islet. There is significant homology across species.

metabolism. R. L. Gingerich and co-workers observed that the amount of PP in the pancreas is elevated in hyperglycemic, obese mice. Subsequently, R. J. Gates and N. R. Lazarus reported reversal of hyperphagia, obesity, and glucose intolerance after administration of PP in obese, hyperphagic mice and concluded that the abnormalities seen were due to a deficiency of PP release.

B. Glaser and colleagues studied the relationship of PP levels with both obesity and glucose intolerance in humans. Obese subjects with normal oral glucose intolerance demonstrated low basal and low stimulated levels of PP compared to age-matched controls. Obese subjects with abnormal oral glucose tolerance demonstrated elevated basal levels of PP, however, and an exaggerated PP response to stimulation. W. B. Zipf and colleagues observed that obese children with Prader-Willi syndrome were PP-deficient and that their hyperphagia was reduced by PP administration. PP administration to normal dogs has also been shown to induce weight loss. A 14-day continuous subcutaneous infusion of PP resulted in a 3.5% loss of body weight in 20 kg dogs allowed normal access to food. The weight loss was reversed within 4 weeks after cessation of the PP infusions. A satiety effect of NPY has also been observed by several investigators and laboratory and clinical studies continue to explore the possible role of PP and NPY as satiety-inducing agents. The role of PP is also being investigated in patients undergoing successful bariatric surgery.

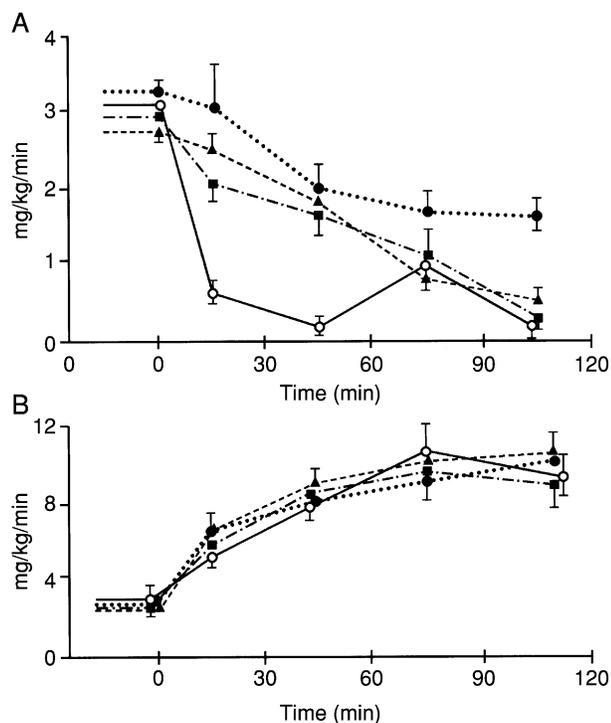
## REGULATION OF INSULIN ACTION

The liver occupies a central role in the regulation of glucose metabolism. The ability of the liver to produce glucose during times of fasting or stress, and to

take up glucose after feeding, is essential for metabolic homeostasis and PP appears to play an important role in the hepatic response to insulin. A study by Y. S. Sun and colleagues examined pancreatogenic diabetes in a canine model of chronic pancreatitis created by pancreatic duct ligation. The hyperinsulinemic–euglycemic glucose clamp technique and radioisotopic methods were used to quantify hepatic and nonhepatic insulin-stimulated glucose turnover. Hepatic glucose production in normal animals was suppressed by insulin, but dogs with chronic pancreatitis demonstrated far less insulin-induced suppression of hepatic glucose production. This insulin resistance could be demonstrated only in the liver and not in nonhepatic tissue. Furthermore, it occurred only in animals with impaired meal-stimulated PP secretory activity. Thus, PP deficiency was uniformly associated with profound hepatic insulin resistance.

In subsequent experiments, the same animals were restudied at varying time points during continuous subcutaneous infusion of bovine PP. The hepatic insulin resistance previously observed in PP-deficient animals was corrected after PP administration (Fig. 3). Serial glucose tolerance tests, performed over 28 weeks after the PP infusion period, demonstrated that glucose tolerance was significantly improved posttreatment, as compared to pretreatment values.

N. E. Seymour and colleagues studied hepatocyte insulin-binding characteristics to determine whether insulin receptor function or expression is altered by PP. Maximal insulin-binding capacity was significantly lower in tissue from animals with chronic pancreatitis than from control, sham-operated animals (Fig. 4). This loss of high-affinity binding sites was observed in the liver but not in skeletal muscle. PP administration in animals with chronic pancreatitis resulted in significantly greater hepatic insulin-binding capacity

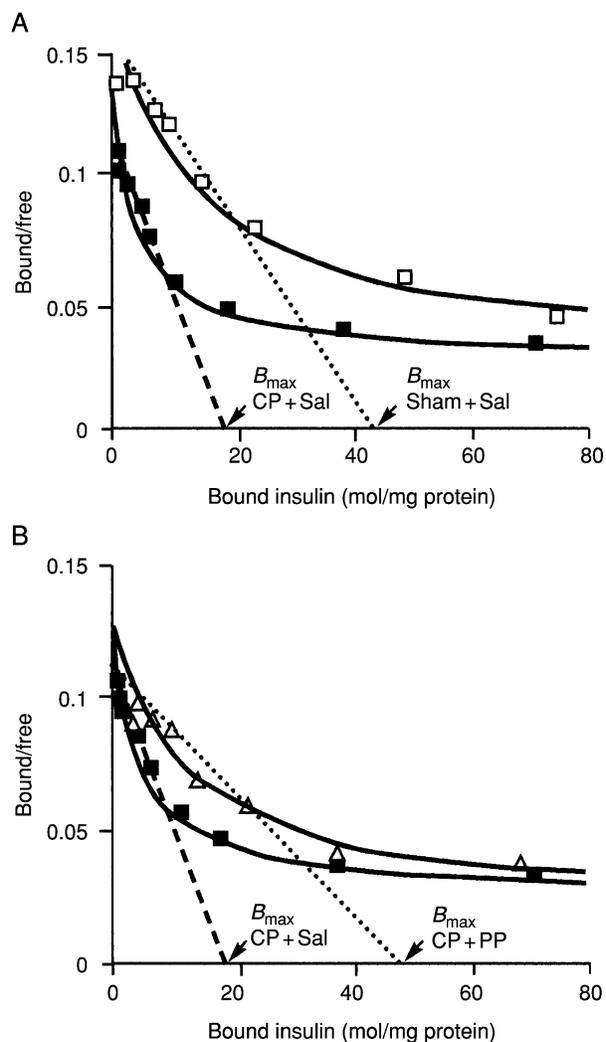


**Figure 3** (A) Hepatic glucose production rates. The responses of dogs with chronic pancreatitis at days 6 and 11 of continuous subcutaneous PP infusion were significantly lower than the pretreatment response by paired analysis ( $P < 0.05$ ). (B) The overall glucose disposal rates achieved by insulin infusion in these studies. Reprinted from Sun *et al.* (1986), with permission.

than was observed after vehicle administration alone (Fig. 4) and approximated that of the control (non-pancreatic) group. During glucose tolerance tests performed immediately prior to the procurement of tissues, the magnitude and duration of the serum glucose elevation following an intraduodenal dextrose challenge were reduced by PP administration in animals with pancreatitis.

Serum immunoreactive insulin levels were not altered by PP administration, but the restoration of high-affinity insulin-binding sites produced by exogenously administered PP suggested alterations in the expression of hepatic insulin receptors. Subsequent studies by this group showed that diminished hepatic insulin receptor gene expression seen in PP-deficient animals is reversed by exogenous PP administration.

The glucoregulatory effect of PP therefore appears to reside in its role in the expression of the hepatic insulin receptor protein. Diminished PP levels are associated with a loss of high-affinity insulin receptor sites on hepatocytes, which creates a state of hepatic insulin resistance. Although other factors may play a



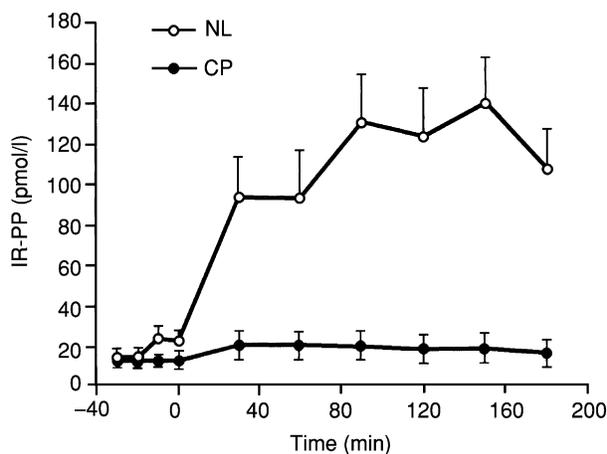
**Figure 4** (A) Effect of chronic pancreatitis on hepatic insulin binding: Scatchard plot of insulin binding in solubilized hepatic membranes from saline-administered, sham-operated rats (sham, open squares) and saline-administered, chronic pancreatic rats (CP, filled squares). (B) Effects of PP administration (PP, 200  $\mu\text{g}/\text{kg}$  per day) on hepatic insulin binding in chronic pancreatitis: Scatchard plot of insulin binding in solubilized hepatic membranes from saline-administered chronic pancreatic rats (CP+Sal, filled squares) and PP-treated rats (CP+PP, open triangles). Reprinted from Seymour *et al.* (1995). Alteration in hepatocyte insulin binding in chronic pancreatitis: Effect of pancreatic polypeptide. *Am. J. Surg.* **169**, 105, with permission.

role in the hepatic insulin resistance that characterizes pancreatogenic diabetes, a loss of insulin-binding sites due to the reduced expression of hepatic insulin receptor protein appears to be the primary mechanism of PP's glucoregulatory role. Replacement of PP by exogenous administration reverses the hepatic insulin resistance and at least partially corrects the abnormal hepatic glucose production.

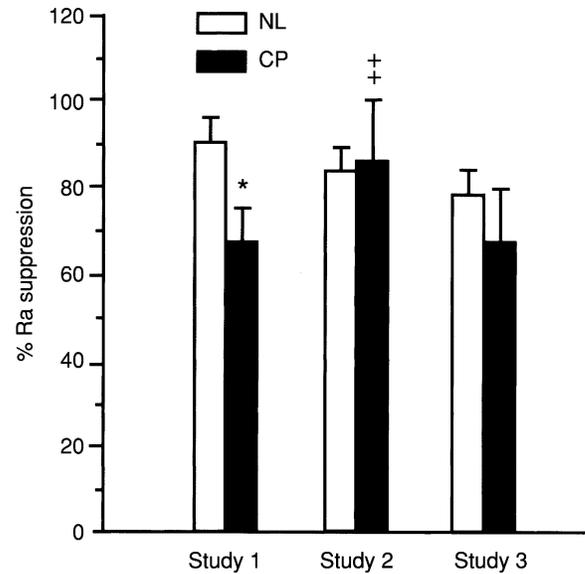
## The Role of PP in Pancreatogenic Diabetes

Patients with severe chronic pancreatitis also have a profound meal-stimulated PP secretory deficiency compared to normal subjects (Fig. 5). During an initial, baseline hyperinsulinemic–euglycemic clamp study, a pattern of profound hepatic insulin resistance similar to that seen in the canine studies was found to occur in these PP-deficient patients (Fig. 6). Glucose clamp studies were repeated 1 to 2 months later during the final 2 hours of an 8 h intravenous infusion of bovine PP, which replicated physiologic serum concentrations of postcibal immunoreactive PP (750–1000 pg/ml). Hepatic responses to insulin in PP-deficient patients were restored to normal by PP (Fig. 6). However, insulin-stimulated glucose disposal in normal control subjects was not altered by the PP infusion. There was no PP-induced alteration in serum insulin or glucagon concentrations in either of the experimental groups. These data indicate that PP replacement was successful in reestablishing the normal hepatic response to insulin. A third set of glucose clamp studies was performed 1 to 2 months after the PP administration. Hepatic glucose output in response to insulin in PP-deficient patients had returned to near-baseline values, which confirmed the etiologic role of PP deficiency in the hepatic resistance to insulin (Fig. 6).

N. E. Seymour and colleagues examined glucose turnover in patients with a remote history of pancreatectomy. In these subjects, proximal pancreatectomy

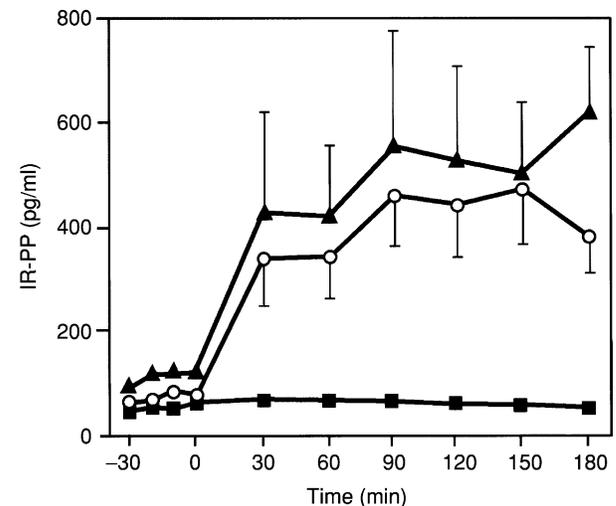


**Figure 5** PP response to test meal. Immunoreactive PP (IR-PP) responses in normal control subjects (NL,  $n = 6$ ) and patients with chronic pancreatitis (CP,  $n = 5$ ) accompanied by PP deficiency. Test meal was administered at 0 min. Reprinted from Brunicaudi *et al.* (1996), with permission.



**Figure 6** Percentage suppression of RA (rate of glucose appearance) during hyperinsulinemic–euglycemic clamps for studies 1, 2, and 3 for normal subjects (NL,  $n = 6$ ) and CP patients (CP,  $n = 5$ ). Reprinted from Brunicaudi *et al.* (1996), with permission.

and distal pancreatectomy were distinguished from one another, since only the former was associated with a loss of PP secretory activity and low circulating PP levels (Fig. 7). Three hyperinsulinemic–euglycemic clamp studies were performed in each subject, separated by at least 1 month. An 8 h intravenous PP infusion



**Figure 7** Immunoreactive PP (IR-PP) response to a test meal for eight control subjects (open circles), four non-PP-deficient resection patients (filled triangles), and six PP-deficient resection patients (filled squares). Reprinted from Seymour *et al.* (1998), with permission.

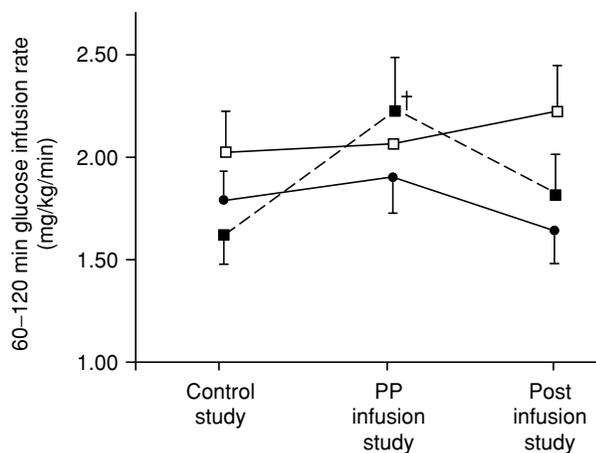
was administered before the second study. PP-deficient, proximal pancreatectomy subjects demonstrated an impairment in hepatic insulin responses in the first study and required significantly smaller glucose infusions than normal control subjects to maintain euglycemia during insulin infusion (Fig. 8). This insulin resistance was completely reversed by PP administration during the second clamp study, but was evident again during the final infusion study in the absence of PP. No alteration in peripheral glucose disposal was observed with either pancreatectomy or PP administration. These studies provide strong evidence for a selective hepatic metabolic effect of PP.

## Clinical Implications of Alterations in PP Secretion

### Diabetes Mellitus

Mean basal concentrations of plasma PP are elevated in normal aging and in type 1 and type 2 diabetes mellitus. PP levels are particularly elevated in type 2 (adult onset, non-insulin-dependent) diabetic patients, which may represent a compensatory response to a loss of insulin synthesis or action. Early type 1 diabetic patients may also exhibit high levels of PP before they experience a complete loss of islet tissue due to the progression of disease.

Destruction of the pancreas by pancreatitis or surgical removal of the pancreas is associated with a form of glucose intolerance termed "pancreatogenic diabetes"



**Figure 8** Glucose infusion rate (mg/kg/min) during the final hours of a 2 h insulin infusion in normal control (open squares), non-PP-deficient resection (filled circles), and PP-deficient resection (filled squares) subjects. The glucose infusion requirement for PP-deficient subjects was significantly greater during the PP infusion study than during the control study ( $P < 0.03$ ). Reprinted from Seymour *et al.* (1998), with permission.

(or type 3 diabetes), which differs from both type 1 and type 2 diabetes in several respects (Table II). Diabetes in these conditions rarely, if ever, results in ketoacidosis, but can be associated with labile blood glucose levels and frequent episodes of iatrogenic hypoglycemia. This specific pattern of abnormal glucose homeostasis is likely caused by the simultaneous deficiencies in insulin, glucagon, and PP secretion.

### Chronic Pancreatitis

Chronic pancreatitis is a complex pathologic state consisting of lifelong exocrine and endocrine derangement. Endocrine dysfunction manifests itself as diabetes mellitus, but multiple hormone deficiencies have been demonstrated. Basal and meal-stimulated levels of circulating immunoreactive PP are especially affected by chronic pancreatitis, so much so that measurement of PP levels has been suggested as a screening procedure for the diagnosis of chronic pancreatitis. In PP-deficient states such as chronic pancreatitis, there is demonstrable hepatic resistance to insulin, which can be reversed by the administration of PP (Fig. 9).

### Pancreatic Surgery

A PP-deficient state frequently arises after surgical resection of the proximal pancreas because the hormone is synthesized predominantly in the head of the gland. K. Inoue and colleagues confirmed the abolishment of PP responses to ingested nutrients in patients after pancreaticoduodenectomy. In the study by N. E. Seymour and colleagues, healthy, young (age less than 30 years) male patients who had undergone pancreatic resection and who had a deficient PP response as a consequence of their pancreatic resection had a measurable impairment of hepatic insulin sensitivity. This effect was not associated with impaired glucose tolerance, however, when insulin and glucagon secretory function was present.

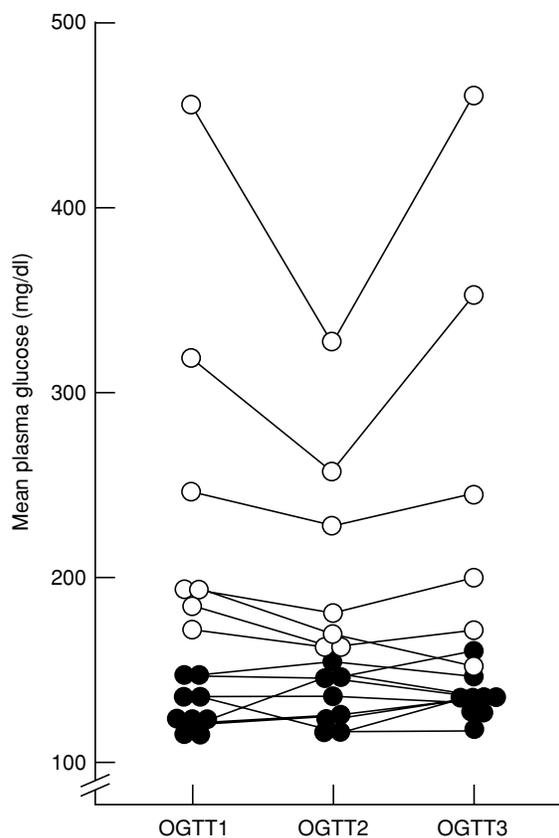
The risk of pancreatogenic diabetes is increased when proximal pancreatectomy is performed in older patients, however, or when near-total or total pancreatectomy is performed. The development of duodenum-preserving pancreatic resection techniques by H. G. Beger and colleagues, and by C. F. Frey, has been shown to be associated with improved glucose tolerance postoperatively. The preservation of glucose tolerance with operative techniques that spare part or all of the PP-rich region of the pancreas further supports the role of PP in glucose homeostasis.

### Artificial Endocrine Pancreas

The artificial endocrine pancreas, used clinically, is simple in its concept. A sensor electrode measures

Table II Types of Diabetes Mellitus

Parameter	Type 1 IDDM juvenile onset	Type 2 NIDDM adult onset	Type 3 pancreatogenic postoperative onset
Ketoacidosis	Common	Rare	Rare
Hyperglycemia	Severe	Usually mild	Mild
Hypoglycemia	Common	Rare	Common
Peripheral insulin sensitivity	Normal or increased	Decreased	Increased
Hepatic insulin sensitivity	Normal	Normal or decreased	Decreased
Insulin levels	Low	High	Low
Glucagon levels	Normal or high	Normal or high	Low
PP levels	High	High	Low
Typical age of onset	Childhood or adolescence	Adulthood	Any



**Figure 9** Mean plasma glucose concentration during 180 min after  $40 \text{ g/m}^2$  oral glucose tolerance test (OGTT) in chronic pancreatitis patients ( $n = 10$ ) and controls ( $n = 6$ ). Open circles indicate impaired or diabetic status during the initial OGTT. Filled circles indicate normal or nondiagnostic status during the initial OGTT. OGTT 2 was performed 18 h after an 8 h infusion of bovine PP ( $2 \text{ pmol/kg min}$ ). OGTT 3 was performed 1 month later. Eighteen hours after PP infusion, every patient with impaired diabetic OGTT showed a lower mean plasma glucose response. Reprinted from Andersen (1990). The role of pancreatic polypeptide in glucose metabolism. In "Gastrointestinal Endocrinology: Receptor and Post-receptor Mechanisms," (J. C. Thompson, ed.), pp. 333–357. Academic Press, San Diego, CA, with permission.

the level of blood glucose. This information is fed into a small computer that energizes an infusion pump, and from a small reservoir, controlled amounts of insulin enter the patient's circulation. K. Hanazaki and colleagues developed a new artificial pancreas with a multiple-chamber injection pump to facilitate infusion of not only insulin, but also PP. The addition of PP decreased insulin requirements in pancreatctomized dogs during a 72 h trial period. This observation suggests that PP, or a PP-like agonist, may become a useful adjunct in the treatment of diabetes.

## SUMMARY

PP release is both hormonally and neurally mediated and both exaggerated or diminished PP responses are valuable indicators of the integrity of the neuro-entero-pancreatic system. Although subtle effects on pancreatic exocrine function and choleresis are associated with PP, its major role appears to be that of a glucoregulatory hormone. PP may play a role in food ingestion as a satiety mediator, but its most consistent function appears to be that of a regulator of hepatic insulin action, through its mediation of hepatic insulin receptor protein synthesis. Therapeutic opportunities, such as replacement of PP in states where the hormone is deficient, are being explored clinically. In the future, PP analogues, as well as PP-receptor agonists or antagonists, may become useful clinical tools.

## See Also the Following Articles

GI Hormones and Endocrine Pancreas: Expressional Regulation • GI Hormones and Endocrine Pancreas: Growth • GI Hormones in Cancer • Hypoglycemia • Neuropeptide Y • Neuropeptide Y, Evolution of • Obesity and Diabetes, Regulation of Food Intake • Pancreatic Islet Cell Tumors • Peptide YY (PYY)

## Further Reading

- Bloom, S. R., Adrian, T. E., Greenburg, G. R., *et al.* (1978). Effects of PP infusion in man. *Gastroenterology* **74**, 1012.
- Brunicaudi, F. C., Chaikin, R. L., Ryan, A. S., Seymour, N. E., Hoffmann, J. A., Lebovitz, H. E., Chance, R. E., Gingerich, R. L., Andersen, D. K., and Elahi, D. (1996). Pancreatic polypeptide administration improves abnormal glucose metabolism in patients with chronic pancreatitis. *J. Clin. Endocrinol. Metab.* **81**, 3566–3572.
- Brunicaudi, F. C., Druck, P., Sun, Y. S., Elahi, D., Gingerich, R. L., and Andersen, D. K. (1988). Regulation of PP secretion in the isolated perfused human pancreas. *Am. J. Surg.* **155**, 63–69.
- Floyd, J. C., Fajans, S. S., and Pek, S. (1978). Physiologic regulation of plasma levels of PP in man. In “Gut Hormones” (S. R. Bloom, ed.), pp. 247–253. Churchill Livingstone, New York.
- Gates, R. J., and Lazarus, N. R. (1977). The ability of PPs (APP and BPP) to return to normal the hyperglycemia, hyperinsulinemia and weight gain of New Zealand obese mice. *Horm. Res.* **8**, 189–202.
- Gersell, D. J., Gingerich, R. L., and Greider, M. H. (1979). Regional distribution and concentration of PP in human and canine pancreas. *Diabetes* **28**, 11–15.
- Glaser, B., Zoghlin, G., Pienta, K., and Vinik, A. I. (1988). PP response to secretin in obesity: Effects of glucose intolerance. *Horm. Metab. Res.* **20**, 288–292.
- Hanazaki, K., Nose, Y., and Brunicaudi, F. C. (2001). Artificial endocrine pancreas. *J. Am. Coll. Surg.* **193**, 310–322.
- Hazelwood, R. L., Turner, S. D., Kimmel, J. R., *et al.* (1973). Spectrum effects of a new polypeptide (third hormone?) isolated from the chicken pancreas. *Gen. Comp. Endocrinol.* **21**, 485–497.
- Lin, T.-M., and Chance, R. E. (1974). Gastrointestinal actions of a new bovine pancreatic peptide (BPP). In “Endocrinology of the Gut” (W. Y. Chey and F. P. Brooks, eds.), pp. 143–145. Charles B. Slack, Thorofare, NJ.
- Orci, L., Baetens, D., Ravazzola, M., *et al.* (1976). PP and glucagon: Non-random distribution in pancreatic islets. *Life Sci.* **19**, 1811–1816.
- Seymour, N. E., Brunicaudi, F. C., Chaikin, R. L., Lebovitz, H. E., Chance, R. E., Gingerich, R. L., Elahi, D., and Andersen, D. K. (1988). Reversal of abnormal glucose production after pancreatic resection by PP administration in man. *Surgery* **104**, 119–129.
- Slezak, L. A., and Andersen, D. K. (2001). Pancreatic resection: Effects on glucose metabolism. *World J. Surg.* **25**, 452–460.
- Spector, S. A., Frattini, J. S., Zdankiewicz, P., Nathan, J., Wang, J. D., Andersen, D. K., and Seymour, N. E. (1997). Insulin receptor gene expression in chronic pancreatitis: Effect of PP. *Surg. Forum* **48**, 168.
- Sun, Y. S., Brunicaudi, F. C., Druck, P., Walfisch, S., Berlin, S. A., Chance, R. E., Gingerich, R. L., Elahi, D., and Andersen, D. K. (1986). Reversal of abnormal glucose metabolism in chronic pancreatitis by administration of PP. *Am. J. Surg.* **151**, 130–140.
- Valenzuela, J. E., Taylor, I. L., and Walsh, J. H. (1979). PP responses in patients with chronic pancreatitis. *Digest. Dis. Sci.* **24**, 862–864.

parathyroidectomy on the basis of local invasion ( $n = 5$ ), tumor pathology ( $n = 4$ ), or distant metastases ( $n = 2$ ). Hypercalcemia was less severe than usually observed in patients with parathyroid carcinoma and parathyroid hormone (PTH) levels were not unusually high for patients on maintenance hemodialysis. Only one patient died of hypercalcemia. The authors concluded that no preoperative features distinguished hemodialysis patients with parathyroid carcinoma from those with parathyroid hyperplasia and that the clinical course may be more benign because of the tendency for renal insufficiency to lower serum calcium levels.

### Familial Hyperparathyroidism

Parathyroid cancer has been reported in patients with familial hyperparathyroidism, particularly those with autosomal dominant isolated hyperparathyroidism that is not part of the multiple endocrine neoplasia type 1 (MEN1) syndrome. In one such family, there was no evidence of antecedent hyperplasia in unaffected glands and chromosomal abnormalities commonly observed in other solid tumors were identified (a reciprocal translocation between chromosomes 3 and 4, trisomy 7, and a pericentric inversion in chromosome 9). Analyses of tumor DNA from one family member with parathyroid carcinoma showed no evidence of *ras* gene mutations, PTH gene arrangement, or allelic loss from chromosome 11q13, the locus of the gene for MEN1. The risk of parathyroid carcinoma is also greatly increased in patients with the hyperparathyroidism–jaw tumor syndrome, which has been localized to chromosome 1q21–q31.

### MOLECULAR PATHOGENESIS

Oncogenes and tumor suppressor genes have been implicated in the development of parathyroid tumors. Cyclin D1 or PRAD1 (parathyroid adenoma 1) is an oncogene located on chromosome 11q13; its protein product is a cell cycle regulator. Rearrangement of the cyclin D1 gene with the regulatory region of the PTH gene has been reported in 5% of parathyroid adenomas and the cyclin D1 oncoprotein is overexpressed in 18–40% of parathyroid adenomas. In one study, overexpression of cyclin D1 protein was found in 90% of parathyroid cancers. However, it is uncertain whether overexpression of cyclin D1 protein is a causative or an associative phenomenon.

There is also strong evidence that inactivation of the tumor suppressor gene RB (retinoblastoma) on

chromosome 13 may contribute to the development of parathyroid carcinoma. Together with cyclin D1, RB is important in cell cycle control. In one study, 11 parathyroid cancers lacked a RB allele and most had a complete absence of nuclear staining for the RB protein. In contrast, only one parathyroid adenoma lacked the allele and none had abnormal staining for the RB protein. Several other investigators have also reported that allelic deletions on chromosome 13 are more common in parathyroid cancer than in benign primary hyperparathyroidism. These data strongly support the presence of a tumor suppressor gene on the long arm of chromosome 13, critical for the development of parathyroid carcinoma. However, as the deleted portion of chromosome 13 is large, it remains to be determined whether RB or a different gene on 13q will prove to be the primary causative tumor suppressor. Another important cell cycle regulator, the p53 tumor suppressor gene, does not appear to be a major contributor to the pathogenesis of parathyroid carcinoma.

Several new locations have been identified for potentially important oncogenes or tumor suppressor genes that seem to be preferentially or exclusively found in parathyroid carcinomas as compared with adenomas. Tumor-specific gains or losses of chromosomal material suggest that oncogenes on chromosomes 1q, 5q, 9q, 16p, 19p, and Xq and tumor suppressor genes on chromosomes 1p, 3q, 4q, 13q, and 21q may be involved in the pathogenesis of parathyroid carcinoma. Moreover, since a number of the regions commonly lost in adenomas (including 11q, the site of the MEN1 gene) were never or rarely lost in carcinomas, these results also support the hypothesis that parathyroid carcinomas tend to arise *de novo* rather than from preexisting adenomas. Clarification of the molecular pathogenesis of parathyroid carcinoma will aid in the diagnosis of difficult cases and may provide important clues or biological targets for future development of new and more effective therapies.

### CLINICAL FEATURES

#### General Considerations

The clinical features of parathyroid cancer are due primarily to the effects of excessive secretion of PTH rather than to infiltration of vital organs by the tumor. Thus, signs and symptoms of hypercalcemia often dominate the clinical picture in addition to symptoms of hyperparathyroid bone disease and renal involvement (renal insufficiency, nephrolithiasis, or nephrocalcinosis). It is critical that parathyroid cancer be

considered in the differential diagnosis of parathyroid hormone-dependent hypercalcemia. All too often the diagnosis of parathyroid carcinoma is made in retrospect when hypercalcemia recurs due to local spread or distant metastasis of the tumor. The best outcomes for patients with parathyroid cancer occur when there has been complete resection of the tumor at the time of the first surgery. Several features of the patient with primary hyperparathyroidism should suggest a malignant rather than a benign disease (Table I).

### Demographic Features

Benign primary hyperparathyroidism is considerably more common in women than in men (3–4:1). In contrast, there is no association of gender with parathyroid carcinoma and the ratio of affected women to men is 1:1 in most reported series. Patients with primary hyperparathyroidism typically present in their fifties or sixties, whereas the average patient with parathyroid carcinoma is in his or her forties, approximately 10 years younger. However, two reviews (the Mayo Clinic experience and the National Cancer Data Base) found that the average age of their patients was in the middle fifties. Thus, although it is reasonable to consider the possibility of parathyroid cancer when primary hyperparathyroidism is diagnosed in a man and or a younger individual, gender and age are of limited assistance in the setting of the individual patient.

### Symptoms and Biochemical Characteristics

Benign primary hyperparathyroidism usually presents with mild hypercalcemia (within 1 mg/dl above the

upper limit of normal) that is often asymptomatic and usually discovered incidentally. In contrast, the serum calcium level in parathyroid cancer is much higher, generally above 14 mg/dl or 3–4 mg/dl above the upper limit of normal. Typical signs and symptoms of hypercalcemia, including fatigue, weakness, weight loss, anorexia, nausea, vomiting, polyuria, and polydipsia, are almost always present. Bone pain, fractures, and renal colic are much more common when primary hyperparathyroidism is due to a parathyroid cancer than an adenoma. Extremely high levels of parathyroid hormone are unusual in benign primary hyperparathyroidism, where circulating concentrations are commonly less than twice the upper limit of normal. In contrast, parathyroid hormone levels in patients with parathyroid cancer are generally much higher, ranging from 3 to 10 times above the upper limit of normal. Serum alkaline phosphatase activity is often frankly elevated in patients with parathyroid carcinoma, whereas levels are usually in the normal or upper-normal range in patients with benign primary hyperparathyroidism. Finally, patients with parathyroid carcinoma may have elevated levels of  $\alpha$ - and  $\beta$ -subunits of human chorionic gonadotropin, whereas patients with primary hyperparathyroidism do not.

### Physical Findings

A palpable neck mass has been reported in 30–76% of patients with parathyroid carcinoma, yet is distinctly unusual in benign primary hyperparathyroidism. This important clinical finding constitutes another striking difference between benign and malignant parathyroid disease. In addition, recurrent laryngeal nerve palsy in a patient with primary hyperparathyroidism who has not had previous neck surgery suggests local spread of

Table I Parathyroid Carcinoma and Benign Primary Hyperparathyroidism: Typical Features

	Parathyroid carcinoma	Primary hyperparathyroidism
Female:Male ratio	1:1	3.5:1
Average age (years)	48	55
Asymptomatic	<5%	>80%
Serum calcium	>14 mg/dl	≤1 mg/dl above upper limit of normal
Parathyroid hormone	Markedly elevated	Mildly elevated
Palpable neck mass	Common	Rare
Renal involvement <sup>a</sup>	32–80%	4–18%
Skeletal involvement <sup>b</sup>	34–91%	<5%
Concomitant renal and skeletal disease	Common	Rare

<sup>a</sup>Includes nephrolithiasis, nephrocalcinosis, and impaired renal function in the absence of any other etiology.

<sup>b</sup>Includes osteitis fibrosa, subperiosteal resorption, “salt and pepper” skull, and diffuse osteopenia on plain radiographs.

the tumor and is also very suggestive of parathyroid cancer.

### Target Organ Involvement

The classical target organs of parathyroid hormone—namely, the kidney and the skeleton—are affected with greater frequency and severity in parathyroid carcinoma than is typically observed in benign primary hyperparathyroidism. The prevalence of renal involvement in benign primary hyperparathyroidism, including nephrolithiasis, nephrocalcinosis, and impaired glomerular filtration, is less than 20%. In contrast, renal colic is a frequent presenting complaint of parathyroid carcinoma. The prevalence of nephrolithiasis was 56% and the prevalence of renal insufficiency was 84% in the Mayo Clinic series, somewhat higher than in previous reports in which the prevalence of renal involvement has ranged from 32 to 60%.

Bone pain and pathologic fractures are also common features of parathyroid cancer. Patients with benign primary hyperparathyroidism rarely have skeletal complaints and specific radiologic signs are found in less than 5% of patients. In contrast, specific radiologic signs of hyperparathyroid skeletal disease are commonly seen in parathyroid carcinoma, with osteitis fibrosa cystica, subperiosteal bone resorption, “salt and pepper” skull, absent lamina dura, and diffuse osteopenia reported in 44 to 91% of cases. Another distinguishing feature is the development of concomitant bone and stone disease in parathyroid cancer, whereas simultaneous and symptomatic involvement of the kidneys and skeleton is distinctly unusual in primary hyperparathyroidism.

Recurrent severe pancreatitis, peptic ulcer disease, and anemia occur with greater frequency in patients with malignant disease than in those with benign primary hyperparathyroidism.

### Acute Primary Hyperparathyroidism

Parathyroid cancer is more difficult to distinguish from acute primary hyperparathyroidism, also known as parathyroid crisis, with which it shares many clinical features. As parathyroid crisis is also characterized by marked elevations of serum calcium and parathyroid hormone, the diagnosis of parathyroid cancer should always be considered. Whereas a preoperative distinction between these two entities is not possible, it is important to bear the potential for cancer in mind since the surgical approaches differ.

### Summary

Features that should lead a physician to suspect that a patient with hypercalcemia and elevated parathyroid hormone levels has parathyroid cancer are shown in [Table I](#). However, some patients with parathyroid cancer may present with signs and symptoms that are quite mild. Whether this is because the disease is diagnosed at an earlier stage is unclear. Conversely, some patients with benign primary hyperparathyroidism may present with more severe disease than is commonly seen. In such patients, the distinction between benign and malignant disease on clinical grounds may be difficult or impossible, since profound hypercalcemia, renal disease, and osteitis fibrosa or diffuse osteoporosis may occur and even concomitant kidney and bone disease may be present. In either situation, it is preferable to have a high index of suspicion for parathyroid carcinoma than to miss the opportunity for surgical cure by failing to consider it in the differential diagnosis.

## PATHOLOGY

### Gross Pathology

Certain operative findings are helpful in distinguishing between benign parathyroid adenomas and parathyroid carcinoma. Parathyroid adenomas are usually soft, round or oval in shape, and reddish-brown in color. In contrast, parathyroid carcinomas are usually lobulated and firm to stony-hard in consistency. Approximately half are surrounded by a dense, fibrous, grayish-white capsule that adheres tenaciously to adjacent tissues and makes the tumor difficult to separate from contiguous structures. If there is gross infiltration of adjacent thyroid, nerve, muscle, or esophagus, or if there are cervical node metastases, the diagnosis of carcinoma is a simple matter. Unfortunately, any one or all of these operative findings may be absent.

### Histology

As is the case with many endocrine neoplasms, the histopathologic distinction between benign and malignant parathyroid tumors is difficult to make. In 1973, Shantz and Castleman, based on an analysis of 70 cases of parathyroid carcinoma, established a set of criteria for the pathologic diagnosis of this malignancy. These histologic features are as follows: (1) uniform sheets of (usually chief but occasionally oxyphil) cells arranged in a lobular pattern and separated by dense fibrous trabeculae; (2) mitotic figures within

tumor parenchymal cells that must be distinguished from endothelial cell mitoses; and (3) capsular or vascular invasion. Unfortunately, neither of the first two features is pathognomonic of parathyroid carcinoma, both having been reported in parathyroid adenomas. However, capsular invasion and vascular invasion correlate well with subsequent tumor recurrence and their presence greatly increases the likelihood that the tumor is malignant. Finally, the overall histologic pattern is probably more useful than any single feature in differentiating parathyroid carcinoma from adenoma. Cellular atypia, including nuclear pleomorphism and enlargement and macronucleoli, has been associated with a greater likelihood of malignancy. The presence of more than one of the above-mentioned characteristics in a lesion should alert the pathologist.

### Other Histologic Markers

Although several histologic techniques have been used to improve the diagnosis of parathyroid carcinoma, none have proved consistently helpful. Electron microscopy of parathyroid cancer reveals nuclear and mitochondrial alterations and evidence of increased secretory activity but is not useful in distinguishing benign from malignant tumors. Nuclear diameter is larger in parathyroid carcinomas than in adenomas, but is not very useful in the individual case. Measurement of nuclear DNA content by flow cytometry is of limited value both in establishing the diagnosis of parathyroid carcinoma and in predicting the invasive potential of the tumor. Although mean nuclear DNA content is greater and an aneuploid DNA pattern is more common in parathyroid carcinoma than in adenomas, aneuploidy occurs too frequently in parathyroid adenomas to be useful in differentiating benign from malignant parathyroid tumors.

Immunohistochemical staining of RB protein with polyclonal antibodies is usually absent in parathyroid carcinomas and almost always present in parathyroid adenomas. However, staining for RB protein with monoclonal antibodies was not useful in distinguishing between benign and malignant parathyroid tumors. Similarly, immunostaining for the cell cycle-associated antigen Ki-67, a marker of proliferative activity, is not consistently useful for distinguishing carcinomas from adenomas.

Invasive growth of various neoplasms may be facilitated by tumoral secretion of proteolytic enzymes, such as gelatinase A. Farnebo *et al.* have reported that gelatinase A mRNA was detected in 14 of 18 unequivocal and 4 of 13 equivocal parathyroid cancers. The

strongest signal was detected in the fibroblasts and macrophages at the tumor border, rather than in the tumor cells. This new technique may prove useful for the diagnosis of malignancy.

### NATURAL HISTORY

Parathyroid carcinoma is an indolent, albeit tenacious, tumor with rather low malignant potential. It tends to recur locally at the operative site and spread to contiguous structures in the neck. Metastases occur late in the course of the disease with spread via both lymphatic and hematogenous routes. Cervical nodes (30%) and lung (40%) are involved most commonly, followed by liver (10%). Occasional involvement of bone, pleura, pericardium, and pancreas has been reported.

### MANAGEMENT

#### Initial Surgery

The most effective therapy for parathyroid carcinoma is complete resection of the primary lesion at the time of the initial operation when extensive local invasion and distant metastases are less likely. Therefore, both preoperative suspicion and intraoperative recognition are of paramount importance in the management of this cancer. This is particularly important in the era of minimally invasive parathyroidectomy, when full neck exploration with identification of all four parathyroid glands is less common than it used to be. Those patients in whom the clinical presentation is suggestive of parathyroid carcinoma warrant thorough exploration of all four parathyroid glands, since parathyroid carcinoma has been reported to coexist with benign adenomas or hyperplasia.

When the gross pathologic findings suggest malignancy, the following steps should be taken:

- The lesion should be removed *en bloc* together with the ipsilateral thyroid lobe and isthmus.
- The trachea should be skeletonized.
- Any contiguous tissues to which the tumor adheres should be resected.
- Great care should be taken to avoid rupture of the capsule of the gland, which increases the likelihood of local seeding of the tumor.
- If the recurrent laryngeal nerve is involved in the tumor, it must be resected.
- Tracheoesophageal, paratracheal, and upper mediastinal lymph nodes should be excised.
- Extensive lateral neck dissection is indicated only when there is spread to the anterior cervical nodes.

The situation is more complex when the diagnosis is made in the early postoperative period on the basis of pathology, particularly in view of the controversy surrounding the histopathology of parathyroid carcinoma. However, a second neck exploration is indicated under the following conditions:

- The gross characteristics of the lesion are typical of a parathyroid cancer.
- The histology appears to be aggressive with extensive vascular or capsular invasion.
- The patient remains hypercalcemic.

If any of these conditions apply, the structures adjacent to the tumor site should be resected in the manner described above. If none of these features are present and the diagnosis was made solely on the basis of the microscopic characteristics, immediate reoperation may not be necessary since a simple complete resection of the tumor may be curative. However, careful observation of the patient is essential and frequent measurement of parathyroid hormone and serum calcium levels is necessary.

### Medical Management of the Postoperative Patient with Parathyroid Cancer

The postoperative management of a patient with parathyroid cancer must include careful attention to the serum calcium level. Prolonged severe elevation of parathyroid hormone is usually associated with increased bone resorption and formation. The amount of unmineralized bone matrix (osteoid) may be greatly increased. Sudden withdrawal of excess parathyroid hormone will permit rapid deposition of calcium and phosphorus into the excess unmineralized osteoid. This process may be associated with severe and symptomatic hypocalcemia ("hungry bone syndrome"). Although hungry bone syndrome should be regarded as a sign that the surgery has been successful, it can be dangerous and must be managed aggressively. The hypocalcemia may be severe and protracted, requiring large doses of intravenous calcium. If hyperparathyroidism was severe and there was evidence of skeletal involvement (elevated alkaline phosphatase, radiographic lesions of hyperparathyroidism), the patient should remain in the hospital until it is clear that hypocalcemia can be managed with oral calcium supplements. Sufficient supplemental calcium and calcitriol should be prescribed to maintain the serum calcium at the low end of the normal range. As the bones heal and the remaining parathyroid glands recover, the requirement for calcium will decrease,

permitting gradual reduction of the doses of calcium and withdrawal of calcitriol. After this point, serum calcium and parathyroid hormone levels should be monitored every 3 months.

### Surgical Management of Recurrent or Metastatic Disease

The management of recurrent or metastatic parathyroid carcinoma reflects the rather indolent biology of this cancer and, in contrast to many other tumors, is primarily surgical. Recurrent carcinoma in the neck should be treated with wide excision of the involved area, including the regional lymph nodes and other involved structures. Since even very small tumor deposits may produce sufficient parathyroid hormone to cause severe hypercalcemia, resection of accessible distant metastases in lymph nodes, lungs, or liver should be performed. Resection, even if incomplete, may provide significant palliation, resulting in periods of normocalcemia that range from months to years. In addition, surgical debulking of tumor deposits may make it easier to control the hypercalcemia medically.

### Preoperative Localization of Recurrent Parathyroid Cancer

In patients with recurrent hypercalcemia, preoperative localization studies should be performed. However, careful palpation of the neck should not be neglected, since recurrence occurs earliest and most often at the original site and such tumors are frequently palpable. Thallium-201–technetium-99m scanning is useful in locating tumors in the neck and upper mediastinum. Technetium-99m–sestamibi used concurrently with a handheld, gamma-detecting probe may also be useful for the intraoperative localization of abnormal parathyroid tissue. Thallium-201 is also helpful for situations in which the thyroid has been partially or completely resected or when pulmonary metastases are suspected. Computerized tomography and magnetic resonance imaging are useful adjuncts to ultrasonography in evaluation of the neck and are superior for detection of distant metastases in the chest or abdomen. Positron emission tomography with a positron-emitting analogue of 2-deoxyglucose ( $^{18}\text{F}$ fluorodeoxyglucose) has also been reported to successfully localize parathyroid cancer deposits. If noninvasive testing does not yield results, arteriography or selective venous catheterization may be useful. Fine-needle aspiration biopsy should be avoided in order to avoid seeding deposits of malignant tissue.

## Radiation Therapy

Parathyroid carcinoma is not a radiosensitive tumor. The use of radiation therapy to control tumor growth and decrease hormone production has been ineffective in the majority of cases in which it has been attempted. In the occasional situation, radiation to the neck after surgery for recurrence may be helpful in preventing tumor regrowth.

## Chemotherapy

Attempts to control tumor burden with chemotherapy have been disappointing. Because of the rarity of parathyroid carcinoma, experience with various chemotherapeutic regimens is limited to scattered case reports. Several regimens (nitrogen mustard; vincristine, cyclophosphamide, and actinomycin D; adriamycin, cyclophosphamide, and 5-fluorouracil; and adriamycin alone) have been ineffective. Two patients have been treated with synthetic estrogens with some success. A single patient with pulmonary metastases responded to treatment with dacarbazine, 5-fluorouracil, and cyclophosphamide with a decrease in PTH and normalization of serum calcium for 13 months. Another patient responded to dacarbazine alone with a brief but significant decline in her serum calcium level. An 18-month remission with regression of a mediastinal mass and pleural effusion was induced in a patient with a nonfunctioning parathyroid carcinoma by a regimen consisting of methotrexate, doxorubicin, cyclophosphamide, and lomustine. Such approaches are probably worth trying if surgical measures fail.

## Management of Hypercalcemia

When parathyroid carcinoma has become widely disseminated and surgical resection is no longer effective, cure is impossible. However, even at this juncture, relatively prolonged survival can be achieved with adequate control of hypercalcemia, although the extremely elevated PTH levels and the intensity of the associated bone resorption often make this goal difficult to realize.

Acute hypercalcemia associated with parathyroid carcinoma is treated in the same way as hypercalcemia due to any other cause. Management includes infusion of saline to restore fluid volume and enhance urinary calcium excretion. After this has been accomplished, loop diuretics are added to further increase calciuresis. Agents that inhibit osteoclast-mediated bone resorption are virtually always necessary.

## Bisphosphonates

The bisphosphonates are a group of drugs that inhibit osteoclast-mediated bone resorption. Several of these drugs have shown some promise in the therapy of parathyroid carcinoma.

Clodronate ( $\text{Cl}_2\text{MDP}$ ) lowers serum calcium in parathyroid carcinoma when administered intravenously. It is widely available in Europe and the United Kingdom, but it is not available in the United States. Pamidronate, when infused for periods ranging from 2 to 24 h and at doses ranging from 45 to 90 mg per day, has been at least transiently effective in lowering serum calcium levels in several patients with parathyroid cancer. A more potent bisphosphonate, zoledronate, has been approved in the United States for the treatment of hypercalcemia. Zoledronate, at doses of 4 and 8 mg, has been shown to be superior to pamidronate (90 mg) in the treatment of hypercalcemia of malignancy, in terms of both the response rate and the duration of response. Though zoledronate has not been specifically evaluated in parathyroid cancer, it is a reasonable approach to controlling hypercalcemia in these patients.

## Plicamycin

Plicamycin (mithramycin), another specific inhibitor of bone resorption, lowers serum calcium levels in parathyroid carcinoma. It is administered intravenously at a dose of 25  $\mu\text{g}/\text{kg}$  over 4 to 8 h and may be repeated at daily intervals for up to 7 days until the serum calcium falls into an acceptable range. Unfortunately, plicamycin has toxic effects on the liver, kidneys, and bone marrow that increase with the number of exposures. Intravenous bisphosphonates have largely supplanted plicamycin in the therapy of hypercalcemia.

## Calcimimetics

In normal circumstances, parathyroid hormone secretion is mediated by a cell surface calcium-sensing receptor and this regulatory response is generally retained in benign parathyroid tumors. An allosteric modulator of the calcium receptor with calcimimetic properties has been shown to lower serum parathyroid hormone and calcium concentrations in patients with primary hyperparathyroidism. This same agent was used to treat a patient with parathyroid carcinoma. Serum calcium was controlled for 2 years without adverse effects. Newer calcimimetic agents are under investigation in parathyroid cancer and may show promise in its management.

### Calcitonin

Although calcitonin inhibits osteoclast-mediated bone resorption, increases urinary calcium excretion, and has been reported to lower serum calcium in two patients with parathyroid cancer, it is not very effective in most cases.

### Immunization

A patient with parathyroid carcinoma metastatic to lungs and pleura had severe hypercalcemia that was resistant to oral clodronate, intravenous pamidronate, octreotide, 5-fluorouracil, and streptozotocin. She was immunized with human and bovine PTH peptides, followed with booster doses at 4 and 11 weeks. Antibodies against PTH were detected at 4 weeks. Before therapy, serum calcium varied between 3.5 and 4.2 mmol/liter. Serum calcium levels remained significantly lower (2.5 to 3.0 mmol/liter) throughout the 6 months of observation. There was rapid improvement in her clinical condition and no significant adverse effects were observed. If confirmed, this would seem to be a novel and relatively simple approach to the control of hypercalcemia in patients resistant to other measures.

### PROGNOSIS

The prognosis of parathyroid carcinoma is quite variable. No single characteristic correlates predictably with outcome. Early recognition and complete resection at the time of the initial surgery offer the best prognosis. The average time between surgery and the first recurrence is approximately 3 years although intervals of up to 20 years have been reported. Once the tumor has recurred, complete cure is unlikely although prolonged survival is still common in these circumstances with palliative surgery. Five-year survival rates vary from 40 to 86%. The National Cancer Database survey reported 10-year survival to be approximately 49%.

### See Also the Following Articles

Bisphosphonates • Hypercalcemia and Hypercalcemia Treatment • Hyperparathyroidism, Primary • Parathyroid Glands, Pathology • Parathyroid Hormone (PTH) • Parathyroid Surgery • Pseudohypoparathyroid States • Thyroid Carcinoma

### Further Reading

- Bradwell, A. R., and Harvey, T. C. (1999). Control of hypercalcaemia of parathyroid carcinoma by immunisation. *Lancet* **353**, 370–373.
- Chow, E., Tsang, R. W., Brierley, J. D., and Filice, S. (1998). Parathyroid carcinoma—The Princess Margaret Hospital experience. *Int. J. Radiat. Oncol. Biol. Phys.* **41**, 569–572.
- Collins, M. T., Skarulis, M. C., Bilezikian, J. P., Silverberg, S. J., Spiegel, A. M., and Marx, S. J. (1998). Treatment of hypercalcemia secondary to parathyroid carcinoma with a novel calcimimetic agent. *J. Clin. Endocrinol. Metab.* **83**, 1083–1088.
- Cordeiro, A. C., Montenegro, F. L., Kulcsar, M. A., Dellanegra, L. A., Tavares, M. R., Michaluart, P., Jr., and Ferraz, A. R. (1998). Parathyroid carcinoma. *Am. J. Surg.* **175**, 52–55.
- Hakaim, A. G., and Esselstyn, C. B., Jr. (1993). Parathyroid carcinoma: 50-year experience at The Cleveland Clinic Foundation. *Cleveland Clin. J. Med.* **60**, 331–335.
- Hundahl, S. A., Fleming, I. D., Fremgen, A. M., and Menck, H. R. (1999). Two hundred eighty-six cases of parathyroid carcinoma treated in the U. S. between 1985–1995: A National Cancer Data Base Report. The American College of Surgeons Commission on Cancer and the American Cancer Society. *Cancer* **86**, 538–544.
- Miki, H., Sumitomo, M., Inoue, H., Kita, S., and Monden, Y. (1996). Parathyroid carcinoma in patients with chronic renal failure on maintenance hemodialysis. *Surgery* **120**, 897–901.
- Sandelin, K., Auer, G., Bondeson, L., Grimelius, L., and Farnebo, L. O. (1992). Prognostic factors in parathyroid cancer: A review of 95 cases. *World J. Surg.* **16**, 724–731.
- Shane, E. (1999). Hypercalcemia: Pathogenesis, clinical manifestations, differential diagnosis and management. In "ASBMR Primer on Metabolic Bone Disease" (M. Favus, ed.), 4th ed., pp. 183–186. Lippincott-Raven, Philadelphia, PA.
- Shane, E. (2001). Clinical Review 122: Parathyroid carcinoma. *J. Clin. Endocrinol. Metab.* **86**, 485–493.
- Shane, E., and Bilezikian, J. (1982). Parathyroid carcinoma: A review of 62 patients. *Endocr. Rev.* **3**, 218–226.
- Wynne, A., Heerden, J. V., Carney, J., and Fitzpatrick, L. (1992). Parathyroid carcinoma: Clinical and pathological features in 43 patients. *Medicine* **71**, 197–205.

parenchymal cells, fat cells, and fibrovascular stroma. The parenchymal cells are usually arranged in nests and cords, nourished by a rich capillary network. The functional cells of the parathyroid are mainly of the chief-cell type although other populations of cells (oxyphil cells, clear cells, and transitional oxyphil cells) are found. Most authorities believe that the different parenchymal cell types reflect variabilities in the morphologic expression of chief cells. Chief cells are polygonal in shape and contain a slightly eosinophilic cytoplasm and small round nucleus. The oxyphil cells show dense eosinophilic cytoplasm and small nuclei. Clear cells are a form of chief cells with abundant cytoplasmic glycogen; these are usually seen in the parathyroids of embryos and fetuses and their numbers significantly decline with age. Cells with cytologic features intermediate between chief cells and oxyphil cells are termed transitional oxyphil cells. The number of oxyphil and transitional oxyphil cells is small at birth; however, it increases with age and in certain pathological conditions.

The stromal fat content of a parathyroid gland is significantly lower than 50%. In an adult normal parathyroid, approximately 17 to 75% of the cells are adipocytes and their number varies with the nutritional status and weight of the individual.

## PATHOLOGY

The clinical syndrome of hypercalcemia, especially if it is asymptomatic or associated with mild symptoms, is due to pathological lesions of the parathyroids. Up to 90% or more of hypercalcemia cases are due to primary hyperparathyroidism.

### Hyperparathyroidism

Hyperparathyroidism can be divided into primary and secondary types. Primary hyperparathyroidism is characterized by inappropriate secretion of parathyroid hormone (PTH) from enlarged parathyroid glands in the absence of a known stimulus for PTH, leading to hypercalcemia. Secondary hyperparathyroidism is an increase in parathyroid hormone most commonly in response to hypocalcemia or hyperphosphatemia associated with renal failure. Tertiary hyperparathyroidism refers to autonomous parathyroid hyperfunction in patients with secondary hyperparathyroidism.

#### **Primary Hyperparathyroidism**

Primary hyperparathyroidism (PHP) is a common disorder of the parathyroid gland and can occur in

all age groups, although it is rare in children. The cause of PHP is unknown; suggested causes include head and neck irradiation, mutations of the multiple endocrine neoplasia type 1 (MEN1) gene, and abnormalities in the genes involving the 11q13 locus and in the retinoblastoma (RB) gene.

The pathologic lesions responsible for primary hyperparathyroidism include hyperplasia, adenoma, and, rarely, carcinoma.

#### **Parathyroid Hyperplasia**

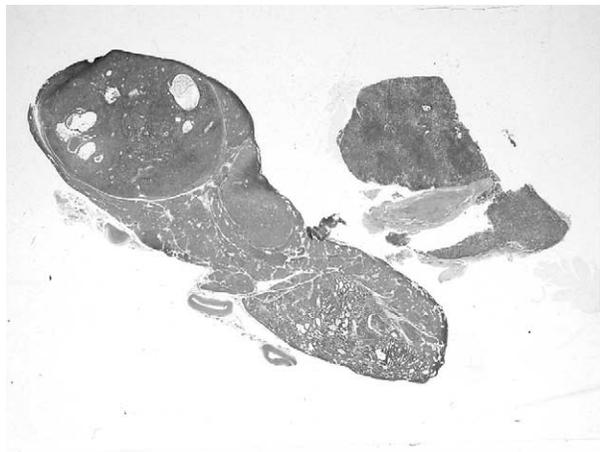
Primary parathyroid hyperplasia is defined as the proliferation of the parenchymal cells, leading to an increase in gland weight in multiple parathyroid glands in the absence of a known stimulus for parathyroid hormone secretion.

Two types of parathyroid hyperplasia are seen: the more common chief cell hyperplasia and the rare water cell or clear cell hyperplasia.

*Chief cell hyperplasia.* It can be estimated from the literature that 15% of hyperparathyroidism cases are caused by parathyroid hyperplasia; however, some reports have indicated that approximately half of the cases of primary hyperparathyroidism are produced by hyperplasia. The stimulus for this disorder is unknown; however, a role for a possible circulating factor that can initiate proliferation of parathyroid cells in culture has been suggested. Familial hyperparathyroidism or MEN syndrome is seen in approximately 30% of patients with chief cell hyperplasia. It has been shown that hyperplasias are monoclonal proliferations. Parathyroid hyperplasia associated with MEN1 involves allelic deletions on chromosome 11 and such lesions are larger than those without deletions. Similar deletions have also been encountered in sporadic parathyroid adenomas. This suggests that the monoclonal proliferations may develop after a phase of polyclonal hyperplasia.

On gross examination, all four glands are enlarged and the combined weight can range from 150 mg to over 20 g, but usually is in the range of 1 to 3 g. Chief cells constitute the dominant cell type; however, one may also observe intermixed oxyphil cells and transitional oxyphil cells. The cells are usually arranged in a solid, trabecular, or nodular pattern (Fig. 1). Nodule formation can also be seen leading to asymmetrical gland enlargement.

The cytoplasmic fat in chief cells is either reduced or absent. The nodular areas are usually devoid of fat, whereas those between the nodules may contain fat. Nuclear pleomorphism and mitoses are rarely found in primary hyperplasia.



**Figure 1** A whole-mount section of a case of nodular hyperplasia of the parathyroid, showing a thinly encapsulated parathyroid nodule surrounded by hypercellular parathyroid.

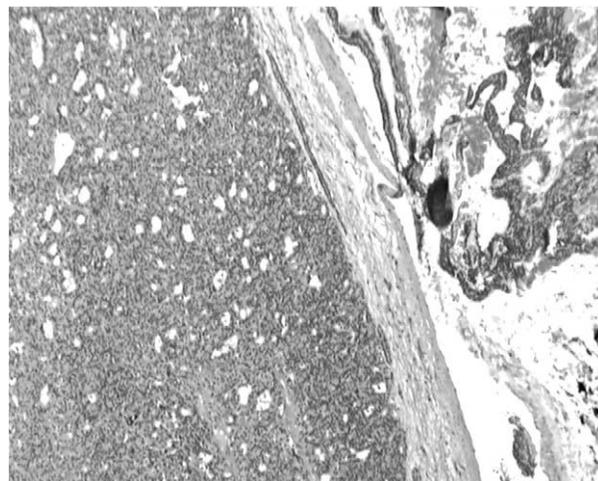
*Water/clear cell hyperplasia.* This rare condition is characterized by the proliferation of vacuolated water/clear cells in multiple parathyroid glands. It is more common in females and causes pronounced hypercalcemia and severe clinical disease. This is the only parathyroid disorder in which the superior glands are larger than the lower pair of glands. The affected glands tend to be larger and irregular in shape and the proliferating cells may infiltrate into the surrounding tissue of the neck.

#### *Parathyroid Adenoma*

Parathyroid adenoma is the single most common cause of PHP. The incidence of parathyroid adenoma is reported to be between 30 and 90%; this variation is due to a lack of application of standardized diagnostic pathologic criteria and differences in the patient populations that were studied. Adenomas are more common in females than in males, with a ratio of approximately 3 to 1.

Parathyroid adenoma can originate in any of the four parathyroid glands; however, it is more common in the lower glands than in the upper glands. Approximately 10% of the adenomas are found at ectopic sites, including the mediastinum, thyroid, and esophagus and within the retroesophageal tissue.

Grossly, adenomas are oval or kidney-shaped, reddish-brown in color, and soft in consistency. They usually replace the entire parathyroid gland or a grossly visible yellowish-brown rim of residual parathyroid may be identified. In some cases, the adenomas are multilobated, which may account for their incomplete excision. Some pathologists believe that the remaining normal parathyroid rim is a reliable criterion for the



**Figure 2** Parathyroid adenoma with a thin fibrous capsule and a rim of compressed normal parathyroid containing fat cells.

diagnosis of adenoma; however, it is seen in only 50 to 60% of cases and its absence does not exclude the presence of a parathyroid adenoma. Parathyroid adenomas can markedly vary in size and weight; the size can range from <1 to >3 cm; similarly, the weight can range from 300 mg to several grams.

As viewed by light microscopy, adenomas are usually encapsulated and can show a rim of compressed normal parathyroid (Fig. 2). The tumor cells are usually arranged in nests and cords surrounded by a rich capillary network. Chief cells are the dominant cell type in the majority of parathyroid adenomas; oxyphil cells and transitional oxyphil cells can also be seen in varying proportions scattered within the collections of chief cells. Some adenomas are composed exclusively of oxyphil cells, the so-called “oxyphil adenomas.” These tumors tend to be larger than the chief cell adenomas and the serum calcium tends to be minimally elevated.

The chief cells in adenomas can exhibit nuclear pleomorphism and giant cell formation. Some authors have indicated that the presence of random nuclear atypia is reassuring that a parathyroid tumor is benign. Mitotic figures are uncommon in adenomas; however, they have been encountered in a few cases. It has been suggested that the presence of mitoses is indicative of malignancy in parathyroid adenoma; however, this observation needs further examination in large cohorts of cases.

Parathyroid adenomas can show prominent cystic degeneration. Other degenerative changes are more commonly seen in large tumors and include fibrosis, hemorrhage, cholesterol clefts, hemosiderin, and calcification.

Adenomas are virtually devoid of adipocytes, which are observed only in the normal rim of the compressed parathyroid. Rarely, some cases of adenomas can show collections of adipocytes, which can be mistaken for normal parathyroid. Parathyroid adenoma is considered a single-gland disease; however, double adenomas have been reported. The proposed criteria for this diagnosis include the presence of two enlarged hypercellular glands and the identification of two other normal-sized parathyroids. Long-term follow-up after excision of double adenomas must demonstrate a lack of recurrence of hypercalcemia to accept this diagnosis; unfortunately, some patients with double adenomas represent metasynchronous quadruple-gland disease; i.e., hyperplasia and the true nature of the cause of hyperparathyroidism in these cases may become apparent only on prolonged follow-up.

#### *Lipoadenoma—The So-Called “Hamartoma of the Parathyroid”*

These are rare parathyroid neoplasms, which were initially described as parathyroid “hamartoma.” They usually occur in the cervical region; however, cases in the superior and posterior mediastinum have been reported. Most of these are functional tumors and are associated with hyperparathyroidism. Grossly, lipoadenomas are circumscribed, rarely encapsulated, and yellowish-tan in color and they show a lobulated cut surface. As viewed by light microscopy, lipoadenomas consist of mature adipose tissue or myxoid stroma and nests of parathyroid parenchyma. In some cases, other mesenchymal elements, including metaplastic bone, can be seen. The parenchymal component of the lipoadenomas includes chief cells and some oncocytic cells, which are usually arranged in thin cords, tubules, and abortive acini.

#### *Parathyroid Carcinoma*

Parathyroid carcinoma is seen in 0.5 to 2% of cases of PHP. There appears to be an equal sex distribution, whereas adenomas are most commonly seen in females. Patients with parathyroid carcinoma are younger than those with adenoma and usually present with pronounced hypercalcemia and/or with systemic manifestations related to hypercalcemia, e.g., nephrolithiasis, renal failure, and bone disease. Parathyroid carcinoma is rarely associated with familial endocrine disorders or secondary parathyroid hyperplasia. The initial clinical presentation of parathyroid carcinoma may be that of a palpable neck mass, which can be mistaken for a tumor originating in thyroid.

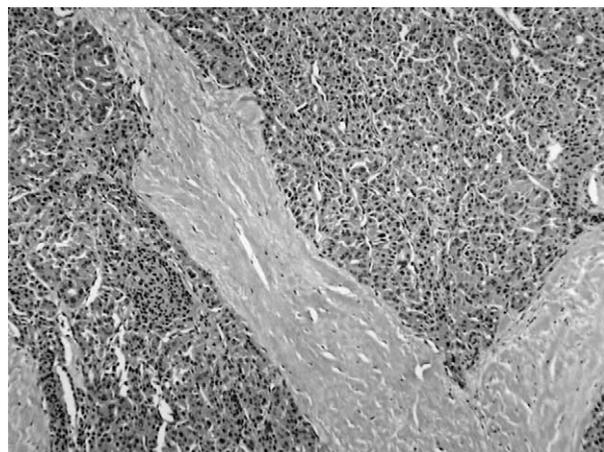
On gross examination, parathyroid carcinomas are large tumors with adherence to and invasion of

the surrounding structures, such as the thyroid and the soft tissues of the neck and peri-esophageal region. This infiltration into the neighboring neck structures serves as an important surgical finding and may initiate an *en bloc* resection of the tumor mass with surrounding adherent structures. However, some tumors may be totally encapsulated, lack gross invasion, and resemble parathyroid adenomas. Parathyroid carcinoma involves only one gland and rarely has been reported to arise in ectopic locations.

As viewed by light microscopy, the entire gland is traversed by broad fibrous bands, which seem to originate from the capsule and extend into the substance of tumor, leading to a lobulated appearance (Fig. 3). Parathyroid carcinomas are usually composed of chief cells; however, oxyphil cells and transitional cells can also be seen. In some cases, the entire tumor is composed of oxyphil cells. The cells can exhibit a bland cytology or show marked anaplasia.

Mitotic figures are observed in most cases; some pathologists have suggested this feature to be of great importance in diagnosing parathyroid carcinoma. However, mitoses can also be seen in parathyroid adenoma and hyperplasia and their absence does not rule out a diagnosis of carcinoma. The authors believe that, although not definite indicators of malignancy, mitoses in parathyroid lesion should be of concern since the follow-up in reported benign cases of parathyroid tumor with mitoses is limited. Marked mitotic activity and abnormal mitotic figures in unequivocal carcinomas are indicators of poor prognosis.

The only reliable indicator of malignancy in parathyroid carcinoma is invasion of the surrounding structures and metastases. Parathyroid carcinomas infiltrate into the surrounding neck structures and



**Figure 3** Parathyroid carcinoma showing broad fibrous bands intersecting the tumor.

soft tissues in the form of “tongue-like” protrusions. However, this should be distinguished from “pseudo-invasion,” which is seen in large adenomas caused by degeneration and subsequent fibrosis and “trapping” of tumor cells. Vascular invasion is seen in 10 to 15% of parathyroid carcinomas and is usually regarded as a reliable indicator of malignancy; however, some authors have observed this feature in large adenomas.

Other microscopic features, such as desmoplastic reaction, mitotic activity, nuclear atypia, and necrosis, are more common in carcinoma than in benign lesions; none of these is diagnostic of malignancy.

Nonfunctioning parathyroid carcinomas are rare. These lesions tend to be large and consist primarily of clear or oxyphil cells. These tumors can be confused with primary thyroid cancers, such as Hürthle cell lesions or medullary carcinoma. The parathyroid origin can be confirmed by positive immunohistochemical staining for parathyroid hormone and a negative reaction for thyroglobulin and calcitonin.

Parathyroid carcinoma usually behaves in an indolent fashion. Metastases can occur in up to one-third of patients and are found in regional lymph nodes, bone, lung, and liver. Multiple recurrences are common and can occur over a 15- to 20-year period. Fatal outcome in patients with parathyroid carcinoma is due mainly to the effects of excessive parathyroid hormone secretion and uncontrolled hypercalcemia rather than to a tumor mass effect. Therefore, surgical excision of recurrences or metastases can provide excellent palliation by lessening the tumor burden and concomitant hormone production.

#### **Atypical Adenoma**

This is a relatively new entity in parathyroid pathology. Some authors have suggested that if some features of malignancy, including mitoses and necrosis, are present without an infiltrative growth pattern or metastases, those cases should be designated as “atypical adenoma.” The few available follow-up studies have shown that these tumors will behave in a clinically benign fashion.

#### **Familial Hyperparathyroidism**

Primary hyperparathyroidism can present as a manifestation of MEN1 and MEN2 or as a familial disease without involvement of other endocrine organs.

#### **Multiple Endocrine Neoplasia Syndromes**

Parathyroids can display pathologic changes in most patients with MEN1 (Wermer’s syndrome). This syndrome is characterized by lesions of the pituitary, parathyroid, and endocrine pancreas, foregut carcinoid

tumors, and adrenocortical tumors. The parathyroid involvement is perceived as the earliest and most prevalent manifestation and is usually detectable at the age of 20 to 30 years in 90% of affected patients. The pathologic changes in the parathyroid are similar to those seen in nodular chief cell hyperplasia. Molecular analysis has shown that MEN1 lesions represent monoclonal proliferations possibly originating in the background of polyclonal hyperplasia.

Primary hyperparathyroidism is seen in 10–20% of patients with MEN2. The other features of this syndrome include medullary thyroid carcinoma (often bilateral), C cell hyperplasia, unilateral or bilateral pheochromocytoma, and adrenal medullary hyperplasia. Usually all four parathyroids are involved but in some cases one gland is affected, suggesting an “adenoma.” The degree of hypercalcemia in MEN2 patients is usually less severe than in those with MEN1; in MEN2, serum calcium and parathyroid hormone levels may be normal despite gland enlargement. Genetic studies have demonstrated the MEN2 locus to be on chromosome 10q11.2, the region of the ret proto-oncogene. Point mutations within the cysteine-rich extracellular region of the ret proto-oncogene have been detected in MEN2 and familial MTC and specific mutations have been revealed in families with parathyroid involvement.

#### **Familial Isolated Primary Hyperparathyroidism**

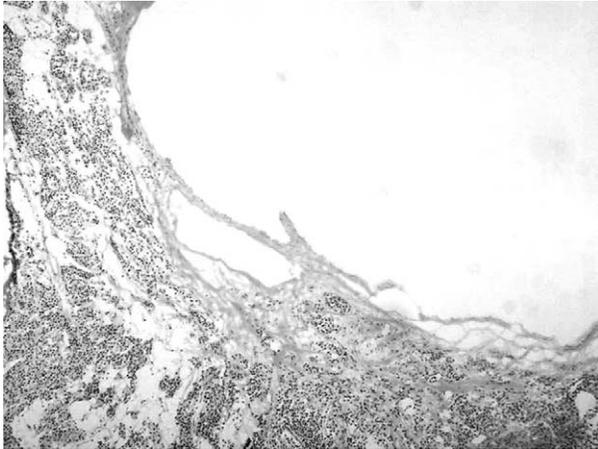
This condition is an autosomal-dominant inherited disease; however, a few families show autosomal-recessive inheritance. Parathyroid glands often show chief cell hyperplasia; some families may show adenomas, an increased risk for parathyroid carcinoma, and other nonendocrine tumors. The autosomal-dominant form has been found to be related to the MEN1 locus; however, other studies have disputed these findings.

### **Unusual Lesions of the Parathyroid**

#### **Parathyroid Cysts**

Parathyroid cysts are rare, commonly occur in the neck, and usually involve the lower parathyroids; however, some have been reported in the mediastinum. Parathyroid cysts are more common in females; these lesions are usually large, ranging in size from 1 to 6 cm. Mediastinal parathyroid cysts can be mistaken for superior/anterior mediastinum tumors. These cysts may also contain fragments of thymic tissue and are sometimes referred to as third pharyngeal pouch cysts.

Grossly, these cysts are unilocular, are smooth-walled, and contain watery clear fluid, which is high



**Figure 4** Light microscopy of a parathyroid cyst, showing a single cyst within the parathyroid lined by a single layer of epithelium.

in parathyroid hormone content. As viewed by light microscopy, the cyst wall is lined by a single layer of glycogenated epithelium and can show nests of normal parathyroid (Fig. 4). Some authors have suggested that these cysts originate due to fusion of the smaller microcysts often seen in normal parathyroids, whereas others believe that these cysts may represent embryologic remnants of pharyngeal pouches, which undergo cystic degeneration with entrapped parathyroid tissue. However, the majority of authors believe that parathyroid cysts represent degenerating adenomas and in fact some of these cases present with or are associated with hyperparathyroidism.

Parathyroid cysts can be encountered on fine-needle aspiration (FNA) and it can be difficult to distinguish them from a cystic thyroid nodule. The aspiration sample almost always consists of clear watery fluid, which can be assayed for parathyroid hormone to confirm the cytologic impression.

### **Parathyromatosis**

In primary and secondary hyperplasia of the parathyroid, small collections of parathyroid cells, mainly chief cells, rarely are encountered embedded within the surrounding soft tissue of the neck and mediastinum outside the confines of the parathyroid gland capsule. Normally, these lesions are not detectable; however, in cases of diffuse hyperplasia of the parathyroids, all functional tissue may become hyperplastic and appear as separate tissue fragments on histologic examination.

Parathyromatosis can occur for two main reasons: seeding of hypercellular parathyroid tissue during

surgical excision of abnormal parathyroid tissue (usually hyperplasia) and overgrowth of embryologic parathyroid rests. Both forms of parathyromatosis can lead to recurrent hyperparathyroidism following excision of abnormal parathyroid glands. These fragments can be confused with local invasion by parathyroid carcinoma. An examination of the parathyroids along with the clinical history always proves to be helpful in arriving at the correct diagnosis.

### **INTRAOPERATIVE ASSESSMENT**

The intraoperative assessment of the parathyroid gland during parathyroidectomy is usually limited to identification of the tissue. This procedure usually involves correct labeling, gross examination measurement, and weighing of the specimen. The representative sample is frozen and stained with hematoxylin and eosin stains (15–16, 147). Usually, the identification of parathyroid tissue is not difficult (the accuracy rate of correct identification of tissue as parathyroid ranges up to 99%); however, in some cases, it may be difficult to distinguish from other neck tissues including lymph node, thyroid, and ectopic thymus. Some authors have also advocated the use of intraoperative cytology preparations in conjunction with frozen section to increase the efficacy of identification of parathyroid tissue.

A frozen section error and failure to localize the abnormal parathyroid gland can lead to failed parathyroidectomy. A rapid parathyroid hormone assay in conjunction with frozen section can be helpful in indicating the successful excision of abnormal parathyroid gland(s). The samples for parathyroid hormone assay are obtained before excision of the parathyroids from the thyroid veins. Follow-up samples are procured again after the removal of the suspected abnormal parathyroid gland(s); a rapid fall in parathyroid hormone (by at least 50%) is indicative of successful removal of abnormal parathyroid.

In normal parathyroid glands, 80% of the cells are nonsecretory and contain intracytoplasmic fat, whereas hyperfunctioning chief cells contain much less or are devoid of intracytoplasmic fat. Therefore, the pathologist can use fat stains (Sudan IV or oil red O) or metachromatic stains (e.g., toluidine blue) to differentiate between adenoma and hyperplasia. However, fat stain is helpful in only 80% of cases and should be interpreted in conjunction with gross findings, gland weight, and size.

Density gradient measurements can also be employed to assess the ratio of parenchymal to fat cells. This rapid technique provides an objective

evaluation of the parenchymal mass. It involves procuring a sample from the abnormal (preferably center) and normal rim of a grossly enlarged gland and determining their densities in a 25% mannitol solution. Abnormal parathyroid tissue will sink due to decreased fat content and high parenchymal mass. The surgeon can perform this test in the operating room to distinguish between normal and abnormal parathyroid glands.

In addition to the above-mentioned techniques of intraoperative assessment of parathyroid pathology, a close communication between the surgeon and the pathologist during surgery is prudent.

In short, the recommended procedure for intraoperative and histologic assessment of the parathyroid gland is as follows: the largest gland, resected first, should be weighed, measured, and examined histologically. In the case of diffuse proliferation of chief cells, presence of normal rim, lack of intracellular lipid, and presence of second normal gland, a diagnosis of adenoma can be rendered. However, another parathyroid gland should always be sampled to distinguish between single-gland and multigland disease, since some hyperplastic glands may resemble adenomas and only up to 70% of adenomas will show a rim of normal tissue.

## SPECIAL STUDIES

### Cytology

A majority of parathyroid lesions are not palpable; therefore, it is unlikely that FNA will be performed on a parathyroid tumor. However, image-guided FNA is helpful in patients undergoing repeat surgical excisions for recurrent/persistent hypercalcemia where the anatomy is distorted. In addition, as mentioned, some parathyroid adenomas, especially those located in the capsule of the thyroid, can be mistaken for thyroid nodules and hence will undergo FNA. The parathyroid FNA samples usually show a monotonous population of small round cells with chromatin arranged in an organoid or a trabecular arrangement. Some cases may also show the presence of vascular cores. Immunostaining for parathyroid hormone may help in distinguishing these lesions from primary thyroid tumors.

### Proliferative Markers

It has been shown that cell proliferation markers (MIB1/Ki67) can be helpful in distinguishing between parathyroid adenomas and hyperplasia. However, other studies have shown similar proliferative indices

for adenomas and hyperplasias. Some authors have reported higher numbers of labeled nuclei for proliferating cell nuclear antigen in parathyroid adenomas than in hyperplasias.

Cyclin-dependent kinase inhibitor p27 helps regulate the transition from the G<sub>1</sub> to the S phase of the cell cycle. Normal tissues show higher levels of this protein than their neoplastic counterparts. It has been reported by performing *in situ* hybridization for p27 mRNA that normal parathyroids express higher levels of p27 than hyperplasia, adenomas, and carcinomas.

### Flow Cytometry

Both image and flow cytometry techniques have been used to assess the nuclear DNA content of normal and abnormal parathyroid glands. Normal parathyroid gland usually shows diploid patterns; however, tetraploidy has been observed associated with adenomas in some normal glands, indicating multiglandular abnormality. Both diploid and tetraploid DNA patterns are observed in adenomas as well as hyperplastic glands. Aneuploidy has been reported in up to 25% of adenomas and in a considerably higher percentage of parathyroid carcinomas. In parathyroid carcinoma, aneuploidy has been shown to be associated with poor prognosis; however, hyperplastic glands can also show aneuploid DNA content. Therefore, the finding of aneuploidy in parathyroid lesions does not indicate a diagnosis of malignancy.

### Genetics

It has been shown that the overexpression of PRAD1 (for parathyroid adenoma)/cyclin D1 induced by a DNA rearrangement of the PTH gene can be seen in parathyroid adenomas. This rearrangement is created by a break in the vicinity of the parathyroid gene on the short arm of chromosome 11 (band 11p15), a second break in the long arm (band 11q13), rotation of the central fragment around the axis of the centromere, and rejoining. However, in addition to adenomas this gene has been found to be overexpressed in nodular hyperplasia as compared to diffuse hyperplasia of parathyroid.

The retinoblastoma gene (Rb) is a tumor suppressor gene. Allelic deletion of the Rb gene on chromosome 13 has been reported in parathyroid tumors. It has been shown that a majority of parathyroid carcinomas show abnormal expression of Rb protein, with a complete or predominant absence of nuclear staining for protein, whereas parathyroid adenoma shows positive nuclear staining for Rb protein. Thus, this differential staining

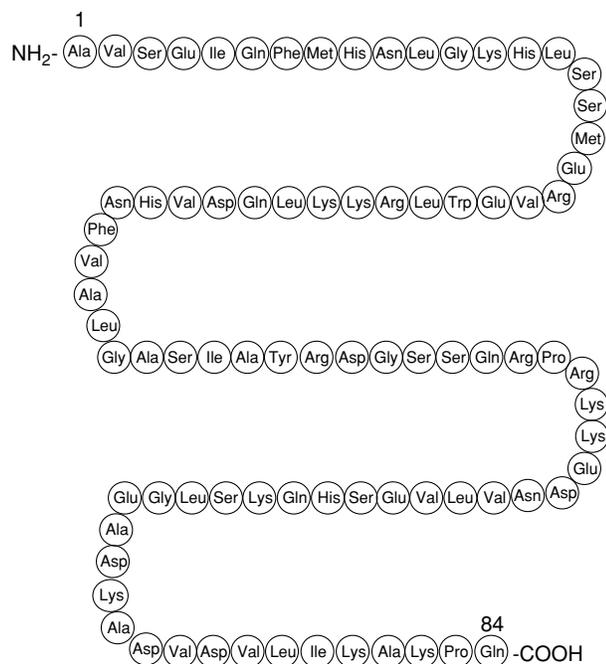
for Rb protein can assist in the distinction between parathyroid adenomas and carcinomas. However, one must be careful in interpreting these results since some parathyroid carcinomas do not show loss of Rb protein and a few adenomas do.

### See Also the Following Articles

Parathyroid Cancer • Parathyroid Hormone (PTH) • Parathyroid Surgery • Thyroid Carcinoma

### Further Reading

- Akerstrom, G., Malmaeus, J., and Bergstrom, S. (1984). Surgical anatomy of human parathyroid glands. *Surgery* **95**, 14–21.
- DeLellis, R. A. (1993). "Tumors of the Parathyroid Glands," Fascicle 6, Series 3. Armed Forces Institute of Pathology, Washington, DC.
- Dekker, A., Dunsford, H.A., and Geyer, S. J. (1979). The normal parathyroid gland at autopsy: the significance of stromal fat in adult patients. *J. Pathol.* **128**, 127–132.
- Dolgin, C., LoGerfo, P., LiVolsi, V., and Feind, C. (1979). Twenty-five year experience with primary hyperparathyroidism at Columbia Presbyterian Medical Center. *Head Neck Surg.* **2**, 92–98.
- Farnebo, F., Teh, B. T., Kytola, S., *et al.* (1998). Alterations of the MEN-1 gene in sporadic parathyroid tumors. *J. Clin. Endocrinol. Metab.* **83**, 2627–2630.
- Ghandur-Mnaymneh, L., and Kimura, N. (1984). The parathyroid adenoma: A histopathologic definition with a study of 172 cases of primary hyperparathyroidism. *Am. J. Pathol.* **115**, 70–83.
- Grimelius, L., and Johansson, H. (1997). Pathology of parathyroid tumors. *Semin. Surg. Oncol.* **13**, 142–154.
- Heppner, C., Kester, M. B., Agaroval, S. K., *et al.* (1997). Somatic mutation of the MEN-1 gene in parathyroid tumors. *Nat. Genet.* **16**, 375–378.
- Marcus, R. (1989). Laboratory diagnosis of primary hyperparathyroidism. *Endocrinol. Metab. Clin. N. Am.* **18**, 647–658.
- Ohye, H., Sato, M., Matsubara, A., *et al.* (1998). Germline mutations of the MEN-1 gene in a family with primary hyperparathyroidism. *Endocr. J.* **45**, 719–723.
- Sandelin, K., Auer, G., Bondeson, L., *et al.* (1992). Prognostic factors in parathyroid cancer: A review of 95 cases. *World J. Surg.* **16**, 724–731.
- Shane, E., and Bilezikian, J. P. (1982). Parathyroid carcinoma: A review of 62 patients. *Endocr. Rev.* **3**, 218–226.
- Verdon, C. A., and Edis, A. J. (1981). Parathyroid "double adenomas": Fact or fiction? *Surgery* **90**, 523–526.
- Wang, C. A. (1976). The anatomic basis of parathyroid surgery. *Ann. Surg.* **183**, 271–275.
- Wynne, A. G., van Heerden, J., Carney, J. A., and Fitzpatrick, L. A. (1992). Parathyroid carcinoma: Clinical and pathologic features in 43 patients. *Medicine* **71**, 197–205.



**Figure 1** Primary structure of bovine PTH. Shown is the complete 84-amino-acid sequence of the mature, secreted form of bovine PTH.

tissues, including the skeleton. The (1-34) region of PTHrP exhibits significant amino acid sequence homology to the (1-34) region of PTH (~35% identity; Fig. 2); like PTH, PTHrP mediates many of its biological actions via the PTH-1 receptor, which is consequently also called the PTH/PTHrP receptor.

Though the calcium homeostatic actions of PTH can be explained by activation of the PTH-1R in bone and kidney, PTH also binds to an as yet uncloned receptor that binds both PTH(1-84) and carboxy-terminal fragments that do not bind to the PTH-1R. These fragments can block the bone-resorbing actions of PTH(1-34) and other stimuli of resorption by mechanisms being explored.

## PTH SYNTHESIS AND SECRETION

In mammals, PTH is synthesized as a 115-amino-acid preprohormone. The 25-amino-acid presequence is removed by signal peptidase during translocation of the nascent polypeptide chain across the endoplasmic reticulum and the 6-amino-acid prosequence is removed during subsequent processing of the peptide through the *trans*-Golgi network. The mature 84-amino-acid hormone is then packaged into storage granules, where it remains until it is secreted or degraded by proteases found in the secretory granules. These proteases include various cathepsins (-B, -D, and -H) and their activities result in the secretion of PTH fragments of various lengths, with most containing carboxy-terminal immunoreactive epitopes. Since the studies mentioned earlier demonstrate activities of carboxy-terminal PTH fragments, intracellular processing of PTH in the parathyroid gland may have great functional significance.

Ambient  $\text{Ca}_o^{2+}$  concentrations regulate the rates at which PTH is synthesized, secreted, and degraded via a negative feedback loop. A calcium-sensing receptor (CSR), which is expressed on the parathyroid cell surface, detects the level of extracellular  $\text{Ca}^{2+}$ . The CSR is a GPCR that contains a large (~600 amino acid) amino-terminal extracellular domain in which resides multiple binding sites for  $\text{Ca}^{2+}$ . The binding of  $\text{Ca}^{2+}$  to the CSR is highly cooperative (Hill coefficient ~5) and this cooperativity results in the capacity to detect extremely small variations in  $\text{Ca}_o^{2+}$ . As  $\text{Ca}_o^{2+}$  binds to the CSR, PTH secretion is repressed immediately and, on a longer time scale, PTH synthesis and parathyroid cell proliferation are repressed. PTH degradation continues when  $\text{Ca}_o^{2+}$  is elevated, so most of the immunoreactive PTH secreted under these conditions consists of carboxy-terminal fragments. All of these activities are reversed in the presence of low concentrations of  $\text{Ca}_o^{2+}$ . Thus, the parathyroid glands adapt to both the immediate (seconds to minutes) and the longer term (days to weeks) changes in

<b>hPTH(1-34)</b>	<b>S V S E I Q L M H N L G K H L N S M E R V E W L R K K L Q D V H N F</b>
cPTH(1-34)	S V S E M Q L M H N L G E H R H T V E R Q D W L Q M K L Q D V H S A
hPTHrP(1-34)	A <b>V S E H Q L L H D K G K S I Q D L R R R F F L H H L I A E I H T A</b>
bTIP39	S L A L A D D A A F R E R A R L <b>L A A L E R R H W L N S Y M H K L L V L D A F</b>

**Figure 2** Sequence alignment of PTH peptides. Shown are the aligned amino acid sequences of the fully bioactive (1-34) regions of human (h) PTH, chicken (c) PTH, hPTHrP, and the entire (1-39) region of bovine (b) TIP39, a PTH-2 receptor selective ligand. The hPTH(1-34) sequence is shown in boldface type, as are residues in the other ligands that are identical to the corresponding hPTH residue.

$\text{Ca}_o^{2+}$  that occur due to variations in nutritional intake, renal function, bone dynamics, and other physiological processes.

## PHYSIOLOGICAL ACTIONS OF PTH IN BONE AND KIDNEY

### Bone Resorption and Calcium Release

The calcium-homeostatic response of bone to PTH is complex but is known to involve two principal cell types: osteoblasts and osteoclasts. Osteoblasts are the bone-building cells of marrow stromal cell origin that express PTH-1 receptors on their cell surface, whereas osteoclasts are multinucleated and mobile bone-resorbing cells of the macrophage lineage that do not express PTH-1 receptors on their cell surface. The osteoclastic cells respond to PTH through indirect mechanisms: PTH stimulates the production of RANK (receptor activator of nuclear factor- $\kappa$ B) ligand by osteoblasts. RANK ligand is a member of the tumor necrosis factor family. As a membrane-spanning protein on the surface of osteoblasts, RANK ligand binds to its receptor, RANK, on osteoclasts and osteoclast precursors to stimulate production of osteoclasts and the activity and survival of mature osteoclasts. PTH can have both acute and long-term effects on release of calcium from bone. Whereas the acute homeostatic release of calcium from bone in response to PTH may require osteoclast activity, the precise cellular and molecular mechanisms are not well understood. The longer term resorptive effects are associated with dramatic increases in osteoclast number and activity and can result in severe reductions in bone mineral density, trabecular number, and bone strength.

### Bone Formation and the Bone Anabolic Response

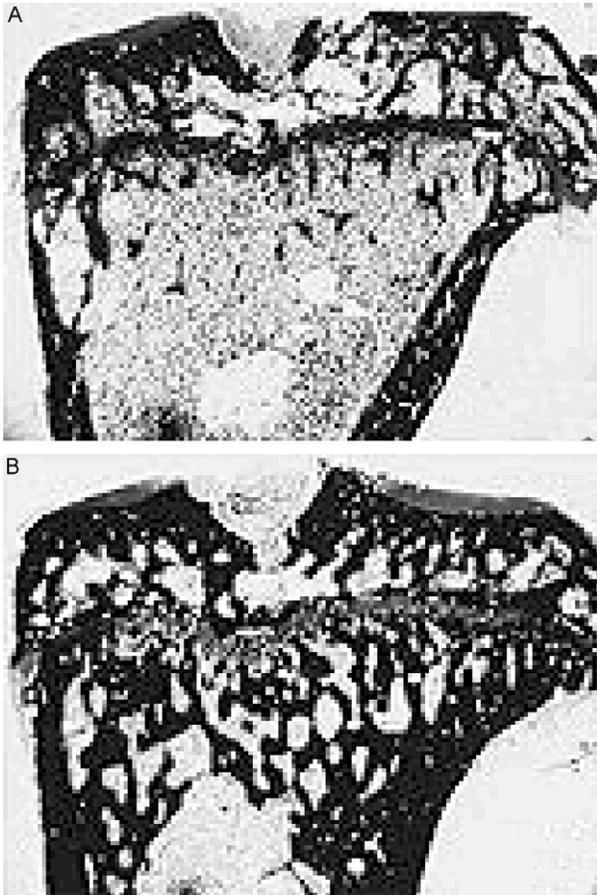
When PTH binds to the PTH-1 receptor, it not only triggers bone resorption through the RANK ligand pathway, but also increases bone formation by osteoblasts. The exact mechanisms for this activation are still poorly understood and likely to be multifactorial. PTH administration increases the number of osteoprogenitor cells, assessed with cell culture assays. PTH also decreases the apoptotic death of cells of the osteoblast lineage and may recruit renewed osteoblast activity from dormant bone-lining cells. Osteoblasts secrete insulin-like growth factor-I and transforming growth factor- $\beta$  in response to PTH and PTH-mediated bone resorption releases a number of growth

factors from bone matrix. Through these and other mechanisms, PTH increases both the number and activity of osteoblasts.

Multiple variables determine whether the bone formation response to PTH or the bone resorption response to PTH will dominate. These variables include the dose and timing of PTH administration and the location of target osteoblasts in trabecular or cortical bone. Whereas continuous overexposure of bone to PTH results in dramatic and severe reductions in bone mineral density (BMD) and structural integrity of the trabecular network, F. Albright and co-workers showed as long ago as 1929 that intermittent exposure of bone to PTH, as achieved through daily subcutaneous injections, produces significant increases in bone mass. Since this discovery, the net anabolic effect of intermittently administered PTH on bone, particularly in the trabecular compartments of long bones and vertebral bodies, has been well documented (Fig. 3). The cellular and molecular mechanisms that underlie this anabolic effect of intermittent PTH on bone are not completely understood, but it is clear that the effect involves an increase in the rate of overall bone turnover, with a net increase in osteoblast-mediated bone formation, relative to osteoclast-mediated bone resorption. The findings have suggested that PTH or a PTH analogue could be used to treat osteoporosis and several clinical trials have been conducted to evaluate this possibility. The largest trial, conducted by Neer and co-workers in 2001, was a double-blind randomized trial involving 1637 postmenopausal osteoporotic women that directly tested the effects of PTH(1-34), administered once daily by subcutaneous injections at doses of 20 or 40  $\mu\text{g}$ , on bone strength. The results showed that relative to the placebo-treated controls, PTH significantly increased BMD (by 9.7% at the lumbar spine) and significantly reduced the risk of vertebral and nonvertebral fractures (by 65 and 53%, respectively) during the 21-month study. PTH(1-34) thus may be more effective than any of the other agents used to treat osteoporosis (calcitonin, bisphosphonates, and estrogen). There is considerable interest in developing other, potentially orally active, PTH-based therapies for the treatment of this common skeletal disease.

### Effects in the Kidney

PTH acts on cells of the distal and proximal portions of the convoluted renal tubules. In response to PTH distal tubular cells mediate the retrieval of calcium from the glomerular filtrate, whereas actions on



**Figure 3** Anabolic effect of PTH in bone. Shown are proximal tibia from adult rats that were ovariectomized to induce an osteoporotic state and treated for 90 days with either (A) vehicle or (B) PTH(1-34). In the PTH-treated tibia, the mineralized surface (stained darkly) and trabecular connectivity are increased, compared to the vehicle-treated control. Reprinted from Stewart *et al.* (2000), with permission.

proximal tubular cells decrease the reabsorption of phosphate from the urine. An important longer term effect of PTH in the kidney is the stimulation of the synthesis of the vitamin D 1- $\alpha$ -hydroxylase in proximal tubule cells. This enzyme catalyzes the synthesis of 1,25-dihydroxyvitamin D<sub>3</sub> [1,25(OH)<sub>2</sub>D<sub>3</sub>], which, in turn, promotes the intestinal absorption of dietary calcium and the release of calcium from bone.

## STRUCTURE–ACTIVITY RELATIONSHIPS IN PTH

### Determinants of Receptor-Binding Affinity

PTH(1-34) contains all the structural information required for biological activity at the PTH-1 receptor;

thus, the peptide binds to this receptor with nanomolar affinity ( $K_d \sim 3$  nM) and stimulates cAMP formation in cells expressing the PTH-1 receptor with nanomolar potency ( $EC_{50} \sim 2$  nM). Structure–activity relationship studies of PTH have revealed that the principal receptor-binding and receptor-activation functions of PTH(1-34) reside largely in the C-terminal (15-34) and N-terminal (1-14) regions of the peptide, respectively. One of the shortest fragments that retains detectable receptor-binding capacity is PTH(15-34) ( $K_d \sim 10$   $\mu$ M). Within the (15-34) segment, residues Trp<sup>23</sup>, Leu<sup>24</sup>, Leu<sup>28</sup>, and Val<sup>31</sup> have been identified by substitution analysis as key determinants of receptor-binding affinity. These residues are thought to form the hydrophobic face of an amphiphilic  $\alpha$ -helix, with Arg<sup>25</sup>, Lys<sup>26</sup>, Lys<sup>27</sup>, and Asp<sup>30</sup> constituting the opposite hydrophilic face (Fig. 2). The hydrophobic face of this helix could play an important role in receptor binding by directly contacting a complementary hydrophobic surface in the receptor. An alternative mechanism, however, is that the hydrophobic surface of this helix mediates interaction with the lipid phase of the cell membrane surrounding the receptor and thereby indirectly facilitates the receptor-binding process. Since the specific mechanism by which PTH binds to its receptor is not known, either mechanism remains plausible and the two, particularly if occurring sequentially, are not mutually exclusive. Whereas the amino acid sequences of the (15-34) regions of PTH and PTHrP molecules from different species are not highly conserved, amphiphilic character is well preserved in this region of each ligand. Though the amino-terminal region of PTH cannot bind to the receptor independently, contacts between this amino-terminal region and the extracellular receptor loops and transmembrane regions of the receptor contribute importantly to the overall binding affinity of PTH to the PTH-1R.

### Determinants of cAMP Signaling

Consistent with their functional importance in activating the AC/cAMP/protein kinase A (PKA) signaling pathway, the N-terminal residues of PTH and PTHrP are highly conserved, with only two positions in the (1-14) regions differing between human and chicken PTH (Fig. 2). Deletion of the first few amino acids from PTH(1-34) results in peptides, such as PTH(3-34) and PTH(7-34), that bind to the receptor with adequate affinities ( $K_d \sim 10$  to 100 nM) but elicit little or no cAMP response. Thus, N-terminally truncated

peptides, such as PTH(3-34) and PTH(7-34), can function as competitive PTH-1 receptor antagonists. One of the most effective and widely used PTH receptor antagonists, [Leu<sup>11</sup>, D-Trp<sup>12</sup>]PTHrP(7-34)NH<sub>2</sub>, also functions as an inverse agonist against constitutively active PTH receptors that have been identified in patients with Jansen's chondrodysplasia. Val<sup>2</sup> is particularly important for receptor activation, as PTH(1-34) analogues having this residue substituted by bulkier amino acids, such as arginine and tryptophan, exhibit severely diminished cAMP-signaling potencies with near normal binding affinities and these analogues also function as PTH-1 receptor antagonists.

N-terminal PTH peptides shorter than PTH(1-28) exhibit little or no biological activity, at least in cells that express the endogenous PTH-1 receptor. In cells that overexpress the cloned PTH-1 receptor via stable or transient DNA transfection, peptides as short as PTH(1-14) exhibit at least some cAMP agonist activity, albeit with potencies that are ~10,000-fold weaker than that of PTH(1-34) ( $EC_{50} = \sim 200 \mu M$  and  $\sim 2 nM$ , respectively). Such weak potency can be attributed to the absence of the principal receptor-binding determinants in the PTH(15-34) domain. Subsequent work on the PTH(1-14) scaffold peptide has resulted in analogues, such as [ $\alpha$ -aminoisobutyric acid (Aib)<sup>1,3</sup>, Gln<sup>10</sup>, Har<sup>11</sup>, Ala<sup>12</sup>, Trp<sup>14</sup>]PTH(1-14)NH<sub>2</sub>, that are as potent as PTH(1-34) in cells overexpressing the cloned PTH-1 receptor and only ~10-fold weaker than PTH(1-34) in cells expressing lower levels of the endogenous receptor. The modifications in these analogues are thought to improve affinity and potency by either providing new side chain contacts with the receptor, as for the Leu<sup>11</sup> → homoarginine substitution, constraining the peptide in the preferred (bioactive) conformation, as for the Ala<sup>1,3</sup> → Aib substitutions, which stabilize  $\alpha$ -helical structure, or both, but the specific mechanisms have not been elucidated. Nevertheless, such minimized PTH analogues could provide clues for the design of novel low-molecular-weight PTH-1 receptor agonists (peptidic or nonpeptidic) that can be used to treat osteoporosis or other diseases of bone and mineral metabolism.

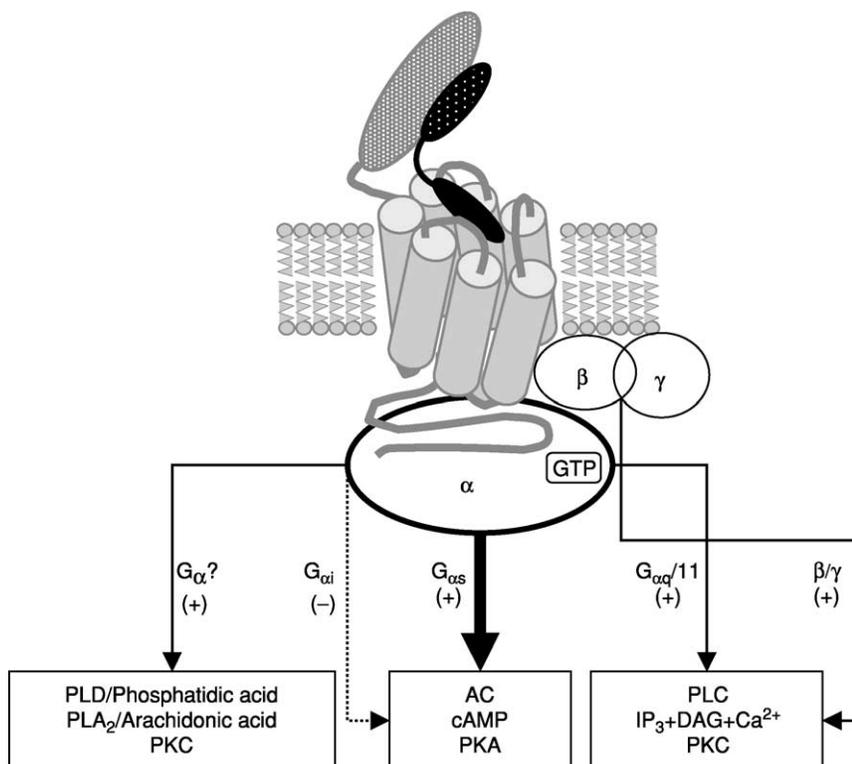
### Determinants of Non-cAMP Signaling Responses

Although in most cell-based PTH assay systems the AC/cAMP/PKA response to PTH predominates, in

certain settings and to varying extents, PTH stimulation of other intracellular signaling cascades is readily observed (Fig. 4). In many cell types, including COS-7 cells transiently transfected with the PTH-1 receptor, PTH stimulates the PLC/IP<sub>3</sub>/diacylglycerol pathway, leading to the mobilization of intracellular calcium and the activation of PKC. In certain cells, such as renal distal tubule cells, PTH stimulates the PLD/phosphatidic acid signaling response, and in renal proximal tubule cells, PTH stimulates the PLA<sub>2</sub>/arachidonic acid signaling response. The residues in PTH(1-34) that are involved in activating the various non-cAMP-mediated signaling cascades are not well defined, although residues 1 and 2 have been shown to be important for stimulating the PLC/IP<sub>3</sub> response in transfected cells, and, interestingly, residues 29-32 have been implicated as PKC activation determinants in rat osteosarcoma cells. This PKC response may involve the activation of PLA<sub>2</sub> or PLD rather than PLC. Little is also known regarding the means by which the various intracellular signaling pathways that are activated determine a given target cell's ultimate response to PTH, but this is an area that clearly warrants intense investigation.

### Three-Dimensional Structure of PTH

The three-dimensional structure of PTH(1-34) has been analyzed by solution-phase nuclear magnetic resonance (NMR) spectroscopy, as well as by X-ray crystallography. The NMR studies, performed in a variety of solvents, including aqueous buffers, buffers containing the helix-promoting solvent, trifluoroethanol, and phospholipid micelles, have generally indicated the presence of two  $\alpha$ -helical segments: a short N-terminal helix extending approximately from Ser<sup>3</sup> to Leu<sup>11</sup> and a longer C-terminal helix extending approximately from residues Asn<sup>16</sup> to Asp<sup>30</sup>. A flexible hinge or turn region generally separates the two helices. It has been proposed based on functional studies and theoretical considerations that PTH(1-34) forms a U-shaped structure with the N- and C-terminal domains interacting. Hydrophobic interactions between Leu<sup>15</sup> and Trp<sup>25</sup> have been observed by NMR methods, but otherwise intramolecular contacts between the N-terminal and C-terminal domains of PTH(1-34) are generally not seen. The crystal structure of PTH(1-34) was reported in 2001 by researchers at Eli Lilly and this study revealed a single linear  $\alpha$ -helix that extended nearly from one end of the molecule to the other (Val<sup>2</sup> to His<sup>32</sup>) and contained only a slight (15°) bend



**Figure 4** The PTH/PTH receptor/G protein complex. The interaction of PTH with the PTH-1 receptor involves binding interactions between residues in the (15–34) region of the ligand (small, stippled oval) and the receptor's amino-terminal extracellular domain (large, stippled oval) and signaling interactions between residues in the (1–14) domain of PTH (small, solid oval) and the portion of the receptor containing the three extracellular loops and seven transmembrane helices (cylinders). The ligand-activated PTH-1 receptor couples to multiple signaling pathways, depending on the cell type and the presence or absence of ancillary cytoplasmic proteins, such as the  $\text{Na}^+/\text{H}^+$  exchanger regulatory factors (NHERFs), as shown by M. Mahon and colleagues. In most cells, the PTH-1 receptor couples strongly (as indicated by the heavy arrow) to the AC/cAMP/PKC pathway via  $G_{\alpha s}$ . It also couples to the PLC/IP<sub>3</sub>/Ca<sup>2+</sup>/PKC pathway via  $G_{\alpha q/11}$ , the PLD/phosphatidic acid/PKC pathway, and the PLA<sub>2</sub>/arachidonic acid/PKC signaling pathway, the latter two pathways involving as yet unidentified G proteins. In certain cells and in the presence of NHERE, the PTH-1 receptor couples to the inhibitory G protein ( $G_{\alpha i}$ ), which results in diminished AC/cAMP signaling, as well as increased PLC/IP<sub>3</sub> signaling, an effect possibly mediated by the  $\beta/\gamma$ -subunits that dissociate from the activated G protein heterotrimer. Positive effects are indicated by a plus sign and solid arrows; a negative effect is indicated by a minus sign and a dashed arrow.

at the midsection. This finding raises the somewhat unexpected possibility that the biologically relevant structure of PTH(1-34) is a continuous  $\alpha$ -helix. The increased potencies seen in PTH analogues containing helix-stabilizing substitutions, such as the Aib substitutions at positions 1 and 3 in PTH(1-14), and side-chain to side-chain lactam bridges between certain pairs of  $i, i+4$  residues (e.g., Lys<sup>13</sup>-Asp<sup>17</sup>, Lys<sup>18</sup>-Asp<sup>22</sup>, Glu<sup>22</sup>-Lys<sup>26</sup>, and Lys<sup>26</sup>-Asp<sup>30</sup>) in PTH(1-31) or PTH(1-34) analogues, support the notion that the N-terminal and C-terminal domains of PTH are  $\alpha$ -helical when the ligand is bound to the receptor. Such studies, however, do not elucidate the tertiary structure of the bound ligand. The question of whether PTH adopts a linear or folded conformation when

binding to the receptor, therefore, remains a matter of uncertainty and debate in the literature.

## PTH-PTH RECEPTOR INTERACTIONS

### Molecular Properties of the PTH-1 Receptor

The primary structure of the PTH-1 receptor was first determined by Jüppner, Abu-Samra, Segre, and colleagues at Massachusetts General Hospital, who isolated cDNA clones encoding the receptor from both kidney and bone cell lines. The human PTH-1 receptor is 593 amino acids in length. Its specific three-dimensional structure is largely unknown, although it

is presumed to follow the general seven-transmembrane domain architecture common to all GPCRs. The amino-terminal extracellular (N) domain of the PTH-1 receptor is approximately 170 amino acids in length and contains four N-linked glycosylation sites and six disulfide-linked cysteine residues. The PTH-1 receptor shares amino acid sequence homology to the receptors for a number of other peptide hormones that are similar in size to PTH(1-34), including calcitonin, secretin, glucagon, vasoactive intestinal polypeptide, growth hormone-releasing hormone, corticotropin-releasing hormone, and several others. These structurally related receptors form the class B subgroup of GPCRs. The PTH-1 receptor mediates the biological actions of both PTH and PTHrP and couples strongly to heterotrimeric G proteins containing the stimulatory variant of  $\alpha$ -subunit ( $G_{\alpha s}$ ) to induce adenylyl cyclase-mediated cAMP signaling and it can also couple to  $G_{\alpha q/11}$ -containing G proteins to induce  $IP_3/Ca_i^{2+}$  signaling, as well as to  $G_{\alpha i}$ -containing G proteins to inhibit adenylyl cyclase activity (Fig. 4).

### TIP39 and the PTH-2 Receptor

Usdin and colleagues at the National Institutes of Health screened a human brain cDNA library for family B-type GPCRs by hybridization methods and thereby isolated a receptor that is 51% identical, at the amino acid level, to the PTH-1 receptor. This PTH-2 receptor responds to PTH but not to PTHrP. The intended ligand for the PTH-2 receptor, however, is most likely not PTH, because the rat PTH-2 receptor responds to neither PTH nor PTHrP. Instead, the intended ligand is likely to be tuberoinfundibular peptide of 39 amino acids (TIP39). This previously unknown peptide was isolated from bovine hypothalamus by Usdin and colleagues and shown to potently activate the PTH-2 receptor (rat or human) but not the PTH-1 receptor. TIP39 bears faint homology to PTH(1-34) (Fig. 2). Although the biological role of this ligand/receptor pair is unknown, Usdin and colleagues have presented evidence suggesting that it plays a role in pain nociception.

### Mode of PTH Binding to the PTH-1 Receptor

The molecular mechanism by which PTH binds to and activates the PTH-1 receptor has been deduced largely from receptor mutagenesis and photochemical cross-linking studies. These data suggest that the interaction is complex and involves multiple points

of contact, with two principal components, as postulated by the "two-domain hypothesis." This involves (1) an interaction between the carboxy-terminal domain of PTH(1-34) and the N domain of the receptor, which provides the majority of binding energy to the complex, and (2) an interaction between the amino-terminal domain of the ligand and the juxta-membrane region of the receptor containing the extracellular loops and seven-transmembrane helices (TMs), which provides the specific contacts required for receptor activation (Fig. 4). Presumably, binding of the ligand to the receptor results in a change in the arrangement of the intracellular loops and helices that results in activation of G proteins and interaction with other intracellular constituents.

## DISEASES OF PTH METABOLISM AND ACTION

Increased secretion of PTH may be an appropriate response to hypocalcemia (secondary hyperparathyroidism) or due to primary abnormalities of the parathyroid glands, which lead to hypercalcemia (primary hyperparathyroidism). Primary hyperparathyroidism is caused by hyperplasia or adenoma affecting one or more of the four parathyroid glands. Most commonly, mutations in parathyroid chief cells lead to clonal expansion of parathyroid cells that do not suppress PTH secretion in response to hypercalcemia as well as normal cells. Inactivating mutations of the CSR cause mild parathyroid hyperplasia when present in one of the gene's alleles. The hypercalcemia in this disorder (familial hypocalciuric hypercalcemia) is caused both by the mild increase in PTH secretion and by the PTH-independent actions of the CSR to increase reabsorption of calcium from the urine. Secondary hyperparathyroidism is caused by renal failure, malabsorption of calcium from the intestine, or other causes of hypocalcemia.

Reduced PTH secretion resulting in hypocalcemia reflects impaired parathyroid function and may be caused by destruction of parathyroid tissue by autoimmune or other processes, by complex congenital disorders, such as the DiGeorge syndrome, or by disorders of PTH secretion caused by hypomagnesemia or activating mutations of the CSR.

Pseudo-hypoparathyroidism (PHP) results from defects not in PTH production but in PTH responsiveness and is of two principal types: PHP1a, which is accompanied by Albright's hereditary osteodystrophy (AHO) and other endocrine system defects, and PHP1b, which is not associated with AHO or other

endocrine defects. PHPIa is caused by inactivating mutations in  $G_{\alpha s}$ ; the gene for PHPIb also maps to the  $G_{\alpha s}$  locus, but specific mutations have not yet been identified.

Genetic abnormalities of the PTH1-R cause diseases that reflect the dual function of the PTH1-R as both a PTH receptor and a PTHrP receptor. Jansen's metaphyseal chondrodysplasia is a rare autosomal dominant disease characterized by dwarfism (reflecting PTHrP-like actions at the growth plates of bones) and hypercalcemia (reflecting PTH-like actions on bone and kidney) and is caused by activating point mutations in the PTH1-R that increase the rate of basal cAMP signaling. Four such mutations have been identified, each of which occurs at the intracellular end of a predicted transmembrane domain segment: His<sup>223</sup> → Arg in TM2; Thr<sup>410</sup> → Pro or Arg in TM6 and Ile<sup>458</sup> → Arg in TM7. The antagonist analogue [Leu<sup>11</sup>, D-Trp<sup>12</sup>]PTHrP(7-34)NH<sub>2</sub> has been shown by *in vitro* studies to function as an inverse agonist against these constitutively active PTH-1 receptors and to thus reduce their basal rate of cAMP signaling. Blomstrand's chondrodysplasia is a rare embryonic lethal condition caused by inactivating mutations in the PTH-1 receptor. The mutations result in a severely overcalcified fetal skeleton with extremely short limbs—a phenotype that closely resembles that seen in “knockout” mice having homozygous deletions of the PTH-1 receptor gene. Three mutations have been identified: a single nucleotide change that results in an 11-amino-acid deletion in TM5 due to an mRNA splicing error, a frameshift mutation in extracellular loop 2, and a missense mutation that changes Pro<sup>132</sup> in the receptor's N domain to leucine.

## CONCLUSIONS

The critical role that PTH plays in regulating the concentrations of extracellular calcium involve direct but complex effects on bone and kidney target cells. These effects are mediated through the PTH-1 receptor, a G protein-coupled receptor, that also regulates the developmental actions of PTHrP. PTH holds promise for the treatment of osteoporosis, since it increases bone mass and strength when administered intermittently. PTH interacts with its receptor through a two-site mechanism that involves interactions between the (15-34)-binding domain of the

ligand and the amino-terminal extracellular domain of the receptor and between the amino-terminal (1-14) signaling portion of the ligand and the juxta-membrane portion of the receptor. Understanding this molecular mechanism further should provide a better appreciation of how alterations in PTH physiology can be interpreted and manipulated in the clinical setting.

## See Also the Following Articles

Bone Turnover Markers • G Protein-Coupled Receptors • Parathyroid Cancer • Parathyroid Glands, Pathology • Parathyroid Surgery • Pseudohypoparathyroid States

## Further Reading

- Brown, E. M., and MacLeod, R. J. (2001). Extracellular calcium sensing and extracellular calcium signaling. *Physiol. Rev.* **81**, 239–297.
- Gardella, T. J., and Jüppner, H. (2001). Molecular properties of the PTH/PTHrP receptor. *Trends Endocrinol. Metab.* **12**, 210–217.
- Hoare, S. R. J., and Usdin, T. B. (2001). Molecular mechanisms of ligand-recognition by parathyroid hormone 1 (PTH1) and PTH2 receptors. *Curr. Pharm. Design* **7**, 689–713.
- Jüppner, H. W., Gardella, T. J., Kronenberg, H. M., and Potts, J. T., Jr. (2001). Parathyroid hormone and parathyroid hormone-related peptide in the regulation of calcium homeostasis and bone regulation. In “Endocrinology” (L. DeGroot and J. Jameson, eds.), Vol. 2, pp. 969–1053. W. B. Saunders, Philadelphia, PA.
- Khosla, S. (2001). Minireview: The OPG/RANKL/RANK system. *Endocrinology* **142**, 5050–5055.
- Kronenberg, H. M., Lanske, B., Kovacs, C. S., Chung, U. I., Lee, K., Segre, G. V., Schipani, E., and Jüppner, H. (1998). Functional analysis of the PTH/PTHrP network of ligands and receptors. *Rec. Prog. Horm. Res.* **53**, 283–301.
- Mahon, M. J., Donowitz, M., Yun, C. C., and Segre, G. V. (2002). Na(+)/H(+) exchanger regulatory factor 2 directs parathyroid hormone 1 receptor signaling. *Nature* **417**, 858–861.
- Morley, P., Whitfield, J. F., and Willick, G. E. (2001). Parathyroid hormone: An anabolic treatment for osteoporosis. *Curr. Pharm. Design* **7**, 671–687.
- Neer, R. M., Arnaud, C., Zanchetta, J. R., Prince, R., Gaich, G. A., Reginster, J. Y., Hodsman, A. B., Eriksen, E. F., Ish-Shalom, S., Genant, H. K., *et al.* (2001). Effect of parathyroid hormone (1-34) on fractures and bone mineral density in postmenopausal women with osteoporosis. *N. Engl. J. Med.* **344**, 1434–1441.
- Stewart, A. F., Cain, R. L., Burr, D. B., Jacob, D., Turner, C. H., and Hock, J. M. (2000). Six month daily administration of parathyroid hormone and parathyroid hormone related protein peptides to adult ovariectomized rats markedly enhances bone mass and biomechanical properties: A comparison of human parathyroid hormone 1-34, parathyroid hormone related protein 1-36, and SDZ parathyroid hormone 893. *J. Bone Miner. Res.* **15**, 1517–1525.

Unfortunately, both men succumbed to their disease and the consequences of its treatment. Solomon Berson and Rosalyn Yalow won the Nobel Prize in 1963 for developing an immunoassay for the measurement of parathyroid hormone (PTH). Adolf Hanson, a medical student in Minnesota, and James Collip, a Canadian biochemist, separately developed parathyroid extract to treat tetany, the symptoms of hypoparathyroidism and osteoporosis.

The introduction of the serum channel autoanalyzer in the mid-1960s issued in a new era of parathyroid surgery in that it facilitated making the diagnosis of primary hyperparathyroidism. There was a precipitous increase in the incidence of the disease and asymptomatic patients became the rule, rather than the exception. Additional technical advances have included the following: improved preoperative localization with sestamibi [ $\pm$  single photon emission computed tomography (SPECT)], intraoperative localization with the gamma probe, the rapid intraoperative parathyroid hormone (intraop PTH) assay, and the use of minimally invasive parathyroidectomy with unilateral neck exploration employing a small incision and regional anesthesia in the ambulatory setting.

## PRIMARY HYPERPARATHYROIDISM

### Anatomy and Pathology

Most humans have four parathyroid glands, which lie on the posterior surface of the thyroid. The average weight of a parathyroid is 35 to 40 mg and in adults their color turns to yellow as the fat content increases. The inferior parathyroids originate from the third branchial pouch, whereas the superior parathyroid glands descend from the fourth branchial pouch. Both the superior and inferior parathyroid glands receive their blood supply from the inferior thyroid artery. They are made up of chief and oxyphil cells, as

well as fibrovascular stroma and adipose tissue. Primary hyperparathyroidism (HPTH) can be produced by three different pathologic lesions (Fig. 1). A parathyroid adenoma is a benign encapsulated neoplasm that is responsible for 80–90% of hyperparathyroidism cases (Fig. 2). It usually affects a single gland, but 2–5% of patients with primary HPTH have adenomas in two glands (“double adenoma”). Hyperplasia is a proliferation of parenchymal cells that affects all of the parathyroid glands, and it accounts for approximately 10% of cases of primary HPTH. It also is associated with multiple endocrine neoplasia (MEN) 1 (primary HPTH combined with lesions of the pancreas and pituitary) and 2A (primary HPTH, medullary thyroid cancer, and pheochromocytoma) syndromes. Surgical treatment for parathyroid hyperplasia is with either subtotal parathyroidectomy or total parathyroidectomy with autotransplantation. Thymectomy also should be performed for patients with the MEN syndromes. Parathyroid carcinoma is a slow-growing neoplasm of the parenchymal cells that is responsible for 0.5–2% of cases of primary HPTH. Unlike adenomas, carcinoma is invasive.

### Diagnosis and Presentation

Primary HPTH is the third most common endocrinologic disorder, after diabetes mellitus and thyroid disease. It is characterized by hypersecretion of PTH from one or more parathyroid glands, leading to hypercalcemia. The diagnosis is made by demonstrating an elevated serum calcium, elevated intact PTH (iPTH) level, and normal or increased urinary calcium concentration. The 24 h urine collection can help to exclude the diagnosis of benign familial hypocalciuric hypercalcemia (BFHH), which results in increased blood calcium and iPTH levels but *low* urinary calcium.

Prior to the advent of the automated serum channel autoanalyzer, patients typically presented with the

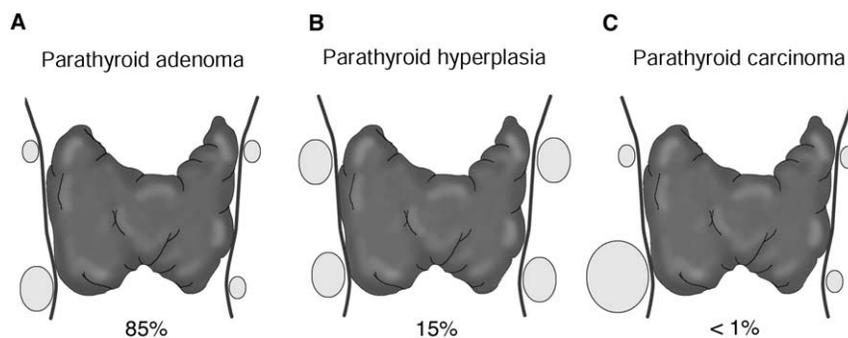
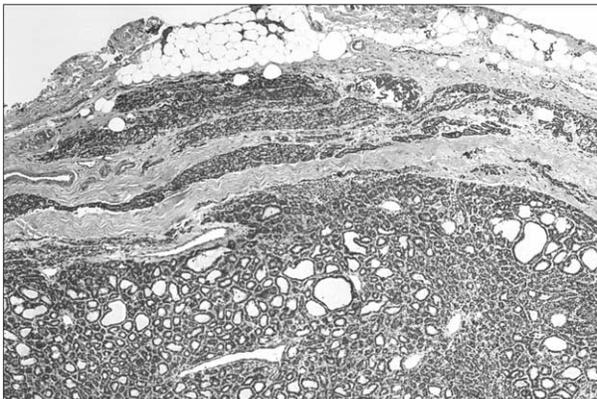


Figure 1 Causes of primary HPTH and their prevalence.



**Figure 2** Parathyroid adenoma. A rim of normocellular parathyroid tissue with adipose surrounds a proliferation of chief and oncocytic cells below.

clinical manifestations of hypercalcemia, which included the classic pentad of painful bones, kidney stones, abdominal groans, psychic moans, and fatigue overtones. Since the autoanalyzer has become available, however, the biochemical diagnosis is usually made prior to the appearance of symptoms, and many patients are asymptomatic or at most minimally symptomatic. Although there was a National Institutes of Health consensus conference in 1990 and another workshop in 2002 about the management of asymptomatic primary HPTH, there is still no consensus among endocrinologists and endocrine surgeons about whether to survey these patients with nonoperative medical therapy or to refer them for early surgery. The value of parathyroidectomy in asymptomatic patients with mild to moderate hypercalcemia has been debated, because the natural history of primary HPTH is still not well understood and rapid increases in the serum calcium level and progression of symptoms or complications are uncommon in patients with borderline hypercalcemia.

## Localization Procedures

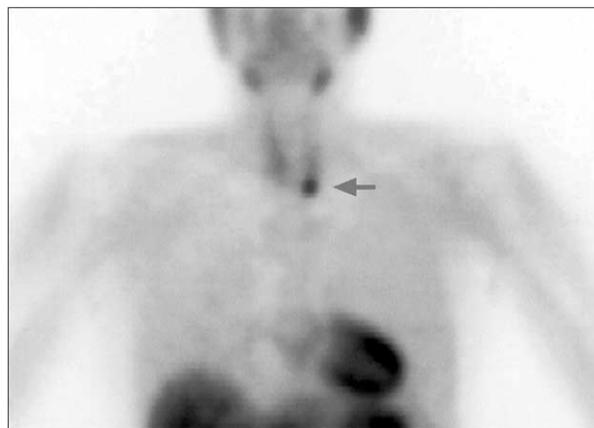
### Preoperative

There is a consensus that preoperative localization is useful prior to primary exploration, since it can allow for unilateral exploration in the setting of a localizing scan. Localization is absolutely essential prior to reoperation. A number of noninvasive preoperative methods exist, including sestamibi technetium (Tc)-99 m scintigraphy, ultrasonography, computed tomography (CT), magnetic resonance imaging (MRI), and thallium (Tl)-201–Tc-99 m pertechnetate scanning. There is general consensus that the single best

study is sestamibi (especially when combined with SPECT), which works by virtue of the fact that mitochondria take up Tc-99 m, and parathyroid cells have a large number of mitochondria (Fig. 3). The sensitivity reported for adenomas is in the 85–95% range, with a specificity of approximately 90%. Its sensitivity is limited in multiglandular disease. SPECT is particularly helpful in detecting smaller lesions and adenomas located behind the thyroid gland.

Ultrasound is easy to perform, well-tolerated, and can be done quickly and inexpensively. However, its accuracy depends on the skill and experience of the ultrasonographer, with sensitivity varying from 22 to 82%. CT is less sensitive than MRI, is relatively expensive, and exposes the patient to radiation. MRI is preferable in that it does not require the administration of contrast, but it is expensive; its sensitivity ranges from 50 to 80%. Tl-Tc uses subtraction imaging dependent on the uptake of thallium by parathyroid tissue. Parathyroids weighing less than 500 mg are usually not detected, and sensitivity ranges dramatically from center to center, with rates as low as 27% reported.

Invasive preoperative localization procedures include parathyroid arteriography, selective venous sampling, and fine-needle aspiration (FNA); these techniques are usually reserved for the reoperative neck. Arteriography, which has a sensitivity of 60%, includes examination of the thyrocervical trunks, the internal mammary arteries, and the carotids to look for parathyroid adenomas, which appear as highly vascularized, oval blushes. Venous sampling lateralizes approximately 80% of tumors by obtaining a blood sample for the PTH assay from the smallest possible



**Figure 3** Sestamibi scan localizing a left-sided, inferior parathyroid adenoma (arrow) in the neck of a patient who subsequently underwent a MIP. Note the physiologic tracer uptake in the heart.

branches of the cervical and mediastinal veins. A gradient between the PTH concentration in peripheral blood and that in the selective venous sample establishes the site of venous drainage of the tumor. FNA is usually combined with ultrasound and can confirm that a cervical mass is indeed a parathyroid gland.

### **Intraoperative**

The rapid intraoperative PTH (quick PTH) assay can be used to confirm the removal of hypersecreting parathyroid glands and predict a curative procedure. Its use is associated with reduced operating time. The first reported use of the assay was in 1988, but it has been refined significantly since then, largely due to the work of George Irvin. A peripheral blood specimen should be obtained prior to surgery. The equipment and technician to run the assay can be stationed inside or just outside the operating room. Repeated blood specimens from a peripheral line (better) or ipsilateral internal jugular vein should be drawn intraoperatively prior to resection of the gland (in order to capture a spike from the preoperative baseline due to manipulation of the gland during exploration) and then approximately 5 and 10 min after the excision. A 50% reduction in the PTH value from the baseline level is used often as an indication that the exploration has been successful. Operative failure rates for initial and reoperative parathyroidectomy have decreased significantly with this adjunct.

Radio-guided parathyroidectomy employs additional intraoperative localization of the adenoma following intravenous injection of sestamibi technetium-99m 2 to 4 h prior to surgery and a quantitative gamma counter with a 9 to 14 mm probe in the operating room. Gamma counts are obtained at the start of surgery in all four quadrants of the neck, through the skin and after incision under the muscles, with care taken not to interpret radioactivity emitted by the heart. Exploration where counts are highest focuses surgery and reduces operative time. The activity of the removed parathyroid gland is checked with the gamma probe to confirm cure. The excised adenoma emits radioactivity that is at least 20%, and usually higher than 50%, of postexcision background. In addition, the postexcision radioactivity in all four quadrants of the neck should equalize.

## **Surgery**

### **Bilateral Neck Exploration**

The standard operative approach to the treatment of hyperparathyroidism is to explore both sides of the

neck and to identify all parathyroid tissue. This is usually performed under general anesthesia, although it can be performed under regional block. Bilateral neck exploration allows accurate intraoperative diagnosis of the pathology causing the patient's disease and individualized treatment based on the number of enlarged glands. Mediastinotomy should not be performed at the primary exploration, but rather should be reserved if necessary for reexploration after further localization studies have been performed.

Parathyroidectomy offers definitive treatment for primary HPTH and the cure rate (defined as normocalcemia 6 months postoperatively) after bilateral neck exploration performed by an experienced surgeon is greater than 95%. There is clear evidence that clinical outcomes are associated with the experience of the surgeon performing the parathyroidectomy, such that high-volume endocrine surgeons have higher cure rates and lower complication rates. The rate of persistent hyperparathyroidism can be as high as 30% in less experienced hands. Perioperative complications include injury to the recurrent laryngeal nerve(s), leading to hoarseness or frank airway compromise if both nerves are damaged (reported rates of <1% to as high as 10%), hypoparathyroidism (2–3%), hematoma, and wound infection. The risks of these are theoretically less when exploration is confined to one side of the neck.

### **Minimally Invasive Parathyroidectomy**

Eighty-five percent of primary HPTH results from a single adenoma, and it is cured by excision of that one gland. Therefore, with better preoperative localization, targeted surgery using unilateral neck exploration under regional or local anesthesia is becoming standard care. MIP involves preoperative localization (by, for example, sestamibi with SPECT) followed by limited exploration with cervical block anesthesia, using the intraoperative PTH assay to confirm the adequacy of resection. Patients with known multigland hyperplasia are not offered MIP. However, if such a patient is encountered during a MIP, bilateral exploration can be accomplished with the technique or the procedure can be converted to general anesthesia.

The skin incision is typically limited to 1 to 4 cm. A superficial cervical block is administered on the ipsilateral side of the sestamibi-localized adenoma. In most patients, approximately 20 cc of 1% lidocaine containing 1:100,000 epinephrine is used and supplemented during surgery as required. It is delivered deep to the posterior border of the sternal head of the sternocleidomastoid muscle on the ipsilateral side

of the adenoma, with care taken not to deliver the anesthetic intravascularly. Propofol is discontinued at least 5 min before PTH sampling because it can interfere with the PTH assay. Sedation is used to minimize patient anxiety while maintaining an awake, conscious patient who can phonate. A focused exploration then is performed based on the preoperative localization study, and the intraoperative PTH assay is used to confirm adequacy of parathyroid gland excision in the operating room as described above. The success of MIP has been confirmed by evidence of cure and complication rates that are at least as good as those achieved by conventional bilateral exploration. Specifically, in a series of 656 parathyroidectomies (of which 401 were performed in the standard fashion and 255 were performed with MIP), there were no significant differences in complication (3.0 and 1.2%, respectively) or cure rates (97 and 99%, respectively).

The first endoscopic parathyroidectomy was reported by Michel Gagner in 1996. Since then, MIP has been modified in a number of ways. In addition to radio-guided parathyroidectomy (described above), endoscopic and video-assisted parathyroidectomy are other variations on the minimally invasive approach. In endoscopic parathyroidectomy, access is obtained at the sternal notch for the endoscope and (generally) two lateral port sites are created for dissecting instruments. Carbon dioxide is insufflated at low pressure (<8 mm Hg) for visualization, but there can be problems when it diffuses in the neck and mediastinum, when operative space is lost during suction, and because there is no tactile assessment during surgery.

Video-assisted parathyroidectomy, pioneered by Paolo Miccoli, does not require steady gas flow but rather a brief insufflation of carbon dioxide to establish the operative space. It requires a 15 mm skin incision 1 cm above the sternal notch, which accommodates tactile assessment, suction-irrigation, dissection, and retraction equipment. A separate 10 mm trochar site is made vertically in the midline to accommodate the insufflator at the start of the case and then a 30° 5 mm endoscope. The largest series showed that mean operative time was less than 1 h. Video-assisted parathyroidectomy has been modified by Brunt, such that it can be performed as a gasless procedure.

### **Reexploration**

Up to 10% of surgical failures may be attributed to an incorrect diagnosis at the time of primary exploration. Whereas multiple-gland disease accounts for 5–15% of patients undergoing initial exploration, up to 37% of patients who go to reexploration have multiple

gland disease rather than a single adenoma. Still, inexperience on the part of the surgeon is a major cause of surgical failure because of the lack of knowledge of the usual and unusual “hiding places” of the parathyroids.

Once the diagnosis of persistent or recurrent primary HPTH has been confirmed, previous operative notes and pathology reports should be obtained and reviewed. There are clear-cut indications for preoperative localizing studies before secondary exploration. Reexploration generally returns the surgeon to the neck, unless there is compelling evidence in the localization workup that the missing gland is located in the mediastinum, because the neck remains by far the most common site of “missed” glands. Furthermore, the majority of ectopic mediastinal parathyroid glands can be extirpated by a cervical approach. The lateral approach should be considered, as it avoids midline scar tissue. If the cervical approach to the mediastinum is inadequate, a partial or complete sternal split can be employed. Intraoperative localization in the form of intraoperative PTH and/or ultrasound is helpful during reoperation.

In the largest series, cure of hypercalcemia at reoperation ranges from 85 to 95%, and it is associated with the experience of the endocrine surgeon. Risks include failure, permanent hypoparathyroidism (10–15%), and recurrent laryngeal nerve injury.

### **Medical Alternatives**

The optimal clinical management of patients with asymptomatic disease has not yet been established. The principal debate is whether patients should be managed with early operation or whether surveillance or medical therapy can be employed safely until the development of clinical sequelae.

Recommendations for referral to surgery were reported at the Workshop on Asymptomatic Primary Hyperparathyroidism: A Perspective for the 21st Century, held at the National Institutes of Health on April 8–9, 2002. They include the following:

- Patients with serum calcium levels greater than 1 mg/dl above the upper limits of normal should be referred for surgery.
- Patients with 24 h urinary calcium greater than 400 mg should be referred for surgery.
- Patients with a creatinine clearance reduced by more than 30% compared with age-matched subjects should be referred for surgery.
- Patients with a bone density measurement at the lumbar spine, hip, or distal radius that is more than

2 standard deviations below peak bone mass ( $T$  score  $< -2.5$ ) should be referred for surgery.

- All individuals with primary HPTH under the age of 50 should be referred for surgery.
- Patients for whom medical surveillance was either undesirable or impossible should be referred for surgery.

There are still no medical therapies for which there are convincing data regarding either their efficacy or their safety. As specific therapy for the hypercalcemia of primary HPTH, estrogen might reduce the serum calcium somewhat, but is not generally effective until levels of estrogen are reached that are higher than most women or their physicians would consider acceptable. Three new classes of agents—raloxifene, bisphosphonates, and calcimimetics—have shown preliminary efficacy on surrogate markers of severity of disease, including serum calcium and bone density, but these effects have not been verified based on clinical outcomes. There is no rationale for diets that are restricted in calcium; vitamin D sufficiency should be maintained.

## SECONDARY AND TERTIARY HYPERPARATHYROIDISM

### Pathology and Clinical Presentation

Secondary HPTH occurs when there are external factors, such as chronic renal failure, that stimulate the parathyroid glands to increase the production of PTH, leading to hyperplastic overgrowth. Renal failure leads to hyperphosphatemia and decreased renal conversion of 25-hydroxycholecalciferol to 1,25-dihydroxycholecalciferol, resulting in diminished intestinal calcium absorption. Both of these effects lead to hypocalcemia, which stimulates PTH secretion and parathyroid gland hyperplasia. Approximately 90% of patients with chronic renal failure have evidence of secondary HPTH.

With prolonged stimulation of the parathyroids, tertiary HPTH may develop. This occurs in patients with longstanding renal insufficiency, when the parathyroids develop autonomous hyperfunction and no longer respond to calcium feedback inhibition, resulting in hypercalcemia. It also occurs in patients with longstanding secondary HPTH following renal transplantation.

Clinical manifestations include persistent or worsening of skeletal symptoms (including the classic osteitis fibrosa cystica), pruritis, extraskelatal calcifications, and calciphylaxis. In most patients, the condition

is managed by reducing phosphate in the diet, administering phosphate binders, dialysis, or treatment with 1,25-dihydroxyvitamin D<sub>3</sub>. Kidney transplantation remains the best option for correction of the metabolic abnormalities.

### Subtotal Versus Total Parathyroidectomy

Parathyroidectomy is indicated when medical therapy fails to control progressive secondary HPTH. The two accepted operations for the management of secondary HPTH are subtotal parathyroidectomy and total parathyroidectomy with heterotopic parathyroid autotransplantation. In subtotal parathyroidectomy, “three and a half” glands are removed and a 40 to 60 mg remnant of the most normal-appearing gland is left in place and marked with a clip. Its theoretical advantage is that it induces less postoperative hypocalcemia because the parathyroid remnant left *in situ* continues to function. This can be particularly important in non-compliant patients who may not take calcium supplementation. The main disadvantage is that remedial cervical operations for recurrent disease are tedious and carry an increased risk of recurrent laryngeal nerve injury. According to the largest series, 10–16% of patients had postoperative hypercalcemia, 8% required reoperation because of remnant growth, and 4–25% had hypocalcemia lasting longer than 1 year.

In total parathyroidectomy with autotransplantation, at least four glands are resected, and the most suitable gland is selected to obtain the autograft. A 40 to 60 mg portion of this gland is sliced into small fragments, and 10 to 15 fragments are placed into separate intramuscular pockets in the forearm (away from present or planned arteriovenous fistula/graft sites). The advantage is that if hyperparathyroidism recurs in the transplanted parathyroid tissue, the graft can be partially resected under local anesthesia. Published series show a 5–38% rate of postoperative hypercalcemia, a 2–6% chance of recurrence requiring graft resection, and a 5–30% chance of hypocalcemia lasting greater than 1 year. Cryotherapy of parathyroid tissue is an important adjunct when autotransplantation is used in the case of parathyroid transplant failure.

## PARATHYROID CANCER

Parathyroid carcinoma accounts for approximately 1% of primary HPTH. Preoperative findings suggestive of the diagnosis include very high calcium and iPTH levels, a large cervical mass, and intraoperatively a

hard grayish-white parathyroid gland that is adherent to surrounding tissue. Still, it may be very difficult to make this diagnosis intraoperatively; in 20% of primary operations in one series, the diagnosis of carcinoma was not expected either preoperatively or during the procedure.

The largest published series included 286 patients collected over a 10-year period and showed a 5-year survival rate of 85.5% with a 10-year survival rate of 49.1%. Like most endocrine tumors, parathyroid carcinoma is slow-growing, and distant metastases occur late in the progression of the disease. To avoid local recurrence, initial *en bloc* resection of parathyroid carcinoma (including any locally invaded tissue such as a contiguous ipsilateral thyroid lobe) is recommended when the diagnosis is suspected at the time of parathyroidectomy. Capsular rupture must be avoided. An aggressive approach to recurrent disease is often beneficial, and surgical excision of metastases is the single most effective treatment in palliating the accompanying hypercalcemia.

### See Also the Following Articles

Hyperparathyroidism, Primary • Parathyroid Cancer • Parathyroid Glands, Pathology • Parathyroid Hormone (PTH) • Pituitary Tumors, Surgery

### Further Reading

- Bilezikian, J. P. (2000). Primary hyperparathyroidism: When to observe and when to operate. *Endocrinol. Metab. Clin. N. Am.* **29**, 465–478.
- Clark, O. H., and Duh, Q.-Y. (1997). “Textbook of Endocrine Surgery.” W. B. Saunders, Philadelphia, PA.
- Dackiw, A. P., Sussman, J. J., Fritsche, H. A., Jr., *et al.* (2000). Relative contributions of technetium Tc 99m sestamibi scintigraphy, intraoperative gamma probe detection, and the rapid parathyroid hormone assay to the surgical management of hyperparathyroidism. *Arch. Surg.* **135**, 550–555.
- Doherty, G. M., and Skogseid, B. (2001). “Surgical Endocrinology.” Lippincott, Williams & Wilkins, Philadelphia, PA.
- Gauger, P. G., and Thompson, N. W. (2001). Early surgical intervention and strategy in patients with multiple endocrine neoplasia type 1. *Best Pract. Res. Clin. Endocrinol. Metab.* **15**, 213–223.
- Howe, J. R. (2000). Minimally invasive parathyroid surgery. *Surg. Clin. N. Am.* **80**, 1399–1426.
- Kebebew, E. (2001). Parathyroid carcinoma. *Curr. Treat. Opt. Oncol.* **2**, 347–354.
- Micoli, P., and Berti, P. (2001). Minimally invasive parathyroid surgery. *Best Pract. Res. Clin. Endocrinol. Metab.* **15**, 139–147.
- Organ, C. H. (2000). The history of parathyroid surgery, 1850–1996: The Excelsior Surgical Society 1998 Edward D. Churchill Lecture. *J. Am. Coll. Surg.* **191**, 284–299.
- Udelsman, R. (2001). Primary hyperparathyroidism. *Curr. Treat. Opt. Oncol.* **2**, 365–372.
- Van Heerden, J. A., Farley, D. R., and Udelsman, R. (1999). “Operative Techniques in General Surgery: Surgical Exploration for Hyperparathyroidism,” Vol. 1, No. 1. W. B. Saunders, Philadelphia, PA.

material varies from 36 to 90%. The pattern of adrenal gland involvement—with the medulla being the site of primary involvement and with areas of focal and generalized hemorrhage affecting both medulla and cortex—initially described in 1983 by Reichert has been repeatedly reported ever since. The adrenal gland is the most common extrapulmonary site of CMV infection.

In 1985, Glasgow and associates found CMV adrenalitis in 21 of 41 patients examined and widespread lipid depletion in most cases (a nonspecific finding on autopsies of critically ill patients). The adrenal cortical involvement was limited to 10% of the cortex in most and was less than 70% of all cases. In an attempt to establish clinicopathologic correlation, 32 patients were also analyzed for clinical signs and symptoms as well as laboratory evidence of adrenal insufficiency. Common findings were hyponatremia (75%), hypotension (34%), hypokalemia (19%), hyperkalemia (16%), vomiting, diarrhea, and fever (percentage not specified). No patient had hyperpigmentation. Morning levels of serum cortisol were normal or elevated in 5 of 5 patients. One of 2 patients tested had a subnormal increase in cortisol after adrenocorticotropic hormone (ACTH) infusion. Despite significant pathologic adrenal abnormalities, no clinical adrenal insufficiency was documented. The degree of adrenal cortical damage seen was considered less than that usually associated with adrenal insufficiency.

In 1990, Pulakhandman and Dincsoy found CMV infections in half of 74 autopsied cases of AIDS. In 37 cases, the adrenal gland was the most commonly affected organ; the lungs were affected in 55% of cases. CMV inclusions were found in endothelial, cortical, and medullary cells. The CMV adrenalitis was diffuse in 10 cases and focal in 20. Analysis of the clinical data from 30 subjects revealed no findings related to the adrenal pathology, with the possible exception of a serum sodium:potassium ratio of less than 30 in those with more severe adrenal pathology.

### Clinical Studies

Adrenal function is among the most commonly reported endocrine parameters in patients with HIV/AIDS. Many of the signs and symptoms characteristic of adrenal insufficiency are frequently seen in severely ill patients with HIV infection. However, definite adrenal insufficiency is rare in patients with AIDS. The results of adrenal function studies in patients with AIDS vary depending on the patients studied, the severity of the disease, the concomitant intercurrent

illnesses and medications to which they are exposed, and the methods of adrenal testing (baseline or stimulated levels of glucocorticoids and/or ACTH).

### Studies in Adults

Most studies of adrenal function have been performed on adults with HIV infection. Early reports documented both normal and diminished cortisol responses to ACTH in patients with AIDS. Membreno *et al.* studied 74 randomly selected hospitalized patients with AIDS and 19 patients with AIDS-related complex (ARC). Based on subnormal cortisol responses to ACTH stimulation, 4 patients with AIDS were diagnosed as having adrenal insufficiency. Mean basal cortisol levels were higher in patients with AIDS than in healthy individuals, but ACTH-stimulated cortisol responses were not different from normal. However, stimulated levels of 17-deoxysteroid levels [corticosterone and 18-OH-deoxycorticosterone (18-OHDOC)] were lower than normal. Patients with ARC responded in a similar manner as those with AIDS. Based on these findings, the researchers suggested that impaired 17-deoxysteroid levels, especially 18-OHDOC, might be a harbinger of progressive adrenal disorder. Plasma ACTH levels were normal, not elevated, as would be anticipated, in patients with adrenal insufficiency, suggesting a possible pituitary defect in these patients. Administration of the hypothalamic factor corticotropin-releasing hormone (CRH) resulted in subnormal 18-OHDOC responses in 2 patients. The researchers advanced the hypothesis that HIV pituitary infection could lead to selective hypopituitarism and hypoadrenalism, with subsequent HIV adrenal infection leading to complete adrenal insufficiency.

Dobs and associates reported normal cortisol responses to ACTH in 36 of 39 ambulatory patients with AIDS. Although no clinical evidence of endocrine disorder was found, Merenich reported lower baseline cortisol and aldosterone and ACTH-stimulated cortisol levels in 40 asymptomatic HIV-infected men compared with 27 HIV-infected age-matched control subjects; 1 patient had low cortisol and also low ACTH-stimulated aldosterone levels.

HIV-infected patients have been reported to have lower responses in cortisol and/or ACTH to cold stress or tetanus toxoid administration. Several investigators, however, have reported elevated baseline levels or 24-h secretion of cortisol and/or ACTH.

In 1992 and 1993, Catania compared proopiomelanocortin-derived peptides and cytokines in 80 patients with AIDS and in 80 normal subjects. Average plasma  $\alpha$ -melanocyte-stimulating hormone levels were higher

in the AIDS patients, but mean levels of cortisol, ACTH,  $\beta$ -endorphins, interleukin-1 (IL-1), IL-6, and tumor necrosis factor were not different between the two groups.

In a prospective study of 98 HIV-infected patients, Raffi *et al.* found only 4 patients with low baseline and 7 with low ACTH-stimulated cortisol levels. Only 2 patients were believed to have adrenal insufficiency.

In 2001, Eledrisi and Verghese described three patients with AIDS who had clinical features suggestive of adrenal insufficiency but normal ACTH stimulation tests. Following repeat testing, the diagnosis of adrenal insufficiency was made in one patient. The other two patients required overnight metyrapone test to confirm the diagnosis. All three patients had improvement in their clinical condition following glucocorticoid therapy.

To further explain the effect of HIV infection on the HPA axis, Azar and Melby administered ovine CRH (oCRH) to 25 non-AIDS ambulatory HIV-infected patients and 10 normal volunteers: 6 patients had diminished cortisol and ACTH responses to CRH, 6 had low cortisol and normal ACTH responses, and 13 had normal cortisol and ACTH responses. They suggested enhanced hypothalamic CRH production in HIV infections as a possible explanation for their results. Complex interactions between the immune and HPA axis, mediated by cytokines and perhaps lymphocyte-produced ACTH, have been postulated to explain the mechanisms by which HIV infections may affect adrenal function.

Hyponatremia has been reported in 30% to more than 50% of hospitalized patients with AIDS. In most patients, however, hyponatremia is thought to be caused by renal and/or gastrointestinal losses and/or the syndrome of inappropriate secretion of antidiuretic hormone. Hyporeninemic hypoaldosteronism and mineralocorticoid deficiency have been reported in a few patients with AIDS. However, the adrenal mineralocorticoid pathway is normal in both the baseline and ACTH-stimulated states in most HIV-infected patients.

Serum levels of dehydroepiandrosterone have been reported to be decreased in HIV-infected patients, to correlate with CD4 levels, and to be a predictor of the progression of HIV infection to AIDS. Norbiato reported nine patients with AIDS and characteristic clinical features of adrenal insufficiency with elevated cortisol levels, suggesting resistance to glucocorticoids by abnormal glucocorticoid receptors on lymphocytes.

The bulk of the data suggest that adrenal insufficiency can be a complication of HIV infection, and

health care professionals should have a high index of suspicion of subtle adrenal dysfunction.

### Studies in Children

In a study of children with symptomatic HIV infections ( $N=28$ ), Rapaport *et al.* measured morning serum cortisol levels and found that the lowest levels occurred in those with the lowest CD4 levels. ACTH tests performed on 2 children with the lowest cortisol levels were normal, suggesting normal adrenal glucocorticoid function. In several ill children with AIDS who were suspected of having adrenal insufficiency, normal baseline cortisol levels were found in all. In 4 subjects, ACTH-stimulated cortisol responses were normal, excluding the diagnosis of adrenal insufficiency.

Oberfield and associates found normal or slightly elevated baseline and ACTH-stimulated cortisol levels, with mildly diminished stimulated serum deoxycorticosterone and corticosterone levels, in 12 HIV-infected children, 2 of whom were receiving ketoconazole, a drug known to inhibit adrenal function. Laue *et al.* reported normal cortisol response to ACTH in 8 of 9 children with AIDS; the 1 patient with a subnormal response was receiving treatment with ketoconazole. Schwartz *et al.* found normal cortisol responses to glucagon stimulation in 12 of 12 HIV-infected children.

## HPA AXIS AND HIV TREATMENT

Treatment of HIV infection has undergone substantial changes during the past two decades. However, several agents still used have been known to affect the HPA axis. Hypoadrenalism may result from ketoconazole or rifampin treatment. Symptoms of adrenal insufficiency have been reported following cessation of megestrol acetate therapy.

Symptoms that may represent HPA axis dysfunction have also been reported. Hypoglycemia has been demonstrated in patients with AIDS treated with pentamidine and trimethoprim sulfamethoxazole. Hyponatremia has been associated with pentamidine and vidarabine therapy.

Yanovski's group postulated that the increase in abdominal adipose tissue found in HIV-infected patients treated with protease inhibitors (PIs) could be due to abnormalities in adrenal function. Patients on a PI regimen had normal diurnal cortisol secretion, cortisol secretory dynamics after oCRH, cortisol-binding globulin levels, and glucocorticoid number and affinity. The following abnormalities were noted: increased basal and CRH-stimulated plasma

ACTH and 24-h urinary 17-OHCS excretion and decreased urinary free cortisol. The study did not include any HIV-infected patients who were not on PI treatment; therefore, the changes observed, although coinciding with PI treatment, can not be solely ascribed to the use of PIs. They concluded that the changes observed seem to represent a distinct but undefined form of hypercortisolism that could not be related directly to the HPA axis.

## CONCLUSION

There have been numerous advances in the areas of epidemiology, management, prevention, and prophylaxis of HIV disease globally. Much effort is currently focused on prevention strategies, with vaccines playing a potential pivotal role. Some phase 3 trials of vaccines are already under way.

Endocrine effects of HIV infection are not easily recognized by clinical signs and/or symptoms, baseline hormonal evaluation, or even stimulation tests. Although a single consistent HIV-related HPA axis dysfunction has not been documented, persistent and careful investigation of the HPA axis may disclose potential abnormalities. Recognition and treatment of these disorders should improve the quality of life of the patients. The evaluation and treatment of endocrine alterations in patients with HIV infection is an increasingly important clinical and investigational area of pediatric endocrinology.

## Acknowledgement

We thank Daphna Sachs for her help in preparing the manuscript.

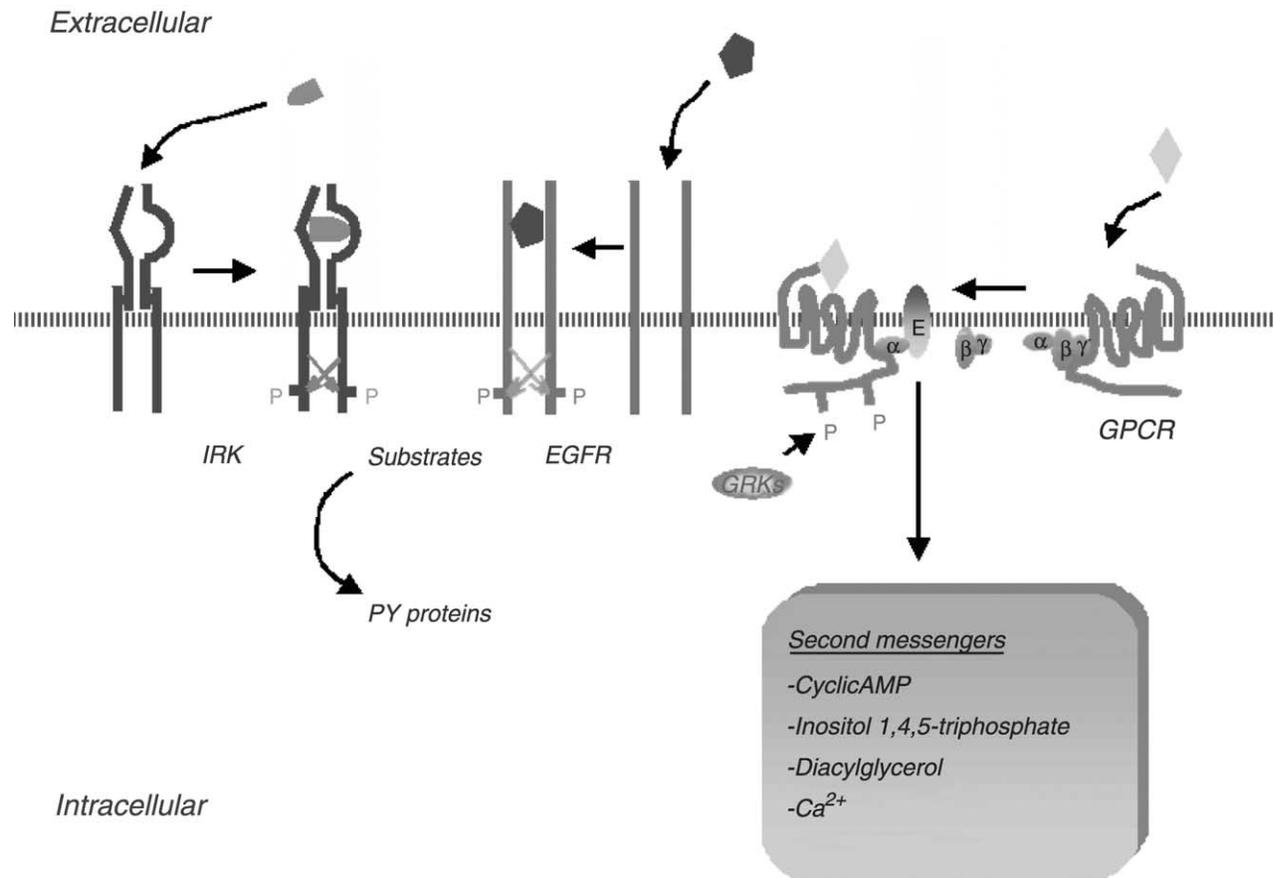
## See Also the Following Article

Immune System, Hormonal Effects on

## Further Reading

- Azar, S. T., and Melby, J. C. (1993). Hypothalamic–pituitary–adrenal function in non-AIDS patients with advanced HIV infection. *Am. J. Med. Sci.* **305**, 321–325.
- Catania, A., Manfredi, M. G., Airaghi, L., *et al.* (1992). Delayed cortisol response to antigenic challenge in patients with acquired immunodeficiency syndrome. *Ann. N. Y. Acad. Sci.* **650**, 202–204.

- Christeff, N., Gharakhanian, S., Thobie, N., Rozenbaum, W., and Nuñez, E. A. (1992). Evidence for changes in adrenal and testicular steroids during HIV infection. *J. Acquired Immune Defic. Syndr.* **5**, 841–846.
- Eledrisi, M. S., and Verghese, A. C. (2001). Adrenal insufficiency in HIV infection: A review and recommendations. *Am. J. Med. Sci.* **321**(2), 137–144.
- Grinspoon, S., and Bilezikian, J. P. (1992). HIV disease and the endocrine system. *N. Engl. J. Med.* **32**, 1360–1365.
- Grubman, S., Conviser, R., and Oleske, J. (1992). HIV infections in infants, children and adolescents. *AIDS and Other Manifestations of HIV Infections*, 2nd ed., pp. 201–216. Raven Press, New York.
- Hauser, L., Sheehan, P., and Simpkins, H. (1991). Pancreatic pathology in pentamidine-induced diabetes in acquired immunodeficiency syndrome patients. *Hum. Patbol.* **22**, 926–929.
- Kumar, M., Kumar, A. M., Morgan, R., Szapocznik, J., and Eisdorfer, C. (1993). Abnormal pituitary–adrenocortical response in early HIV-1 infection. *J. Acquired Immune Defic. Syndr.* **6**, 61–65.
- Malone, J. L., Oldfield, E. C., Wagner, K. F., *et al.* (1992). Abnormalities of morning serum cortisol levels and circadian rhythms of CD4 lymphocyte counts in human immunodeficiency virus type 1-infected adult patients. *J. Infect. Dis.* **156**, 185.
- Mulder, J. W., Jos Frissen, P. H., Krijen, P., *et al.* (1992). Dehydroepiandrosterone as predictor for progression to AIDS in asymptomatic human immunodeficiency virus-infected men. *J. Infect. Dis.* **165**, 413–418.
- Norbato, G., Bevilacqua, M., Vago, T., *et al.* (1992). Cortisol resistance in acquired immunodeficiency syndrome. *J. Clin. Endocrinol. Metab.* **74**, 608–613.
- Pulakhandam, U., and Dincsoy, H. P. (1990). Cytomegaloviral adrenalitis and adrenal insufficiency in AIDS. *Am. J. Clin. Patbol.* **93**, 651–656.
- Raffi, F., Brisseau, J. M., Planchon, B., Remi, J. P., Barrier, J. H., and Grolleau, J. Y. (1991). Endocrine function in 98 HIV-infected patients: A prospective study. *AIDS* **5**, 729–733.
- Schwartz, L. J., Louis, Y., Wu, R., Wiznia, A., Rubinstein, A., and Saenger, P. (1991). Endocrine function in children with human immunodeficiency virus infection. *Am. J. Dis. Child.* **145**, 330–333.
- Tang, W. W., Kaptein, E. M., Feinstein, E. I., and Massry, S. G. (1990). Hyponatremia in hospitalized patients with acquired immune deficiency syndrome. *J. Acquired Immune Defic. Syndr.* **3**, 949–953.
- UNAIDS, WHO (2000). *AIDS epidemic update: December 2000*. Joint United Nations Programme on HIV/AIDS, Geneva.
- Vitting, K. E., Gardenswartz, M. H., Zabetakis, P. M., *et al.* (1990). Frequency of hyponatremia and nonosmolar vasopressin release in the acquired immunodeficiency syndrome. *J. Am. Med. Assoc.* **263**, 973–978.
- Yanovski, J. A., Miller, K. D., Kino, T., Friedman, T. C., Chrousos, G. P., Tsigos, C., and Falloon, J. (1997). Endocrine and metabolic evaluation of human immunodeficiency virus-infected patients with evidence of protease inhibitor-associated lipodystrophy. *J. Clin. Endocrinol. Metab.* **84**(6), 1925–1931.



**Figure 1** Signaling from RTKs and GPCRs. Insulin and EGF bind to their respective receptors (IRK and EGFR), leading to the activation of the intrinsic kinase activity and autophosphorylation of the receptor on tyrosine residues. This promotes the tyrosine phosphorylation and the activation of downstream effectors, which in turn bind to SH2-containing proteins and thus entrain downstream signaling steps. The binding of agonist to GPCRs facilitates the association between receptor and trimeric G proteins. This association promotes the dissociation of GDP and association of GTP with the  $G_{\alpha}$ -subunit of the trimeric complex.  $G_{\alpha}$ -GTP is then released from the trimeric complex and interacts with cell surface enzymes (E), leading to the activation of downstream signaling. PY proteins, phosphotyrosine proteins.

domain of the  $\beta$ -subunit, in which circumstance it manifested augmented tyrosine kinase activity against exogenous substrates. Studies using scanning electron microscopy and computer modeling have established that one molecule of insulin binds within a pocket formed by the two  $\alpha$ -chains of the IR and that there is a subsequent conformational change that approximates the  $\beta$ -subunits, thus facilitating autophosphorylation on contralateral tyrosine residues (i.e., transphosphorylation).

An analogous process is observed with the EGFR and other receptor tyrosine kinases (RTKs) in that one ligand effects dimerization by bifacially binding separate RTK subunits, in effect promoting their association and facilitating their autophosphorylation on tyrosine residues (Fig. 1). Of significance was the recognition that autophosphorylation effected RTK

activation and hence the tyrosine phosphorylation of other cellular substrates—a key step in the signaling mechanism for this family of receptors. Rather uniquely, it was found that the level of tyrosine kinase activity of the IRK was sustained in the absence of insulin binding to the  $\alpha$ -subunit as long as the  $\beta$ -subunit remained tyrosine-phosphorylated (Table I). Of importance was the recognition that the tyrosine kinase activity of the IRK and EGFR as well as other RTKs is necessary for the realization of hormone action. Thus, mutating the ATP-binding site of the IRK completely inactivated kinase activity and abolished the insulin response without impairing insulin binding. Interestingly, it was soon apparent that IRK activation was also sufficient for the insulin response in that activating the IRK in the absence of insulin entrained the full insulin response.

## Postreceptor Cascade

The activation of RTKs effects signaling in two general ways. One is to provide specific phosphotyrosine motifs in the cytosolic domain of the receptor itself, the motifs of which bind downstream signaling molecules via their intrinsic SH2 domains. This signaling pattern is well exemplified by the platelet-derived growth factor (PDGF) receptor. Alternatively, the RTK promotes the tyrosine phosphorylation of downstream docking proteins, which in turn bind SH2-containing molecules, and thus entrains downstream signaling steps. This is exemplified by the IRK and its insulin receptor substrate (IRS) substrates. Depending on the nature of the SH2-containing molecule, there is an activation of an inherent enzyme activity [e.g., phosphatidylinositol 3-kinase (PI3K)] or the facilitation of its association with other molecules (e.g., Grb2, Shc), eventually effecting the activation of distal kinase(s) and/or phosphatases.

The serpentine receptors contain no intrinsic enzymatic activity. The binding of hormone in this receptor family appears to alter the conformation of its intracellular loop(s) to facilitate the association of the hormone-bound receptor with trimeric G proteins (Fig. 1). This association promotes the dissociation of GDP and the correlative association of GTP with the

$G_{\alpha}$ -subunit of the trimeric G protein. Subsequently,  $G_{\alpha}$ -GTP is released from the trimeric complex to interact with and alter the activity of various cell surface enzymes [viz. adenylate cyclase, phospholipase C (PLC)]. This alteration may be either an augmentation or a restraint of the production of intracellular second messengers (viz. cyclic AMP, diacylglycerol) (Fig. 1).

## INTERNALIZATION AND SIGNALING

### The Insulin Receptor Kinase (IRK)

Following the demonstration that receptors could be identified in non-plasmalemma-containing cell fractions, it was quickly found that  $^{125}\text{I}$ -labeled peptide hormones (viz. insulin and lactogens) were rapidly internalized into nonlysosomal vesicular elements in rat liver. Of particular interest was the appreciation that these internalized ligands retained their integrity. These observations led to the proposal that the internalized hormones were acting to effect signaling from these intracellular structures, which were originally appreciated as "unique vesicles" and subsequently named endosomes (ENs). The demonstration that activated IRKs and EGFRs were internalized into endosomes strengthened the view that signal

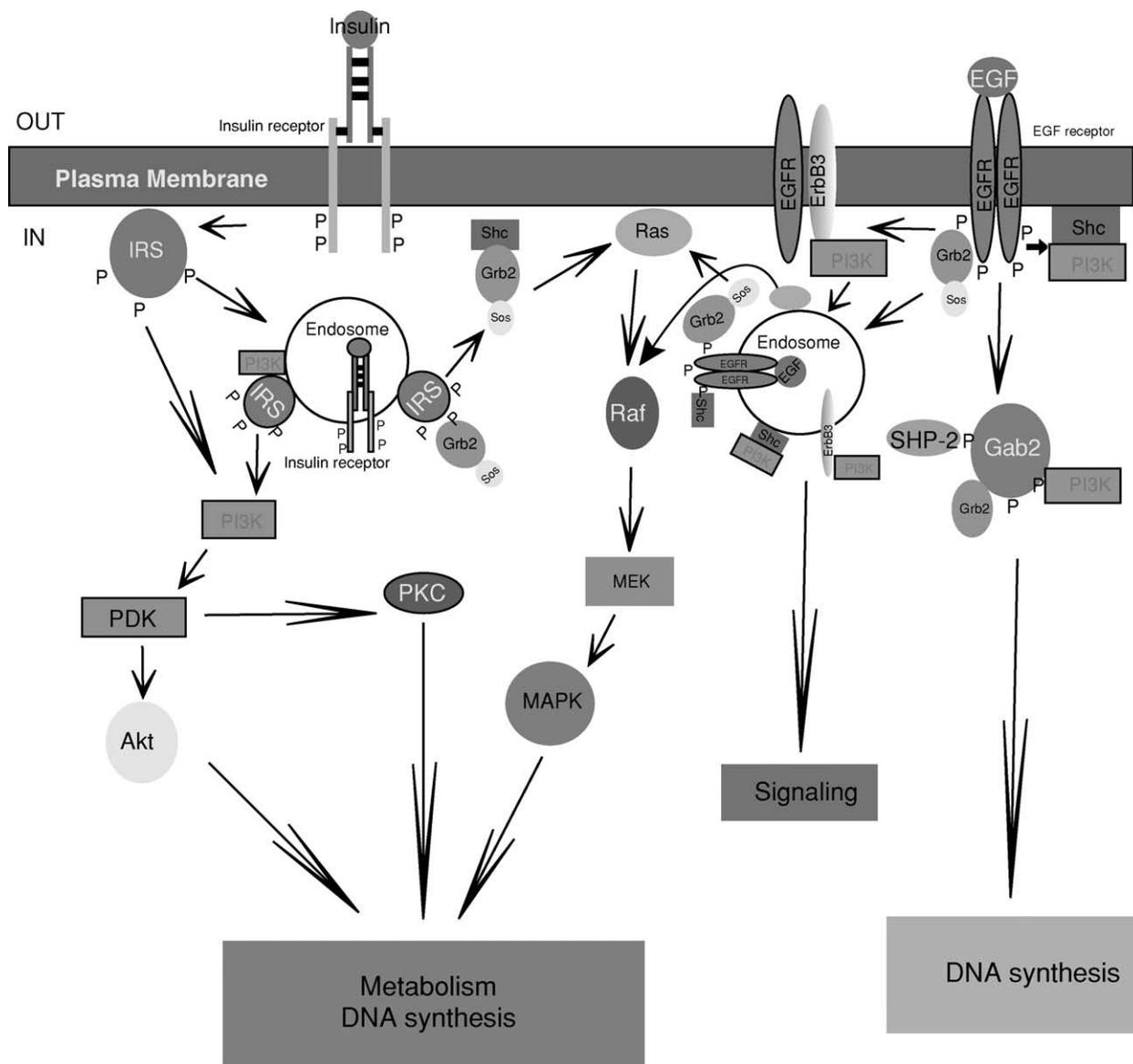
**Table I Evidence for IRK-Mediated Signaling in ENs**

Effect following hormone administration	References
Uptake of intact $^{125}\text{I}$ -insulin into lipoprotein containing vesicles (LPVs) of liver	Bergeron <i>et al.</i> (1979) <i>J. Cell Biol.</i> <b>80</b> , 427 Posner <i>et al.</i> (1980) <i>J. Biol. Chem.</i> <b>255</b> , 735
Increased autophosphorylation and kinase activity of IRK associated with ENs	Khan <i>et al.</i> (1986) <i>J. Biol. Chem.</i> <b>261</b> , 8462 Klein <i>et al.</i> (1987) <i>J. Biol. Chem.</i> <b>261</b> , 10557
Specific activity of IRK is higher in ENs than in PM	Khan <i>et al.</i> (1989) <i>J. Biol. Chem.</i> <b>264</b> , 12931 Burgess <i>et al.</i> (1992) <i>J. Biol. Chem.</i> <b>67</b> , 10077
Phosphorylation of IRS1 and activation of PI3K in ENs	Kelly (1993) <i>J. Biol. Chem.</i> <b>268</b> , 4391
Phosphorylated IRK and IRS1 are detected in ENs	Kublaoui <i>et al.</i> (1995) <i>J. Biol. Chem.</i> <b>270</b> , 59 Smith <i>et al.</i> (1997) <i>Int. Rev. Cytol.</i> <b>173</b> , 243
Specific activation of endosomal IRK by bpV(phen) results in phosphorylation of IRS1 and hypoglycemia	Bevan <i>et al.</i> (1995) <i>J. Biol. Chem.</i> <b>270</b> , 10784
90% of the IRK-associated phosphate is located in ENs	Wang <i>et al.</i> (1996) <i>Biochem. Biophys. Res. Commun.</i> <b>227</b> , 27
In 3T3-L1 adipocytes, insulin treatment increased PI3K activity in the microsomal fraction	Nave <i>et al.</i> (1996) <i>Biochem. J.</i> <b>318</b> , 55
In 3T3-L1 adipocytes, insulin induced redistribution of IRS1 and PI3K between LDM and cytosol	Ricort <i>et al.</i> (1996) <i>Eur. J. Biochem.</i> <b>239</b> , 17
In ENs, insulin induced association of IRS1/2 to p85 and recruitment of Akt1 with a higher activity than in PM	Balbis <i>et al.</i> (2000) <i>Endocrinology</i> <b>141</b> , 4041
ENs contained an insulin-regulated pool of Cdk2-cyclin E complexes	Gaulin <i>et al.</i> (2000) <i>J. Biol. Chem.</i> <b>275</b> , 16658
Activated IRK is localized in GLUT4 vesicles of skeletal muscle	Dombrowski <i>et al.</i> (2000) <i>Diabetes</i> <b>49</b> , 1772

transduction was not limited to the cell surface but occurred within the endosomal system as well (Fig. 2).

With respect to the IRK, activated receptor has been shown to accumulate in ENs from liver, adipocytes, and skeletal muscle. The discovery and use of

the powerful insulin mimetic peroxovanadium (pV) compounds has contributed to the understanding of the significance of internalized IRKs. In particular, Bevan *et al.* showed that *in vivo* administration of pV compounds selectively activated the endosomal IRK



**Figure 2** Signaling from endocytosed receptors. Insulin binding leads to autophosphorylation of IRK $\beta$  subunits, increased IRK exogenous kinase activity, and rapid internalization of the activated hormone–receptor complex into ENs. Tyrosine phosphorylation of downstream adapter proteins (such as IRS1) by PM and endosomal IRK leads to the recruitment of SH2-containing proteins (such as PI3K and Grb2). Recruitment of PI3K to IRS1 effects the activation of its kinase activity entraining downstream events, leading to metabolic changes and DNA synthesis. Recruitment of Grb2 to IRS1 leads to activation of the Ras–Raf–MEK–MAPK cascade. EGF binding to its PM receptor leads to receptor homo- or heterodimerization, autophosphorylation, and activation. This is followed by a rapid internalization of the receptor. EGFR or ErbB3 phosphorylation enables recruitment of proteins via their SH2 domain (Shc, Grb2, PI3K) or phosphotyrosine binding (PTB) domain (Shc) at both the PM and ENs, leading to the activation of Raf-1 and the MAPK cascade. In liver, activated EGFR leads to the accumulation of cytosolic tyrosine-phosphorylated Gab2, which recruits and activates PI3K, as part of a multimeric complex, leading to DNA synthesis. PKC, protein kinase C; PDK, phosphoinositide dependent kinase.

and that IRS1 tyrosine phosphorylation ensued as a consequence.

There has also been a growing appreciation that compartmentalization of downstream signaling molecules may play a significant role in cell signaling. Thus, insulin was found to induce a decrease in hepatic plasma membrane (PM) content of IRS1 and IRS2 with a corresponding increase in their concentration in ENs. In 3T3-L1 adipocytes, substantial quantities of IRS1 and IRS2 were found in internal membranes. Similar findings were made in low-density membranes (LDM) from adipocytes, a fraction representing endosomal elements in this cell type. Furthermore, insulin administration effected the phosphorylation of IRS1 in LDM followed by its dissociation into the cytosol. These events in adipocytes appear to exclude interactions at the PM, which appears to differ from observations in rat liver. It has also been observed that insulin induces a rapid increase in the levels of p85 in ENs, in association with IRS1 and IRS2, with IRS1 accounting for the bulk of the recruitment. The kinetics of association are consistent with the formation of these complexes at the cell surface and their internalization into ENs. These findings compare to those in other insulin target tissues, such as 3T3-L1 adipocytes, where insulin induced the association of PI3K mainly with LDM or microsomes. In contrast, PDGF stimulated the recruitment and activation of PI3K exclusively in PM but unlike insulin PDGF did not induce GLUT4 translocation from the LDM to the PM. Interestingly, it was found that, in 3T3-L1 adipocytes, impaired insulin-stimulated GLUT4 translocation induced by oxidative stress is associated with the inhibition of redistribution of both IRS1 and activated PI3K between LDM and cytosol, whereas the activity of this enzyme in total lysates was not affected. Taken together, these data support the concept that insulin-induced compartmentalization of key signaling molecules is important for transducing at least some aspects of the insulin response. These studies and related observations are summarized in [Table I](#).

### The Epidermal Growth Factor Receptor Kinase (EGFRK)

Ligand binding to the EGFR initiates dimerization, receptor autophosphorylation, and activation of signal transduction pathways as well as trafficking events that relocalize the receptor from the cell surface to intracellular compartments. There are data indicating the involvement of specialized membrane domains in the initial steps leading to EGFR internalization. These

include caveolae and clathrin-coated pits but probably involve noncoated domains as well. The initial endocytic vesicles fuse to endosomes that deliver the receptor and their ligands to various intracellular destinations.

EGF-induced receptor endocytosis plays at least two functions. It was first demonstrated as a mechanism for attenuating the signal generated at the cell surface. Indeed, various studies have shown that an increase in mitogenic responses induced by EGF binding to the receptor correlates with the decrease of receptor internalization. In contrast, in rat liver, reports have described the presence of activated receptors in endosomes and the increasing association over time of Shc, Grb2, and mSos with the internalized receptor whose tyrosine phosphorylation state correspondingly increased. There has been increasing evidence indicating a role for internalized activated EGFR in influencing downstream signaling molecules. Of particular interest are the well-characterized data linking endocytosis of EGFR with ERK1/2 activation. In fibroblasts, internalized EGFRs participate in the activation of p21/Ras and in normal rat kidney (NRK) cells intact actin cytoskeleton has been reported to be necessary for EGF-mediated transport of caveolin from the cell surface into early endosomes as well as activation of the MAPK pathway. Other evidence has been provided by the work of Viera and colleagues in HeLa cells overexpressing dominant-negative (K44A) dynamin. In such cells, which are specifically defective in receptor-mediated endocytosis, a decrease in EGFR tyrosine phosphorylation was found, consistent with earlier observations showing the hyperphosphorylation of endosomal EGFR versus PM EGFR. Furthermore, in these cells a decrease in MAPK phosphorylation and activation was seen along with augmented phosphorylation of PLC- $\gamma$  and Shc. Other evidence showing the importance of receptor internalization in aspects of EGF signaling comes from two additional studies. Molecules involved in the mitogen-activated protein kinase (MAPK) cascade were localized in highly purified endosomes from rat liver. Using the same purified fraction, the authors also showed that administration of EGF led to the redistribution of Raf-1 from the plasma membrane into endosomes. Finally, using combined imaging microscopy and fluorescence resonance energy transfer, Sorkin elegantly demonstrated that activated EGFR interacts with Shc and Grb2 in membrane ruffles and endosomes. These studies, which provide evidence that EGFR endocytic trafficking is required to trigger MAPK signaling pathways, and others implicating endosomal EGFR in cell signaling are summarized in [Table II](#).

Table II EGFR-Mediated Signaling: Evidence for Role of ENs

Effect following hormone administration	References
Tyrosine-phosphorylated EGFR persists after internalization	Cohen et Fava (1985) <i>J. Biol. Chem.</i> <b>260</b> , 12351 Kay et al. (1986) <i>J. Biol. Chem.</i> <b>261</b> , 8473
EGFR phosphorylation content is higher in ENs than in PM	Wada et al. (1992) <i>J. Cell Biol.</i> <b>116</b> , 321
Grb2 and Shc associate with EGFR in ENs, leading to MAPK activation	Di Guglielmo et al. (1994) <i>EMBO J.</i> <b>13</b> , 4269
Mutant dynamin decreases EGFR phosphorylation, activation of MAPK, and phosphorylation of Shc and PLC- $\gamma$	Vieira et al. (1996) <i>Science</i> <b>274</b> , 2086
Endocytosed EGFR induced a sustained activation of p21/CIP1	Skarpen et al. (1998) <i>Exp. Cell Res.</i> <b>243</b> , 161
Activation of ERK2 by internalized EGFR	Xue et al. (1998) <i>Cell. Signal.</i> <b>10</b> , 339
EGF treatment led to the redistribution of activated Raf-1 from calveolin-enriched PM to ENs	Pol et al. (1998) <i>FEBS Lett.</i> <b>441</b> , 34
Internalization of EGFR led to the activation of p21(Ras) in fibroblasts	Haugh et al. (1999) <i>J. Biol. Chem.</i> <b>274</b> , 34350
Activated MAPK associates with EGFR in ENs	Faure et al. (1999) <i>Mol. Cell. Biol. Res. Commun.</i> <b>1</b> , 132
Intact cytoskeleton is required for EGFR endocytosis and ERK activation	Pol et al. (2000) <i>J. Biol. Chem.</i> <b>275</b> , 30566
Localization of activated EGFR and Shc in early ENs	Oksvold et al. (2000) <i>J. Histochem. Cytochem.</i> <b>48</b> , 21
Interaction of EGFR/Grb2/Shc and Eps8 in ENs	Sorkin (2001) <i>Biochem. Soc. Trans.</i> <b>29</b> , 480 Burke et al. (2001) <i>Mol. Biol. Cell</i> <b>12</b> , 1897

Work from the authors' laboratory has shown that the activation of PI3K is a necessary step in EGF-induced mitogenesis of primary rat hepatocytes in culture. Following EGFR activation, three different molecules become tyrosine-phosphorylated and recruit PI3K to phosphotyrosine motifs. These three molecules were shown to be ErbB3, Shc, and Gab2. The last molecule was cloned and shown to be the essential docking molecule involved in EGF-induced PI3K-dependent mitogenesis in rat liver. In addition, it has been shown that, whereas ErbB3 and Shc are tyrosine-phosphorylated and internalized into endosomes, tyrosine-phosphorylated Gab2 is largely a cytosolic entity. These relations are also noted in Fig. 2.

### G Protein-Coupled Receptors

In addition to their classical responses (Fig. 1), many G protein-coupled receptors (GPCRs) activate MAPK cascades. As with the EGFR, internalization seems to be crucial for ERK1/2 activation. Following ligand binding, the activation of GPCRs leads to the recruitment of  $\beta$ -arrestin to the activated receptor. This step often terminates or attenuates signaling by blocking the interaction of G proteins with the receptor (desensitization). However, various studies have shown that  $\beta$ -arrestins also target the receptors to clathrin-coated vesicles for subsequent internalization and act as scaffold proteins, which contribute to the modulation of responses under the control of GPCRs.  $\beta$ -Arrestins form complexes with signaling molecules

such as Src. The association of  $\beta$ -arrestins and Src modulates GPCR endocytosis, leading to MAPK activation. The downstream mechanism of ERK1/2 activation by GPCRs appears to involve the same intermediates as those activated by RTKs. It is thought that  $G_{\beta\gamma}$ -subunits stimulate the tyrosine phosphorylation of the adapter protein Shc and the recruitment to the plasma membrane of the Ras guanine exchange factors such as Grb2-mSos complex. Recruitment of mSos facilitates Ras GDP/GTP exchange, leading to the recruitment of Raf into the complex with activated Ras. Subsequent signal transduction involves the sequential phosphorylation of MEK (MAPK kinase) and MAPK. Studies using various inhibitors of endocytosis are consistent with the view that GPCR-mediated ERK1/2 activation is endocytosis dependent and that different inhibitors of internalization specifically block Raf-mediated activation of MEK. The endocytosis of GPCRs is also required for CXCR2-mediated chemotaxis and somatostatin-mediated inhibition of growth hormone expression. These and related observations are summarized in Table III.

In contrast, the classical  $\beta$ 2-AR (adrenergic receptor) signaling pathway, mediated by G proteins and leading to the activation of adenylate cyclase and phospholipase C, is unaffected by dominant-negative (K44A) dynamin, which significantly inhibits endocytic vesicular trafficking. This demonstrates that in contrast to ERK1/2 activation, these signaling events occur entirely at the plasma membrane as depicted in Fig. 1.

Table III GPCR-Mediated Signaling: Evidence for Role of ENs

Receptor	Evidence for signaling from endosomes	References
$\beta$ 2-AR	$\beta$ 2-AR-mediated ERK1/2 activation involves receptor sequestration DN cSrc blocks $\beta$ 2-AR internalization and receptor-stimulated tyrosine phosphorylation of dynamin	Daaka <i>et al.</i> (1998) <i>J. Biol. Chem.</i> <b>273</b> , 685 Miller <i>et al.</i> (2000) <i>J. Biol. Chem.</i> <b>275</b> , 11312
<b>LPA and thrombin receptors</b>	Shc phosphorylation and receptor endocytosis correlate with an increase of ERK kinase activities	Luttrell <i>et al.</i> (1997) <i>J. Biol. Chem.</i> <b>272</b> , 31648
<b>5-HT1A</b>	5-HT1A receptor-mediated ERK activation requires calcium/calmodulin-dependent receptor endocytosis	Della Rocca <i>et al.</i> (1999) <i>J. Biol. Chem.</i> <b>274</b> , 47491
<b>m1 muscarinic receptor <math>\mu</math> and <math>\gamma</math> opioid receptors</b>	DN dynamin blocks ERK activation mediated by these receptors	Vogler <i>et al.</i> (1999) <i>J. Biol. Chem.</i> <b>274</b> , 12333 Ignatova <i>et al.</i> (1999) <i>J. Neurosci.</i> <b>19</b> , 56
<b>L1 receptors</b>	Activated ERK localizes with endocytosed vesicular L1 rather than cell surface L1 Mutant dynamin blocks L1-induced neurite outgrowth	Schaefer <i>et al.</i> (1999) <i>J. Biol. Chem.</i> <b>274</b> , 37965 Schmid <i>et al.</i> (2000) <i>J. Neurosci.</i> <b>20</b> , 4177
<b>Sst2</b>	Inhibition of growth hormone expression depends on the receptor internalization	Sarret <i>et al.</i> (1999) <i>J. Biol. Chem.</i> <b>274</b> , 19294
<b>CXCR2</b>	Receptor endocytosis is required for CXCR2-mediated chemotaxis	Yang <i>et al.</i> (1999) <i>J. Biol. Chem.</i> <b>274</b> , 11328
<b>ET receptor</b>	Internalized ET receptor interacts with Tip60 and HDAC7 potentiating significantly ERK phosphorylation	Lee <i>et al.</i> (2001) <i>J. Biol. Chem.</i> <b>276</b> , 16597

In some systems, GPCRs also stimulate tyrosine kinase signaling cascades by inducing transactivation of RTKs. An example is the agonist activation of the  $\beta$ 2-AR in COS-7 cells inducing the formation of a multireceptor complex containing both  $\beta$ 2-AR and EGFR in which the EGFR becomes activated along with the activation of ERK1/2. These transactivations are also sensitive to inhibitors of endocytosis, consistent with a key role for the endocytic process in activation of the ERK cascade by GPCRs via this mechanism.

### Other Receptor Systems

Table IV summarizes observations about other receptors that indicate that there is internalization of activated receptors in aspects of cell signaling. With time it has become increasingly clear that peptide hormones effect signal transduction both at the cell surface and within the endosomal system as well. A number of observations supporting this model of signaling are summarized in Tables I to IV. In Fig. 2, signal transduction is depicted as deriving from both the cell surface and the earlier components of the endosomal system. Though Fig. 2 deals with insulin and EGF, comparable patterns clearly apply to other hormones and growth factors. In ENs, in addition to insulin- and EGF-induced concentration of activated IRKs and EGFRs, respectively, there was a concentration of

key signaling molecules involved in the propagation of insulin and EGF signaling.

### SUMMARY

Different patterns of cell signaling have been distinguished. In one instance, a signal initiated at the cell surface might be continued within endosomes and hence be augmented in both intensity and duration. This would appear to be the case for IRK activation and tyrosine phosphorylation of IRS1 and the molecular events that derive from the activation of this docking protein. Alternatively, certain signaling events might be restricted to one or another of the cellular compartments traversed by ligand-receptor complexes. Thus, the activation of the MAPK pathway by EGFR exemplifies the requirement for ligand-receptor internalization.

The availability of new techniques for manipulating molecular structure and for visualizing molecular localization has accelerated the elucidation of cell signaling. Ever more has been learned about the complexity of the signaling process with respect to the number of molecules involved in mediating and modulating the signal cascade and with respect to the requirement that there be a redistribution of activated receptors and signal transducing molecules into different cellular compartments for signaling to be fully realized. In the next few years, it is anticipated

Table IV Other Receptors Signaling via the Endosomal System

Receptor	Evidence for signaling from endosomes	References
<b>IL-2 receptor</b>	IL-2 remains associated with its receptor following internalization	Duprez <i>et al.</i> (1992) <i>J. Biol. Chem.</i> <b>267</b> , 18639
<b>CD2 receptor</b>	Recruitment of activated p56 <sup>lck</sup> and ZAP-70 in endosomes of CD2-triggered T cells	Marie-Cardine <i>et al.</i> (1996) <i>J. Biol. Chem.</i> <b>271</b> , 20734
<b>IGF-I receptor</b>	MAPK pathway is activated by endosomal receptor and IRS1 pathway by both surface and endosomal receptors	Chow <i>et al.</i> (1998) <i>J. Biol. Chem.</i> <b>273</b> , 4672
<b>Angiotensin II receptor</b>	Requirement of receptor internalization for Angiotensin II-induced generation of phosphoinositols and increment in intracellular Ca <sup>2+</sup> concentration	Jimenez <i>et al.</i> (1999) <i>Biochem. Pharmacol.</i> <b>57</b> , 1125
<b>PDGF receptor</b>	Inhibitors of PI3 kinase- and clathrin-mediated endocytosis reduced PDGF receptor-dependent activation of ERK	Rakhit <i>et al.</i> (2000) <i>Mol. Pharmacol.</i> <b>58</b> , 413
<b>TNF receptor</b>	Blocking TNF endocytosis impaired activation of endosomal A-SMase, JNK	Schutze <i>et al.</i> (1999) <i>J. Biol. Chem.</i> <b>274</b> , 10203
<b>NGF receptor Trk</b>	Endocytosed Trks activate ERK5 pathway Neuronal differentiation is promoted by endosomal Trks Endosomal Trka is tyrosine-phosphorylated and binds PLC-γ	Grimes <i>et al.</i> (1996) <i>J. Neurosci.</i> <b>16</b> , 7950 Zhang <i>et al.</i> (2000) <i>J. Neurosci.</i> <b>20</b> , 5671 Watson <i>et al.</i> (2001) <i>Nat. Neurosci.</i> <b>4</b> , 981

Note. IGF-I, insulin-like growth factor-I; IL-2, interleukin 2; NGF, nerve growth factor; TNF, tumor necrosis factor.

that there will be a considerable increase in the knowledge of the three-dimensional patterns of signal transduction entrained by hormone and growth factors.

### See Also the Following Articles

EGF and Related Growth Factors • G Protein-Coupled Receptors • Growth Factor Receptors • Insulin-like Growth Factors • Peptide Hormones, Intracellular Transport • Peptide Hormones, Regulated Secretion • Peptide Hormones, Regulation and Gene Expression • Peptide Hormones, Subcellular Structure • Receptor Tyrosine Kinase

### Further Reading

- Barash, I., and Posner, B. I. (1989). Homologous induction of growth hormone receptors in cultured rat hepatocytes. *Mol. Cell. Endocrinol.* **62**, 281–286.
- Bergeron, J. J., Posner, B. I., Josefsberg, Z., and Sikstrom, R. (1978). Intracellular polypeptidehormone receptors. The demonstration of specific binding sites for insulin and human growth hormone in Golgi fractions isolated from the liver of female rats. *J. Biol. Chem.* **253**, 4058–4066.
- Bevan, A. P., Burgess, J. W., Yale, J. F., Drake, P. G., Lachance, D., Baquiran, G., Shaver, A., and Posner, B. I. (1995). *In vivo* insulin mimetic effects of pV compounds: Role for tissue targeting in determining potency. *Am. J. Physiol.* **268**, E60–E66.
- Bevan, A., Drake, P. G., Bergeron, J. J., and Posner, B. I. (1996). Intracellular signal transduction: The role of endosomes. *Trends Endocrinol. Metab.* **7**, 13–21.
- Carpenter, G. (2000). The EGF receptor: A nexus for trafficking and signaling. *BioEssays* **22**, 697–707.
- Fantl, W. J., Escobedo, J. A., Martin, G. A., Turck, C. W., del Rosario, M., McCormick, F., and Williams, L. T. (1992). Distinct phosphotyrosines on a growth factor receptor bind to

specific molecules that mediate different signaling pathways. *Cell* **69**, 413–423.

- Ferguson, S. S. (2001). Evolving concepts in G protein-coupled receptor endocytosis: The role in receptor desensitization and signaling. *Pharmacol. Rev.* **53**, 1–24.
- Khan, M. N., Posner, B. I., Verma, A. K., Khan, R. J., and Bergeron, J. J. (1981). Intracellular hormone receptors: Evidence for insulin and lactogen receptors in a unique vesicle sedimenting in lysosome fractions of rat liver. *Proc. Natl. Acad. Sci. USA* **78**, 4980–4984.
- Kong, M., Mounier, C., Wu, J., and Posner, B. I. (2000). Epidermal growth factor-induced phosphatidylinositol 3-kinase activation and DNA synthesis: Identification of Grb2-associated binder 2 as the major mediator in rat hepatocytes. *J. Biol. Chem.* **275**, 36035–36042.
- Maudsley, S., Pierce, K. L., Zamah, A. M., Miller, W. E., Ahn, S., Daaka, Y., Lefkowitz, R. J., and Luttrell, L. M. (2000). The beta(2)-adrenergic receptor mediates extracellular signal-regulated kinase activation via assembly of a multi-receptor complex with the epidermal growth factor receptor. *J. Biol. Chem.* **275**, 9572–9580.
- McPherson, P. S., Kay, B. K., and Hussain, N. K. (2001). Signaling on the endocytic pathway. *Traffic* **2**, 375–384.
- Posner, B. I., Bergeron, J. J., Josefsberg, Z., Khan, M. N., Khan, R. J., Patel, B. A., Sikstrom, R. A., and Verma, A. K. (1981). Polypeptide hormones: Intracellular receptors and internalization. *Rec. Prog. Horm. Res.* **37**, 539–582.
- Posner, B. I., Faure, R., Burgess, J. W., Bevan, A. P., Lachance, D., Zhang-Sun, G., Fantus, I. G., Ng, J. B., Hall, D. A., Lum, B. S., *et al.* (1994). Peroxovanadium compounds: A new class of potent phosphotyrosine phosphatase inhibitors which are insulin mimetics. *J. Biol. Chem.* **269**, 4596–4604.
- Rosen, O. M. (1987). After insulin binds. *Science* **237**, 1452–1458.
- Sorkin, A. (1998). Endocytosis and intracellular sorting of receptor tyrosine kinases. *Front. Biosci.* **3**, D729–D738.
- Yarden, Y., and Ullrich, A. (1988). Growth factor receptor tyrosine kinases. *Annu. Rev. Biochem.* **57**, 443–478.
- Yip, C. M., Yip, C. C., and Ward, M. D. (1998). Direct force measurements of insulin monomer–monomer interactions. *Biochemistry* **37**, 5439–5449.

insertion occur while the growing peptide chain is still attached to the ribosome. After the peptide chain is completely inserted into the ER, the signal sequence is removed by the signal cleavage enzyme (signalase) located on the inner surface of the ER. With the assistance of chaperones, the prohormone is correctly folded and disulfide bonds are formed. Some prohormones are modified by N-linked glycosylation in which sugar residues are added to asparagines and glutamines.

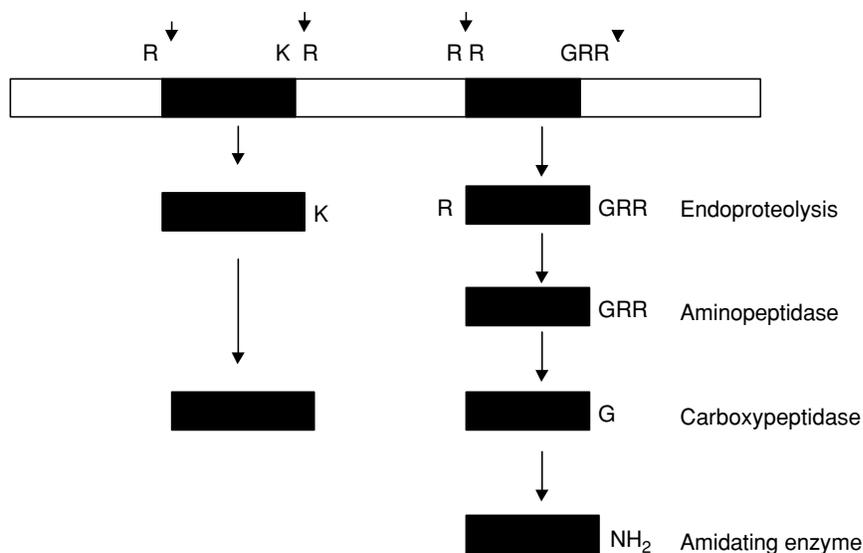
The prohormones are transported by specific vesicles to the Golgi apparatus, where they are further modified in different Golgi compartments. Additional carbohydrates are added to serine and threonine residues (O-linked glycosylation). Existing carbohydrate residues are further modified. In the *trans*-Golgi network (TGN), some tyrosine residues are sulfated on their hydroxyl group. This enzymatic reaction is performed by protein tyrosine sulfotransferase. The tyrosines that are sulfated have multiple adjacent acidic residues that constitute the consensus sequence for sulfation. This modification may be important for the biological activity of the peptide.

In the TGN, prohormones are sorted into regulated secretory granules with other proteins destined for secretion, including the enzymes responsible for their processing. The nature of the molecular recognition of prohormones as material for these granules is still a mystery. Whether there are specific sorting receptors that recognize prohormone or whether some structural feature is recognized is unknown. It

is not known whether prohormones are cleaved entirely within secretory granules or whether the cleavage starts earlier in the ER or Golgi. This may differ for different prohormones. Inside the secretory granule, there is a concentration step, wherein the internal calcium concentration increases while the pH drops. Drugs that disrupt the Golgi (Brefeldin) or prevent the acidification of secretory granules (chloroquine and ammonium chloride) greatly decrease the processing of prohormones. A typical example of the modifications that occur during the production of several neuropeptides from one prohormone is shown in Fig. 1.

Prohormones for peptide hormone and peptides undergo limited and specific endoproteolysis during their passage through the regulated secretory pathway. The biologically active elements in the prohormone are flanked by single or double basic residues. Not all such residues are sites for cleavage; somehow the endoproteases detect which sites are to be cleaved and ignore the others. This probably involves the three-dimensional structure of these sites and how it relates to the active site of the enzyme.

The enzyme(s) responsible for the endoproteolytic cleavages during processing has not been completely established, although considerable progress has been made in this area. The most likely candidates are members of a family of enzymes called the prohormone convertases (PCs). They are calcium-dependent serine proteases that operate at acidic pH and are



**Figure 1** Schematic diagram of the sequential modifications required to produce an amidated peptide (on the right) and a non-amidated peptide (on the left). The active peptide is indicated by the black boxes. The cleavage sites are indicated by the down arrows. R, arginine; K, lysine; G, glycine.

related to the enzyme subtilisin from the bacteria *Bacillus subtilis*. A large body of experimental evidence suggests that three members of this family are the most important for endoproteolysis in endocrine and neuronal cells that possess the regulated secretory pathway PC1, PC2, and PC5. These enzymes are widely distributed in the brain and in endocrine tissues and are present in a number of endocrine and neuronal tumor cell lines. They appear to have the correct catalytic activity and inhibition or deletion of their expression has been shown to alter the processing of prohormones that they express.

After endoproteolysis, exoproteases, such as carboxypeptidases and aminopeptidases, remove basic amino acids that may have been left on the peptides that were excised by endoproteases (see Fig. 1). These residues are not needed for biological activity and in some cases need to be removed to expose a glycine residue, so that it can be converted to a carboxy-terminal amide by the amidating enzyme. A number of carboxypeptidase enzymes that have the correct activity have been identified; carboxypeptidase E appears to be responsible for the processing of many, but not all, peptides.

The amidation reaction is probably the last modification and it is performed by a copper- and ascorbate-requiring enzyme. This amidation modification is found in approximately one-half the biologically active peptides and, in these peptides, it is essential for biological activity. In all cases, the immediate

precursor of the amide is a glycine residue on the carboxyl terminus. This reaction is catalyzed by a multifunctional enzyme called peptidyl- $\alpha$ -amidating monooxygenase (PAM).

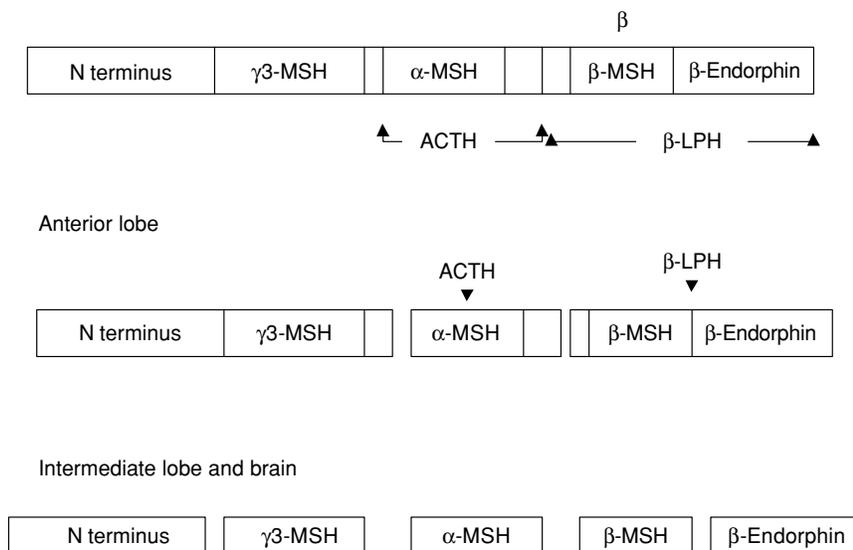
The vesicles containing the processed peptides are transported to the site of release and wait until an appropriate signal arrives, such as depolarization or increased external calcium, to cause them to fuse with the plasma membrane, releasing their contents outside the cell.

## TWO EXAMPLES OF DIFFERENTIAL PROCESSING

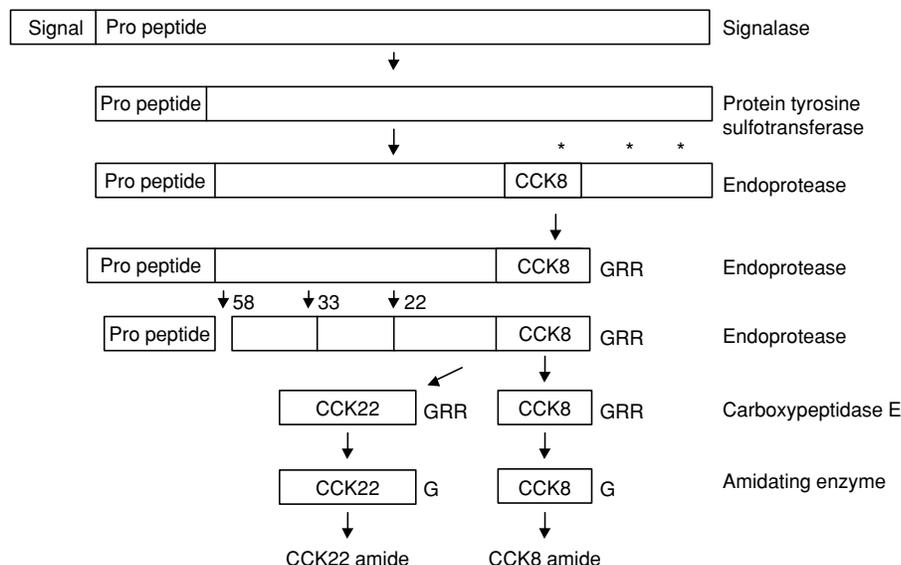
Most peptide prohormones do not display differential processing; however, there are a number of them that do. Two examples are described.

The proopiomelanocortin precursor is one of the first examples of differential processing to be discovered. As shown in Fig. 2, POMC mRNA is expressed in the anterior and intermediate lobes of the pituitary and in the brain of the rodent. The products that are isolated from these tissues are quite different and that is a result of differential tissue-specific processing. The anterior lobe makes larger forms, whereas the intermediate lobe and the brain make smaller, more highly processed forms.

The other example is pro-cholecystokinin processing. The biosynthesis of cholecystokinin (CCK) is



**Figure 2** Schematic diagram of differential processing of POMC in the anterior, intermediate lobes of the pituitary and the rodent brain. Intact POMC is shown at the top of the figure. The major products found in the anterior lobe of the pituitary are shown in the middle and the major forms found in the intermediate lobe and brain are shown at the bottom. ACTH, adrenocorticotrophic hormone;  $\beta$ -LPH,  $\beta$ -lipotropic hormone; MSH, melanocyte-stimulating hormone.



**Figure 3** Schematic diagram of the temporal order and the enzymes involved in pro-CCK processing and how different forms of CCK are produced. This diagram shows the production of CCK22 and CCK8. Larger forms are cleaved at sites indicated by the down arrows.

shown in Fig. 3. This case is different and simpler than that for POMC. The forms of CCK that are found are different in size, but they share the same CCK8 amide moiety and have similar biological activity. CCK8 is the major form found in mammalian brain, whereas mammalian intestine makes mainly larger forms of CCK, such as CCK22, CCK33, and CCK58. All of the CCK found is tyrosine sulfated.

A related peptide, gastrin, which shares some homology with CCK, is found in both sulfated and unsulfated forms. Like CCK, it comes in different sizes (gastrin 14, 17, and 34) that have the same biological activity. The degree of sulfation varies with the age of the animal. This represents an additional aspect of differential processing.

### THE MECHANISM OF DIFFERENTIAL PROCESSING

The mechanism of differential processing is not completely understood. The most likely explanation is tissue-specific differences in the distribution and activity of the endoproteases.

### THE PHYSIOLOGICAL SIGNIFICANCE OF DIFFERENTIAL PROCESSING

Differential processing happens for a reason. The process of producing peptides is very energy-dependent

and represents a major commitment on the part of the cell producing the peptides. In both of the examples described, some of the final peptides have different physiological roles. One of the major products of POMC processing is adrenocorticotrophic hormone, the peptide responsible for stimulating corticosterone secretion from the rodent adrenal gland. It is made in the anterior pituitary, where it is destined for secretion into the circulation. It would not be useful to make it in the brain as it would reach the adrenal with difficulty. In contrast, other forms, such as the melanocyte-stimulating hormone peptides and  $\beta$ -endorphin, are made in the brain, where they serve neurotransmitter or neuromodulator roles.

In the case of CCK, a similar reasoning applies. CCK is a major gastrointestinal hormone that is released into the bloodstream from the intestine after a meal to regulate food intake, gallbladder contraction, and pancreatic enzyme secretion. The larger forms of CCK made and released by the intestine are more stable in circulation than the smaller forms, so they make better hormonal agents. CCK8, the major form found in brain, is a neurotransmitter or neuromodulator. It is released in small amounts at the synaptic cleft, excites its receptor, and is rapidly degraded.

Another interesting aspect of the differential processing of pro-CCK and gastrin involves the existence of glycine-extended peptides. Different tissues contain varying amounts of glycine-extended CCK and gastrin peptides. These peptides have not been converted to

amides by the amidating enzyme. Interestingly, these peptides are reported to have their own receptors and biological activity distinct from the receptors that bind amidated CCK and gastrin peptides. In this case, regulation of PAM activity may be controlling the production of these peptides to subserve specific physiological functions.

### See Also the Following Articles

Osteoporosis in Reproductive Age Women: Impact of Sex Steroid Replacement • Peptide Hormones and Growth Factors: Cellular Signaling Mechanisms •

Peptide Hormones, Intracellular Transport • Peptide Hormones, Regulated Secretion • Peptide Hormones, Regulation and Gene Expression • Peptide Hormones, Segregation Mechanism • Peptide Hormones, Subcellular Structure • Prohormone Convertases

### Further Reading

- Beinfeld, M. C. (2003). Biosynthesis and processing of pro CCK: Recent progress and future challenges. *Life Sci.* **72**, 747–757.
- Hook, V. Y. H. (ed.) (1998). “Proteolytic and Cellular Mechanisms in Prohormone and Proprotein Processing.” R. G. Landes, Austin, TX.
- Strand, F. L. (1999). “Neuropeptides.” MIT Press, Cambridge, MA.

maintaining homeostasis, enabling the organism to develop, grow, and reproduce normally and to respond effectively to a wide range of environmental demands and stresses. By definition, they are released into the circulation and thereby reach distant targets, but many are also released locally where they influence neighboring cells via paracrine mechanisms or signal across synaptic junctions in the case of neurons in the central nervous system (CNS).

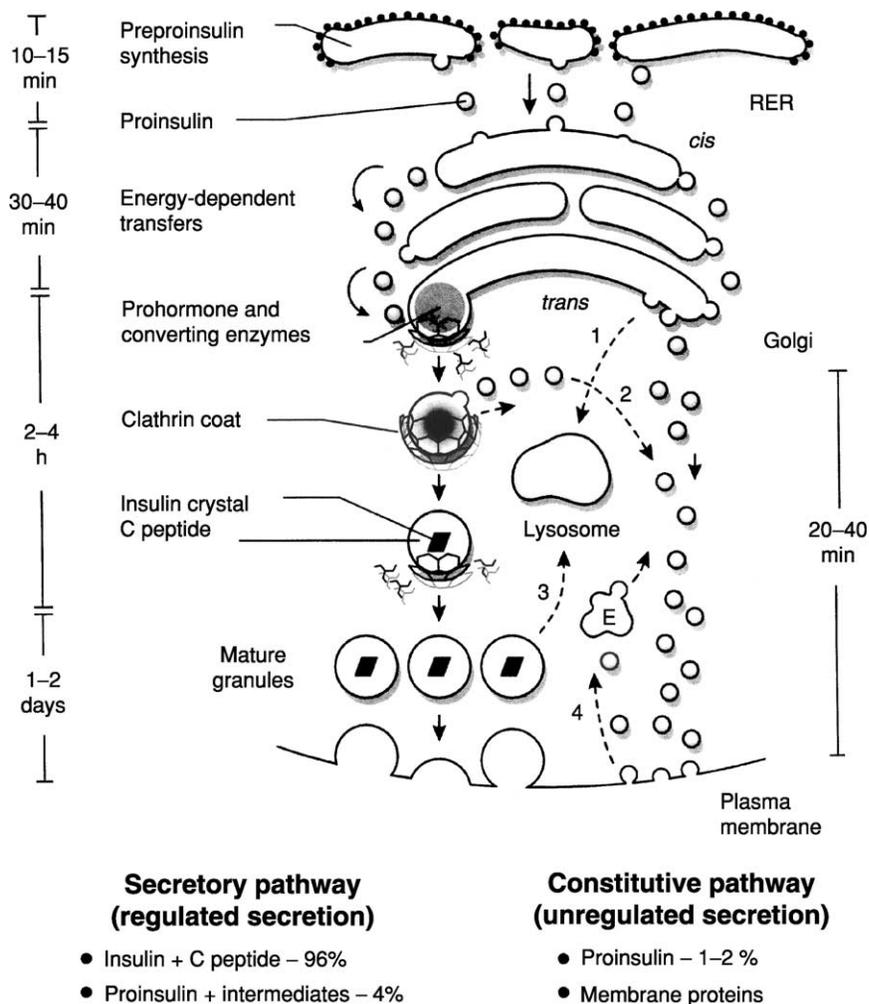
These important regulatory peptides or their precursors are all encoded in typical genes that are expressed via the conventional transcriptional and ribosomal mechanisms of protein biosynthesis. However, they are translated initially as presecretory proteins that are directed into the secretory pathway at the level of the endoplasmic reticulum (ER) via N-terminal signal peptides that interact with SRP (signal recognition particle) and SRP receptors. During or shortly after translocation, they fold and undergo signal peptide cleavage in the ER cisternae, are transported to the Golgi apparatus, and are then stored intracellularly in mature secretory vesicles (storage granules) awaiting secretory signals or, in some instances, are secreted constitutively in an unregulated manner (Fig. 1). In rare instances, proteins destined for secretion may lack a typical N-terminal signal peptide, achieving entry into the secretory pathway or direct exit from the cell by alternative intrinsic motifs acting via mechanisms that are not well understood. Important requirements for most peptide hormones and neuropeptides are that their sorting and retention within secretory granules are highly efficient and that they are discharged by exocytosis only in response to nutrients, metabolites, neurotransmitters, or other hormonal signals, i.e., that their secretion is tightly regulated. Achieving this requires that neuroendocrine cells express certain gene combinations, or modules, that encode components of a regulated secretory pathway to receive and store the secreted products, as well as mechanisms for directing the newly synthesized hormonal peptides, often as precursors, into this highly developed arm of the secretory pathway. This sorting mechanism must either actively or passively allow these products to gain access to secretory granules as these form in the *trans*-Golgi network (TGN). Once formed, secretory granules mature in the cytosol and become competent for secretion. They then are released in response to signals by exocytotic fusion of the granule membrane with the plasma membrane.

## REGULATION OF PEPTIDE HORMONE BIOSYNTHESIS

The biosynthesis of most peptide hormones is regulated at multiple levels. The rate of transcription of the gene encoding the hormone is usually responsive to the level of demand for the hormone. The transcriptional complex is usually made up of combinations of various transcription factors, coactivators, repressors, and elements of the basal transcriptional machinery. Individual components may also be regulated by phosphorylation/dephosphorylation, proteolysis, or other rapidly acting modulatory signals. The stability of the mRNA is also regulated in many instances. However, these regulatory actions can act only at a relatively slow pace over periods of hours or even days to alter the levels of mRNA for the hormone, as it usually constitutes a very major proportion of the total mRNA of the endocrine cell. For example, insulin mRNA makes up approximately 10% of the total pancreatic beta cell mRNA population. Hence, it is difficult to rapidly modulate proinsulin synthesis on a minute-to-minute basis via this mechanism, whereas long-term adjustments to increased demand rely on this level of control.

The rate of translation of the particular mRNA encoding a hormone is also often closely controlled, independently of other cellular mRNAs, leading to rapid changes in the rate of hormone biosynthesis. This modulation occurs largely at the level of initiation of translation through mechanisms that vary the number of mRNA molecules being translated at any given time. The precise mechanism of regulation of translation remains unclear, but clearly involves signal transduction pathways that can act within seconds to minutes to modulate both initiation and elongation rates, but only the former is known to be specific for a particular (hormonal) product. As a result, the rate of synthesis can be greatly accelerated or shut down very rapidly in response to signals such as the glucose level in the pancreatic beta cell.

The result of the operation of the above-described mechanism regulating hormone biosynthesis is to maintain a relatively constant large pool of hormone in the secretory granules (for example, there are as many as 13,000 granules containing insulin, approximately  $5 \times 10^{-15}$  mol in each beta cell). Active bursts of secretion can lead to rapid fluctuations (5–20%) in the size of the granule pool, which is then replenished more slowly, by synthesis, usually over a period of several hours. Prolonged or excessive stimulation of secretion, however, can



**Figure 1** Subcellular organization of the biosynthetic pathway of a protein/peptide hormone within a typical neuroendocrine cell—the pancreatic islet beta cell. The preprohormone is synthesized by ER-associated ribosomes (RER) and then is cotranslationally sequestered and cleaved to proinsulin within the lumen of the ER. The prohormone chain folds and forms the correct disulfide bonds of insulin within this compartment and then the fully structured proinsulin molecules are transferred to the *cis*-Golgi region in small coated vesicles. After traversing the Golgi, the prohormone is sorted from the *trans*-Golgi network (TGN) into secretory granules, processed by convertases to insulin, and ultimately released from the cell by exocytosis, as shown. Arrows indicate the direction of movement via subcellular compartments. The regulated secretory pathway is shown on the left and the unregulated, or constitutive, pathway is shown on the right. Timescales indicate intracompartamental residence times. Dashed arrows indicate alternative or minor pathways of (pro)hormone sorting as follows: (1) from the TGN to the lysosomes via the mannose-6-phosphate pathway for degradation; (2) from the early immature secretory granules via small clathrin-coated vesicles that link to the constitutive pathway, i.e., “constitutive-like” secretion; (3) from mature granules to lysosomes, a pathway of intracellular degradation of unused mature secretory granules via autophagy, and (4) a proposed endocytic pathway for recycling excess membrane from the cell surface after exocytosis, along with membrane-associated processing enzymes, such as furin, via endosomes (E), ultimately to the Golgi apparatus (see text for details). Modified from Steiner, D. F. (2001). Biosynthesis of insulin. In “Handbook of Physiology” (H. M. Goodman, ed.), Section 7, “The Endocrine System,” Vol. 11 (L. S. Jefferson and A. D. Cherrington, eds.), pp. 49–78. Oxford University Press, New York.

deplete the granule pool and may lead to abnormal states of glandular hyperplasia, associated with chronic increased production combined with diminished stores due to accelerated turnover of secretory granules.

### POSTTRANSLATIONAL PROCESSING OF HORMONE PRECURSORS

Most small peptide hormones, growth factors, and neuropeptides are derived from larger precursor

proteins (proproteins), which are synthesized and transported, as outlined in Fig. 1. After reaching the TGN, these precursor forms begin to interact with specialized endoproteolytic enzymes, the prohormone or proprotein convertases, commonly referred to as PCs, which carry out the proteolytic excision of their product hormone or hormones. The convertases are serine proteases, related to bacterial subtilisin and the yeast convertase, kexin (Fig. 2). Some members of the nine-member convertase family (furin, PC6B, and PC7) are tethered by C-terminal transmembrane domains to the TGN membrane and cycle to the cell surface and endosomes in constitutive vesicles, whereas others, such as PC2, PC1/3, and PC5/6A, are efficiently sorted into newly forming secretory granules along with newly synthesized precursor proteins and carry out their proteolytic processing within this subcellular compartment. The processing of a much-studied typical peptide hormone precursor, proinsulin, is illustrated schematically in Fig. 3. The convertase PC1/3 acts first to cleave on the C-terminal side of the pair of arginine residues spanning the junction between the insulin B chain and C peptide (B–C junction), creating an intermediate form: 32,33 split proinsulin. A second enzyme, carboxypeptidase E (CPE), an exoprotease, then rapidly removes the basic residues that remain attached to the C terminus of the B chain. PC2 then preferentially recognizes this intermediate and cleaves at the basic pair linking the C peptide to the A chain (C–A) junction, followed by removal of the basic residues from the C peptide by CPE. This sequence of four reactions gives rise to mature insulin and C peptide, the two major secreted products of the beta cell. These products remain stored in mature secretory granules pending their release by exocytosis in response to glucose or other signals or until their ultimate intracellular turnover if they remain unused.

Proinsulin conversion provides a paradigm for the proteolytic processing of most, but not all, of the many other neuroendocrine precursor proteins. Intragranular processing is also an appropriate mechanism for the processing of neuropeptide precursors. These are sorted into small, dense-core vesicles in the neuronal cell body and undergo maturation as the vesicles migrate along microtubular networks within axons to distal synaptic junctions, where they are stored and secreted. Proteolytic processing thus typically begins 15–30 min after biosynthesis of the precursor, as it initially must traverse the secretory pathway from the ER via the Golgi apparatus and into the granules, and it then continues over a period of several hours. The overall time-course exhibits

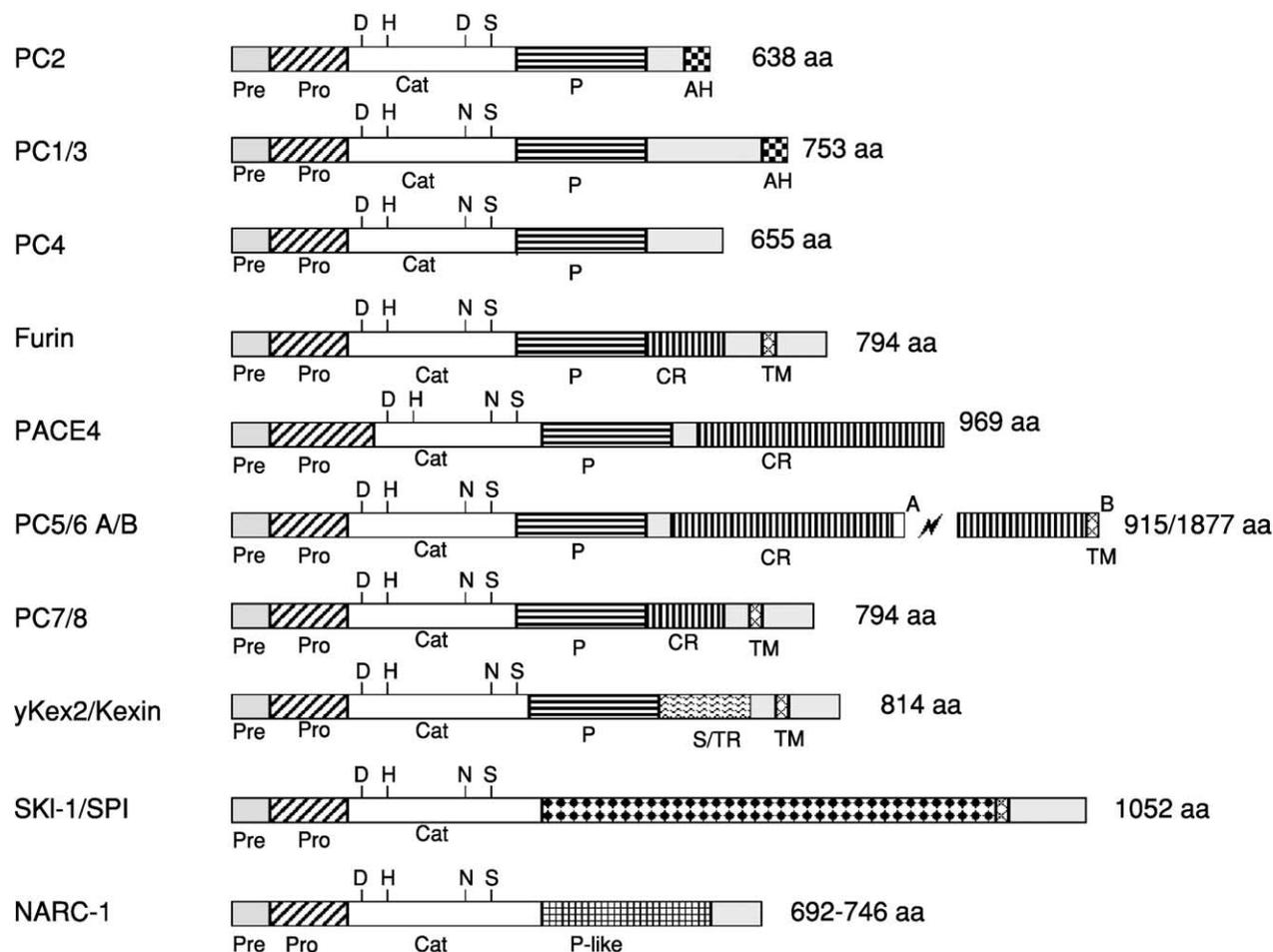
pseudo-first-order kinetics with half-times for precursor to product disposition of 30 min to 1 h in warm-blooded species. This relatively slow time-course of maturation of secretory products is thus well adapted to the kinetics of axonal transport, but also is well suited to most endocrine cells, which usually contain large granule stores that are only partially depleted by normal secretory demands.

## SPECIFIC EXAMPLES OF PROHORMONE PROCESSING

### Islet Hormones

Whereas both PC2 and PC1/3 are expressed at high levels in the insulin-producing pancreatic beta cells, the islet alpha, delta, and gamma cells that produce glucagon, somatostatin-14 (SS-14), and pancreatic polypeptide, respectively, express only PC2 and this convertase acts alone to cleave these hormones from their precursors. PC2 null mice have been produced and are essentially completely blocked for glucagon production, exhibiting a chronic mild hypoglycemia and alpha cell hyperplasia that can be corrected by chronic glucagon administration. The processing of proglucagon is more complex, however, as this multifunctional precursor also contains the additional biologically active peptides, glucagon-like peptide-1 (GLP-1) and GLP-2. These peptides are excised from proglucagon when it is expressed in intestinal L cells that contain only, or mainly, PC1/3. This convertase processes proglucagon at several distinct sites to release GLP-1 (7–37) and GLP-2, whereas glucagon remains in an inactive form as part of larger N-terminal fragments (glycetin and/or oxyntomodulin), as illustrated in Fig. 4. Note that in pancreatic alpha cells, GLP-1 and GLP-2 are released only as inactive components of the larger, unprocessed C-terminal fragment called the “major proglucagon fragment.” Prosomatostatin processing to SS-14 in the islet cells, as noted above, requires PC2 and this is also the case in brain. In PC2 null mice, SS-28 accumulates instead of SS-14 through the action of an as yet unidentified convertase (Fig. 4).

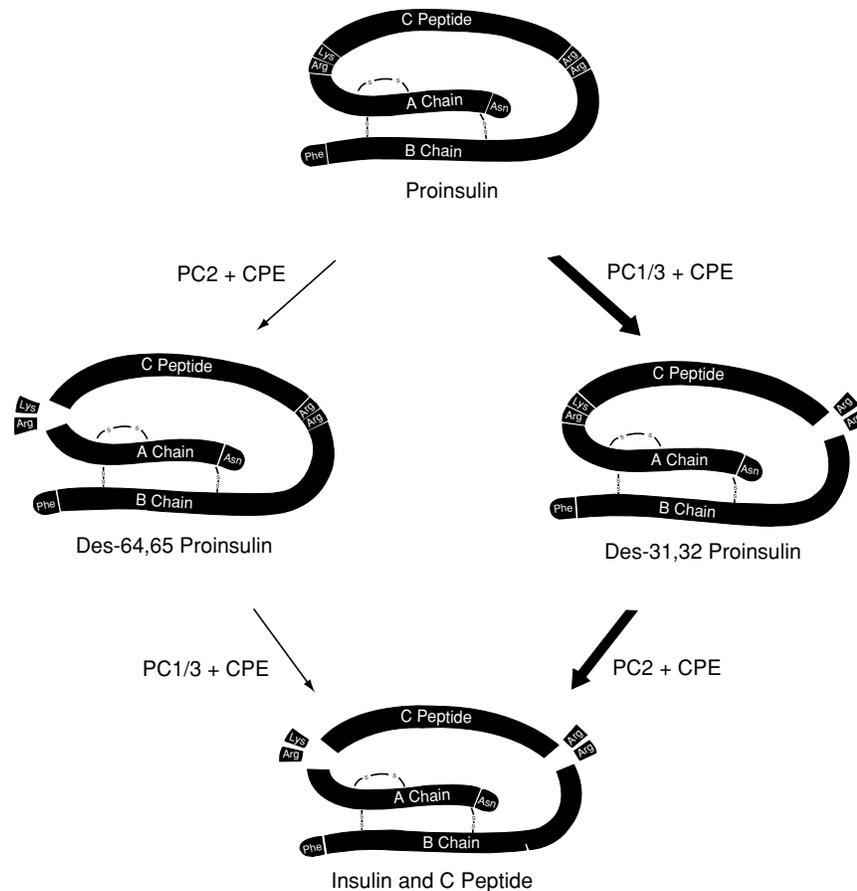
In the beta cells, PC2 ablation leads to significant (30–35%) accumulations of proinsulin, including elevated levels of des-31,32 proinsulin intermediate (normal value less than 5%), whereas PC1/3 ablation causes a more severe proinsulin processing defect, associated with high levels of intact proinsulin and elevated des-64,65 proinsulin intermediate (85–90% of total insulin-related material). Thus, PC1/3 appears to be more important than PC2 for proinsulin processing



**Figure 2** Structural organization of the eukaryotic prohormone convertase family. All members have a preregion (Pre) and a proregion (Pro) followed by a subtilisin-like catalytic domain (Cat) having conserved Asp (D), His (H), Asn (N), and Ser (S) residues that form the catalytic triad (D, H, S) and the transition state-stabilizing oxyanion residue (N). Note that the latter is (D) rather than (N) in all PC2s examined in both vertebrates and invertebrates. The catalytic domain is followed by the conserved and essential P domain, which is necessary for the folding and stability of the enzymes. The variable C-terminal region determines the subcellular localization and sorting of the convertases. Those with transmembrane (TM) and/or cysteine-rich (CR) regions (e.g., furin, PC5/6B, PC7) are usually found in the distal Golgi or TGN, where they become active at the slightly acidic pH of this compartment and function mainly in the constitutive pathway. The convertases usually undergo autocatalytic propeptide cleavage as they exit the ER, but do not release the propeptide or become active on substrates until they reach their Golgi/TGN destination. Pro-PC2 is again exceptional in that propeptide cleavage occurs in a later compartment (TGN/secretory granules) and activation follows this cleavage only if pro-PC2 has interacted with NE protein 7B2 during intracellular transit to the Golgi. (See text for more details.) All the convertases and the yeast furin homologue kexin require micro- to millimolar calcium concentrations for activity. Kexin has a serine/threonine-rich region following the P domain (S/TR). The convertases lacking TM domains, especially PC2, PC1/3, PC4, and PC5/6A (A), are not retained in the Golgi, but instead enter the regulated secretory granules where their more acidic enzyme optimum (pH 5–6) enhances their processing. Most remain intact, except for PC1/3, which loses its C-terminal 137-residue tail via cleavage at an intrinsic dibasic site either by its own action or by that of other convertases within the secretory granules. This truncated form has enhanced processing activity on some PC1/3 substrates. Shown below kexin are two novel convertases, SKI-1/SPI and NARC-1. These more distantly related forms are also expressed in the secretory pathway and process a variety of precursors, but have significantly altered substrate specificity. Whereas most convertases generally cleave only at sites containing K/R-R↓, R-X-K/R-R↓, or R-X-X-R↓, the novel convertases prefer sites having a hydrophobic residue in both the P1 and P2 positions, e.g., R-X-V-L↓ or Y-V-L-L↓. They also lack other features, having either a highly modified P-like domain (NARC-1) or an unrelated sequence (SKI-1) (A = PC5/6A; B = PC5/6B). AH, amphipathic helix.

and, though either convertase alone is capable of producing some fully processed insulin in the absence of the other (Fig. 4), both are required for normal highly

efficient processing. In contrast, the precursor of islet amyloid polypeptide (amylin), a peptide related to calcitonin/calcitonin gene-related peptide (CGRP), which



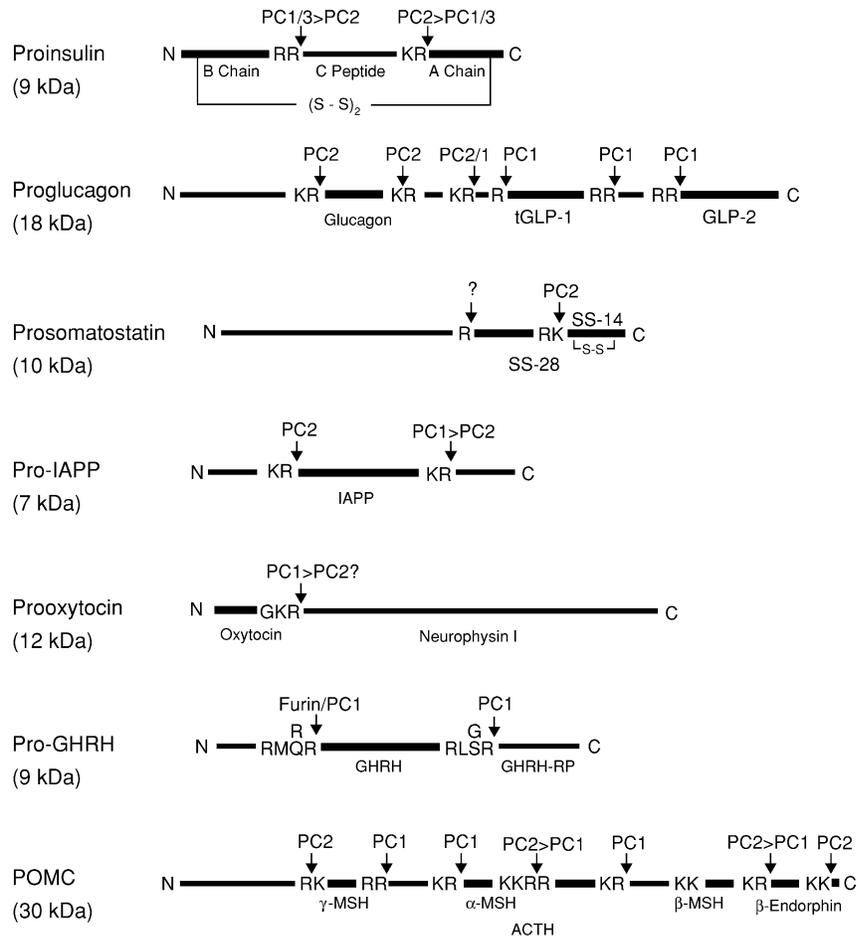
**Figure 3** Processing of proinsulin by PC1/3, PC2, and CPE in the pancreatic  $\beta$  cell. Although both convertases PC1/3 and PC2 are able to process either of the two sites that must be cleaved to remove the C peptide, PC1/3 prefers the B/C junction and PC2 prefers the C/A site. Moreover, cleavage of the B/C site generates a preferred substrate for PC2, the des-31,32 proinsulin intermediate. Consequently, the pathway shown on the right-hand side is predominant and the pathway on the left probably occurs only to a limited extent. In mice lacking PC2, some insulin is generated (approximately 65% of total insulin-like material) along with elevated levels of des-31,32 proinsulin and intact proinsulin. However, as predicted from this formulation, PC1/3 null mice have a much more severe block in proinsulin processing (less than 10–15% of insulin is generated) associated with a small buildup of des-64,65 proinsulin with large amounts of intact proinsulin. CPE acts after each convertase to remove the two C-terminal basic residues in a stepwise fashion from the split products generated by the initial PC1/3 or PC2 cleavages, giving rise to the two naturally occurring major stored products in beta cells, insulin and C peptide (bottom structure). The fate of the basic residues released in precursor processing is not known.

is also expressed mainly in the islet beta cells, is more dependent on PC2 than PC1/3 for its processing. In PC2 nulls, only N-terminally extended pro-islet amyloid polypeptide (pro-IAPP) intermediate material is produced, whereas PC1/3 nulls exhibit only a modest accumulation of unprocessed material (see Fig. 4).

### Gastrointestinal Hormones

Whereas PC1/3 clearly plays an essential role in proglucagon processing in the intestinal L cells to generate GLP-1 and GLP-2, as mentioned above, it is also

necessary for cholecystokinin (CCK) maturation in intestinal I cells, as well as a major portion of gastrin maturation in antral G cells. However, neural processing of CCK in PC1/3 null mice is unaffected, indicating the likely involvement of other convertases in the CNS processing of CCK precursors. The effects of PC2 deficiency on gastrin processing are also partial, indicating that other convertases expressed in the antral gastrin-producing G cells (e.g., PC1/3 and PC5/6) may also be capable of cleaving at specific K-K sites to generate mature gastrin. Such sites, which have a P1 (first upstream position) lysine (K) rather than an arginine (R) residue, are frequently substrates for PC2.



**Figure 4** Selected examples of peptide prohormone precursor structure illustrate the diversity in size and structural organization of this subset of the many kinds of proproteins that are processed in the secretory pathway by the PC family of convertases. The N-terminal signal peptides are not shown here, as these are usually cleaved off by signal peptidases in the ER before synthesis of the entire peptide chain is completed (i.e., within the first 1–2 min postinitiation of translation *in vivo*). The active regions in these precursors are all represented by thicker lines and the surrounding inactive prosequences are represented by thin lines. The lysine (K) and/or arginine (R) residues at cleavage sites are indicated and usually occur as pairs. However, in some precursors, upstream basic residues are sometimes present, as indicated in the GHRH sequence. Above each site is indicated the convertase or convertases that cleave at that site in order of their processing efficiency at that site, as shown by *in vivo* or *in vitro* studies, as discussed in the text. Superscripts R and G above the pro-GHRH sequence denote substitutions found in the human vs mouse proteins.

Other gut hormones, such as motilin, ghrelin, secretin, peptide YY, vasoactive intestinal polypeptide, and glucose-dependent insulinotropic peptide/gastric inhibitory peptide, are also derived from precursors, but little is known about their processing requirements.

### Hypothalamic and Pituitary Hormones

Adrenocorticotrophic hormone (ACTH) is the only major anterior pituitary hormone that is derived from a proprotein precursor. Its precursor, proopiomelanocortin (POMC), is perhaps the most remarkable

example of a large multifunctional precursor (Fig. 4). POMC contains within its ~30 kDa sequence a large N-terminal proregion that is believed to be involved in its sorting into the regulated secretory pathway followed by a melanocyte-stimulating hormone (MSH)-like sequence,  $\gamma$ -MSH, that has been found to play a role in blood pressure regulation via the melanocortin 3 receptor (MCR3). A joining peptide is then followed by the 39-residue ACTH sequence, the first 13 amino acids of which also can be cleaved out, acetylated, and amidated to produce  $\alpha$ -MSH (melanocortin), another important MSH-like sequence that acts via MCR4 to regulate

food intake and through MCR1 to influence pigmentation. The ACTH sequence is followed by that of  $\beta$ -lipoprotein, a large C-terminal peptide that contains another MSH-like sequence ( $\beta$ -MSH) and ends in  $\beta$ -endorphin, an opioid peptide family member that contains the 5-residue Met enkephalin sequence at its N terminus.

The processing of POMC differs significantly between the anterior and intermediate lobes of the pituitary. Anterior pituitary corticotrophs express mainly PC1/3 and this convertase is responsible for selected cleavages in POMC that produce mainly ACTH and smaller amounts of  $\beta$ -endorphin (Fig. 4). These peptides are secreted in response to stressful stimuli under the influence of hypothalamic corticotrophic-releasing hormone/factor (CRH/CRF) transported to the pituitary directly via the hypothalamic–hypophyseal portal vein system. Humans or mice lacking PC1/3 are unable to produce mature ACTH, but exhibit only a mild degree of adrenocortical insufficiency. In the neurointermediate pituitary lobe (more prominent in rodents, absent in human), both PC1/3 and PC2 are expressed and the added PC2 acts mainly to increase the production of the MSH-like peptides ( $\alpha$ ,  $\beta$ , and  $\gamma$ ) from POMC (Fig. 4). CPE or its homologues, N-acetylases and the bifunctional peptidyl amidating monooxygenase (PAM), also act on many of these products to acetylate N termini and/or remove C-terminal basic residues to expose C-terminal glycine residues that are then converted to carboxamides by the action of PAM.

The posterior pituitary hormones oxytocin and vasopressin are derived from homologous precursors that also contain their corresponding neurophysins or binding proteins (Fig. 4). The nonapeptide hormones lie at the amino terminus in their precursors and thus only a single cleavage is required for their maturation (after signal peptide excision). It appears that either PC1/3 or PC2 or both of these convertases may be involved in the processing of provasopressin and prooxytocin in their cells of origin—the magnocellular neurons of the supraoptic nuclei at the base of the hypothalamus.

The hypothalamus is also the source of several releasing factors that control the secretion of anterior pituitary hormones, as well as various other peptide hormones/neuropeptides that regulate appetite, food intake, and energy balance. The four major hypothalamic releasing factors are CRH, gonadotropin-releasing hormone (GnRH), growth hormone-releasing hormone (GHRH), and thyrotropin-releasing hormone (TRH). The first three are derived from linear proproteins having an AB (CRH), a BC (GnRH), or

an ABC structure (GHRH), where B represents the active sequence and A and C are propeptides. Processing on the N-terminal side of GHRH can be accomplished *in vitro* by furin, PC1/3, and PC5A, whereas the C-terminal site is cleaved preferentially by PC1/3 (Fig. 4). In mice lacking PC1/3, GHRH is not processed normally, leading to the accumulation of higher molecular weight forms that lack activity. As a consequence, the mice exhibit postnatal growth retardation accompanied by low levels of growth hormone and insulin-like growth factor-I. The mode of processing of the other releasing factor precursors has not yet been investigated in detail. However, GnRH function is relatively normal in both PC1/3 and PC2 null mice, whereas TRH production appears to be reduced in PC1/3 nulls. *In vitro* studies of TRH processing indicate that both PC1/3 and PC2 are required for efficient processing of its multicopy precursor, which contains at least five copies of the sequence Lys-Arg-Gln-His-Pro-Gly-Lys-Arg. After PC cleavage, CPE and PAM then process it to give rise to Gln-His-Pro-NH<sub>2</sub>, which is then cyclized at position 1 to pyroGlu-His-Pro-NH<sub>2</sub>, the mature releasing hormone.

Peptides involved in hypothalamic regulation of food intake include leptin, agouti-related protein (AGRP),  $\alpha$ -MSH, neuropeptide Y (NPY), GLP-1, cocaine and amphetamine related transcript (CART), orexin-A, and orexin-B. As noted earlier,  $\alpha$ -MSH processing requires PC2 and PC1/3, whereas GLP-1 processing requires only PC1/3. The CART precursor, pro-CART, also requires both PC1/3 and PC2 to produce the two major active isoforms CART I (55–102) and CART II (62–102). However, PC2 is the major convertase producing CART II. The CART precursor exists in two isoforms differing in length in the N-terminal propeptide region due to alternative splicing of the mRNA. Both PC2 and PC1/3 have been reported to be able to process pro-NPY, but *in vivo* PC2 may predominate in sympathetic neurons. The requirements for processing pro-NPY are not well defined. Leptin, AGRP, and the orexins are not processed by convertases.

The opioid peptides arise mainly from three major precursors; proenkephalin, prodynorphin, and POMC. Both PC1/3 and PC2 appear to be involved in the full maturation of these multicassette precursors, as noted earlier in the case of POMC. The PC2 null mice, however, exhibit major deficits in enkephalins and dynorphins, suggesting that this convertase plays a major role in processing *in vivo*. A dual role for both convertases has been found in processing the proorphalin/nociceptin precursor and also the promelanin-concentrating hormone precursor. The neurotensin/

neuromedin precursor is processed by PC2 together with PC1/3 or PC5/6A.

### Parathyroid Hormone and Calcitonin

The biosynthesis of parathyroid hormone (PTH) differs significantly from that of most of the others discussed above because of its unusual structure, mode of processing, and secretory regulation. Pro-PTH bears only a hexapeptide extension at its N terminus, analogous to that in proalbumin (Fig. 4). Proalbumin is processed by furin and is secreted constitutively from hepatocytes. However, PTH secretion is regulated primarily by circulating calcium levels. The PTH prohormone is also processed by furin and/or PC7 and is then either secreted or degraded intracellularly by lysosomal cathepsins, dependent on the ambient calcium concentrations, rather than being stored. This unique mechanism seems to represent a special adaptation to regulation via calcium sensing. Less is known regarding the biosynthesis and processing of procalcitonin (Fig. 4), aside from the derivation of its mRNA via alternative splicing of a single gene that encodes both calcitonin and the homologous CGRP on separate interchangeable exons. CGRP is expressed mainly in the brain and autonomic systems, whereas calcitonin is secreted mainly from the thyroid medullary C cells. Both PC1/3 and PC2 have been identified in calcitonin/CGRP-producing cells. IAPP is also a member of this family (Fig. 4).

## ADDITIONAL POSTTRANSLATIONAL MODIFICATIONS

### Glycosylation

Both N-glycosylation and Ser/Thr glycosylation occur in the secretory pathway. N-Glycosylation involves the addition of preformed mannose-rich immature polysaccharides to asparagine residues. It usually occurs early in the secretory pathway in the ER, where it assists in peptide folding and ER quality control. The glycoprotein hormones thyroid-stimulating hormone (TSH), follicle-stimulating hormone (FSH), and luteinizing hormone (LH) are the most prominent examples in this category, but prohormones, such as POMC, also are N-glycosylated. In some instances, correct N-glycosylation influences biological activity, e.g., TSH glycosylation is regulated by TRH. Rarely, O-glycosylation of serine or threonine residues takes place during later stages of transit from the ER to the Golgi or in the *cis* or medial Golgi cisternae after peptides have exited the ER. This type of modification has been described in POMC and is also

postulated to occur in proglucagon, but its role remains obscure.

### Other Modifications

#### C Terminus

Many neuropeptides are modified by amidation to enhance their stability and/or receptor-binding potency. This modification usually results whenever a C-terminal glycine is present or is exposed by the actions of a carboxypeptidase E-like activity during the maturation of a precursor. The glycine then becomes a substrate for the dual-activity PAM complex, a single polypeptide having two enzymatic activities. Many examples of such modifications exist, including GLP-1 (7–37)-amide, vasopressin, oxytocin, and  $\alpha$ -MSH.

#### N Terminus

Peptide hormones are sometimes modified by N-acetylation, as in the case of  $\alpha$ -MSH. This modification results from the transfer of an acetyl group from acetyl coenzyme A to the N-terminal amino group. Other N-terminal modifications result from cyclization of N-terminal glutamine to pyroglutamic acid. Such modifications enhance stability by preventing the action of aminopeptidases. In very rare instances, amino mono- or dipeptidases act on precursor proteins following their maturation to remove interfering N-terminal residues. Examples are the activation of the yeast  $\alpha$  mating factor and honey bee promellitin by an aminodipeptidase, which removes several successive Glu-Ala dipeptides as secondary cleavages during the maturation of these biologically active peptides.

#### Sulfation/Phosphorylation

Phosphorylation of serine or threonine residues occasionally occurs, e.g., in POMC. However, the role, if any, of this modification is not apparent. Sulfation of tyrosine residues, however, is found in gastrin and some other peptide hormones, where it often enhances biological activity.

#### Lipid Modifications

A number of lipid modifications of proteins are well known, including prenylation, palmitoylation, and cholesterol addition; these modifications have not yet been described in peptide hormones. However, it has been shown that ghrelin, a hormone produced in the stomach that regulates appetite and also stimulates growth hormone release, is octanoylated at position 3,

a serine. This modification is necessary for biological activity and also might be expected to influence the behavior of the peptide in the circulation due to its added lipophilicity. The mechanism of addition of octanoate is not known.

## DEFECTS IN PEPTIDE HORMONE BIOSYNTHESIS

### Inherited Defects

Hormone deficiencies can arise from defects at many levels in the processes that underlie the production and/or action of hormones. Genetic defects play a prominent role by affecting the structure, synthesis, folding, intracellular transport, processing, and release of hormonal peptides. A number of inherited defects in proinsulin structure that influence the folding, sorting, processing, and/or biological activity of insulin have been described. The Akita mouse is diabetic due to a point mutation that changes the A7 cysteine to arginine in proinsulin II, resulting in misfolding of proinsulin, severe ER stress, and ultimately beta cell failure and cell death, leading to hyperglycemia. Another point mutation at position B10 (His → Asp) in human proinsulin causes selective mis-sorting of some of the mutant proinsulin into the constitutive secretory pathway, causing elevated circulating proinsulin levels. Other insulin gene mutations have less dramatic phenotypes, but are often associated with mild familial MODY (maturity-onset diabetes of the young)-like diabetes (autosomally dominant forms of diabetes). Some of these single point mutations affect proinsulin cleavage sites, especially at the C peptide–A chain junction, and lead to elevated circulating levels of partially cleaved proinsulin intermediates, whereas those of another group affect insulin structure, lowering its biological activity by factors of up to 1000-fold. Defects of a similar nature that influence the biosynthesis, translocation, folding, processing, or receptor binding and activity have also been described in a number of other protein hormones or their precursors, including growth hormone, leptin, POMC, PTH, TSH, LH, FSH, and vasopressin in both human and experimental animals (e.g., vasopressin in the Brattleboro rat and leptin in the ob/ob mouse). Such mutations cover the gamut of possibilities, including disordered splicing, frame-shifts, deletions, and point mutations. Gene disruptions of a number of hormones or their precursors, including growth hormone, prosomatostatin, proenkephalin, POMC, pro-TRH, and pro-NPY, have also been carried out in the laboratory.

Inherited defects in transcription factors that can influence hormone gene expression and secretion have also been described, in addition to the large numbers of such mutations that have already been described in the membrane and nuclear receptor families that influence hormone action. Mutations in PDX-1, a beta cell-specific transcription factor that influences both proinsulin and prohormone convertase gene transcription, lead to diminished insulin production, giving rise to the MODY4 syndrome. Other forms of MODY are caused by mutations of the beta cell glucose-sensor protein, glucokinase (MODY2), or in several hepatic nuclear factor genes, including HNF4 $\alpha$  (MODY1), HNF1 $\alpha$  (MODY3), and HNF1 $\beta$  (MODY5), that regulate beta cell glucose metabolism and insulin secretory responses to glycemic stimuli.

### Processing Enzyme Defects

It is clear that defects in the prohormone convertases can produce multifaceted endocrine/metabolic disorders in human subjects, as well as in knockout mouse models. Some of the consequences of gene disruption of PC1/3 and PC2 have already been described above. PC2 null mice are surprisingly vigorous despite the lack of glucagon, SS-14,  $\alpha$ -MSH, and  $\gamma$ -MSH and relatively severe deficits of opioid and several other neuropeptides. Interestingly, many of the features of PC2 ablation are reproduced in the 7B2 null mouse. The multifunctional 28 kDa neuroendocrine protein 7B2 contains a potent PC2 inhibitory region at its C terminus, but also contains another domain that plays an essential role in pro-PC2 activation. Thus, 7B2 null mice have no active PC2 and share many aspects of the PC2 null phenotype, but unlike PC2 nulls they tend to oversecrete ACTH, which accumulates in the pituitary intermediate lobe in the absence of PC2, and this leads to the development of fulminating Cushing's-like syndrome early in life, leading to death at 5–6 weeks. Altered dopaminergic regulation of interleukin secretory activity seems to underlie this phenomenon, but strain differences in mice with this mutation also appear to contribute to its severity.

Mice lacking PC1/3, as noted earlier, exhibit abnormalities in growth, due to a lack of mature GHRH, but also have a complex polyendocrinopathy resulting in part from disordered hypothalamic/pituitary processing of POMC, as well as of gastrointestinal (GI) processing of proglucagon to GLP-1 and GLP-2, and the production of other GI peptide hormones, as well. These animals exhibit a mild chronic diarrhea with bulky, soft stools, without evident GI mucosal

pathology, suggestive of a motility disorder. In human, PC1/3 deficiency has been described in a 43-year-old woman of normal stature, but with severe obesity that began as early as age 4. This patient also exhibited mild hypogonadism and ACTH deficiency with elevated circulating POMC-like material. Like the PC1/3 null mice, this patient also had a very high proportion of circulating intact proinsulin associated with increased levels of des-64,65 proinsulin intermediate. Two very young subjects lacking PC1/3 have also been studied in the laboratory of Stephen O'Rahilly and associates at Addenbrooke Hospital (Cambridge, UK). Interestingly, these patients suffered from diarrhea and in this respect resembled the PC1/3 null mice.

Another convertase null, the PC4<sup>-/-</sup> mouse, exhibits defects only in reproductive functions, as this convertase is expressed mainly in testis and ovary.

A spontaneous mutation causing obesity in a mouse strain first identified at The Jackson Laboratory in the 1970s and initially called the "fat" mouse has been shown to be due to an inactivating point mutation in the CPE gene (CPE<sup>fat</sup>). These animals develop obesity and hyperglycemia, which in males progresses to diabetes. The processing of a large number of prohormones, including proinsulin, is disrupted by a lack of CPE in many parts of the neuroendocrine system. A surprising consequence is the powerful product inhibition of the endoproteolytic activity of the prohormone convertases by the buildup within secretory granules of C-terminally extended intermediates generated by their action. These animals exhibit a polyendocrinopathy with obesity as a major consequence, suggesting that CPE plays an important role in the processing of many key hypothalamic peptides involved in food intake and energy balance, acting in concert with the PCs. Interestingly, neither the PC1/3 or PC2 null mice exhibit obesity, for reasons that are not yet clear.

## Conclusions

It is clear that defects are possible at many levels in the biosynthesis and processing of peptide hormones, ranging from their structural genes, their regulatory elements, the specialized transcription factors that regulate the differentiation of their cells of origin or the expression of their genes per se, the regulation of their translation, folding, posttranslational processing, sorting, and retention in secretory granules, and the regulation of the signal transduction pathways and

ion channels leading to secretory granule release. Increased knowledge of such defects can contribute in major ways to the understanding of many common endocrine disorders, including diabetes, obesity, and other disturbances of growth, reproduction, food intake, energy metabolism, stress responses, and aging.

## See Also the Following Articles

Peptide Hormones and Growth Factors: Cellular Signaling Mechanisms • Peptide Hormones, Intracellular Transport • Peptide Hormones, Regulated Secretion • Peptide Hormones, Regulation and Gene Expression • Peptide Hormones, Segregation Mechanism • Peptide Hormones, Subcellular Structure • Prohormone Convertases

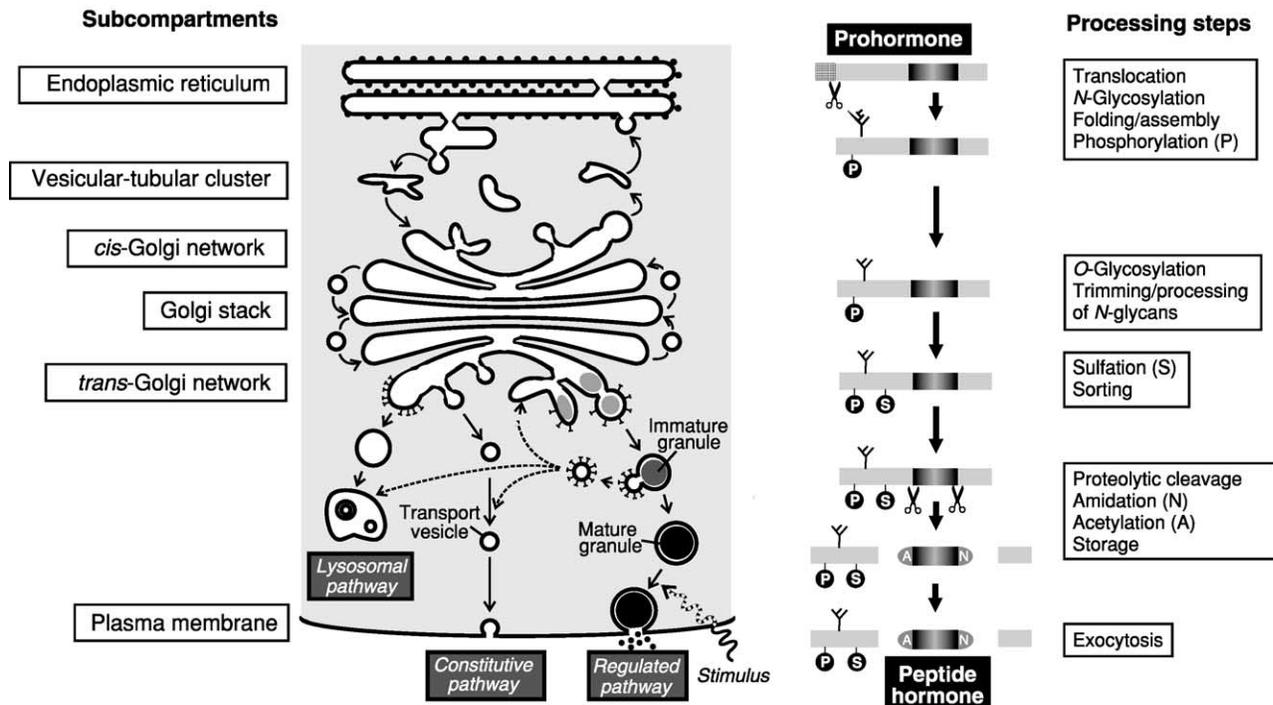
## Further Reading

- DeGroot, L. and Jameson, J. (eds.) (2001). "Endocrinology," 4th ed. Vols. 1–3, W. B. Saunders, Philadelphia, PA.
- Furuta, M., Yano, H., Zhou, A., Rouille, Y., Holst, J., Carroll, R., Ravazzola, M., Orci, L., Furuta, H., and Steiner, D. (1997). Defective prohormone processing and altered pancreatic islet morphology in mice lacking active SPC2. *Proc. Natl. Acad. Sci. USA* **94**, 6646–6651.
- Molloy, S., Anderson, E., Jean, F., and Thomas, G. (1999). Bicycling the furin pathway: From TGN localization to pathogen activation and embryogenesis. *Trends Cell Biol.* **9**, 28–35.
- Muller, L., and Lindberg, I. (2000). The cell biology of the prohormone convertases PC1 and PC2. In "Progress in Nucleic Acid Research and Molecular Biology," (K. Moldave, ed.), Vol. 63, pp. 69–108. Academic Press, San Diego, CA.
- Nilni, E. A. (1999). Neuroregulation of ProTRH biosynthesis and processing. *Endocrine* **10**, 185–199.
- Prigge, S. T., Mains, R. E., Eipper, B. A., and Amzel, L. M. (2000). New insights into copper monooxygenases and peptide amidation: Structure, mechanism and function. *Cell. Mol. Life Sci.* **57**, 1236–1259.
- Pritchard, L. E., Turnbull, A. V., and White, A. (2002). Pro-opiomelanocortin processing in the hypothalamus: Impact on melanocortin signalling and obesity. *J. Endocrinol.* **172**, 411–421.
- Seidah, N. G., and Prat, A. (2002). Precursor convertases in the secretory pathway, cytosol and extracellular milieu. *Essays Biochem.* **38**, 79–94.
- Westphal, C., Muller, L., Zhou, A., Zhu, X., Bonner-Weir, S., Steiner, D., Lindberg, I., and Leder, P. (1999). The neuroendocrine protein 7B2 is required for peptide hormone processing *in vivo* and provides a novel mechanism for pituitary Cushing's disease. *Cell* **96**, 689–700.
- Zhou, A., Webb, G., Zhu, X., and Steiner, D. F. (1999). Proteolytic processing in the secretory pathway. *J. Biol. Chem.* **274**, 20745–20748.
- Zhou, X., Zhou, A., Dey, A., Norrbom, C., Carroll, R., Zhang, C., Laurent, V., Lindberg, I., Ugelholdt, R., Holst, J., and Steiner, D. (2002). Disruption of PC1/3 expression in mice causes dwarfism and multiple neuroendocrine peptide processing defects. *Proc. Natl. Acad. Sci. USA* **99**, 10293–10298.

nuclei, lysosomes, and peroxisomes. Proteins that are destined to be secreted, such as (precursors for) peptide hormones, contain a signal peptide at their amino terminus and are translocated to the lumen of the rough ER after cleavage of the signal sequence. Once a secretory protein has reached the lumen of the ER, it does not have to cross an intracellular membrane any longer and moves toward the Golgi complex (Fig. 1). This complex consists of three functionally distinct compartments (*cis*, medial, and *trans*) that correspond to sequential cisternae (rigid lamellae) in its stack. Secretory proteins enter the Golgi on the *cis* side and exit it from the *trans* side to reach the *trans*-Golgi network (TGN) of the secretory pathway. In the TGN, regulated secretory proteins are sorted away from the constitutively secreted proteins, packaged, and concentrated into electron-dense, granular storage vesicles called secretory granules. The number of secretory granules may vary considerably within a cell depending on its functional state. The granules that bud off the TGN are immature and gradually develop to mature secretory granules that are stored in the cytoplasm until the proper and specific stimulus triggers the cell to release the regulated product by the process of exocytosis (fusion of the membrane of the secretory granule with the plasma membrane).

**PROCESSING STEPS IN THE ER**

Initially, the folding and assembly of newly synthesized proteins into their proper tertiary and quaternary structures were thought to be a spontaneous process. However, it is clear that protein folding in the cell needs to be controlled by helper proteins, termed molecular chaperones, which reduce the probability of formation of incorrect structures. Molecular chaperones are structurally unrelated proteins (e.g., heat shock proteins of the hsp90, hsp70, and hsp60 families) that are part of the cell's protein folding and degradation machinery and that mediate correct protein folding/assembly (without being themselves components of the final structures) or prevent aggregation reactions and premature folding in a number of compartments in a variety of cell types. Their expression is specifically induced by temperature shifts or other cellular stress conditions, but most chaperones are also expressed under normal growth conditions. Examples of ER molecular chaperones include the hsp70 family member immunoglobulin-binding protein (BiP, also known as GRP78), the calcium-binding, integral type I single-spanning membrane protein calnexin, the luminal calnexin-homologue calreticulin, and the soluble acidic protein disulfide isomerase



**Figure 1** The secretory pathway in an endocrine cell. Biologically inactive peptide hormone precursor proteins (prohormones) are transported through the various subcompartments of the regulated secretory pathway and undergo a number of processing steps, eventually leading to the regulated secretion of bioactive peptide hormones.

(PDI) that catalyzes thiol:protein-disulfide interchange reactions. Proteins that fold properly in the ER are efficiently transported through the secretory pathway. The export of proteins that fail to fold is prevented by a stringent process of conformation-based ER quality control. These unfolded proteins are rapidly degraded in a pre-Golgi degradation compartment, presumably often involving their reverse translocation back to the cytosol and degradation by a cytosolic proteolytic machinery (the ubiquitin–proteasome system). Accumulation of misfolded proteins in the ER may induce the so-called unfolded protein response, a mechanism protecting the cell against damage caused by improperly folded proteins. In addition to folding, most secretory proteins undergo a number of post-translational modifications in the ER, including disulfide bond formation (facilitated by PDI), glycosylation (proceeding to the TGN), phosphorylation, and oligomerization.

## TRANSPORT FROM ER TO GOLGI AND THROUGH THE GOLGI COMPLEX

Once secretory proteins are properly folded and modified in the ER, they move to specialized regions, the ER export sites, and are packaged in coated, irregularly shaped transport vesicles. Following uncoating, the transport vesicles form vesicular–tubular clusters [VTCs; also called the ER–Golgi intermediate compartment (ERGIC)] that move along microtubules to the Golgi. ER-resident proteins (such as the folding enzymes) contain retention and retrieval signals to hold them in this compartment. For example, lumenal ER proteins contain the tetrapeptide retrieval sequence Lys-Asp-Glu-Leu, in single-letter code KDEL, at their carboxyl terminus and these proteins are retrieved by the KDEL receptor. Although it is not clear whether a signal is needed for secretory proteins to leave the ER, receptors have been proposed to be involved in their exit from the ER. One example concerns ERGIC-53, a recycling type I transmembrane protein with an ER retention motif that functions as a mannose-binding lectin to extract secretory glycoproteins from the ER. Furthermore, soluble secretory proteins may be transported via the p24 family of type I transmembrane putative cargo receptors that are major constituents of the transport vesicles and continuously recycle between the ER and Golgi with their short, carboxy-terminal cytoplasmic domains acting as ER export or retrieval signals by binding cytoplasmic coat protein

complexes (COP-I and -II). Each of the transport steps connecting the ER and the Golgi, and connecting the Golgi compartments to one another, is unidirectional and energy- and GTP-dependent. Small GTP-binding proteins (20–30 kDa, e.g., the Rab proteins) are distributed throughout the secretory pathway with distinctive subcellular locations and regulate vesicular trafficking between the subcompartments. In addition, the traffic is regulated by integral membrane proteins [vesicular soluble *N*-ethylmaleimide-sensitive attachment protein receptors (v-SNAREs) and target (t)-SNAREs] for docking the vesicle to the membrane of the acceptor compartment and soluble proteins (NSF and SNAPs) to initiate vesicle fusion. In this vesicular model of protein transport, all anterograde (forward) and retrograde (backward) intra-Golgi steps involve transport only via vesicles. Alternatively, a nonvesicular transport mechanism may be present in which Golgi cisternae form at the *cis*-face of the stack, probably by VTC fusion, and then progressively mature into *trans*-cisternae (cisternal maturation model). In this model, cisternae (or intermittent tubular continuities) carry secretory cargo through the stack in the anterograde direction, and vesicles transport Golgi enzymes in the retrograde direction, allowing cisternal maturation to occur by progressive uptake of material from older stacks. At the TGN, the cisternae ultimately disintegrate and evolve into a collection of secretory vesicles, including immature secretory granules.

## SORTING IN THE TGN

If a protein does not possess any sorting signal other than the signal peptide sequence, this protein will be vectorially transported from the ER to the Golgi to the cell surface. Such proteins follow the constitutive pathway, a default route of nonselective bulk flow. The constitutive pathway is a basic feature of all secretory cells and is also responsible for the delivery of integral membrane components (e.g., receptors) to the cell surface. Signal-dependent protein sorting is thought to occur in the TGN. Proteins en route to lysosomes (e.g., lysosomal hydrolases) are modified by the addition of high-mannose chains (mannose-6-phosphate) to their attached N-linked oligosaccharides and this modification allows the lysosomal proteins to bind to mannose-6-phosphate receptors in the Golgi complex (lysosomal pathway). In addition to the constitutive pathway, specialized eukaryotic secretory cells (exocrine, endocrine, and neuronal cells) possess a regulated pathway for the delivery of regulated secretory

products, such as prohormones and prohormone processing enzymes, to the extracellular space; the regulated exocrine pathway is, however, clearly distinct from the (neuro)endocrine route. Sorting signals are likely present to direct (neuro)endocrine proteins to secretory granules and to separate them from constitutively secreted proteins. It is not clear whether such a sorting signal (1) would be recognized by a membrane-bound receptor (with carboxypeptidase E being a proposed but hotly debated receptor candidate); (2) is responsible for the observed spontaneous aggregation of (neuro)endocrine proteins in the mildly acidic, high- $\text{Ca}^{2+}$  milieu of the TGN; or (3) is necessary to retain the regulated proteins in the immature granule, away from the budding constitutive vesicle (sorting by retention). Also, lipid microdomains (rafts) in the TGN may be involved in the sorting process by interacting with regulated proteins. Furthermore, members of the granin family of acidic neuroendocrine-specific proteins (chromogranin A, chromogranin B/secretogranin I, secretogranins II and III, and the neuroendocrine polypeptide 7B2) may play a role in the sorting event by acting as an aggregation vehicle for other regulated proteins; granins are major matrix constituents of secretory granules and thus markers for these granules. Finally, when subsets of soluble proteins are packaged into different ER-derived vesicles that, according to the cisternal maturation model, lead to distinct VTCs and Golgi cisternae, presorting of secretory proteins at the level of the ER would occur, thus increasing the efficiency of the final sorting step in the TGN.

## BIOGENESIS OF SECRETORY GRANULES

At the TGN, secretory proteins are packaged in immature secretory granules (ISGs), the formation of which may involve lipid rafts, similar to the situation for apical targeting in epithelial cells. ISGs undergo homotypic (ISG–ISG) fusion and gradually develop to a mature secretory granule (MSG) with a change in the size, membrane composition, and content of the granule, and an increased concentration (as much as 200-fold) of the regulated secretory proteins. Furthermore, maturation involves progressive acidification and loss of the clathrin coat because small regions of the maturing granules pinch off as clathrin-coated vesicles that remove soluble proteins for constitutive-like secretion, transport to endosomes, or retrieval to the TGN. The MSGs thus contain only regulated proteins as any constitutive protein is

removed from the ISG by its packaging in the constitutive transport vesicles that readily fuse with the plasma membrane.

## POSTTRANSLATIONAL MODIFICATIONS OF PROHORMONES

In addition to glycosylation and phosphorylation in the early stages of the secretory pathway, a secretory protein can undergo sulfation at carbohydrate side chains or tyrosine residues in the late pathway (by carbohydrate and tyrosylprotein sulfotransferases, respectively). The first step in proprotein proteolytic processing is usually an endoproteolytic cleavage in the TGN or ISG on the carboxy-terminal side of a recognition site, often a pair of basic amino acids and mostly Lys-Arg or Arg-Arg. The proprotein processing enzymes are calcium- and pH-dependent serine endoproteases related to the bacterial proteolytic enzyme subtilisin and are called proprotein convertases (PCs). These enzymes are structurally related (most strongly at their active sites) and form a gene family consisting of at least seven members, including furin (also called paired basic amino acid residue cleaving enzyme), the prohormone convertases PC1/PC3 and PC2, and (in yeast) KEX2. Furin and KEX2 are active in constitutively secreting cells. The neuroendocrine-specific enzymes PC1/PC3 and PC2 cleave prohormones and are selectively present in cells equipped with the regulated secretory pathway. Differential expression of PC1/PC3 and PC2 may result in a different secretory output from endocrine and neuronal cells. Furthermore, a single prohormone may give rise to multiple peptide hormones with a variety of bioactivities. For example, processing of proopiomelanocortin (POMC) can result in  $\alpha$ -,  $\beta$ -, and  $\gamma$ -melanocyte-stimulating hormones (MSHs), the stress hormone adrenocorticotropin (ACTH), and the endorphins, peptides with endogenous opiate-like activity. POMC cleavage by the PC1/PC3 enzyme generates ACTH, whereas PC2 produces the MSHs and endorphins. PCs become active by autocatalytic cleavage of an amino-terminal propeptide that may act as an intramolecular chaperone for proenzyme folding. The 7B2 protein has been found to act as a chaperone specific for PC2 by transiently interacting with the proenzyme form, facilitating pro-PC2 transport and activation. The pro-SAAS protein, with a structural organization similar to that of 7B2, appears to have PC1/PC3 as its major intracellular binding target. The proper acidic environment in the subcompartments of the secretory pathway, essential for optimal PC cleavage activity, is

supplied by a H<sup>+</sup>-pumping vacuolar-type ATPase. Following cleavage by PCs, exoproteolytic removal of the exposed carboxy-terminal basic residues occurs by the enzyme carboxypeptidase E (or KEX1 in yeast). Finally, the generated peptide may undergo one or two modifications that are crucial for its biological activity, namely, acetylation at the amino terminus and amidation at the carboxy terminus if the peptide ends in glycine (by the enzyme peptidyl-glycine- $\alpha$ -amidating monooxygenase).

## ENDOCRINE DISEASES LINKED TO THE SECRETORY PATHWAY

A detailed understanding of the processes by which peptide hormones are produced and released is desirable, as a growing number of endocrine disorders have been linked to malfunctioning transport, sorting, and processing mechanisms in the secretory pathway. For instance, mutations in the thyroid prohormone thyroglobulin may lead to defective prohormone folding and assembly, reduced ER export, and an ER storage disease (hypothyroidism). In patients with familial hyperproinsulinemia, a mutation in the B-chain of proinsulin probably causes improper prohormone folding and sorting with diversion of the mutant molecule from the regulated to the constitutive pathway. Alternatively, in such patients an altered prohormone processing site in proinsulin may occur, resulting in partially cleaved proinsulin intermediates. Familial neurohypophyseal diabetes insipidus may be caused by mutations in the vasopressin–neurophysin II prohormone gene, leading to inefficient signal peptide cleavage, impaired ER export, or improper prohormone folding and resulting in reduced vasopressin production. Familial hypoparathyroidism is due to a mutation in the gene region encoding the signal peptide of preproparathyroid hormone. A mutated exon–intron splice junction gives rise to exon skipping during growth hormone gene expression, resulting in

misfolding of the mutant growth hormone that causes Golgi fragmentation, disrupted ER-to-Golgi traffic, and familial growth hormone deficiency. Another growth disorder (Laron-type dwarfism) can be caused by a mutation in the growth hormone receptor gene, leading to defective membrane expression of the receptor. A form of diabetes mellitus appears to be due to a mutant insulin receptor, leading to disrupted receptor folding and transport. Finally, PC1/PC3-inactivating gene mutations lead to multiple endocrine deficits resulting from defective prohormone processing (obesity, hyperproinsulinemia, and hypogonadism).

## See Also the Following Articles

Peptide Hormones and Growth Factors: Cellular Signaling Mechanisms • Peptide Hormones, Regulated Secretion • Peptide Hormones, Regulation and Gene Expression • Peptide Hormones, Segregation Mechanism • Peptide Hormones, Subcellular Structure • Prohormone Convertases • Prohormones

## Further Reading

- Barlowe, C. (2000). Traffic COPs of the early secretory pathway. *Traffic* **1**, 371–377.
- Dannies, P. S. (1999). Protein hormone storage in secretory granules: Mechanisms for concentration and sorting. *Endocr. Rev.* **20**, 3–21.
- Golgi Centenary Issues (1998). *Histochem. Cell Biol.* **109**, 423–600
- Trends Cell Biol.* **8**, 1–49.
- Halban, P. A., and Irminger, J.-C. (1994). Sorting and processing of secretory proteins. *Biochem. J.* **299**, 1–18.
- Kim, P. S., and Arvan, P. (1998). Endocrinopathies in the family of endoplasmic reticulum (ER) storage diseases: Disorders of protein trafficking and the role of ER molecular chaperones. *Endocr. Rev.* **19**, 173–202.
- Kuiper, R. P., and Martens, G. J. M. (2000). Prohormone transport through the secretory pathway of neuroendocrine cells. *Biochem. Cell Biol.* **78**, 289–298.
- Zhou, A., Webb, G., Zhu, X., and Steiner, D. F. (1999). Proteolytic processing in the secretory pathway. *J. Biol. Chem.* **274**, 20745–20748.

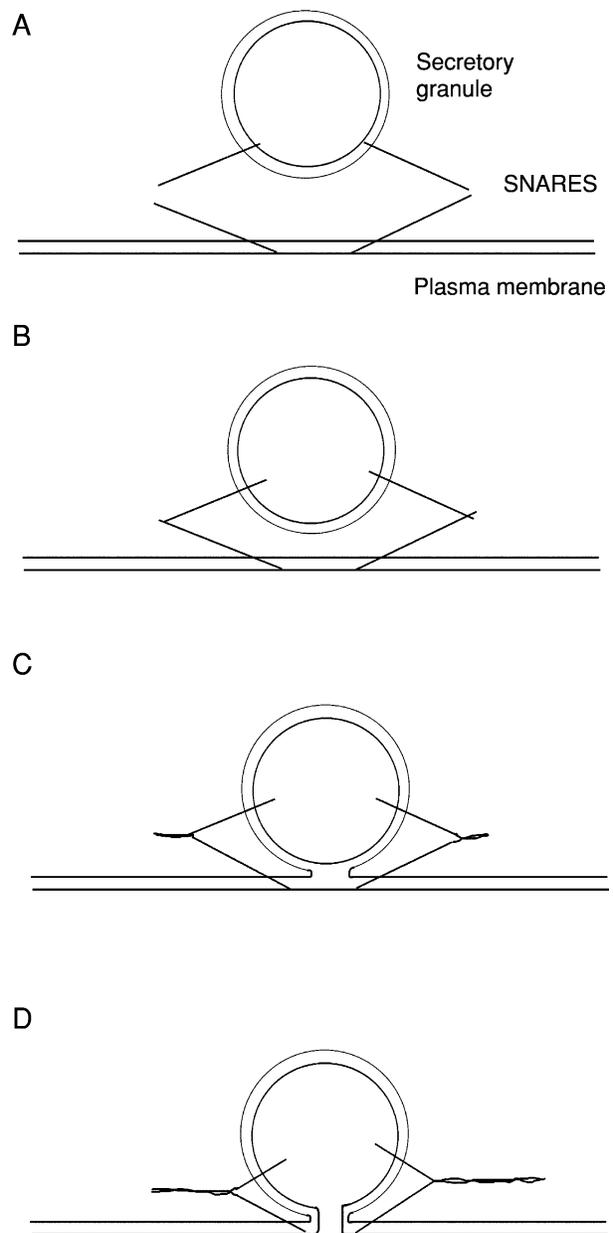
proteins identified as being necessary for secretion in yeast would be related to those necessary for neurotransmitter release. Combining these genetic techniques with biochemical techniques, reconstitution assays, and effects on morphology has led to the identification of proteins involved in vesicle-mediated transport from the endoplasmic reticulum to the Golgi complex, for sorting vacuolar proteins from secretory proteins and for transporting proteins from the Golgi complex to the plasma membrane. Many of the proteins identified in yeast turned out to be identical to those involved in vesicle transport in mammalian cells and in neurotransmitter release from synaptic vesicles. In addition, the same proteins were found in neuroendocrine cells and hormone release had the same sensitivity to cleavage of the toxin-sensitive proteins as neurotransmitter release had. These results demonstrated that there are many membrane fusion events in cells with the same underlying mechanisms.

### PROTEINS INVOLVED IN SECRETORY GRANULE EXOCYTOSIS AND THEIR POSTULATED FUNCTIONS

The following classes are families of proteins for which there is good evidence that they are involved in membrane fusion.

#### Soluble *N*-Ethylmaleimide-Sensitive Factor Receptor Proteins

The soluble *N*-ethylmaleimide-sensitive factor receptor (SNARE) proteins came from the original assay for these proteins, which was the ability to bind NSF (*N*-ethylmaleimide-sensitive factor). The best characterized SNAREs are those that were identified as being important for membrane fusion in neurons, based on their sensitivity to toxins: synaptobrevin 2 (or VAMP), syntaxin Ia, and SNAP-25. All SNAREs are membrane-bound and are weakly homologous to one another in sequences of approximately 60 amino acids called the SNARE motif. The SNARE motifs are functionally important for membrane fusion as these sequences mediate the formation of a core complex that initiates the start of membrane fusion (Fig. 1). The core complex has been well described for the three SNAREs mentioned above and consists of a stable complex of synaptobrevin 2, syntaxin Ia, and SNAP-25, composed of four  $\alpha$ -helices of SNARE motifs, one each from the C- and N-terminal ends of SNAP-25, one from synaptobrevin, and one from syntaxin. Formation of this complex is necessary for



**Figure 1** Steps in exocytosis. (A) A secretory granule is transported to the plasma membrane and is attached to specific sites in ways that are not yet identified. This step is often called docking and may involve munc18a. (B) A second step, sometimes called priming, may involve the start of the formation of the core complex by SNARE proteins. (C) Partial fusion causes merging of one layer of each membrane. (D) Fusion, caused by complete formation of a fusion pore that subsequently fully opens, is driven by complete formation of the core complex of SNARE proteins.

the fusion process to occur and may begin fusion by forcing the two membranes into close proximity (Fig. 1). The SNARE motifs are unstructured when not assembled in the core complex. An exocytotic

fusion complex made from yeast SNARE proteins has similar properties, indicating that vesicle fusion occurs by similar mechanisms in yeast and human. Mutations that cause single amino acid substitutions in SNARE proteins that occur at residues in the core of the  $\alpha$ -helix bundle result in severe impairment of the SNARE function of membrane fusion, providing further proof of the importance of these proteins in the process.

Core complex formation among SNAREs is a property with little specificity, so that many SNAREs will bind to one another to make the complex. Therefore, formation of this complex is important for membrane fusion, but does not confer the specificity necessary to ensure that vesicles fuse in the correct locations.

SNAREs are widely distributed in all cells and each SNARE has a characteristic subcellular distribution, although they may be present in more than one intracellular location. Neuroendocrine cells have SNAREs mediating secretory granule exocytosis that are the same as those mediating synaptic vesicle exocytosis, including syntaxin 1A, syntaxin 1B, and SNAP-25, found primarily in the plasma membrane, and synaptobrevin 2 (or VAMP 2), found in the secretory granule membrane.

## NSF

The function of this protein is to dissociate the core complex formed by SNAREs. NSF was originally isolated because it is necessary for membrane fusion in reconstituted systems, where it acts with SNAPs (soluble NSF attachment proteins; no relation to SNAP-25) and ATP to allow the core complex of SNAREs to dissociate so that they may recycle to cause more fusion. ATP is hydrolyzed in the process, which makes the action of NSF similar to that of chaperones in the cells that facilitate folding. NSF does not participate in the membrane fusion process directly and appears not to be necessary except to put SNAREs into a state from which they can form a complex.

## Sec1/Munc18 Proteins

Sec1/Munc18 proteins (or SM proteins) were originally discovered by genetic screens for secretion in yeast and *Caenorhabditis elegans* and were identified biochemically in mammalian cells as proteins present in nerve endings associated with synaptic vesicles. Investigations with mice lacking munc18a have also shown that the protein is necessary for exocytosis in

mammals. Munc18a binds the SNARE protein syntaxin 1 and it appears that most SM proteins interact with members of the syntaxin family with some degree of specificity. When munc18a is bound to syntaxin, syntaxin is unable to form the core complex with other SNAREs, which suggested a role as a negative regulator for exocytosis, but such a role is not consistent with the data indicating that it is essential for membrane fusion. In mouse adrenal chromaffin cells lacking Munc18a (or Munc18-1), exocytosis of large dense-core vesicles containing catecholamines was reduced 10-fold and the number of vesicles found next to the plasma membrane was reduced 10-fold, but overexpressing munc18a increased the number of releasable vesicles from adrenal chromaffin cells. These results have been interpreted to indicate that SM proteins function in the initial attachment of vesicles to membranes, before the actual process of membrane fusion begins.

## Rab Proteins

The Rab family members are relatively small proteins that bind and hydrolyze GTP, initially identified in yeast as being necessary for secretion by genetic screening. There are over 10 Rab genes in yeast and over 60 in mammals; each gene is associated with specific subcellular organelles. Rab3a is one of the best investigated proteins and it is present in large amounts in presynaptic nerve endings. When GTP is bound, Rab3a is associated with synaptic vesicles, and when the GTP is hydrolyzed to GDP, Rab3a dissociates. Dissociation of GDP and rebinding of GTP allow Rab3a to rebind to synaptic vesicles. Rab3a and a similar member, Rab3b, are also present in neuroendocrine cells associated with secretory granule membranes.

Although Rab proteins are implicated in secretion, the exact role has been difficult to define, possibly because each Rab may have more than one role or because Rab proteins from one set of subcellular organelles may be able to substitute when needed for Rab proteins with other distributions. There is evidence that Rab proteins both facilitate and inhibit  $\text{Ca}^{2+}$ -stimulated exocytosis.

GTPases represent a superfamily of proteins whose mode of action in general is to bind to and activate effector proteins when they themselves have GTP bound, but this property is lost after GTP hydrolysis. Characterization of Rab3 effectors appears to be a way to learn more of the function of Rab3. Several such proteins have been identified; these include rabphilin, a peripheral membrane protein associated with

secretory vesicles, Rim (Rab-interacting molecule), identified in a yeast 2 hybrid screen, and calmodulin, a protein that changes conformation when it binds  $\text{Ca}^{2+}$ , which allows it to activate many proteins after increases in cytosolic  $\text{Ca}^{2+}$ .

### Synaptotagmins

Both exocytosis from synaptic vesicles and exocytosis from secretory granules occur in response to increases in cytosolic  $\text{Ca}^{2+}$  and in this respect differ from most fusion of other vesicles involved in intracellular transport. Synaptotagmins are a family of proteins; synaptotagmin I is associated with synaptic vesicles and secretory granules and has been proposed to be the molecule that mediates the  $\text{Ca}^{2+}$  responsiveness. *C. elegans*, *Drosophila melanogaster*; and mice that lack synaptotagmin have been produced and in all these organisms, neurotransmitter release evoked by  $\text{Ca}^{2+}$  was reduced. The exact way in which synaptotagmin mediates  $\text{Ca}^{2+}$  responsiveness is not known; the amino acid residues identified as binding  $\text{Ca}^{2+}$  in solution using the isolated protein are not necessary for  $\text{Ca}^{2+}$  responsiveness in neurons. It seems logical that synaptotagmin would exert an inhibitory effect on fusion processes until  $\text{Ca}^{2+}$  concentrations increase, but a specific mechanism has not yet been elucidated.

### STEPS IN FUSION

Although exocytosis from synaptic vesicles and that from secretory granules appear very similar at the final stages, the initial mechanisms for locating the vesicle near the membrane are quite different. Synaptic vesicles form by invaginations from the plasma membrane and so are near the plasma membrane from the start. They fill by pumping neurotransmitters into the vesicle, using membrane transport proteins. Secretory granules, however, form around aggregates of protein hormones in the *trans*-Golgi lumen and then must be transported from that region to the plasma membrane, a process that includes penetrating a network of actin filaments near the plasma membrane in neuroendocrine cells. Two proteins with homology to receptor-type protein tyrosine phosphatases are enriched in neuroendocrine cells and associated with secretory granules. These membrane proteins, known as ICA512 (or IA2) and phogrin, are not associated with synaptic vesicles. Their preferential location on secretory granules indicates that they have a specific function and evidence is consistent with these proteins playing a role in the transport of secretory granules to the plasma membrane.

Once at the plasma membrane, both synaptic vesicles and secretory granules attach to the membrane, a process that is sometimes referred to as docking. Attaching to the membrane is not the same as initiating fusion and the process appears to involve munc18a in mammalian cells. It is possible that Rab3s may play a role in this step. Docking occurs at specific sites on the plasma membrane and the specificity of different Rabs for intracellular organelles makes them attractive candidates for providing specificity, but if they do provide such specificity, the mechanisms by which they do so are unknown.

After docking, there is an additional step before fusion, which has in some cases been called priming, which occurs both to synaptic vesicles and to secretory granules. A candidate for this step may be partial assembly of the SNARE core complex and may also involve partial fusion of the membranes by fusion of one layer of the bilayer of each membrane (Fig. 1). The fusion reaction itself involves opening of a fusion pore. Opening the pore is reversible; it may flicker open and close so quickly that vesicle contents are not released. Once the fusion pore widens, however, the process goes to completion, the contents are released, and exocytosis has occurred. This step is mediated by complete formation of the core complex of SNARE proteins. It has not yet been completely demonstrated how formation of the core complex causes fusion nor whether there are other as yet unidentified proteins that are actively involved after the core complex forms.

For secretory granules and synaptic vesicles in neuroendocrine cells and neurons, the increase in cytosolic  $\text{Ca}^{2+}$ , caused by entry through  $\text{Ca}^{2+}$  channels, triggers exocytosis of vesicles that have docked at the plasma membrane and that have undergone a subsequent step, possibly partial fusion. The exact mechanisms by which  $\text{Ca}^{2+}$  triggers opening of the fusion pore remain to be identified.

### See Also the Following Articles

Peptide Hormones and Growth Factors: Cellular Signaling Mechanisms • Peptide Hormones, Intracellular Transport • Peptide Hormones, Regulation and Gene Expression • Peptide Hormones, Segregation Mechanism • Peptide Hormones, Subcellular Structure

### Further Reading

Avery, J., Jahn, R., and Edwardson, J. M. (1999). Reconstitution of regulated exocytosis in cell-free systems: A critical appraisal. *Annu. Rev. Physiol.* **61**, 777–807.  
 Bruns, D., and Jahn, R. (2002). Molecular determinants of exocytosis. *Pflugers Arch.* **443**, 333–338.

- Darchen, F., and Goud, B. (2000). Multiple aspects of Rab protein action in the secretory pathway: Focus on Rab3 and Rab6. *Biochimie* **82**, 375–384.
- Jahn, R., and Südhof, T. C. (1999). Membrane fusion and exocytosis. *Annu. Rev. Biochem.* **68**, 863–911.
- Lin, R. C., and Scheller, R. H. (2000). Mechanisms of synaptic vesicle exocytosis. *Annu. Rev. Cell Dev. Biol.* **16**, 19–49.
- Zenisek, D., Steyer, J. A., and Almers, W. (2000). Transport, capture and exocytosis of single synaptic vesicles at active zones. *Nature* **406**, 849–854.

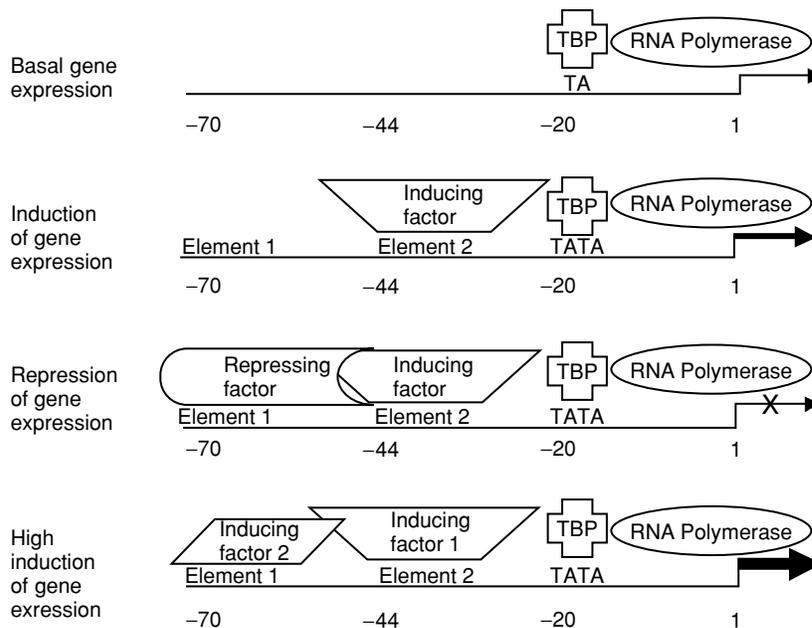
### GENE EXPRESSION

A gene has two parts: (1) the section of DNA on a chromosome that contains a region that is transcribed into an RNA chain, which in turn is translated into the protein encoded thereon, and (2) a DNA region that is involved in controlling or regulating the expression (transcription) of the gene. The regulatory region is often called the promoter region and is located adjacent to and just before (usually referred to as being upstream or 5' of) the transcribed region. The other end of the gene is called the 3'-end.

The DNA sequence is numbered starting at +1, corresponding to the first nucleotide of the region of the gene that is transcribed into RNA. The nucleotide located immediately 5' of the start of transcription is numbered -1 and the nucleotide positions in the DNA sequence of the promoter region of the gene are described by increasing negative numbers as one counts further upstream from -1. A specific DNA sequence "element" known as the TATA-box is located at approximately -20 in the promoter region of most mammalian protein-encoding genes (see Fig. 1). Certain nuclear proteins bind specifically to this TATA-box and are involved in the positioning of the RNA polymerase enzyme at the correct starting point of transcription. Other regulatory DNA sequence elements are usually located 5' to the TATA-box. These other elements are sites where proteins called

transcription factors can specifically bind and cause an increase, or sometimes a decrease, in the amount of expression observed for the gene being controlled (Fig. 1). Most promoters have several different elements where transcription factors can bind. It is these transcription factors that are involved in the regulation of the expression of the gene. In some situations, a transcription factor can bind to an element and have a negative influence on the effect of another (different) protein binding to a second element, but in a different situation that factor may have a positive influence on the second factor and result in an even higher level of gene expression.

An example of such a DNA element capable of specifically binding to certain transcription factors is the cyclic AMP (cAMP) regulatory element (CRE). This site has a consensus DNA sequence consisting of eight nucleotides; TGACGTCA. The CRE binds to a transcription factor called CREB (for CRE-binding protein) and also to other factors, such as activating transcription factor 1 and AP-1 (activating protein 1). A protein that can bind to a given DNA sequence element may require different nucleotides within the element for specific binding compared to those nucleotides required by another protein that can bind to the same DNA sequence. For example, when binding to the CRE, the CREB and AP-1 proteins both touch some, but not all, of the nucleotides within the DNA



**Figure 1** Examples of general molecular mechanisms of gene expression regulation showing the relative positions of some regulatory DNA elements and their associated transcription factor proteins aligned along a gene's promoter region.

sequence of the element, but the actual nucleotides that each of these proteins interacts directly with are not the same.

CREB activity is regulated by phosphorylation at specific amino acids within its protein sequence. Phosphorylation of transcription factors is performed by a variety of protein kinases that are intermediaries in many of the signaling pathways that transmit signals (such as growth- or differentiation-related signals) from the cell surface membrane to the nucleus. Protein kinase A (PKA) is involved in phosphorylating CREB to stimulate its binding to the CRE, whereas another enzyme, protein kinase C, is involved in activating AP-1. Other enzymes, such as calcium-calmodulin-dependent protein kinases (CaM kinases) I, II, and IV, can also phosphorylate CREB to activate it, as can enzymes in the Ras-mitogen-activated protein kinase (MAPK) pathway.

Nuclear hormones represent another major pathway through which growth and differentiation signals are transmitted from the cell surface to the nucleus, where they affect gene expression. Nuclear hormones function by binding to a family of transcription factors known as the nuclear hormone receptors. These include receptors for sex steroids such as the estrogen receptor (ER) and those for corticosteroids such as the glucocorticoid receptor (GR) and they also include receptors for vitamins, such as the retinoids [retinoic acid receptor (RAR) and retinoid X receptor (RXR)], and for lipid metabolites, which are the ligands for a group of proteins called the peroxisome proliferator receptors (PPAR). Each member of the nuclear hormone receptor family exists in several subtypes usually indicated by the Greek letters  $\alpha$ ,  $\beta$ , and  $\gamma$ . The nuclear hormone receptor family of proteins can regulate the expression of genes by binding to specific DNA elements in the promoter of the gene being controlled or by interacting with other transcription factors through protein-protein interactions.

The following sections provide several detailed examples of the molecular mechanisms regulating peptide hormone gene expression.

## CORTICOTROPIN-RELEASING HORMONE

CRH is a hypothalamically derived hormone that orchestrates the stress response. Stressor-induced activation of the hypothalamus-pituitary-adrenal axis results in a series of neural and endocrine adaptations known as the "stress response." The stress response

is responsible for allowing the body to make the physiological and metabolic changes necessary to cope with the demands of a homeostatic challenge. The challenge to homeostasis by any stressor initiates the release of CRH from the hypothalamus, which in turn results in the release of adrenocorticotrophic hormone (ACTH) from the pituitary into the general circulation. ACTH then acts on the adrenal cortex, resulting in the release of steroid hormones, glucocorticoids, into the blood. Glucocorticoids act in a negative feedback fashion to terminate the release of CRH from the hypothalamus. CRH has also been identified in many tissues outside the hypothalamus, in particular, in the human placenta during pregnancy. Placental CRH becomes detectable in maternal plasma at approximately 16–20 weeks gestation and the plasma CRH concentration increases exponentially throughout pregnancy and peaks during labor. Placental production of CRH has been linked to the length of gestation in humans and several groups have reported that increased concentrations in maternal blood are associated with preterm delivery.

The promoter region of the human CRH (hCRH) gene contains consensus DNA sequences for several known transcription factors (see Fig. 2). There is complete conservation among four species (human, ovine, mouse, and rat) of several of these consensus regulatory elements, including a CRE, estrogen regulatory element half-sites (1/2ERE), an ecdysone regulatory element (EcRE), two TATA-boxes (TBP), and a CDXA regulatory element, suggesting the importance of these sites in hCRH gene expression.

There is a consensus CRE located at position –221 to –228 in the hCRH promoter, and activators of the PKA signaling pathway, such as cAMP, increase the expression of CRH both in the placenta and in the hypothalamus. Stress can rapidly lead to phosphorylation of CREB and to induction of CRH in the hypothalamus. CaM kinases are involved in the regulation of CRH in the hypothalamus, but this has not been explored in the placenta. The Ras-MAPK pathway is also present in the placenta but it has not been shown to be involved in the regulation of placental CRH production.

Different proteins bind to the CRE in placental versus hypothalamic cell models. The CREB protein is present in both the hypothalamus and the placenta. However, component proteins of the AP-1 transcription factor (Jun and Fos) differ between these two tissues. The Jun protein is present in the nucleus of placental cells and binds to the CRE, whereas the nuclear protein from a mouse pituitary cell line (AtT20) shows the presence of the Fos protein. The transcription factor AP-2 (activating protein 2) has

```

AAAGCAGATA GACGTTTAAA AGCTGGATGT TCCTGCACACC CCTCTCCTG
      GATA                ½GRE          MBFI

ATGCTTCATT CTTTTCCAGG CAGAAAGATG GTGGGACTCT GTCTCTAGCA

AAGGGATATT TCCAGATACT GAGGTGTTGT CAGAGACACC TGGTCAGGGA
      GATA

GGTTAGGAGA AGGGGCATCC AGGTCCACCC CCTCCAACCTG GCTGCTGCTT

TCCTGGCAGG GCTGCACTGG GACACCTCAC TTCCTTCCCA CTTCCCCTTC

CTCCTCCCAT TCGCTGTCTC TTTGCACACC CCTAATATGG CCTTTCATAG

TAAGAGGTCA ATATGTTTTT ACACCTGGGA AATCTCATT C AAGAATTTTT
      ½ERE                NF-IL-6

GTCAATGGAC AAGTCATAAG AAGCCCTTCC ATTTTAGGGC TCGTTGACGT
      EcRE                CRE

CACCAAGAGG CGATAAATAT CTGTTGATAT AATTGGATGT GAGATTCAGT
      TBP (2)

GTTGAGATAG CAAAATTCTG CCCCTCGTTC CTTGGCAGGG CCCTATGATT
      GATA

TATGCAGGAG CAGAGGCAGC ACGCAATCGA GCTGTCAAGA GAGCGTCAGC
      CDXA

TATTAGGCAA ATGCTGCGTG GTTTTTGAAG AGGGTCGACA CTATAAAATC
      TBP

CCACTCCAGG CTCTGGAGTG G
      -1

```

**Figure 2** The DNA sequence of the proximal section of the hCRH gene promoter. The regulatory elements discussed in the text are indicated.

also been reported to interact with transcription factors binding to the CRE and to regulate CRH gene expression in placenta.

Deletion or mutation of the CRE in the hCRH promoter results in markedly reduced hCRH gene expression in placental cells in response to cAMP but does not completely abolish it. Further analysis in human choriocarcinoma (JEG) cells and rat choriocarcinoma cells has identified another region, located between nucleotides  $-128$  and  $-109$ , that is responsive to cAMP. This region contains a consensus element for the homeobox transcription factor CDXA

(see Fig. 2). Mutation of this CDXA-response element (CDXRE) also reduces but does not eliminate induction of the hCRH promoter by cAMP in AtT20 cells. Therefore, cAMP-mediated signaling pathways use both the CRE and the CDXRE to stimulate hCRH gene expression in placenta and hypothalamic cells.

Glucocorticoids mediate inhibition of hCRH gene expression in the hypothalamus, whereas these steroid hormones stimulate the expression of hCRH in the placenta. The hCRH promoter does not contain a consensus glucocorticoid-response element (GRE) although there is a half-site at  $-600$  to  $-593$  (Fig. 2).

Assays *in vitro* identified three regions of the promoter that can bind GR (Fig. 2, heavily overlined sequence). These sites are not involved in the glucocorticoid-mediated stimulation of hCRH gene expression in the placenta and glucocorticoids stimulate hCRH gene expression in the placenta through the CRE, presumably through protein–protein interactions with proteins that bind at the CRE. In other cells, the hormone-activated GR negatively or positively modulates the expression of AP-1-dependent genes, depending on which subunit of the AP-1 dimeric protein complex is present. For example, homodimers of Jun can interact with GRs to stimulate promoter activity, whereas heterodimers between Jun and Fos interact with GRs to repress promoter activity.

A region of the hCRH promoter, between nucleotide positions –278 and –246, has been identified as being responsible for negative regulation by glucocorticoids in cAMP-stimulated AtT20 cells. This negative GRE (nGRE) corresponds to one of the DNA regions mentioned above that can bind the GR *in vitro*. Surprisingly, in the absence of this nGRE, the hCRH promoter is stimulated by glucocorticoids in AtT20 cells in a manner similar to that seen in placenta cells. The nGRE colocalizes to a consensus EcRE (see Fig. 2). Mutation of this EcRE in the hCRH promoter results in a doubling of hCRH gene expression in primary placental cell cultures. This mutation of the EcRE did not cause any change in hCRH expression in AtT20 cells. Therefore, this region of the promoter appears to act as a general inhibitor in placental cells and as a nGRE in pituitary cells. Ecdysone is an insect steroid hormone that has a role in insect molting and metamorphosis. Mammals do not have ecdysone receptors and ecdysone as such has no known role in mammalian gene regulation. Ecdysone and its receptor (EcR) do not tend to form homodimers but tend to associate with other members of the nuclear receptor superfamily. RAR and RXR can form functional dimers with EcR in mammalian cells and it is possible that PPAR may have a role in regulating hCRH through the

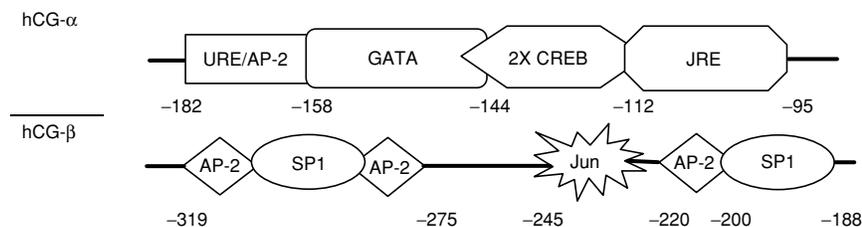
EcRE since it often forms heterodimers with RXR and RAR.

In placenta, hCRH gene expression is repressed by estrogen, whereas the estrogen antagonist ICI 182780 stimulates hCRH gene expression. Although the hCRH promoter contains five 1/2 EREs, no binding of estrogen receptor to the 1/2 EREs has been observed. In endometrial cells, estrogen has also been found to inhibit hCRH gene expression. The placental estrogen effect involves ER- $\alpha$  and apparently occurs through protein–protein interactions with other transcription factors or cofactors.

### CHORIONIC GONADOTROPIN

The human reproductive hormone chorionic gonadotropin (hCG) is a key messenger in embryonal/fetal to maternal communication during embryonic development and plays an important role in the establishment and maintenance of pregnancy. hCG is a member of the heterodimeric glycoprotein hormone family. The members of this hormone family share a common  $\alpha$ -subunit and have unique  $\beta$ -subunit peptides. The  $\alpha$ -subunit gene is regulated in three different cell types and two tissue types: the trophoblasts of the placenta and the gonadotropes and thyrotropes of the anterior pituitary. In addition, the  $\alpha$ -subunit gene needs to be coordinated with the expression of a specific  $\beta$ -subunit gene. The hCG- $\beta$  gene has a TATA-less promoter and is expressed almost exclusively in the placenta.

A summary of the regulatory sites identified in the hCG- $\alpha$  and hCG- $\beta$  genes is shown in Fig. 3. In addition to these, there is a pituitary-specific region located between nucleotide positions –500 and –200. Within that region, two smaller regions have been identified as being specific for expression in cells of gonadotrope lineage and another region is specific for expression in thyrotropes. The junction regulatory element (JRE) of the hCG- $\alpha$  gene appears to bind a trophoblast-specific protein. Cellular signaling



**Figure 3** A schematic representation of the hCG- $\alpha$  and hCG- $\beta$  gene promoters showing the regulatory elements discussed in the text.

pathways utilizing cAMP are able to activate hCG- $\alpha$  and hCG- $\beta$  genes through two tandem CREs in the hCG- $\alpha$  promoter and through the several sites that constitute a composite regulatory element in the hCG- $\beta$  promoter, as shown in Fig. 3. In hCG- $\beta$ , the three AP-2 sites and the Jun-binding site are all involved in activation through cAMP pathways. The two adjacent CREs in the hCG- $\alpha$  promoter and the Jun site of hCG- $\beta$  bind the Jun protein in addition to the CREB protein.

AP-2 is involved in regulating both the  $\alpha$ - and  $\beta$ -subunit genes, the hCG- $\alpha$  gene through a region also referred to as the upstream regulatory element (URE) and the hCG- $\beta$  gene through three AP-2-binding sites. AP-2 and Sp1 (selective promoter factor 1) work through complex cooperative mechanisms to control basal as well as cAMP-mediated gene regulation. AP-2 is regulated by retinoic acid and this may be a mechanism whereby retinoic acid stimulates hCG secretion from JEG3 cells. However, RXR $\alpha$  and PPAR $\gamma$  are able to bind directly to specific sites within the hCG- $\beta$  promoter and stimulate gene expression, and ligands for these two receptors have an additive effect on hCG synthesis. Thus, retinoic acid may have a direct (RXR/PPAR) or an indirect (through AP-2) effect on hCG production.

## PLACENTAL LACTOGEN

The hPL/growth hormone (hGH) gene locus consists of five structurally related genes clustered on human chromosome 17. hGH-N is a pituitary-specific growth hormone, hGH-V is a placental-specific growth hormone variant, hPL-A and hPL-B are identical proteins found in the fetal and maternal circulations, and hPL-L was long thought to be a pseudo-gene (nonfunctional gene) until the discovery of novel peptides encoded by it. Although the promoter regions of the hGH/hPL gene family share 91–99% sequence identity over the proximal 500 bp region (–500 to –1), it is evident that the factors that regulate hPL gene expression are largely different from those that regulate the hGH gene. These conserved 500 bp sequences located immediately 5' to the coding regions in this gene family are sufficient to permit pituitary-specific gene

expression. Expression in placental trophoblast cells requires a Sp1 site for basal expression and two AP-2 sites for maximal expression (see Fig. 4). Trophoblast-specific expression also requires the placental-specific enhancer region [or chorionic somatomammotropin enhancer (CSE)] located ~2.2 kb 3' of the gene cluster. There is also a pair of pituitary-specific silencers (PSF) located ~2 kb 5' of each placentally expressed gene. Thus, tissue-specific regulation of the hPL gene is controlled by a combination of enhancer and silencer elements.

There are also two composite sites identified that are important for activation of hPL gene expression by nuclear hormone receptors (NHR) (see Fig. 4). Site A is transactivated by RAR $\alpha$ , whereas site B is transactivated by RAR $\alpha$ , by T<sub>3</sub>R $\beta$  (thyroid hormone receptor), and by vitamin D receptor. The RAR- and T<sub>3</sub>R-stimulated expression is repressed by the orphan receptor ARP-1, which also binds to the composite NHR. As was the case for hCG, it has been shown that ligands for RXR and for PPAR are able to stimulate hPL synthesis. Although not yet thoroughly defined, it is clear that several different nuclear hormone receptors may function to control hPL gene expression in the placenta.

The hPL-B promoter has two potential binding sites for the transcription factor NF-IL-6 [nuclear factor–interleukin-6; also known as CCAAT/enhancer-binding protein (C/EBP)]. Interleukin-1 acts on cell surface receptors, which activate the Ras-dependent MAPK cascade, and this, in turn, phosphorylates NF-IL-6. As mentioned earlier, this pathway is functional in placental cells and so it is interesting to note that a NF-IL-6 regulatory element is also present in the CRH promoter (see Fig. 2).

## LEPTIN

Leptin is a hormone most commonly thought of in terms of its role in regulating body weight. Leptin performs this function through feedback control of energy balance to increase metabolism and depress appetite. Secreted by adipose cells, its concentration in the circulation correlates with body adiposity. Investigators have found that leptin is also expressed

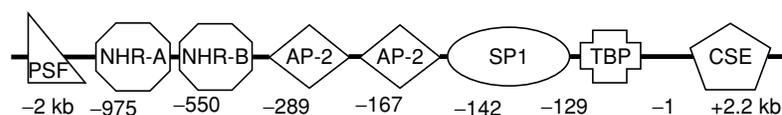
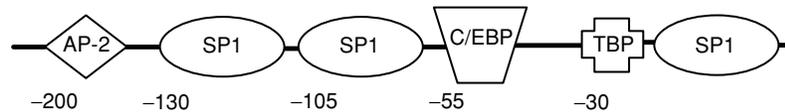


Figure 4 A schematic representation of the hPL gene promoter showing the regulatory elements discussed in the text.



**Figure 5** A schematic representation of the LEP gene promoter showing the regulatory elements discussed in the text.

in other tissues, including the brain, pituitary, and placenta. A correlation has been reported between umbilical cord leptin and conceptus size, indicating that leptin may also be involved in regulating conceptus growth or placental function.

A number of differences have already been recognized for leptin gene regulation in these different tissues. Expression of the leptin gene in adipocytes requires only the proximal 200 bp of the 5'-flanking DNA sequence. This region of the leptin promoter contains a TATA-box, a C/EBP site (or a NF-IL-6 site), an AP-2 site, and three Sp1-binding sites (Fig. 5). Placenta-specific leptin gene expression is enhanced by an upstream region of the promoter between positions -1951 and -1546. This upstream placenta-specific enhancer (PSE) requires the proximal region of the promoter to function as it has no promoter activity on its own. The transcription factor Sp1 can bind to sites within the PSE, although how these Sp1 sites might function as a tissue-specific enhancer in placentally derived cell lines has not yet been determined.

The leptin gene promoter is stimulated by estrogen and inhibited by anti-estrogens through ER- $\alpha$ , but in cells transfected with ER- $\beta$  the leptin gene is stimulated by anti-estrogens. The molecular mechanism of this estrogen-mediated regulation has not as yet been clearly identified. Tissue-specific differences in leptin gene regulation are further evident from studies that show that cAMP inhibits leptin expression and secretion in adipocytes, whereas it stimulates leptin expression and secretion in placental trophoblast and trophoblast-derived cells and in rat C6 glioma and GH3 pituitary tumor cells.

hCG and hPL are stimulated by PPAR $\gamma$  ligands and this is also true for leptin gene expression in placental cells. In adipocytes, however, PPAR $\gamma$  ligands inhibit leptin gene expression. This inhibition occurs through the C/EBP site but does not involve binding of the PPAR $\gamma$  to the DNA directly.

## DIGEST

Although the peptide hormones discussed here represent proteins with a variety of distinctly different

physiological functions, they are all similar in that they are expressed in brain, placenta, and several other tissues. The promoter regions of each of the genes encoding these peptide hormones are also organized in ways distinctly different from the others. Nevertheless, it can be seen that there are also some striking similarities that warrant mention. cAMP-based signaling systems play significant roles in regulating these genes, often through complex interactions involving several transcription factors including CREB, AP-1, and AP-2. Nuclear hormone receptors regulate these genes through similarly complex molecular mechanisms involving the interaction of receptors such as GR, ER, RXR, and PPAR with transcription factors such as CREB, AP-1, AP-2, and C/EBP.

## See Also the Following Articles

Alternative Promoters • Peptide Hormones and Growth Factors: Cellular Signaling Mechanisms • Peptide Hormones, Intracellular Transport • Peptide Hormones, Regulated Secretion • Peptide Hormones, Segregation Mechanism • Peptide Hormones, Subcellular Structure

## Further Reading

- Altucci, L., and Gronemeyer, H. (2001). Nuclear receptors in cell life and death. *Trends Endocrinol. Metab.* **12**, 460–468.
- Anthony, R. V., Limesand, S. W., and Jeckel, K. M. (2001). Transcriptional regulation in the placenta during normal and compromised fetal growth. *Biochem. Soc. Trans.* **29**, 42–48.
- Barrera-Saldana, H. A. (1998). Growth hormone and placental lactogen: Biology, medicine and biotechnology. *Gene* **211**, 11–18.
- Considine, R. V. (2001). Regulation of leptin production. *Rev. Endocr. Metab. Disord.* **2**, 357–363.
- King, B. R., Nicholson, R., and Smith, R. (2001). Placental corticotrophin-releasing hormone, local effects and fetomaternal endocrinology. *Stress* **4**, 219–233.
- King, B. R., Smith, R., and Nicholson, R. C. (2001). The regulation of human corticotrophin-releasing hormone gene expression in the placenta. *Peptides* **22**, 1941–1947.
- Kurtzman, J. T., Wilson, H., and Rao, C. V. (2001). A proposed role for hCG in clinical obstetrics. *Semin. Reprod. Med.* **19**, 63–68.
- Nicholson, R. C., and King, B. R. (2001). Regulation of CRH gene expression in the placenta. *Front. Horm. Res.* **27**, 246–257.
- Papavassiliou, A. G. (1995). Molecular medicine: Transcription factors. *N. Engl. J. Med.* **332**, 45–47.

An implication of aggregation occurring before secretory granules form is that protein aggregation may both concentrate and segregate secretory granule proteins. Aggregation of secretory granule proteins in the *trans*-Golgi layer results in sorting because the small vesicles that bud off carry away soluble proteins, leaving behind the hormone aggregates, which are too large to be removed and which therefore stay as the core of the granule (Fig. 1). The ultimate proof that aggregation is sufficient for segregation will come with mutants that show a correspondence between aggregation in cells and storage in secretory granules.

There are proteins in secretory granules that do not form large aggregates and may remain soluble, such as galanin. It is not known whether these proteins are in fact efficiently stored in granules; the bulk of the newly synthesized protein may be constitutively secreted or degraded and small amounts of soluble proteins may end up in granules because they are incompletely taken away by the budding small vesicles. If they remain soluble and are efficiently segregated to granules, there must be an additional sorting mechanism besides aggregation.

Prolactin is a relatively simple hormone that is not processed beyond cleavage of the signal sequence responsible for its transport into the endoplasmic reticulum during synthesis. It is not yet known whether aggregates of prohormones exist in such an early stage as those of prolactin, because three-dimensional views of granules forming are available only for prolactin, but dense masses of proinsulin and proopiomelanocorticotropin have been detected by electron microscopy in the *trans*-Golgi lumen of the cells that produce each hormone, consistent with an early aggregation step.



**Figure 1** The *trans*-layer of the Golgi complex in cells producing secretory granules. Protein hormones aggregate (large black shape) in the lumen of the *trans*-layer. The excess membrane and soluble proteins are removed by budding of small vesicles, until only membrane-enclosed aggregates are left. In order for this membrane-enclosed aggregate to function properly, the appropriate membrane proteins (shown as small black shapes in the membrane around the aggregates) must accumulate in the secretory granule membrane and not be removed by budding vesicles.

## SEGREGATION OF SECRETORY GRANULE MEMBRANE PROTEINS

Formation of an aggregate is not sufficient for the formation of a normally functioning secretory granule. For example, Von Willebrand factor, a secretory protein that forms aggregates, forms dense-core membrane-enclosed structures that look like secretory granules when expressed in neuroendocrine cells, but release of the factor cannot be regulated as release of endogenous secretory granule proteins is. After formation, secretory granule proteins are transported to the plasma membrane where they undergo regulated exocytosis. These functions are mediated by membrane proteins including synaptobrevin, synaptotagmin, ICA512, and phogrin. For secretory granules to function properly, proteins that mediate the functions must accumulate in appropriate amounts in the granule membranes as they form. Because they do not accumulate around all aggregates, there must be some form of recognition between the aggregates of secretory granule proteins and membrane proteins necessary for secretory granule function (Fig. 1). Recognition of surface motifs of the aggregates may occur through protein, lipid, or a combination of the two. The features that are recognized and what causes the recognition have not yet been elucidated.

### See Also the Following Articles

Peptide Hormones and Growth Factors: Cellular Signaling Mechanisms • Peptide Hormones, Intracellular Transport • Peptide Hormones, Regulated Secretion • Peptide Hormones, Regulation and Gene Expression • Peptide Hormones, Subcellular Structure

### Further Reading

- Baertschi, A. J., Monnier, D., Schmidt, U., Levitan, E. S., Fakan, S., and Roatti, A. (2001). Acid prohormone sequence determines size, shape, and docking of secretory vesicles in atrial myocytes. *Circ. Res.* **89**, E23–E29.
- Dannies, P. S. (1999). Protein hormone storage in secretory granules: Mechanisms for concentration and sorting. *Endocr. Rev.* **20**, 3–21.
- Rambourg, A., Clermont, Y., Chretien, M., and Olivier, L. (1992). Formation of secretory granules in the Golgi apparatus of prolactin cells in the rat pituitary gland: A stereoscopic study. *Anat. Rec.* **232**, 169–179.
- Sankoorikal, B. J., Zhu, Y. L., Hodsdon, M. E., Lolis, E., and Dannies, P. S. (2002). Aggregation of human wild-type and H27A-prolactin in cells and in solution: Roles of Zn(2+), Cu(2+), and pH. *Endocrinology* **143**, 1302–1309.
- Tooze, S. A., Martens, G. J., and Huttner, W. B. (2001). Secretory granule biogenesis: Rafting to the SNARE. *Trends Cell Biol.* **11**, 116–122.

function, at least in part, by preventing unfolded proteins from forming irreversible aggregates instead of folding. Some proteins are also covalently modified in the endoplasmic reticulum during and after folding by glycosylation. Proteins that normally exist as oligomers must assume the correct quaternary structure in order to be transported further along the secretory pathway. Proteins that do not fold properly are normally degraded. There are proteases, enzymes that cleave proteins, in the endoplasmic reticulum; in addition, proteins that do not fold properly may be transported back through the membrane of the endoplasmic reticulum, where they are usually degraded by complexes, known as proteasomes, made up of proteases and other proteins. Neuroendocrine cells that synthesize large amounts of protein tend to have large numbers of proteasomes.

After folding and assembling correctly, proteins are transported further along the secretory pathway; all transport of these proteins is carried out in vesicles (Fig. 1). Transfer from the endoplasmic reticulum occurs through irregularly shaped vesicles that emerge from specific areas of reticulum membrane. The means by which some proteins are carried forward and others left behind is an area of intense investigation; although all the details are not clear, it appears that some proteins are actively carried forward to the next stage and other proteins are actively retained or retrieved from later stages in the secretory pathway.

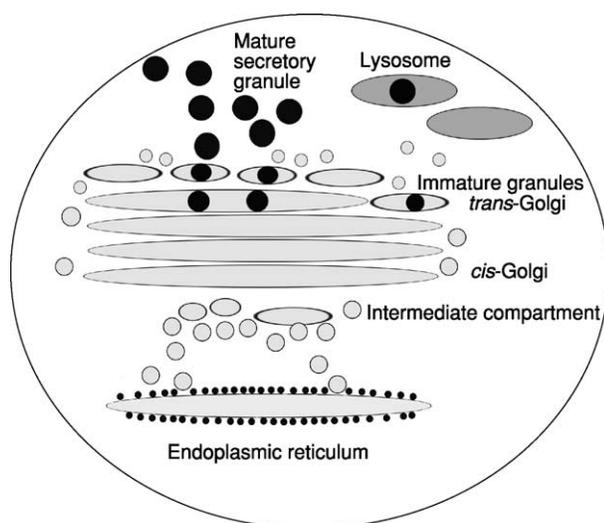
The next stage is near the *cis*-Golgi layer; the vesicles that have left the endoplasmic reticulum fuse in the *cis*-Golgi region. This region of actively fusing

vesicles or vesicular-tubular clusters has been called the pre-Golgi intermediate compartment or the endoplasmic reticulum-Golgi intermediate complex. Sorting goes on in this region as proteins that escaped surveillance and left the endoplasmic reticulum inappropriately are deposited into vesicles to be returned there. The remaining vesicles eventually fuse to form the *cis*-layer of the Golgi complex.

The endoplasmic reticulum has other functions in addition to folding proteins. One important function essential for all cells is that it serves as a storage reservoir for  $\text{Ca}^{2+}$ , which facilitates the use of changes in the concentrations of intracellular  $\text{Ca}^{2+}$  as a signaling mechanism. In neuroendocrine cells, the release of intracellular  $\text{Ca}^{2+}$  is part of the signaling that controls the release of secretory granules. The endoplasmic reticulum also produces most of the lipids needed by cells.

## THE GOLGI COMPLEX

The Golgi complex is a stack of flattened membrane cisternae (Fig. 1) usually located near the nucleus of the cell. In three dimensions, it resembles a broad ribbon-like structure. In cells secreting large amounts of proteins, there may be more than one such stack. Proteins arrive at the Golgi complex on the *cis* side and process to the *trans* side. It had long been assumed that proteins were carried forward from one layer of the complex to the next by small vesicles, but it has become generally accepted that each of the layers of the Golgi complex progresses through the stack, so that the *cis*-layer eventually becomes the *trans*-layer. Each layer metamorphoses in turn from a *cis* to a medial to a *trans* cisterna as small vesicles transport the enzymes that characterize the different layers of the Golgi complex back to the layer behind. A *cis*-layer takes on the characteristics of later compartments sequentially as vesicles or possibly tubular structures derived from later compartments cycle back the appropriate enzymes. The enzymes in the different layers further covalently modify some of the proteins as they process through the Golgi complex; possible modifications include further glycosylation, as well as phosphorylation and sulfation. The concept that it is the layers that process through the stack provides the simplest explanation why some proteins process through the stack in forms that are too large to be included in the small vesicles seen around the stacks by electron microscopy. Atrial natriuretic peptide, for example, appears in a dense aggregated form too large to be carried in vesicles in all layers of the Golgi complex in atrial cardiomyocytes.



**Figure 1** The subcellular structures involved in the secretory pathway of protein hormone-producing cells.

When a Golgi layer reaches the *trans* side of the Golgi complex, small vesicles still bud to return enzymes and other constituents to the more newly formed layers; in addition, clathrin-coated vesicles bud off to transport vesicles to the lysosome and other vesicles bud to take soluble proteins to the plasma membrane, so that the entire *trans*-layer of the Golgi complex is consumed by budding of small vesicles to be replaced by the layer that formed after it. Proteins clearly have been sorted when they leave the *trans*-Golgi region, but the sites of sorting of different proteins may vary, as suggested by the finding of Howell and colleagues, who examined the Golgi complex in three dimensions, that only one type of vesicle is seen budding from regions of the *trans*-Golgi layer.

## LYSOSOMES

Lysosomes are subcellular organelles that are full of hydrolytic enzymes called acid hydrolyases, which are active at pH 5, the pH of the lysosomal lumen. The function of lysosomes is to degrade unwanted structures in the cell. Unlike proteasomes, which are complexes of proteins that degrade cytosolic proteins one at a time, lysosomes are much larger and may not only digest single proteins, but may engulf entire structures in the cells, such as mitochondria and, in neuroendocrine cells, secretory granules (Fig. 1), resulting ultimately in their destruction.

The sorting of soluble acid hydrolases is the best understood of the sorting that is required in the secretory pathway. Lysosomal hydrolases have mannose-6-phosphate groups attached to their carbohydrate moieties and these groups are recognized by a transmembrane protein in the *trans*-Golgi layer called the mannose-6-phosphate receptor. Lysosomal enzymes bind to this transmembrane protein in the lumen of the Golgi cisternae and a protein called adaptor protein 1 (AP-1) binds to this protein on the cytosolic side. Clathrin, a protein identified years ago as a component of a caged structure around certain small membrane vesicles, binds to AP-1 and the membrane in this area begins to invaginate and forms a vesicle that carries the enclosed enzymes to the endosomes, a compartment en route to lysosomes. The two processes, sorting and transport, are linked, because the mannose-6-phosphate receptor that binds the vesicular cargo of lysosomal enzymes on the lumen side also binds the proteins necessary to form the vesicle on the cytosolic side. It seems likely that this mode of

sorting linked to transport may be a model for sorting soluble proteins in transport systems in many parts of the cell and not just to lysosomes.

## SECRETORY GRANULES

The hallmark of the cytoplasm of neuroendocrine cells is large dense-core secretory granules, 100 to 1000 nm in diameter, containing concentrated hormone. The protein is so concentrated that it stains darkly, giving the granules their characteristic dense cores. The process of concentrating protein hormones has been called condensation; in the case of prolactin and growth hormone, in which it has been investigated biochemically, condensation is the formation of a large aggregate of protein hormone that is insoluble under the conditions present in the secretory granule. The aggregate is different from that which forms from unfolded proteins in the endoplasmic reticulum in that the hormone aggregate in the granules will rapidly resolubilize, producing native folded protein, when released from the cell, whereas the aggregates of unfolded proteins that sometimes accumulate in the endoplasmic reticulum are very difficult to resolubilize.

These aggregates form in the *trans*-Golgi layer before secretory granules form; this has been seen clearly by electron microscopy for proteins such as prolactin and atrial natriuretic peptide and detected biochemically for prolactin and growth hormone. As the *trans*-Golgi layer is consumed by budding of small vesicles, the aggregates, too large to be removed, remain, so that the dense core is left with less and less excess membrane around it; at this stage it is termed an immature secretory granule (Fig. 1). Eventually all the excess membrane is completely removed, leaving the mature secretory granule. In cells that make more than one secretory granule protein, the granules may not contain all proteins. Gonadotrophs, for example, produce both luteinizing hormone and follicle-stimulating hormone, but the secretory granules of these cells contain primarily one or the other hormone. In bag cells of *Aplysia californica*, proegg-laying hormone is cleaved into two portions and secretory granules form containing primarily one or the other part of the prohormone. The mechanism that causes these separations appears to be primarily self-association of each protein separately in the Golgi complex.

After they form, secretory granules are transported through the cell to the plasma membrane, where they

are released by exocytosis, when cells are stimulated. These functions of transport and stimulated release are mediated by membrane proteins in the secretory granules, such as synaptobrevin and synaptotagmin, that are known to mediate exocytosis. In addition, there are proteins that are specifically located in dense-core vesicles, such as IA2 (islet antigen 2), also called ICA512, and phogrin. These proteins may play a role in the correct transport of the secretory granules. Therefore, a second step in forming granules is the correct localization of proteins necessary for their proper function.

Regulated secretion of polypeptide hormones occurs through secretory granules formed from the Golgi complex. There is also another type of vesicle causing regulated secretion in neuroendocrine cells, synaptic vesicle-like microvesicles, which are small vesicles, less than 100 nm in diameter, the release of which is also stimulated by the same signals that regulate exocytosis. These vesicles have some of the same proteins that mediate stimulated release, including synaptotagmin and synaptobrevin. They differ from secretory granules, however, in that they do not have dense cores containing protein hormones and are not derived from the Golgi complex. In some cases, their contents are known; the microvesicles in beta cells of pancreatic islets contain  $\gamma$ -aminobutyric acid and those in adrenal cells contain acetylcholine. These vesicles appear closely related to synaptic vesicles in neurons

and, as synaptic vesicles, are formed at the plasma membrane by a process resembling endocytosis.

### See Also the Following Articles

Peptide Hormones and Growth Factors: Cellular Signaling Mechanisms • Peptide Hormones, Intracellular Transport • Peptide Hormones, Regulated Secretion • Peptide Hormones, Regulation and Gene Expression • Peptide Hormones, Segregation Mechanism

### Further Reading

- Dannies, P. S. (1999). Protein hormone storage in secretory granules: Mechanisms for concentration and sorting. *Endocr. Rev.* **20**, 3–21.
- Lee, M. S., Zhu, Y. L., Chang, J. E., and Dannies, P. S. (2001). Acquisition of lubrol insolubility, a common step for growth hormone and prolactin in the secretory pathway of neuroendocrine cells. *J. Biol. Chem.* **276**, 715–721.
- Marsh, B. J., Mastronarde, D. N., Buttle, K. F., Howell, K. E., and McIntosh, J. R. (2001). Organellar relationships in the Golgi region of the pancreatic beta cell line, HIT-T15, visualized by high resolution electron tomography. *Proc. Natl. Acad. Sci. USA* **98**, 2399–2406.
- Marsh, B. J., Mastronarde, D. N., McIntosh, J. R., and Howell, K. E. (2001). Structural evidence for multiple transport mechanisms through the Golgi in the pancreatic beta-cell line, HIT-T15. *Biochem. Soc. Trans.* **29**, 461–467.
- Rambourg, A., Clermont, Y., Chretien, M., and Olivier, L. (1992). Formation of secretory granules in the Golgi apparatus of prolactin cells in the rat pituitary gland: A stereoscopic study. *Anat. Rec.* **232**, 169–179.

by release of inhibitory neurotransmitters [vasoactive intestinal peptide (VIP) and nitric oxide (NO)] from noncholinergic neurons.

Motilin is the only hormonal or neural peptide whose main function is interdigestive. It is released in cycles that coincide with the start of the migrating myoelectric complex. Peaks of circulating motilin coincide with phase 3 of the cycle, a short period of intense motor activity that sweeps residual contents of the intestine caudad. Immunoneutralization of circulating motilin disrupts phase 3 activity; conversely, infusion of motilin in concentrations that mimic circulating levels triggers premature phase 3 activity. Neural activity can modulate motilin levels and motor activity, but the precise mechanisms are not known.

## PEPTIDE NEUROTRANSMITTERS AND SMOOTH MUSCLE ACTIVITY

### Neural Organization

Neurons of the enteric nervous system, especially neurons of the myenteric plexus, modulate the intrinsic electrical and mechanical properties of smooth muscle. They constitute the final relay to smooth muscle cells and interstitial cells of Cajal; the latter appear to be interposed between motor nerve terminals and smooth muscle cells in some regions of the gut. Extrinsic adrenergic and peptidergic neurons in pre- and paravertebral ganglia synapse with and modulate the activity of enteric neurons.

Neurons of the myenteric plexus located between the circular and longitudinal smooth muscle layers fall into two broad categories. Approximately 25% contain VIP and/or its homologue, pituitary adenylate cyclase activating peptide (PACAP), often found together with neuronal nitric oxide synthase (nNOS). More than 60% contain ACh, often found together with the tachykinins, substance P (SP) and neurokinin A (NKA); the latter are derived from the same precursor,  $\beta$ -protachykinin. VIP neurons also contain a homologous peptide derived from the same precursor designated PHI in animals or PHM in humans. There is little or no overlap between neurons of these two categories. Acetylcholine and the tachykinins are the major excitatory motor neurotransmitters, whereas VIP and its homologs and NO are the major inhibitory neurotransmitters. The term inhibitory refers to their ability to inhibit (i.e., relax) smooth muscle tone and rhythmic electrical and contractile activity.

## Pharmacology of Neurotransmitters

### VIP, PACAP, and PHI/PHM

Receptors for these peptides are present on both smooth muscle cells and neurons. VIP and PACAP interact with VPAC<sub>2</sub> receptors on smooth muscle cells, which are coupled via G<sub>s</sub> to stimulation of cAMP formation and inhibition of muscle tone and rhythmic activity. Neural VIP/PACAP receptors are present in two locations: on nerve terminals, where they act as autoinhibitory receptors to suppress further neurotransmitter release, and on cholinergic/tachykinin motor neurons that innervate the longitudinal muscle layer. Activation of these receptors upon release of VIP and PACAP from a subpopulation of myenteric interneurons stimulates release of ACh and tachykinins and induces longitudinal muscle contraction.

### Nitric Oxide

Two types of NOS are found in smooth muscle of the gut: nNOS in neurons and eNOS in smooth muscle cells. The release of VIP and its homologues from nNOS neurons is mediated by NO: Suppression of NO release inhibits VIP release but not vice versa, and addition of NO or NO donors to isolated ganglia or synaptosomes stimulates VIP release. NO released from nerve terminals diffuses to smooth muscle cells to participate in muscle relaxation. There is no molecular or biochemical evidence that VIP and PACAP act on nerve terminals to stimulate NO release.

A distinct interplay occurs between VIP/PACAP and NO in smooth muscle cells. VIP and PACAP interact not only with their cognate VPAC<sub>2</sub> receptors but also with a single-transmembrane receptor, the natriuretic peptide receptor C (NPR-C), which is widely expressed in smooth muscle cells of the gut and is coupled to G<sub>i1</sub> and G<sub>i2</sub>. Detailed cellular and molecular studies have shown that interaction of VIP or PACAP with NPR-C leads to G<sub>i1/2</sub>-dependent activation of eNOS and stimulation of NO formation in smooth muscle cells. Cyclic GMP formed in smooth muscle cells derives from activation of soluble guanylyl cyclase by NO diffusing from nerve terminals and NO formed by eNOS in smooth muscle cells. Thus, NO regulates VIP/PACAP release in neurons; in turn, VIP and PACAP regulate NO formation in smooth muscle cells.

It is worth noting that smooth muscle cells of *Tenia coli* are unique in two respects. The cells do not express eNOS, but they express a VIP-specific receptor that does not recognize PACAP and a PACAP-specific receptor that does not recognize VIP. Only the

VIP-specific receptor is coupled to cAMP formation. The PACAP-specific receptor may be a splice variant of the PAC<sub>1</sub> receptor, which causes relaxation by activating apamin-sensitive K<sup>+</sup> channels.

### **SP and NKA**

SP and NKA interact preferentially with NK<sub>1</sub> and NK<sub>2</sub> receptors on smooth muscle cells and nerve terminals. NKB, the preferred ligand for NK<sub>3</sub> receptors, is the product of a precursor that is virtually absent from the gut. NK<sub>1</sub> and NK<sub>2</sub> receptors are coupled via G<sub>q</sub> to IP<sub>3</sub>-dependent Ca<sup>2+</sup> release and contraction in circular muscle and to arachidonic acid-mediated Ca<sup>2+</sup> influx and Ca<sup>2+</sup>-induced Ca<sup>2+</sup> release in longitudinal muscle. The receptors stimulate ACh release from cholinergic neurons and induce a contraction that is superimposed on the direct contraction.

### **Enkephalin and Dynorphin**

Opioid neurons of the myenteric plexus release a variety of peptides derived from two precursors: proenkephalin, which yields [Met]enkephalin and C-terminally extended derivatives, and prodynorphin, which yields α- and β-endorphin and dynorphin-17; the latter is processed to smaller dynorphin fragments and eventually to [Leu]enkephalin. [Met]enkephalin and its derivatives are much more abundant than dynorphin and its derivatives in the gut. Three opioid receptor types (μ, δ, and κ) are present on circular smooth muscle cells only where they mediate contraction. In addition, μ receptors on cholinergic neurons mediate inhibition of ACh release, and δ receptors on inhibitory neurons mediate inhibition of NO, VIP, and related peptides. The neural receptors are most relevant functionally, particularly δ receptors on inhibitory neurons. When opioids are injected *in vivo* or added to smooth muscle strips *in vitro*, they produce a transient increase in muscle tone by direct action on smooth muscle cells and a sustained increase in rhythmic, nonpropulsive contractile activity that reflects the suppression of a predominant inhibitory neural input.

### **Somatostatin**

Somatostatin is present exclusively in a small population of myenteric interneurons, where it serves as a relay in the peristaltic reflex. Although somatostatin receptors that mediate contraction are present on circular muscle cells, they lack functional relevance in the absence of muscle innervation by somatostatin neurons.

### **CGRP**

In the stomach, CGRP-containing nerve fibers are extrinsic primary afferent fibers, whereas in the intestine and colon they are both extrinsic and intrinsic; the latter mediate the peristaltic reflex induced by mucosal stimuli. CGRP is present in other neurons, but its function is unknown. CGRP receptors on muscle cells are reported to cause direct relaxation.

In summary, neurotransmitters of functional importance are the excitatory motor neurotransmitters, ACh and tachykinins; the inhibitory motor neurotransmitters, NO and VIP and its homologues; the modulatory neurotransmitters, opioids and somatostatin; and the sensory neurotransmitter, CGRP. Excitatory motor neurotransmitters depolarize and inhibitory motor neurotransmitters hyperpolarize membrane potential. The changes in potential are known as excitatory junction potentials and inhibitory junction potentials, respectively. At an appropriate threshold, depolarization opens voltage-sensitive Ca<sup>2+</sup> channels, resulting in Ca<sup>2+</sup> influx, an increase in tone, and increase in the amplitude or frequency of rhythmic contractions. Hyperpolarization affects mainly rhythmic activity, which is either decreased or suppressed.

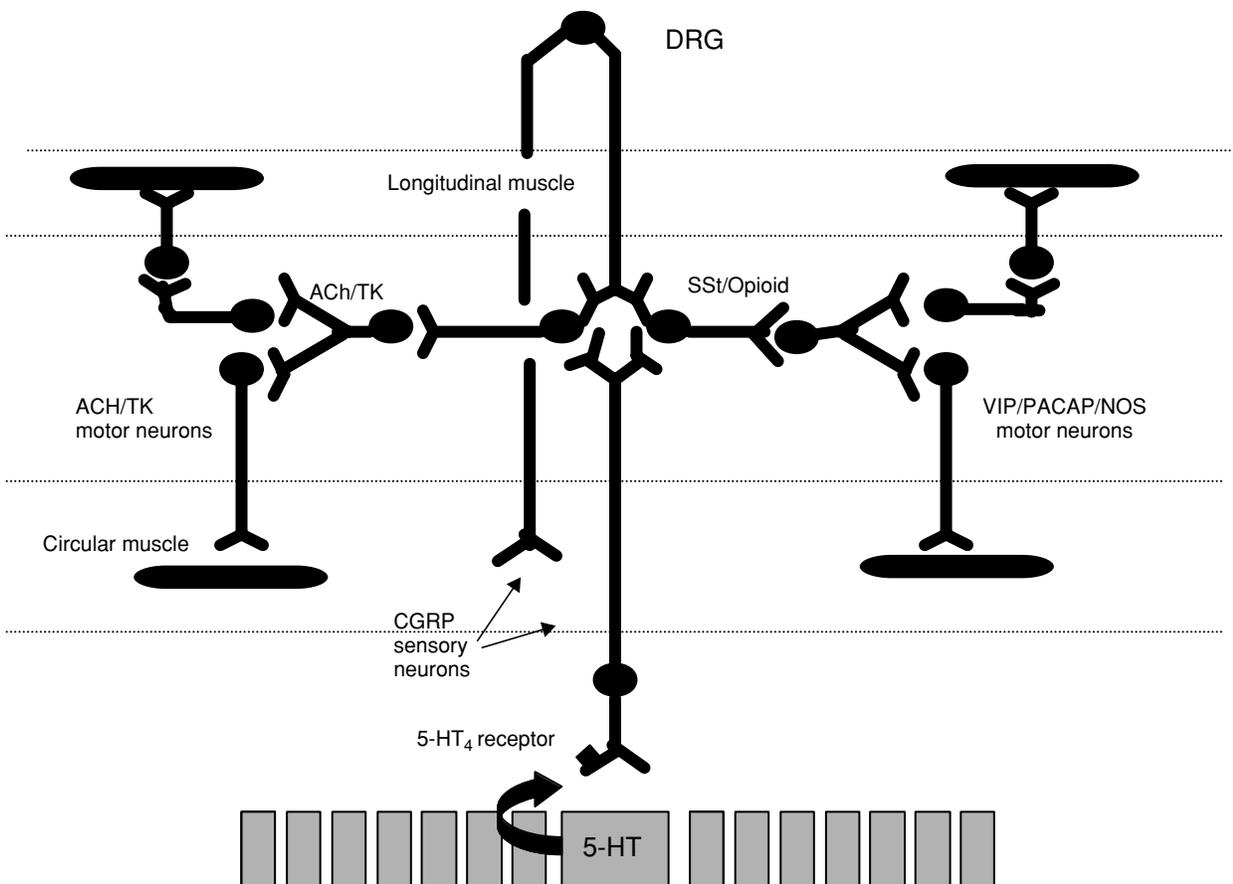
## **THE PERISTALTIC REFLEX: INTERPLAY OF MOTOR, MODULATORY, AND SENSORY NEUROTRANSMITTERS**

The functional role of peptide and nonpeptide neurotransmitters is best exemplified by their coordinated release during the intestinal peristaltic reflex (Fig. 1). The reflex can be evoked by mechanical or chemical stimulation of the mucosa (stroking or pH change) and by radial muscle stretch. The reflex consists of a descending phase, during which circular muscle relaxes and longitudinal muscle contracts caudad to the site of stimulation, and an ascending phase, during which circular muscle contracts and longitudinal muscle relaxes. Two types of *in vitro* preparations have been used to study the reflex that permit simultaneous measurement of mechanical response and neurotransmitter release. In one preparation, a hollow segment of intestine is stretched at the caudad or orad end, and mechanical response is measured as the medium is sampled for neurotransmitter release. In the other, the segment is opened and pinned as a flat sheet, separated into three compartments. The reflex is evoked by stimulation in the central compartment, while mechanical response and neurotransmitter release in each compartment is measured.

### Neurotransmitter Release during the Ascending and Descending Phases of the Reflex

A precise pattern of neurotransmitter release occurs with each phase. During the descending phase, there is an increase in the release of VIP, PACAP, NO, and somatostatin and a decrease in the release of [Met]enkephalin. The pattern of release reflects the role of these neurotransmitters in mediating the descending phase. Thus, release of VIP, PACAP, and NO from the nerve terminals of motor neurons innervating circular muscle mediates descending relaxation of this

layer. Release of the same neurotransmitters from interneurons that synapse with and activate ACh/tachykinin motor neurons that innervate longitudinal muscle mediates reciprocal contraction of this layer. The increase in the activity of VIP/PACAP/NOS motor neurons and interneurons is caused by a decrease in the tonic inhibitory influence of opioid ([Met]enkephalin) interneurons, which in turn is caused by an increase in the activity of somatostatin interneurons. Thus, as the reflex is evoked sensory input is relayed to somatostatin interneurons coupled to opioid interneurons, which in turn are coupled to VIP/PACAP/NOS motor neurons and interneurons (Fig. 1).



**Figure 1** Model illustrating the regulation of intestinal peristaltic reflex. Mucosal stimulation causes the release of 5-HT, which acts on a 5-HT<sub>4</sub> receptor on intrinsic primary afferent CGRP neurons. Mechanical distension of the gut wall activates extrinsic primary afferent CGRP neurons with cell bodies located in the dorsal root ganglion (DRG). Both afferent neurons activate the same reflex pathway. The descending or caudad pathway is mediated by somatostatin (SSt) and opioid peptide interneurons coupled in series to (i) VIP/PACAP/NOS motor neurons, which innervate the circular muscle layer and cause descending relaxation of circular muscle, and (ii) VIP interneurons coupled to ACh/tachykinin (TK) motor neurons, which innervate the longitudinal muscle layer and cause descending contraction of longitudinal muscle during circular muscle relaxation. The ascending or orad pathway is less well defined but involves cholinergic interneurons and results in activation of ACh/TK motor neurons, which innervate the circular muscle layer and cause ascending contraction of circular muscle, and inhibition of ACh/TK motor neurons, which innervate the longitudinal muscle layer and cause ascending relaxation of longitudinal muscle during circular muscle contraction.

A reverse pattern of neurotransmitter release accompanied by an increase in SP and NKA release occurs during the ascending phase and determines the contraction of circular muscle and relaxation of longitudinal muscle. There is little direct inhibitory input to longitudinal muscle, at least in small animals. Contraction and relaxation of this muscle reflect, respectively, the increase and decrease in ACh and tachykinin release from motor neurons (Fig. 1).

### The Sensory Limb of the Peristaltic Reflex

Two distinct populations of sensory neurons mediate the peristaltic reflex evoked by muscle stretch and mucosal stimulation. After nerve degeneration following surgical or chemical (using capsaicin) ganglionectomy, the reflex can be evoked by mucosal stimulation but not by muscle stretch. Conversely, after the mucosa is removed while maintaining intact extrinsic innervation, the reflex can still be evoked by muscle stretch. Thus, mucosal stimulation activates intrinsic sensory neurons with terminals in the mucosa and cell bodies in the myenteric plexus, whereas circular muscle stretch activates sensory neurons with nerve terminals in circular muscle and cell bodies in the dorsal root ganglia. The main sensory neurotransmitter in extrinsic and intrinsic neurons is CGRP, and its effect is relayed to the same circuits of interneurons and motor neurons that mediate the ascending and descending phases of the reflex.

The mechanism by which CGRP-containing intrinsic sensory neurons are activated is unique in that it involves release of 5-hydroxytryptamine (5-HT) from the mucosa. In both humans and experimental animals, mucosal stimulation releases 5-HT from enterochromaffin cells, one of two large stores of 5-HT in the body. 5-HT acts on 5-HT<sub>4</sub> receptors located on mucosal sensory nerve terminals, causing release of CGRP (Fig. 1). The addition of 5-HT or a selective 5-HT<sub>4</sub> agonist to the mucosal surface releases CGRP from capsaicin-sensitive sensory neurons and triggers the peristaltic reflex. Selective 5-HT<sub>4</sub> antagonists block CGRP release and the peristaltic reflex evoked by mucosal stimulation or by the addition of 5-HT to the mucosa. Both 5-HT<sub>4</sub> and CGRP antagonists suppress the ascending and descending phases of the reflex and the release of corresponding neurotransmitters. No release of 5-HT is seen when the peristaltic reflex is evoked by radial muscle stretch. Although the reflex has been traditionally viewed as a stereotypical response to intestinal distension (radial muscle stretch), it is unlikely that this represents the

normal physiological modality. The mere passage and chemical composition of digesta appear to be the main physiological triggers of the reflex and propulsion of intestinal contents.

### Propagation of the Reflex

The caudad propagation of the reflex leads to propulsion of intestinal contents. The velocity of propulsion is readily measured *in vitro* using synthetic fecal pellets inserted into the oral end of the segment without distending the lumen. 5-HT<sub>4</sub> agonists added to the lumen of the segment accelerate the velocity of propulsion, whereas 5-HT<sub>4</sub> antagonists have the opposite effect. Elimination of the inhibitory influence of opioid neurons on VIP/PACAP/NOS neurons with opioid antagonists (particularly  $\delta$  receptor antagonists) accelerates propulsion. The suppression of opioid activity greatly potentiates the ability of 5-HT to accelerate propulsion, such that a combination of near-threshold concentrations of a 5-HT<sub>4</sub> agonist and an opioid  $\delta$  antagonist produces near-maximal stimulation of propulsion. The remarkable effect of this combination indicates its therapeutic potential, and highlights the importance of understanding the physiological interplay of peptide neurotransmitters and their regulators.

### Acknowledgment

This work was supported by Grants DK 15564, DK 28300, and DK 34153 from the National Institute of Diabetes and Digestive and Kidney Diseases.

### See Also the Following Articles

Calcitonin Gene-Related Peptide (CGRP) • CCK (Cholecystokinin) • Neurokinins • Neurotransmitters, Overview • Pituitary Adenylate Cyclase-Activating Polypeptide (PACAP)/Glucagon Superfamily • Somatostatin Analogs • Substance P

### Further Reading

- Costa, M., Brookes, S. J. H., Steele, P. A., Gibbins, I., Burcher, E., and Kandiah, C. J. (1996). Neurochemical classification of myenteric neurons in the guinea pig ileum. *Neuroscience* **75**, 949–967.
- Foxx-Orenstein, A. E., Jin, J.-G., and Grider, J. R. (1998). 5-HT<sub>4</sub> receptor agonists and delta-opioid receptor antagonists act synergistically to stimulate colonic propulsion. *Am. J. Physiol.* **275**, G979–G983.
- Grider, J. R. (1994). CGRP as a sensory transmitter in the sensory pathway that mediates the peristaltic reflex. *Am. J. Physiol.* **266**, G1139–G1145.

- Grider, J. R. (2003). Reciprocal activity of longitudinal and circular muscle during intestinal peristalsis. *Am. J. Physiol.* **284**, 768–775.
- Grider, J. R., and Foxx-Orenstein, A. E. (1998). Neuroendocrine regulation of intestinal peristalsis. In *Endocrinology of the Gastrointestinal System* (G. H. Greeley, ed.), pp. 295–315. Humana, Totowa, NJ.
- Grider, J. R., and Jin, J.-G. (1994). Distinct populations of sensory neurons mediate the peristaltic reflex elicited by muscle stretch and mucosal stimulation. *J. Neurosci.* **14**, 2854–2860.
- Grider, J. R., Katsoulis, S., Schmidt, W. E., and Jin, J.-G. (1994). Regulation of the descending relaxation phase of intestinal peristalsis by PACAP. *J. Auton. Nervous System* **50**, 151–159.
- Grider, J. R., Foxx-Orenstein, A. E., and Jin, J.-G. (1998). 5-Hydroxytryptamine<sub>4</sub> receptor agonists initiate the peristaltic reflex in human, rat, and guinea pig intestine. *Gastroenterology* **115**, 370–380.
- Itoh, Z. (1997). Motilin and clinical application. *Peptides* **18**, 593–608.
- Lu, Y. X., and Owyang, C. (1999). Duodenal acid-induced gastric relaxation is mediated by multiple pathways. *Am. J. Physiol.* **276**, G1501–G1506.
- Makhlouf, G. M. (1997). A century of regulatory peptides: From discovery to function. *Gastroenterology* **112**, 2111–2113.
- Makhlouf, G. M. (2003). Smooth muscle of the gut. In *Textbook of Gastroenterology* (T. Yamada, D. H. Alpers, N. Kaplowitz, L. Laine, C. Owyang, and D. W. Powell, eds.), pp. 92–116. Lippincott, Philadelphia.
- Murthy, K. S., Teng, B. Q., Jin, J.-G., Grider, J. R., and Makhlouf, G. M. (2000). G<sub>i1</sub>/G<sub>i2</sub>-dependent signaling by single-transmembrane natriuretic peptide clearance receptor. *Am. J. Physiol.* **278**, G974–G980.
- Poitras, P. (1994). Motilin. In *Gut Peptides: Biochemistry and Physiology* (J. Walsh and G. J. Dockray, eds.), pp. 261–304. Raven Press, New York.
- Raybould, H. E. (1999). Nutrient tasting and signaling mechanisms in the gut. I. Sensing of lipid by the intestinal mucosa. *Am. J. Physiol.* **277**, G751–G755.
- Teng, B.-Q., Murthy, K. S., Kuemmerle, J. F., Grider, J. R., and Makhlouf, G. M. (1998). Selective expression of VIP<sub>2</sub>/PACAP<sub>3</sub> receptors in rabbit and guinea pig gastric and *Tenia coli* smooth muscle cells. *Regul. Peptides* **77**, 127–134.
- Teng, B.-Q., Grider, J. R., and Murthy, K. S. (2001). Identification of a VIP-specific receptor in guinea pig. *Tenia coli*. *Am. J. Physiol.* **281**, G718–G725.

and a subtle change from full-length PYY stimulating both  $Y_1$  and  $Y_2$  receptors to a predominant PYY(3-36) activation of only  $Y_2$  receptors occurs. This change in peptide pharmacology may be important in modulating feeding behavior and initiating satiety after eating a meal and as a result of postprandial increases in plasma PYY and its conversion to PYY(3-36).

## DISTRIBUTION OF Y RECEPTORS ACTIVATED BY PYY OR PYY(3-36)

Each of the five different types of mammalian Y receptor known to exist ( $Y_1$ ,  $Y_2$ ,  $Y_4$ ,  $Y_5$ , and  $y_6$ ; the last type is written in lowercase because it is not expressed fully in human cells) is most commonly coupled to the inhibitory G protein,  $G_i$ . PYY and its C-terminal product PYY(3-36) exhibit high affinities for  $Y_1$  and  $Y_2$  receptors, respectively, and these receptor types are considered in detail in this article. The  $Y_4$  receptor is not discussed in detail since this receptor exhibits a marked preference for PP and is relatively insensitive to PYY, PYY(3-36), and NPY. It should also be noted that the  $Y_3$  receptor, which has not been cloned but exhibits an apparently unique pharmacology, is most probably an overtly NPY-preferring,  $Y_1$  receptor.

### $Y_1$ Receptors Activated by PYY

NPY and circulating PYY, but not physiological circulating levels of PYY(3-36) (or other C-terminal fragments of either peptide), will activate  $Y_1$  receptors. Synthetic Pro<sup>34</sup>-substituted analogues of NPY and PYY also have a high affinity for this receptor type. The  $Y_1$  receptor is extensively expressed in the brain (e.g., cerebral cortex, dentate gyrus of the hippocampus, and certain thalamic and hypothalamic nuclei), the vascular system (where  $Y_1$  receptors on vascular smooth muscle mediate increased tone), enteric neurons, and epithelia in mammalian intestine. In human and murine intestinal tissues, PYY (NPY and their Pro<sup>34</sup> analogues) activates epithelial  $Y_1$  receptors to cause prolonged anti-secretory effects (a  $G_i$  protein-coupled mechanism). However, the  $Y_1$  receptor is not the only Y receptor of importance in the gut. Both the  $Y_2$  and  $Y_4$  receptors also exhibit functional roles (and all three Y receptors will inhibit mucosal vectorial ion and fluid transport, i.e., anti-secretory effects). The different expression patterns of Y receptors have been described in rat intestine by Laburthe and co-workers and from these detailed studies likely functions may be postulated.

### $Y_2$ Receptors Activated by PYY and PYY(3-36)

In the 1980s, Håkanson and colleagues presented the first pharmacological data indicating that a Y receptor type was preferentially activated by C-terminal fragments of either NPY or PYY [e.g., NPY(3-36), NPY(13-36) or PYY(3-36)] and that this receptor type was located prejunctionally at a number of sympathetic neuro-effector junctions. Activation of these receptors resulted in inhibition of noradrenaline (NA) and NPY release and the  $Y_2$  receptor was therefore proposed to mediate neuronal feedback inhibition. This has subsequently been found to be the case in both the peripheral and central nervous systems in a wide range of mammalian tissues. Such a mechanism in the hypothalamus has also been suggested to underlie the anorexic effects of elevated postprandial PYY(3-36) in human subjects. But it would be wrong to presume that  $Y_2$  receptors are exclusively prejunctional, as this receptor has been found frequently to mediate direct actions on postjunctional target cells, e.g., intestinal epithelia, where  $Y_2$  receptors mediate anti-secretory ion transport effects.

### $Y_5$ Receptors and Their Activation by PYY, PYY(3-36), NPY, and PP

PYY and PYY(3-36) can also stimulate  $Y_5$  receptors and their localization in specific CNS regions outside the blood-brain barrier, known to be involved in appetite control, stimulated great research efforts aimed at discovering novel centrally active anti-obesity drugs. This receptor type is rarely expressed in peripheral tissues and thus circulating PYY and PYY(3-36) are predicted to have little if any peripheral  $Y_5$ -mediated side effects. Initial hopes of a novel Y-related anti-obesity drug have faded as the complexity of multiple peptide pathways regulating appetite has been elucidated.

### Other Y Receptor Types

At physiological concentrations, neither PYY nor PYY(3-36) has a significant effect on the PP-preferring  $Y_4$  receptor. It is also debatable whether the  $Y_3$  receptor actually exists. Subsequent to the sequencing of the human genome, it appears likely that this Y receptor type is probably a  $Y_1$  receptor in tissues that are for some reason significantly less sensitive to PYY.

In conclusion, it can be presumed that elevated circulating levels of PYY will activate  $Y_1$  and  $Y_2$  receptors, whereas its metabolite will preferentially

stimulate  $Y_2$  receptors. Neither peptide is predicted to cross the blood–brain barrier to any significant extent. Studies by Bloom and colleagues show that PYY(3–36) may be an important satiety factor acting on hypothalamic mechanisms in human.

## REGULATION OF PYY SECRETION FROM INTESTINAL ENDOCRINE CELLS

Whereas NPY is an important neuropeptide that acts centrally to initiate feeding, PYY acts to ensure more efficient digestion of the ingested meal. Plasma PYY levels are raised within 30 min of ingestion of a meal, long before the luminal contents have reached the terminal intestine where there is a high density of PYY<sup>+</sup> endocrine cells. PYY release is regulated by and also regulates the parasympathetic nervous system, specifically vagal nerve activity. Luminal nutritional factors, particularly fat and short-chain fatty acids (produced through fermentation of carbohydrates by colonic microflora), have also been found to stimulate PYY release and the peptide subsequently alters intestinal motility and electrolyte secretion/absorption. PYY may thus provide a negative feedback mechanism, inhibiting various mechanisms including neurally mediated pancreatic exocrine secretion after eating a meal.

## FUNCTIONAL CONSEQUENCES OF PYY RELEASE

### Ileal Brake and Epithelial Ion Transport

Two major effects of elevated postprandial PYY are reduced gastric emptying and delayed intestinal transit; both contribute to the ileal brake mechanism (together with reduced exocrine pancreatic secretion). PYY and PYY(3–36) exert long-lasting inhibition of ion and fluid transport, promoting a proabsorptive state in most mammalian intestines, including human isolated intestinal preparations and also in human subjects infused with PYY. These inhibitory effects can be mediated by  $Y_1$  or  $Y_2$  receptor or by a combination of the two together and (depending on the species under study) these effects may be either direct or indirect epithelial mechanisms.

### Central Effects That Alter Intestinal Function

PYY, either when infused intravenously or when injected directly into the dorsovagal complex, inhibits

vagally stimulated gastric acid secretion and motility (a  $Y_2$  receptor-mediated effect).

## Pancreatic Exocrine Secretion

In human, PYY may be involved in stimulating the transition from the digestive to the interdigestive phase. Depending on the type of stimulant, PYY has inhibitory effects on exocrine pancreatic secretion in most species that have been tested. This is likely to be a  $Y_1$  receptor-mediated, indirect effect, as few if any  $Y_1$  receptors could be identified immunochemically on pancreatic duct cells. The mediator(s) of such indirect PYY effects is unknown. The  $Y_1/Y_2$  agonist PYY can also inhibit postprandial cholecystokinin release. Duodenal fat-initiated release of secretin is also attenuated by PYY. Depending on the species of interest, endocrine pancreatic secretion, e.g., insulin and glucagon release, can be inhibited by PYY. Whether these inhibitory effects are direct or indirect also varies according to the prevailing stimulus. Evidence therefore indicates that PYY has the potential to inhibit pancreatic secretions at multiple sites and that the peptide is one of the major regulators that ultimately slow normal intestinal function.

## Fatty Acid-Induced Differentiation

Aponte and colleagues have suggested that PYY and free fatty acids (FFA) may act in a concerted manner to change mucosal cell differentiation. Continuous replacement of mucosal cells by differentiating crypt stem cells is a process that is essential for maintaining normal intestinal function. The fact that FFA stimulate PYY release and that the peptide may then alter epithelial cell adhesion and differentiation provides another mechanism by which PYY could alter intestinal function over the longer term.

## Growth Effects and Cancer

*In vivo* studies show that PYY infused to mimic postprandial levels has protective effects on gastric erosion stimulated by ethanol. This protective mechanism appears to be independent of vagal input and is likely to be  $Y_1$  receptor-mediated. Colonic PYY has been shown to have local trophic effects in rat and mouse large bowel; thus, the peptide has a role in intestinal development and dietary adaptation. However, PYY expression is reduced in carcinomas compared with surrounding tissue areas. Specific adenocarcinoma cell lines constitutively express  $Y_1$  and  $Y_4$  receptors (and

would therefore be expected to be sensitive to circulating PYY and PP, respectively). Whether PYY has a role in the dysregulation of cell growth and the development and progression of adenocarcinoma is not clear. Its growth-promoting effects certainly require further investigation before stable analogues of PYY may be considered as novel therapies, e.g., to enhance the nutritional status of cancer patients, in combination with other agents.

## PYY AND INTESTINAL DISORDERS

Elevated plasma levels of PYY have been reported in several diseases including steatorrhea, inflammatory bowel diseases, and acute infective diarrhea. More commonly, the changes in plasma PYY appear to be an adaptive response to alterations in the pathophysiology of gut function. However, in the condition of chronic idiopathic slow transit constipation, PYY appears to be a primary cause and may be one of the etiological factors of this particular syndrome. For these patients, who are insensitive to laxatives or bulking agents, only an enema will promote defecation. Although basal and peak circulating levels of PYY appear unchanged, the tissue levels of PYY are raised in these patients. Individuals show disruption of their neuroendocrine peptides and an increase in the number of PYY-containing endocrine cells in the ascending colon compared with controls. Increased PYY levels could therefore be responsible for the increased absorption, reduced secretion of water and electrolytes, and reduced motility (together with enhanced ileal brake), all of which would be proconstipatory.

Surgical removal of large intestine as a treatment for gastrointestinal disease, such as colorectal carcinoma or chronic slow transit constipation, also significantly removes the main source of PYY from the body. However, rapid adaptive changes appear to occur post-operatively so that these patients show either unaltered or elevated postprandial PYY levels and increased tissue PYY levels adjacent to the anastomosis site.

In conclusion, PYY exhibits multiple actions, many of them inhibitory and mediated by at least two of the five cloned Y receptors, depending on their discrete and differential expression. The PYY-synthesizing endocrine L cells of the large bowel play a pivotal role, not only in the regulation of digestive processes, but also in the adaptive responses to a range of luminal cues and changing dietary components.

## See Also the Following Articles

GI Hormones Outside the Gut: Central and Peripheral Nervous System • Neuropeptide Y • Neuropeptide Y, Evolution of • Pancreatic Polypeptide (PP)

## Further Reading

- Batterham, R. L., Cowley, M. A., Small, C. J., Herzog, H., Cohen, M. A., Dakin, C. L., Wren, A. M., Brynes, A. E., Low, M. J., Ghatei, M. A., Cone, R. D., and Bloom, S. R. (2002). Gut hormone PYY(3-36) physiologically inhibits food intake. *Nature* **418**, 650–654.
- Böttcher, G., Sjölund, K., Ekblad, E., Håkanson, R., Schwartz, T. W., and Sundler, F. (1984). Co-existence of PYY and glicentin in endocrine cells of the gut. *Regul. Pept.* **8**, 261–266.
- Cox, H. M. (1998). Peptidergic regulation of intestinal ion transport: A major role for pancreatic polypeptide family. *Digestion* **59**, 395–399.
- Larhammar, D. (1996). Evolution of neuropeptide Y, peptide YY and pancreatic polypeptide. *Regul. Pept.* **62**, 1–11.
- Michel, M. C., Beck-Sickinger, A., Cox, H. M., Doods, H. N., Herzog, H., Larhammar, D., Quirion, R., Schwartz, T. W., and Westfall, T. (1998). XVI. IUPHAR recommendations for the nomenclature of neuropeptide Y, peptide YY and pancreatic polypeptide receptors. *Pharmacol. Rev.* **50**, 143–150.
- Playford, R. J., Domin, J., Beecham, J., Parmar, K. B., Tatemoto, K., Bloom, S. R., and Calam, J. (1990). Peptide YY: A natural defence against diarrhea. *Lancet* **335**, 1555–1557.
- Sundler, F., Böttcher, G., Ekblad, E., and Håkanson, R. (1993). PP, PYY and NPY, occurrence and distribution in the periphery. In "The Biology of Neuropeptide Y and Related Peptides" (W. F. Colmers and C. Wahlestedt, eds.), pp. 157–183. Humana Press, Totowa, NJ.
- Wahlestedt, C., Yanaihara, N., and Håkanson, R. (1986). Evidence for different pre- and post-junctional receptors for neuropeptide Y and related peptides. *Regul. Pept.* **13**, 307–318.

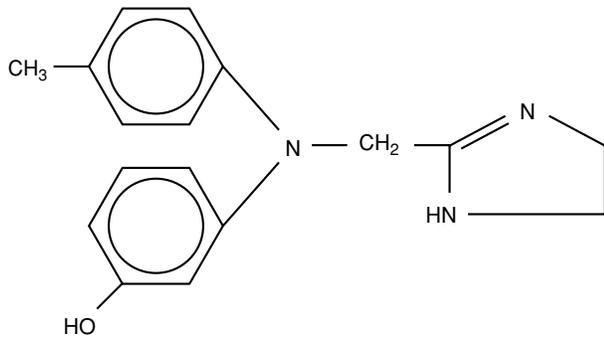


Figure 1 The chemical structure of phentolamine.

doxazosine, is generally preferred in the prevention of hypertensive paroxysms or in the preoperative medical treatment of patients with pheochromocytoma.

Phentolamine is also used in the prevention or treatment of dermal necrosis and sloughing following intravenous administration or extravasation of norepinephrine.

Phentolamine has several other uses as well.

Phentolamine may be useful in relieving pseudo-obstruction of the bowel in patients with pheochromocytoma. This condition is dependent on the inhibitory effect of catecholamines on intestinal smooth muscle.

Phentolamine may be useful in the treatment of hypertensive crises caused by the abrupt withdrawal of clonidine or by the ingestion of tyramine-containing foods in patients treated with nonselective inhibitors of monoamine oxidase.

The use of phentolamine as a test for the diagnosis of pheochromocytoma is considered obsolete. In fact, the test, which is based on the degree of blood pressure decrease after intravenous phentolamine administration, is characterized by low sensitivity and specificity.

### See Also the Following Articles

Adrenergic Mechanisms • Adrenergic Receptors • Antiadrenergic Agents • Catecholamines • Phenylethanolamine *N*-Methyltransferase • Pheochromocytoma

### Further Reading

- Hofmann, B. B. (2001). Catecholamines, sympathomimetic drugs and adrenergic receptor antagonists. *In* "Goodman & Gilman's The Pharmacological Basis of Therapeutics" (J. G. Hardman and L. E. Limbird, eds.), pp. 215–268. McGraw-Hill, New York.
- Young, J. B., and Landsberg, L. (1998). Catecholamines and the adrenal medulla. *In* "Williams' Textbook of Endocrinology" (J. D. Wilson, D. W. Foster, H. M. Kronenberg, and P. R. Larsen, eds.), pp. 665–728. W. B. Saunders, Philadelphia, PA.

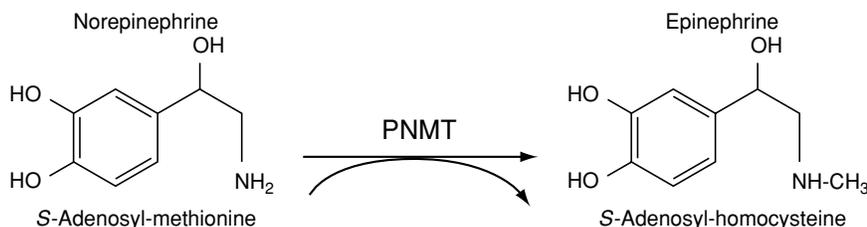


Figure 1 Action of PNMT.

PNMT mRNA is also analyzed by Northern blot analysis using a PNMT cDNA probe.

## LOCALIZATION

In mammalian tissues, only the heart, adrenal medulla, brain, and brainstem have measurable amounts of PNMT.

PNMT is found mainly in the adrenal medulla and in a few neurons in the CNS. The enzyme is found in the cytoplasm of chromaffin cells.

PNMT activity exists in human lung, kidney, heart, liver, spleen, and pancreas. There is a significant correlation between tissue PNMT and epinephrine levels. PNMT in human lung and that in human bronchial epithelial cells were indistinguishable from adrenal PNMT. PNMT activity is present in red blood cells and cancer cell lines. Human kidney, lung, and pancreas show immunohistochemical staining with an antibody to adrenal PNMT.

Radiotracer studies indicate that nearly half of the epinephrine in urine may be made by the kidney.

## PNMT GENE, mRNA, AND PROTEIN

The PNMT gene is located on chromosome 17. PNMT is encoded by a single gene that consists of three exons and two introns, spanning approximately 2100 bp.

The 5'-flanking regions of the gene contain possible transcription regulatory elements, such as the cyclic AMP-response element, glucocorticoid-response element, and Sp1-binding site.

In addition to the major human PNMT mRNA (type A, 1.0 kb), a minor mRNA (type B, 1.7 kb) was observed. The two types of mRNA are thought to be produced from a single gene through the use of alternative promoters.

Determination of the nucleotide sequence revealed that human PNMT consists of 282 amino acid residues with a predicted molecular weight of 30,853 kDa, including the initial methionine.

## REGULATION OF PNMT

In embryonic adrenal gland, development of the enzyme activity is regulated by glucocorticoids and in cultured chromaffin cells and superior cervical ganglia, the activity is induced by nerve growth factor.

PNMT activity in a human bronchial epithelial cell line is dramatically increased by incubation with dexamethasone. Red blood cell PNMT activity is lower in males than in females, is increased in hyperthyroidism, and is decreased in hypothyroidism.

Pituitary adenylate cyclase-activating polypeptide, a member of the secretin/glucagon/vasoactive intestinal peptide family, induces the expression of PNMT mRNA. Egr-1 also increases the expression of the PNMT gene.

## DISORDER IN DISEASES

### Cushing's Syndrome

A possible contributing mechanism to the elevated cardiac output in Cushing's syndrome may be the increased epinephrine formation resulting from enhanced PNMT activity in the adrenal medulla and also in the myocardium.

### Parkinson's Disease

Parkinson's disease primarily affects the dopaminergic nigrostriatal system; however, various cells in brain and adrenal medulla that express PNMT or exist in proximity to PNMT-expressing cells are also affected in the disease.

### Multiple Sclerosis

PNMT is known to map to a region identified in two genome screens for multiple sclerosis (MS) and it directly regulates the amounts of norepinephrine and epinephrine, both of which play a significant role in the modulation of the innate immune response. The frequencies of two promoter polymorphisms of the

PNMT gene showed genetic association with MS in a case control study.

### Pheochromocytoma

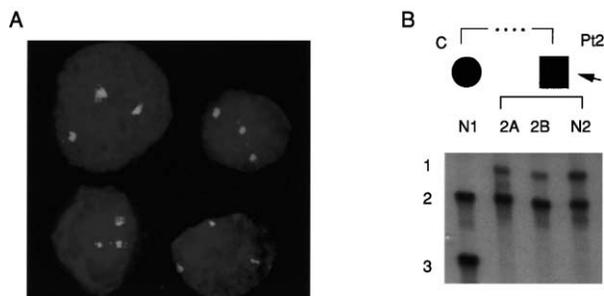
PNMT mRNA expression is greatly enhanced in epinephrine-secreting pheochromocytomas and gene expression may be enhanced by both cortisol and Egr-1.

### See Also the Following Articles

Catecholamines • Norepinephrine Receptors • Norepinephrine Transporter • Phentolamine • Pheochromocytoma

### Further Reading

- Gearhart, D. A., Neafsey, E. J., and Collins, M. A. (2002). Phenylethanolamine N-methyltransferase has  $\beta$ -carboline 2N-methyltransferase activity: Hypothetical relevance to Parkinson's disease. *Neurochem. Int.* **40**, 611–620.
- Isobe, K., Nakai, T., Yashiro, T., Nanmoku, T., Yukimasa, N., Ikezawa, T., Suzuki, E., Takekoshi, K., and Nomura, F. (2000). Enhanced expression of mRNA coding for the adrenaline-synthesizing enzyme phenylethanolamine-N-methyltransferase in adrenaline-secreting pheochromocytomas. *J. Urol.* **163**, 357–362.
- Isobe, K., Tatsuno, I., Yashiro, T., Nanmoku, T., Takekoshi, K., Kawakami, Y., and Nakai, T. (2003). Expression of mRNA for PACAP and its receptors in intra- and extra-adrenal human pheochromocytomas and their relationship to catecholamine synthesis. *Regul. Peptides* **110**, 213–217.
- Kaneda, N., Ichinose, H., Kobayashi, K., Oka, K., Kishi, F., Nakazawa, A., Kurosawa, Y., Fujita, K., and Nagatsu, T. (1988). Molecular cloning of cDNA and chromosomal assignment of the gene for human phenylethanolamine N-methyltransferase, the enzyme for epinephrine biosynthesis. *J. Biol. Chem.* **263**, 7672–7677.
- Kennedy, B., Bigby, T. D., and Ziegler, M. G. (1995). Nonadrenal epinephrine-forming enzymes in humans. *J. Clin. Invest.* **95**, 2896–2902.
- Mann, M. B., Wu, S., Rostamkhani, M., Tourtellotte, W., MacMurray, J. P., and Comings, D. E. (2002). Association between the phenylethanolamine N-methyltransferase gene and multiple sclerosis. *J. Neuroimmunol.* **124**, 101–105.
- Nagatsu, T. (1991). Genes for human catecholamine-synthesizing enzymes. *Neurosci. Res.* **12**, 315–345.
- Ziegler, M. G., Kennedy, B., and Elayan, H. (1988). A sensitive radioenzymatic assay for epinephrine forming enzymes. *Life Sci.* **43**, 2117–2122.



**Figure 1** Duplication of mutant *RET* in trisomy 10 in MEN2-related pheochromocytoma. (A) Fluorescence *in situ* hybridization analysis showing trisomy for chromosome 10 in pheochromocytoma cells. (B) C, cousin; Pt2, patient; N, normal tissue; 2A and 2B, pheochromocytoma; 1–3 are alleles, with allele 2 being the inherited, mutant *RET* allele shared by cousin C and patient Pt2. Modified from Huang *et al.* (2000).

corneal nerve fibers, and marfanoid habitus. The *RET* mutation affects exons 15 and 16, coding for the intracellular tyrosine kinase domain of RET protein.

### Von Hippel–Lindau Syndrome

Von Hippel–Lindau (VHL) syndrome is another autosomal-dominantly inherited tumor syndrome associated with pheochromocytoma. Apart from pheochromocytoma, major tumors in VHL disease include renal cell carcinoma, hemangioblastoma, neuroendocrine pancreatic tumors, and endolymphatic sac tumors. Pheochromocytoma occurs in approximately 10–20% of VHL patients and is the presenting manifestation in approximately 5% of cases. In contrast to pheochromocytomas of patients with MEN2 (“adrenergic”), VHL-associated pheochromocytomas have a more “noradrenergic” phenotype, pointing to their different developmental stage; i.e., MEN2-related pheochromocytomas frequently possess all the enzymes to synthesize catecholamines from tyrosine to epinephrine, whereas less differentiated pheochromocytomas including those in VHL syndrome lack the enzymes involved in the final catecholamine biosynthesis pathway. Thus, symptoms and signs of patients with VHL-related pheochromocytoma are related to norepinephrine excess including, e.g., hypertension. The VHL tumor suppressor gene is located on chromosome 3p25–p26. A “second hit” is required in patients with *VHL* germ-line mutations to develop pheochromocytoma.

### Neurofibromatosis Type 1

Neurofibromatosis type 1 (NF1) is the most common familial cancer syndrome predisposing to

pheochromocytoma and affects approximately 1 in 3000 individuals. The risk of pheochromocytoma in NF1, however, is small, approximately 1%. NF1 is inherited in an autosomal-dominant manner. Only 12% of NF1 patients are diagnosed with bilateral and multifocal pheochromocytomas and less than 5% of patients have metastatic pheochromocytoma. The *NF1* gene is a tumor suppressor gene mapping to chromosome 17q11.2.

### BENIGN VERSUS MALIGNANT PHEOCHROMOCYTOMA

Although mostly benign, approximately one-third of pheochromocytomas are malignant. Distinguishing benign from malignant pheochromocytoma is not possible based on histopathological features. Malignant pheochromocytoma is diagnosed based on the presence of metastatic lesions at sites where chromaffin cells are not usually present (e.g., liver, lungs, lymphatic nodes, and bones). Therefore, all pheochromocytomas, whether benign or not, must be considered to have the potential for malignancy. Whether benign or malignant, pheochromocytomas are potentially lethal tumors that require prompt diagnosis, localization, and treatment.

### DIAGNOSIS

Biochemical evidence of excessive catecholamine production is crucial for the diagnosis of pheochromocytoma. Commonly used tests include 24h urinary outputs of catecholamines (norepinephrine and epinephrine), metanephrines (normetanephrine and metanephrine), and vanillylmandelic acid (VMA). Spectrometric assays of catecholamines, total metanephrines, and VMA have been superseded by high-performance liquid chromatography (HPLC) assays that allow diagnostically more sensitive measurements of fractionated norepinephrine and epinephrine or normetanephrine and metanephrine. HPLC assays have also been adapted for measurements of the much lower plasma concentrations of catecholamines and metanephrines. The latter can be measured in either the free or the much higher deconjugated forms (i.e., free plus sulfate-conjugated metanephrines). Urinary assays involve measurements of deconjugated metanephrines. Pheochromocytomas continuously produce free metanephrines by a process that is independent of catecholamine release. Measurements of free metanephrines in plasma are therefore proving to offer the most sensitive test for the diagnosis of pheochromocytoma. Since familial pheochromocytoma

has variable expression and age- and tumor-dependent penetrance, patients with familial predisposition to develop pheochromocytoma should be followed up on an annual basis, including the measurement of plasma metanephrines.

## LOCALIZATION

Computed tomography (CT) and magnetic resonance imaging (MRI) are the most appropriate imaging modalities for initial localization of pheochromocytoma and both offer high sensitivity for detection of adrenal tumors. MRI is superior to CT for detecting extra-adrenal tumors. Both CT and MRI, however, have inadequate specificity to positively identify a mass as a pheochromocytoma. Metaiodobenzylguanidine (MIBG) scintigraphy provides an imaging modality that offers high specificity and is particularly useful for detecting extra-adrenal tumors and metastases. Sensitivity of MIBG scintigraphy is not as high as that of CT or MRI, but is higher for  $^{123}\text{I}$ -labeled than for  $^{131}\text{I}$ -labeled MIBG. Other promising methods for specific localization of pheochromocytoma include positron emission tomography coupled with agents such as [ $^{11}\text{C}$ ]hydroxyephedrine or 6-[ $^{18}\text{F}$ ]fluorodopamine. In rare cases, where imaging studies are all negative, but where suspicion of a pheochromocytoma remains high, vena cava sampling may be useful for localizing the source of the high circulating levels of catecholamines or free metanephrines.

## TREATMENT

Surgery provides the only effective curative treatment for pheochromocytoma. Because of the potentially fatal consequences of surgical anesthesia-induced release of catecholamines by a tumor, it is imperative that patients with pheochromocytoma be appropriately prepared for surgery. Maintenance of adequate blood pressure control using  $\alpha$ -adrenergic blockers (e.g., phenoxybenzamine) or calcium channel blockers before surgery is important. Administration of  $\alpha$ -methyl-*para*-tyrosine (metyrosine) to block synthesis of catecholamines may also be appropriate in patients with larger active tumors. There is no effective treatment for malignant pheochromocytoma. Chemotherapy with cyclophosphamide, vincristine, and dacarbazine may produce partial remission. Radiotherapies using  $^{131}\text{I}$ -labeled MIBG or indium- or yttrium-labeled octreotide provide benefit in some patients with malignant pheochromocytoma, but again have not proven curative.

## See Also the Following Articles

Adrenal Tumors, Molecular Pathogenesis • Catecholamines • Hypertension, Endocrine • Incidentaloma, Adrenal • Multiple Endocrine Neoplasia (MEN) Type 2 • Neurofibromatosis • Von Hippel-Lindau Syndrome

## Further Reading

- Bender, B. U., Gutsche, M., Glasker, S., Muller, B., Kirste, G., Eng, C., and Neumann, H. P. (2000). Differential genetic alterations in von Hippel-Lindau syndrome-associated and sporadic pheochromocytomas. *J. Clin. Endocrinol. Metab.* **85**, 4568–4574.
- Bravo, E. (1994). Evolving concepts in the pathophysiology, diagnosis, and treatment of pheochromocytoma. *Endocr. Rev.* **15**, 356–368.
- Eisenhofer, G., Keiser, H., Friberg, P., Mezey, E., Huynh, T. T., Hiremagalur, B., Ellingson, T., Duddempudi, S., Eijsbouts, A., and Lenders, J. W. (1998). Plasma metanephrines are markers of pheochromocytoma produced by catechol-*O*-methyltransferase within tumors. *J. Clin. Endocrinol. Metab.* **83**, 2175–2185.
- Eisenhofer, G., Lenders, J. W., Linehan, W. M., Walther, M. M., Goldstein, D. S., and Keiser, H. R. (1999). Plasma normetanephrine and metanephrine for detecting pheochromocytoma in von Hippel-Lindau disease and multiple endocrine neoplasia type 2. *N. Engl. J. Med.* **340**, 1872–1879.
- Huang, S. C., Koch, C. A., Vortmeyer, A. O., Pack, S. D., Lichtenauer, U. D., Mannan, P., Lubensky, I. A., Chrousos, G. P., Gagel, R. F., Pacak, K., and Zhuang, Z. (2000). Duplication of the mutant RET allele in trisomy 10 or loss of the wild-type allele in multiple endocrine neoplasia type 2-associated pheochromocytomas. *Cancer Res.* **60**, 6223–6226.
- Lenders, J. W., Keiser, H. R., Goldstein, D. S., Willemsen, J. J., Friberg, P., Jacob, M. C., Kloppenborg, P. W., Thien, T., and Eisenhofer, G. (1995). Plasma metanephrines in the diagnosis of pheochromocytoma. *Ann. Intern. Med.* **123**, 101–109.
- Margioris, A., Chrousos, G. P. (eds.) (2001). "Adrenal Disorders." Humana Press, Totowa, NJ.
- Pacak, K., Eisenhofer, G., Carrasquillo, J. A., Chen, C. C., Li, S. T., and Goldstein, D. S. (2001). 6-[ $^{18}\text{F}$ ]Fluorodopamine positron emission tomographic (PET) scanning for diagnostic localization of pheochromocytoma. *Hypertension* **38**, 6–8.
- Pacak, K., Linehan, W. M., Eisenhofer, G., Walther, M. M., and Goldstein, D. S. (2001). Recent advances in genetics, diagnosis, localization, and treatment of pheochromocytoma. *Ann. Intern. Med.* **134**, 315–329.
- Sisson, J. C., and Shulkin, B. L. (1999). Nuclear medicine imaging of pheochromocytoma and neuroblastoma. *Q. J. Nucl. Med.* **43**, 217–223.
- Walther, M., Reiter, R., Keiser, H., et al. (1999). Clinical and genetic characterization of pheochromocytoma in von Hippel-Lindau families: Comparison with sporadic pheochromocytoma gives insight into natural history of pheochromocytoma. *J. Urol.* **162**, 659–664.
- Xu, W., Mulligan, L. M., Ponder, M. A., Liu, L., Smith, B. A., Mathew, C. G., and Ponder, B. A. (1992). Loss of NF1 alleles in pheochromocytomas from patients with type I neurofibromatosis. *Genes Chromosomes Cancer* **4**, 337–342.

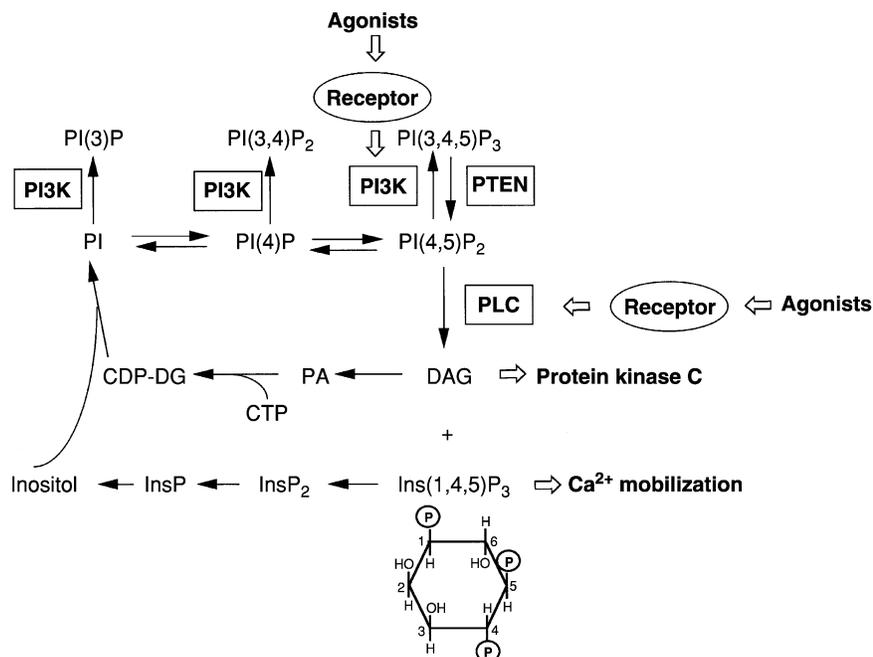
different types of cell-surface receptors that are activated by extracellular signaling molecules (agonists), including hormones, neurotransmitters, cytokines, growth factors, and antigens.

### METABOLIC PATHWAYS FOR PHOSPHATIDYLINOSITOL TURNOVER AND ROLES OF INTERMEDIATES

These PLC products, DAG and Ins(1,4,5)P<sub>3</sub>, are further metabolized by numerous enzymes and finally resynthesized to the PLC precursor PI(4,5)P<sub>2</sub> through many intermediate steps (Fig. 1). The entire receptor agonist-sensitive lipid pool (estimated at 80% of the total) is metabolized several times per minute. Thus, during receptor stimulation, PI(4,5)P<sub>2</sub> is continuously synthesized and supplied to PLC. This explains why the amount of Ins(1,4,5)P<sub>3</sub> produced often exceeds the decrease in the amount of PI(4,5)P<sub>2</sub> and, in some cases, the precursor lipid fails to decrease under receptor stimulation. This also explains why receptor stimulation resulted in an increase in <sup>32</sup>P incorporation

in phospholipids such as phosphatidylinositol (PI) and phosphatidic acid, which was reported in the early studies. Therefore, the PI response is also called PI turnover or PI cycle.

PI(4,5)P<sub>2</sub> is a substrate for not only PLC but also PI 3-kinase (PI3K), which catalyzes the phosphorylation of PI(4,5)P<sub>2</sub> to produce PI(3,4,5)P<sub>3</sub>. There are several isoforms of PI3K and certain forms catalyze the conversion of PI to PI 3-phosphate [PI(3)P]. PI3K is also regulated by several extracellular stimuli through cell-surface receptors. The PI3K products PI(3,4,5)P<sub>3</sub> and PI(3)P as well as PI(4,5)P<sub>2</sub> serve as a membrane docking site for PLCs and other signaling molecules through their pleckstrin homology (PH) domains. PI(3,4,5)P<sub>3</sub> is also an important signaling molecule for cell survival through the activation of Akt/protein kinase B. The role of the cell-survival activity of PI(3,4,5)P<sub>3</sub> is supported by the observation that the loss of function mutation of the phosphatase and tensin homologue, which catalyzes the reverse reaction of PI3K, is found in certain cancer cells. PI(4,5)P<sub>2</sub> also plays critical roles as regulators for several enzymes such as protein kinase C, phospholipase



**Figure 1** Metabolic pathway for inositol lipids and regulation by receptor stimulation. Receptor stimulation induces the activation of phosphoinositide-specific phospholipase C (PLC), which catalyzes the cleavage of phosphatidylinositol 4,5-bisphosphate [PI(4,5)P<sub>2</sub>] into two intracellular second messengers: diacylglycerol (DAG), an activator for protein kinase C, and inositol 1,4,5-trisphosphate [Ins(1,4,5)P<sub>3</sub>], a Ca<sup>2+</sup> mobilizer from intracellular stores. The PLC substrate, PI(4,5)P<sub>2</sub>, is also metabolized to phosphatidylinositol 3,4,5-trisphosphate [PI(3,4,5)P<sub>3</sub>] by phosphatidylinositol 3-kinase (PI3K). The enzyme also produces phosphatidylinositol 3-phosphate [PI(3)P]. The enzyme activity is also regulated by receptors. These phosphoinositides play crucial roles in the recruitment of PLC from cytosol to membrane. PI, phosphatidylinositol; PI(4)P, phosphatidylinositol 4-phosphate; PA, phosphatidic acid; PTEN, phosphatase and tensin homologue.

D, and phospholipase A2. In addition, this phosphoinositide modulates the actin cytoskeleton through interaction with a variety of actin regulatory proteins.

## FOUR TYPES OF PLC AND THEIR DOMAIN STRUCTURE

Eleven isoforms of PLC have been isolated and are divided into four types based on their structure:  $\beta$  ( $\beta 1$ ,  $\beta 2$ ,  $\beta 3$ ,  $\beta 4$ );  $\gamma$  ( $\gamma 1$ ,  $\gamma 2$ );  $\delta$  ( $\delta 1$ ,  $\delta 2$ ,  $\delta 3$ ,  $\delta 4$ ); and  $\epsilon$  (Table I). Among them,  $\epsilon$ -type was identified in 1998 and its characteristics are being studied. In addition, it has been reported that numerous spliced variants are present in mammals. PLCs are also found in simple organisms such as yeasts, slime molds, filamentous fungi, and plants and these PLCs resemble mammalian  $\delta$ , which is only  $\sim 85$  kDa and is the smallest of the

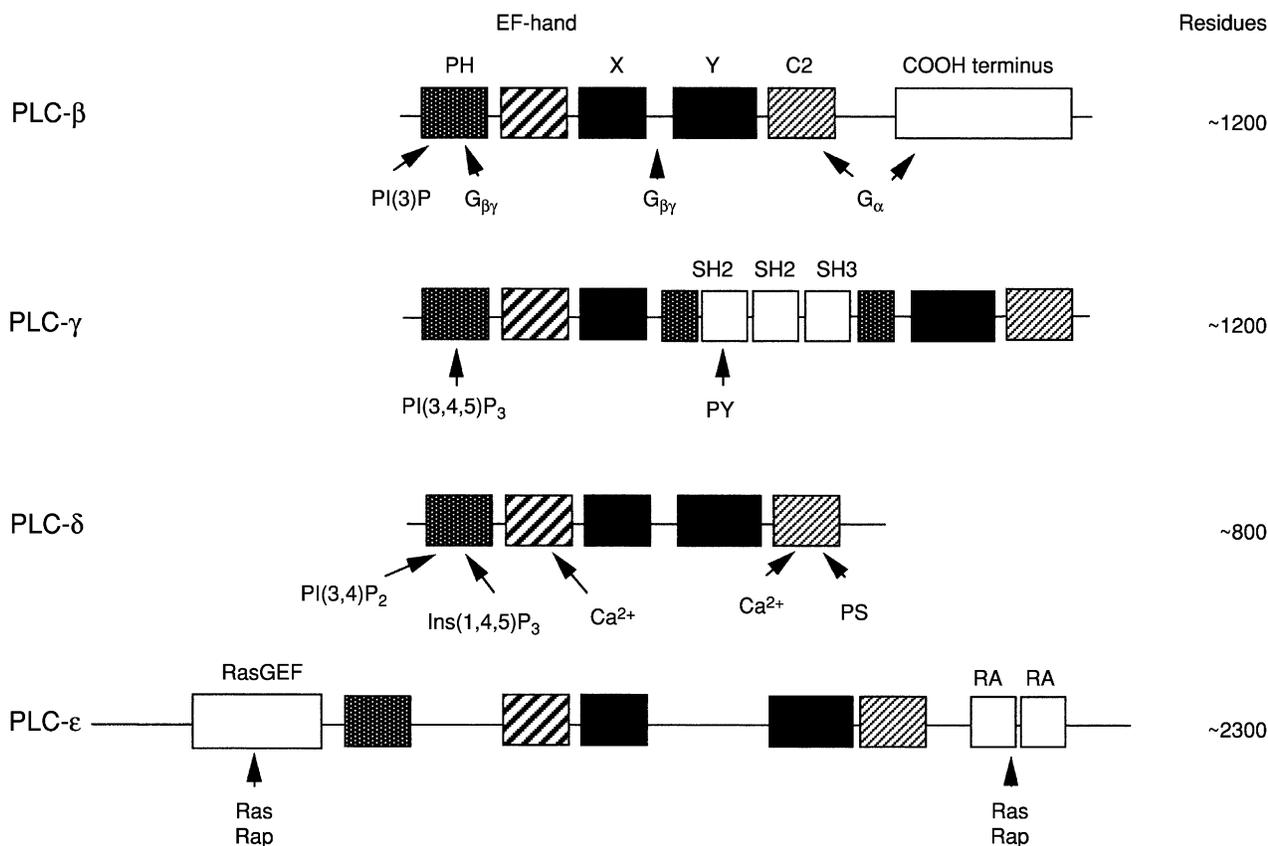
PLCs. Thus, PLC- $\beta$ , PLC- $\gamma$ , and PLC- $\epsilon$ , which are present in higher eukaryotes, are thought to have arisen from an evolutionarily primitive PLC- $\delta$ .

The sequences of PLC contain a string of molecular domains (Fig. 2). All isoforms contain catalytic domains (X and Y) and regulatory domains that serve to recruit PLC to the vicinity of their substrate or activators through protein-protein or protein-lipid interactions. These regulatory domains include a PH domain, an EF-hand domain, and a C2 domain. The PH domains are composed of  $\sim 100$  amino acid residues and have been identified in more than 100 other signaling proteins. These domains bind to polyphosphoinositides (or inositol phosphate) and  $\beta\gamma$ -subunits of heterotrimeric G proteins ( $G_{\beta\gamma}$ ) and appear to function as adapters or tethers, linking their host proteins to the membrane surface. A single PH

**Table I** Receptors and Regulators Linked to PLC Isoforms and Their Distribution

Isoforms	Receptors	Regulators	Distribution	
PLC- $\beta 1$	GPCRs	$G_{\alpha q}$	PI(3)P	Widely expressed, especially in cerebral cortex and hippocampus in brain
PLC- $\beta 2$	GPCRs	$G_{\beta\gamma}$	PI(3)P	Hematopoietic cells
PLC- $\beta 3$	GPCRs	$G_{\alpha q}$ , $G_{\beta\gamma}$	PI(3)P	Widely expressed
PLC- $\beta 4$	GPCRs	$G_{\alpha q}$	PI(3)P	Retina and cerebellum
PLC- $\gamma 1$	RTKs (PDGF, EGF, FGF, NGF, VEGF, HGF, etc.) Cytokines and immunoglobulin receptors (TCR, BCR, Fc $\epsilon$ RI, Fc $\gamma$ Rs etc.)	NRTKs (Src, Syk, Tec, etc.)	PI (3,4,5) P <sub>3</sub>	Widely expressed, especially in brain and lung
PLC- $\gamma 2$	GPCRs (acetylcholine, angiotensin II, thrombin, PAF, ATP, etc.)	(Src/RTK)	PI(3,4,5)P <sub>3</sub>	Hematopoietic cells
PLC- $\delta 1$	$\alpha 1$ -Adrenergic, oxytocin, thromboxan A <sub>2</sub> , bradykinin	$G_{\text{H}}$	Ca <sup>2+</sup> , PI(4,5)P <sub>2</sub>	Most abundant among $\delta$ -types and widely expressed
PLC- $\delta 2$			Ca <sup>2+</sup> , PI(4,5)P <sub>2</sub>	Cerebral cortex and cerebellum Purkinje cells
PLC- $\delta 3$			Ca <sup>2+</sup> , PI(4,5)P <sub>2</sub>	Kidney, cardiac muscle, aorta, brain
PLC- $\delta 4$			Ca <sup>2+</sup> , PI(4,5)P <sub>2</sub>	Brain, skeletal muscle, testis, kidney
PLC- $\epsilon$	EGF, $\beta$ -adrenergic, prostaglandin E <sub>2</sub> , muscarinic M <sub>3</sub> -acetylcholine	Ras, G <sub>12</sub> (G <sub>13</sub> ), G $\beta\gamma$		Widely expressed, especially in heart in human and liver in rat

*Note.* The representative receptors and regulators involved in 11 isoforms of PLC and their distribution in tissues or cells in mammals are shown. GPCRs, G protein-coupled receptors; RTKs, receptor tyrosine kinases; NRTKs, nonreceptor tyrosine kinases; PDGF, platelet-derived growth factor; EGF, epidermal growth factor; FGF, fibroblast growth factor; NGF, nerve growth factor; VEGF, vascular endothelial growth factor; HGF, hepatocyte growth factor; TCR, T cell antigen receptor; BCR, B cell antigen receptor; Fc $\epsilon$ RI, high-affinity IgE receptor; Fc $\gamma$ Rs, IgG receptors; PAF, platelet-activating factor.



**Figure 2** Domain structure of four types of PLC and their regulatory molecules. Structural domains present within four types of PLC are shown. PH, EF-hand, catalytic (X and Y), and C2 domains are present in all types of PLCs. In addition, a regulatory COOH terminus is present in PLC- $\beta$ ; Src homology 2 (SH2) and Src homology 3 (SH3) domains are present within the second PH domain in PLC- $\gamma$ ; and the Ras guanine nucleotide exchange factor (RasGEF)-like and Ras-associating (RA) domains are present in PLC- $\epsilon$ . Major regulatory molecules interacting with these domains are also shown.  $G_{\beta\gamma}$ ,  $\beta\gamma$ -subunits of heterotrimeric G proteins;  $G_{\alpha}$ ,  $\alpha$ -subunits of heterotrimeric G proteins; PY, phosphotyrosine; and PS, phosphatidylserine.

domain appears to be present in all PLCs and an additional PH domain is split by two Src homology (SH) 2 (SH2) domains and one SH3 domain in PLC- $\gamma$ . To the  $NH_2$ -terminal PH domains, PI(3)P and  $G_{\beta\gamma}$  are bound in PLC- $\beta$ , PI(3,4,5)P<sub>3</sub> is bound in PLC- $\gamma$ , and PI(3,4)P<sub>2</sub> and Ins(1,4,5)P<sub>3</sub> are bound in PLC- $\delta$ . In the case of PLC- $\epsilon$ , the role of this domain has not been characterized. The roles of the EF-hand domain and C2 domain are unclear. These domains are known to bind  $Ca^{2+}$ . The C2 domain of PLC- $\delta$ 1 has been estimated to contain three  $Ca^{2+}$ -binding sites and their binding is thought to increase enzyme activity by forming a ternary complex with the enzyme and phosphatidylserine. In contrast, C2 domains of PLC- $\beta$  and PLC- $\gamma$  do not seem to bind  $Ca^{2+}$ , but this domain of PLC- $\beta$ 1 interacts with the  $\alpha$ -subunit of  $G_q$  protein ( $G_{\alpha q}$ ).

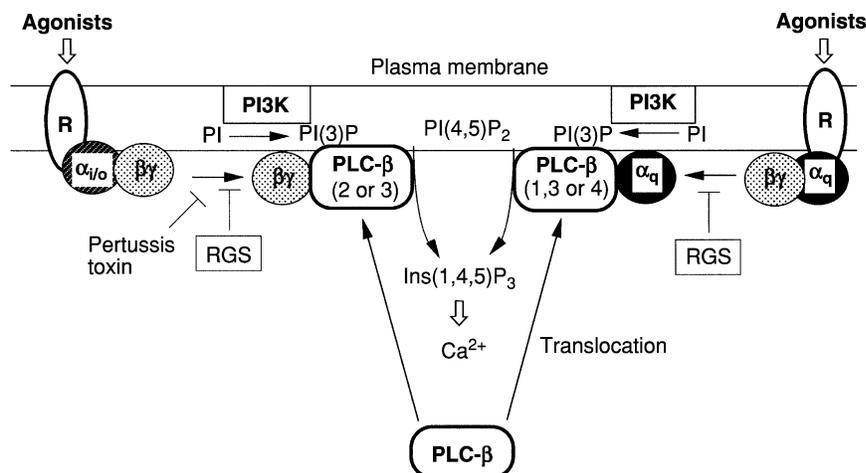
Additional regulatory domains or motifs are present in PLC- $\beta$ , PLC- $\gamma$ , and PLC- $\epsilon$ . PLC- $\beta$  contains a

long COOH-terminal region; PLC- $\gamma$  contains two SH2 domains and one SH3 domain, which bind phosphotyrosine-containing sequence and proline-rich sequences, respectively; and PLC- $\epsilon$  contains Ras-associating (RA) domains and a Ras guanine nucleotide exchange factor (RasGEF)-like domain. The different combinations of these regulatory domains in each PLC type determines their regulatory mechanism of enzyme activity.

## REGULATION OF PLC ACTIVITY

### PLC- $\beta$

PLC is usually present in cytosol. The PLC- $\beta$  type interacts with PI(3)P, a PI3K product, through its PH domain, and this interaction is responsible for the recruitment of the enzyme from the cytosol to the membrane. All PLC- $\beta$  isoforms are regulated by G



**Figure 3** Regulation of PLC- $\beta$ . PLC- $\beta$  types are regulated by both G<sub>q</sub> and G<sub>i</sub> subfamilies. On stimulation of G<sub>q</sub>-coupled receptors, G<sub>αq</sub> subunits are dissociated from the complex and activate PLC- $\beta$ 1, PLC- $\beta$ 3, and PLC- $\beta$ 4 by interacting with the C2 domain and regulatory COOH terminus. On the other hand, when G<sub>i</sub>- or G<sub>o</sub>-protein-coupled receptors are stimulated, G<sub>βγ</sub> subunits are dissociated and activate PLC- $\beta$ 2 and PLC- $\beta$ 3 by interacting with the PH domain. In both cases, PI(3)P produced by PI3K may play a role in the recruitment of the enzyme from cytosol to membrane. See text for further details.

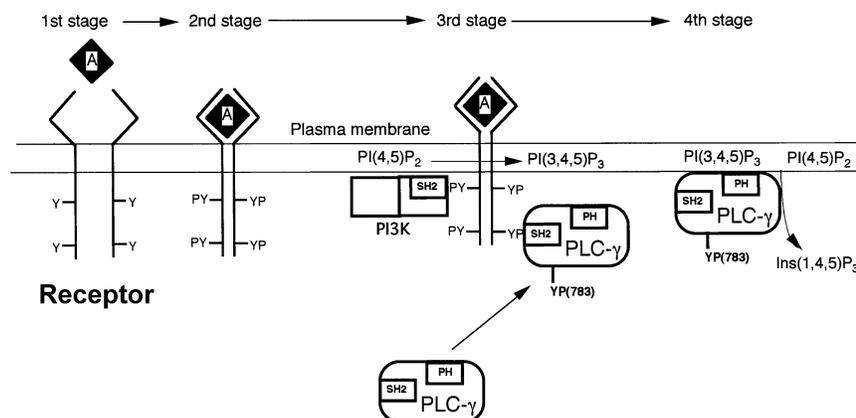
protein-coupled receptors (GPCRs), such as muscarinic acetylcholine receptors and adrenergic receptors, through heterotrimeric G proteins that are composed of  $\alpha$ -,  $\beta$ -, and  $\gamma$ -subunits (Table I and Fig. 3). The  $\beta$ - and  $\gamma$ -subunits are tightly associated under physiological conditions. The heterotrimeric G proteins are divided into four subfamilies based on their amino acid sequences and effector interactions of  $\alpha$ -subunits: G<sub>s</sub>, G<sub>i</sub>, G<sub>q</sub>, and G<sub>12</sub>. Receptor stimulation induces an exchange of GDP to GTP on the  $\alpha$ -subunits and thereby the dissociation of GTP-bound  $\alpha$ -subunits and  $\beta\gamma$ -subunits. All members ( $\alpha_q$ ,  $\alpha_{11}$ ,  $\alpha_{14}$ , and  $\alpha_{16}$ ) of the  $\alpha$ -subunits of the G<sub>q</sub> subfamily can activate PLC- $\beta$ 1, - $\beta$ 3, and - $\beta$ 4 through their C2 domains and the regulatory COOH terminus (~400 amino acid residues) sequences unique to PLC- $\beta$  (Fig. 2). This terminus is also thought to be involved in membrane binding and nuclear localization. On the other hand, G<sub>βγ</sub> activates  $\beta$ 2,  $\beta$ 3, and, to a lesser extent,  $\beta$ 1 through their PH domains, although the concentrations of G<sub>βγ</sub> required for maximal enzyme activation are much higher than those of G<sub>αq</sub>. Pertussis toxin-sensitive G proteins, G<sub>i</sub> or G<sub>o</sub> proteins, are usually expressed at a higher level in the cells than other heterotrimeric G proteins. Thus, G<sub>βγ</sub>, which activates PLC- $\beta$ , may be derived from the toxin-sensitive G proteins. The activation of PLC is turned off by the intrinsic GTPase of G<sub>αq</sub>. The hydrolysis of GTP to GDP on G<sub>αq</sub> leads to the reassociation of G<sub>αq</sub> with G<sub>βγ</sub>. The GTPase activity of G<sub>αq</sub> is stimulated by PLC itself and also by a group of regulatory

proteins called regulators of G protein signaling (RGS). Approximately 20 species of RGS have been identified and have been shown to stimulate the GTPase activity of  $\alpha$ -subunits of the G<sub>i</sub> and G<sub>q</sub> subfamilies.

### PLC- $\gamma$

Except for the differences in their distribution and hence the receptor species coupled, there seems to be no fundamental difference in the regulatory mechanisms between two isoforms of the enzyme (PLC- $\gamma$ 1 and - $\gamma$ 2). This type of PLC contains two important domains, the PH domain and the SH2 domain, which serve to recruit the enzymes to membranes and activate them (Fig. 2). Representative receptors and important regulators involved in the regulation of PLC- $\gamma$  are summarized in Table I. Tyrosine kinase, regardless of whether the receptors themselves possess intrinsic activity, is very important for the activation of PLC- $\gamma$ . Receptors coupled to this type of PLC can be divided into three groups: receptor tyrosine kinases, which contain intrinsic tyrosine kinase activity; cytokine and immunoglobulin receptors; and GPCRs. In the latter two cases, receptors have no intrinsic tyrosine kinase activity and must activate nonreceptor tyrosine kinases such as Src and Syk.

The regulatory mechanisms of PLC- $\gamma$  by receptor tyrosine kinases (RTKs) such as platelet-derived growth factor (PDGF) receptor and epidermal growth factor (EGF) receptor are illustrated in Fig. 4. These



**Figure 4** Regulation of PLC- $\gamma$ . The mechanisms of PLC- $\gamma$  activation by receptor tyrosine kinases (RTKs), such as platelet-derived growth factor (PDGF) receptor, are shown. See text for further details.

receptors generally dimerize after the binding of these growth factors (first stage to second stage) and thereby autophosphorylate tyrosine residues (second stage). These phosphotyrosines provide docking sites for signaling molecules including PLC- $\gamma$  and PI3K through their SH2 domains (third stage). The docking of PLC- $\gamma$  results in the phosphorylation of the tyrosine residue (Tyr-783) of the enzyme that is required for enzyme activation. In concert with these events, PI3K is also activated through the binding of the SH2 domain of PI3K to another phosphotyrosine site of receptors, which in turn supplies PI(3,4,5)P<sub>3</sub>. This phosphoinositide serves to anchor the enzyme through the PH domain of PLC- $\gamma$  and the enzyme is then activated to cleave PI(4,5)P<sub>2</sub> (fourth stage). Thus, tyrosine kinase activation and subsequent protein–protein and lipid–protein interactions play pivotal roles in the activation of PLC- $\gamma$ .

PLC- $\gamma$  activation by immunoglobulin receptors, which have no intrinsic tyrosine kinase activity, requires the involvement of nonreceptor tyrosine kinases (NRTKs) (Table I). In the case of T cell antigen receptor (TCR), ligation of the TCR triggers the activation of Src-related NRTKs Lck and Fyn, which in turn results in phosphorylations of tyrosine residues in one chain of the receptor complex and ZAP-70 (a member of the Syk family of NRTK). The activated ZAP-70 then phosphorylates the downstream substrates and finally leads to the activation of PLC- $\gamma$ . The activation of GPCRs sometimes induces tyrosine phosphorylation of RTKs such as PDGF receptor and EGF receptor. As a result, PLC- $\gamma$  is activated as shown in Fig. 4. The phosphorylation of RTKs appears to be mediated by Src. Both G $_{\beta\gamma}$  subunits and G $_{\alpha}$  subunits including G $_{\alpha i}$  and G $_{\alpha s}$  have been suggested to be involved in the activation of Src.

### PLC- $\delta$

Four isoforms of PLC- $\delta$  have been identified in mammals. This type of PLC exhibits much higher sensitivity to Ca<sup>2+</sup> than PLC- $\beta$  and PLC- $\gamma$ . This high Ca<sup>2+</sup> sensitivity is attributable to the C2 domain specific to PLC- $\delta$ . It is therefore suggested that the enzyme activity of PLC- $\delta$  may be regulated by an intracellular Ca<sup>2+</sup> increase that was attained by the activation of other PLC types. The PH domain of PLC- $\delta$  binds PI(4,5)P<sub>2</sub> and serves to anchor the enzyme on the membrane. The binding of PI(4,5)P<sub>2</sub> to the PH domain is promoted by Ca<sup>2+</sup> binding to the EF-hand of the enzyme. Thus, intracellular Ca<sup>2+</sup> is crucial for the regulation of the enzyme activity of PLC- $\delta$ . Although the physiological role is unclear, certain GPCRs are shown to couple to PLC- $\delta$ 1 through an atypical G protein, G<sub>h</sub> (molecular weight of 74–80 kDa) (Table I). G<sub>h</sub> also possesses transglutaminase activity. Unlike  $\alpha$ -subunits of heterotrimeric G proteins, the GDP form of G<sub>h</sub> tightly binds to PLC and GTP promotes their dissociation. Thus, G<sub>h</sub> might be a negative regulator of PLC- $\delta$ 1. It is not known whether PLC- $\delta$  isoforms other than  $\delta$ 1 are regulated by receptor stimulation.

### PLC- $\epsilon$

This type of PLC was discovered in 1998 and its regulatory mechanism has not been well characterized as yet. It is the largest (230–260 kDa) of the PLCs, and PLC- $\epsilon$  contains extra domains called RasGEF and RA domains specific to this type in addition to the catalytic (X and Y) domain and the regulatory C2 domain common to other types of PLC. The RasGEF domain is thought to function as a GDP–GTP exchange factor for Ras and Ras-related proteins. GTP-bound forms of

Ras and Rap bind to the RA domain and activate PLC- $\epsilon$ . EGF-induced Ras activation may be linked to the activation of PLC- $\epsilon$  by this mechanism. Cyclic AMP sometimes up-regulates PLC activity. This appears to be mediated by the nucleotide-dependent activation of Epac1 (GEF for Rap) and the subsequent activation of Rap2B, which may interact with the RA domain of the enzyme. This mechanism may partly explain cyclic AMP-dependent  $\beta$ -adrenergic and PGE1-receptor-mediated PLC activation. Thus, PLC- $\epsilon$  seems to function not only as a regulator of Ras and Ras-related proteins through the RasGEF domain but also as a regulator of their targets through the RA domain. In addition to the mechanisms related to Ras, PLC- $\epsilon$  is shown to be activated by  $G_{\alpha 12}$  and  $G_{\beta\gamma}$ .

## DISTRIBUTION AND PHYSIOLOGICAL ROLES OF PLC ISOFORMS

The major distribution sites of PLC isoforms are shown in Table I. PLC- $\beta$ 1 and PLC- $\beta$ 3 are widely expressed, especially in specific regions of the brain such as the cerebral cortex and hippocampus for PLC- $\beta$ 1 and in the brain and liver for PLC- $\beta$ 3, whereas PLC- $\beta$ 2 is expressed abundantly in hematopoietic cells such as leukocytes and PLC- $\beta$ 4 is expressed in the retina and cerebellum. As for PLC- $\gamma$  types, PLC- $\gamma$ 1 is widely expressed in several types of cells, whereas PLC- $\gamma$ 2 expression is high in hematopoietic cells. Among PLC- $\delta$  types,  $\delta$ 1 is most abundant and widely expressed. PLC- $\delta$ 2 is detected especially in the brain, whereas PLC- $\delta$ 3 and  $\delta$ 4 seem to be widely distributed. The expression of PLC- $\delta$ 4 in testis is noteworthy.

The pattern of the expression of PLC- $\gamma$  types is reflected in their isoform-specific functions, which have been confirmed from studies of knockout mice. Mice lacking PLC- $\beta$ 1 die suddenly due to epileptic-like seizures that are associated with a decrease in inhibitory interneuron firing mediated by muscarinic acetylcholine receptors. Thus, PLC- $\beta$ 1 is suggested to be involved in muscarinic acetylcholine receptor signal transduction in the cerebral cortex and hippocampus. On the other hand, mice lacking PLC- $\beta$ 4 show ataxia, which is associated with retardation of the development of the cerebellum. The lower response to the metabotropic glutamate receptor in the cerebellum is suggested to be involved in ataxia. PLC- $\beta$ 4 has also been shown to be involved in type-1 metabotropic glutamate receptor-mediated long-term

depression. In addition, PLC- $\beta$ 4 is abundantly expressed in the retina and is shown to play crucial roles in visual responses. In the case of PLC- $\beta$ 2 knockout mice, unexpectedly from their distribution, hematopoiesis was normal. Furthermore, even though PLC activation and superoxide production were inhibited, chemotaxis was barely affected or rather was enhanced when chemoattractant effects were examined in neutrophils from these animals. Thus, PLC- $\beta$ 2 is necessary for superoxide production but not for chemotaxis. A crucial role of PI3K in chemotaxis has been demonstrated in mice lacking PI3K $\gamma$ . Mice lacking PLC- $\beta$ 3 exhibited an approximately 10-fold decrease in the ED<sub>50</sub> value for morphine in producing antinociception. Thus, PLC- $\beta$ 3 appears to be involved in the negative modulation of  $\mu$ -opioid-induced antinociception. As expected from the finding that PLC- $\gamma$ 1 is coupled to various growth factor receptors, the homozygous disruption of PLC- $\gamma$ 1 resulted in embryonic lethality at embryonic day 9.0. Thus, this enzyme is essential for normal development and growth. PLC- $\gamma$ 2-deficient mice showed a decrease in mature B cells, a decrease in IgM and IgG levels, a loss of IgM receptor-induced Ca<sup>2+</sup> response, and defective Fc $\gamma$ R and Fc $\epsilon$ R functions. Thus, PLC- $\gamma$ 2 plays important roles in the signaling linked to immunoglobulin superfamily receptors. Among PLC- $\delta$  types, PLC- $\delta$ 4-deficient mice were studied and showed slightly smaller litters. The sperm function of the male animals was depressed; their sperm were unable to initiate the acrosome reaction, an exocytotic event required for fertilization. Thus, PLC- $\delta$ 4 is important for the acrosome reaction during fertilization.

## See Also the Following Articles

G Protein-Coupled Receptors • Receptor Tyrosine Kinase • Receptor-Regulated Phospholipases

## Further Reading

- Berridge, M. J. (1987). Inositol triphosphate and diacylglycerol: Two interacting second messengers. *Annu. Rev. Biochem.* **56**, 159–193.
- Hokin, L. E. (1985). Receptors and phosphoinositide-generated second messengers. *Annu. Rev. Biochem.* **54**, 205–235.
- Rebecchi, M. J., and Pentylala, S. N. (2000). Structure, function, and control of phosphoinositide-specific phospholipase C. *Physiol. Rev.* **80**, 1291–1335.
- Rhee, S. G. (2001). Regulation of phosphoinositide-specific phospholipase C. *Annu. Rev. Biochem.* **70**, 281–312.
- Williams, R. L. (1999). Mammalian phosphoinositide-specific phospholipase C. *Biochim. Biophys. Acta*, **1441**, 255–267.

fibers whose perikarya are located in the superior cervical ganglia. The preganglionic cell bodies in this chain of neurons are located in the intermediolateral cell column of the upper thoracic cord. These neurons receive synaptic input from cell bodies located in the higher brainstem, most prominently in the paraventricular nuclei (PVN) and suprachiasmatic nuclei (SCN) of the anterior hypothalamus. In turn, the SCN receive a robust neural input via the retino-hypothalamic tract from the ganglion cells of the retina. Thus, the lateral eyes have a multisynaptic connection to the pineal gland and, as noted above, when the neural pathway between the SCN and pineal gland is interrupted, the pineal becomes functionally inept.

Via the neural connections between the eyes and the pineal gland, the prevailing photoperiod exerts a major regulatory influence on the endocrine function of the gland. In general, light perceived by the lateral eyes inhibits the pineal gland, whereas during darkness the gland is endocrinologically active. The perception of light at the retinal level for regulation of the pineal gland relies on different cells and photopigments than does the visual system; that is, the pineal is not regulated by the three cone photopic ocular cells. The predominant wavelengths of light detected by the retinas, which suppress nighttime melatonin production in humans, are in the range of 446–477 nm; the peak absorbance of the photopigment appears different from those in the classic rod and cone cells of the retinas.

## BLOOD SUPPLY

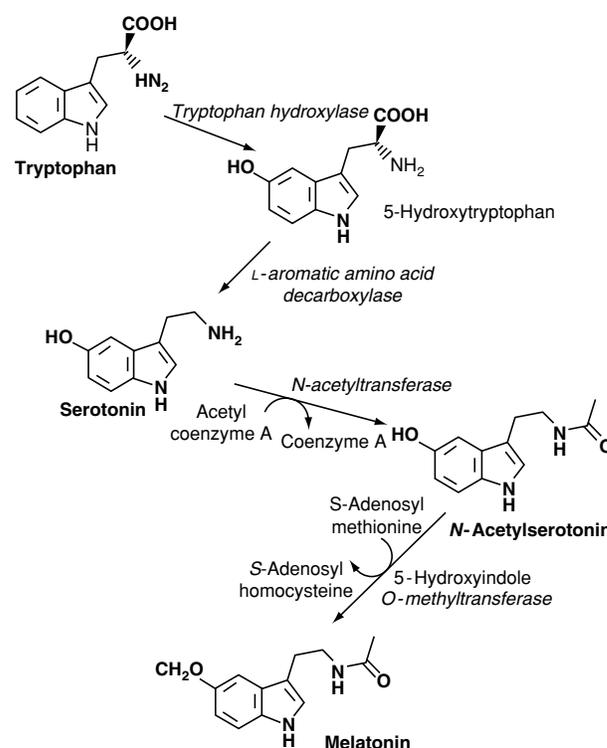
The pineal gland, like other endocrine organs, has a perfuse blood supply. This is derived from the posterior choroidal arteries, which are tributaries of the terminal branches of the basilar artery, the posterior cerebrals. Within the gland, the posterior choroid arteries break up into a rich capillary network. Venous blood drains into the large cerebral sinuses that surround the pineal gland.

## MELATONIN SYNTHESIS

Melatonin, i.e., *N*-acetyl-5-methoxytryptamine, is the chief secretory product of the pineal gland. The details of the synthesis of melatonin within the gland have been for the most part identified. Melatonin is a metabolite of the amino acid tryptophan, which is taken into the pinealocytes from the blood after its consumption in the diet. Tryptophan is quickly converted

to 5-hydroxytryptamine (serotonin) in pinealocytes; the concentration of this monoamine in the pineal exceeds that of any other organ or group of neurons in the body. The conversion of serotonin to melatonin is a two-step process; both steps have been studied in detail. The synthetic pathway for melatonin is summarized in Fig. 1.

During the night, norepinephrine is released from the postganglionic fibers that end in the vicinity of the pinealocytes. The neurotransmitter then interacts with primarily  $\beta$ -adrenergic receptors, and to a lesser degree with  $\alpha$ -receptors, in the pinealocyte membrane to stimulate a cascade of events that promotes the conversion of serotonin to *N*-acetylserotonin (NAS), the immediate precursor of melatonin. This metabolic cascade includes the stimulation of the intracellular second-messenger cyclic AMP (cAMP) and protein synthesis, which are required for the activation of the enzyme serotonin *N*-acetyltransferase. The acetylation of serotonin to NAS is the rate-limiting step in melatonin formation (Fig. 1). NAS is then *O*-methylated to form melatonin. Unlike other endocrine organs that typically store large amounts of the hormone(s)



**Figure 1** Synthesis of melatonin (*N*-acetyl-5-methoxytryptamine) from the essential amino acid tryptophan. The activity of the enzyme *N*-acetyltransferase is generally considered rate-limiting in melatonin production.

that they produce, the pineal gland lacks a storage mechanism for melatonin. Rather, once generated, melatonin is quickly released from the pinealocyte.

The route of release of melatonin has been traditionally thought to be directly into the rich capillary plexus that pervades the pineal gland. Studies in sheep, the pineal gland of which is attached to the posterodorsal aspect of the diencephalon like that in human, have shown that melatonin is at least in part discharged from the pineal directly into the ventricular fluid of the third ventricle. This has implications for melatonin's functions in the central nervous system. Simultaneously, melatonin may also be discharged into the blood.

### Melatonin Concentrations in Body Fluids

Within the blood and cerebrospinal fluid (CSF), melatonin levels closely track its production in the pineal gland such that at night, when melatonin synthesis within the gland is elevated, concentrations of melatonin in the serum and CSF are likewise high. Typical nighttime serum melatonin concentrations vary from 50 to 150 pg/ml, whereas during the day, serum levels are usually less than 10 pg/ml. In the third ventricular CSF, the melatonin concentrations may be several orders of magnitude, up to 90 nM, higher than in the blood, indicating a direct release of the indole from the pinealocytes into the CSF. Once in bodily fluids, melatonin has access to every cell in the body and is probably taken up by many cells.

### Photoperiod and Melatonin Synthesis

The mammalian pineal gland is unique in that its biosynthetic activity is controlled by an environmental variable, namely, the prevailing photoperiod. Pineal melatonin production occurs exclusively during the dark phase of the daily photoperiodic cycle. In modern societies, daily periods of environmental light and darkness have been markedly subverted by the introduction of artificial light sources. If of sufficient intensity and proper wavelength, high nighttime melatonin levels are quickly suppressed by perception of artificial light by the eyes at night. Likewise, during transmeridian travel, the normal 24 h circadian melatonin rhythm is disturbed due to changes in the light/dark cycle.

The SCN are a critical relay in the neural connections between the eyes and the pineal gland. A variety of clock genes in the SCN ensure that these neurons are inherently active during the dark portion of every

24 h period. The active/inactive period of the SCN neurons has a duration of approximately 25 h; thus, it is not normally in synchrony with the prevailing 24 h light/dark cycle. The natural light/dark cycle acting via the retina entrains the circadian SCN rhythm to precisely 24 h.

The ability of light to regulate melatonin synthesis is readily apparent when individuals are exposed to artificial light at night; this exposure is followed by a rapid drop in circulating melatonin levels. The influence of light on the pineal gland is also seen in individuals living at the extremes of latitude. During the winter months, the daily period of light can be very short, in which case the pineal produces more melatonin for a longer duration than during the summer months when day lengths are exceptionally long. The duration of elevated melatonin production is typically proportional to the duration of dark exposure. The circadian melatonin cycle with high levels at night and low values during the day may be disrupted in some individuals, e.g., in blind individuals incapable of light perception and in Smith-Magenis syndrome, where the melatonin cycle free-runs.

### Effects of Age on Melatonin Production

Increased age is another factor that significantly changes the amount of melatonin produced in the pineal gland. Pineal melatonin production gradually wanes during aging such that elderly individuals may have only a fraction of the circulating melatonin levels that they had when they were younger. There may be a correlation between health status and melatonin levels during aging, with the healthy elderly seemingly having a better preserved melatonin synthesis cycle.

## FUNCTIONS OF MELATONIN

The effects of melatonin in the organism are mediated via direct actions, e.g., free radical scavenging, and indirectly through receptors. Several membrane melatonin receptors have been identified and cloned and they are denoted MT<sub>1</sub>, MT<sub>2</sub>, and Mel<sub>1c</sub>, although there are subtypes of each. These receptors (or binding sites) have seven transmembrane-spanning domains, are G protein-coupled, have a nanomolar affinity for melatonin, and have a wide distribution in organisms, with significant between-species variations. Two major locations of membrane melatonin receptors are the SCN, the so-called biological clock, and the pars tuberalis (PT) of the anterior pituitary gland. The signal transduction pathways linking

melatonin receptors to cellular physiology include a pertussis toxin (PTX)-sensitive G protein, which decreases intracellular cAMP. There are available a number of pharmacological agonists and antagonists of the membrane receptors that may prove useful in identifying the functions of the subtypes of melatonin-binding sites. Intracellular binding sites, e.g., to cytoplasmic calmodulin and also molecules in the nucleus, have been uncovered but their role in cellular function remains to be clarified.

### Melatonin and Seasonal Reproduction

It is via the daily and seasonal changes in the duration of elevated nocturnal melatonin levels that the brain of vertebrates integrates photoperiodic information into the functioning of the organism. In photosensitive mammals, perhaps the most clearly defined function known to be controlled by the integrated action of the prevailing photoperiod and the changing melatonin rhythm is seasonal changes in reproductive capability. Although this relationship has been unequivocally established, the site at which melatonin intervenes to regulate the neuroendocrine-reproductive axis, e.g., at the SCN, the PT, other hypothalamic nuclei, and the gonadotropes of the anterior pituitary, remains unidentified as do the specific receptors involved. These seasonal reproductive cycles may have a corollary in humans, in which the changing melatonin cycle may drive the annual cycle in seasonal affective disorders. A common misconception is that in terms of seasonal fluctuations in reproduction, melatonin is inhibitory to sexual physiology. In reality, the changing melatonin cycle is merely a signal to identify the season of the year with the status of the reproductive organs, i.e., functional or atrophic, responding according to the physiological needs of the specific organism. Thus, some animals are sexually competent during long days and others are reproductively active during the short days of the year. In both cases, the time of delivery of the young is usually closely correlated with the season that ensures maximal survival of the offspring, i.e., spring and early summer.

### Melatonin and Circadian Functions

The daily melatonin rhythm is one of the efferent signals of the biological clock. Whereas the loss of the rhythmic melatonin cycle does not significantly disrupt circadian organization, subtle desynchrony of several physiological functions as well as reentrainment of activity rhythms may occur after surgical removal of the pineal gland.

Melatonin, when administered exogenously, can influence either directly or indirectly the phase and/or period of the circadian clock. This chronobiotic property of melatonin has been used to manipulate the sleep-wake cycle, to synchronize the activity pattern of free-running profoundly blind humans, and to reduce the severity of jet lag. In these situations, the circadian time at which melatonin is administered is critical to the action of the indole. This presumably relates to a changing sensitivity of the melatonin receptor to its ligand.

### Anticancer Effects of Melatonin

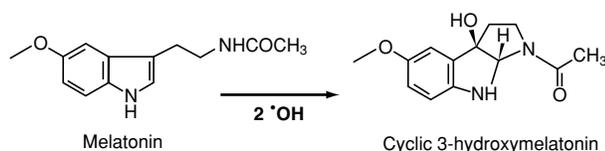
The anticancer effects of melatonin, which have been well documented at several laboratories, are at least in part mediated via membrane receptors. Thus, the inhibition of growth and proliferation of cancer cells includes changes in intracellular signaling pathways that involve a drop in cAMP, a reduction in the uptake of a key growth factor for cancer cells (linoleic acid) followed by a reduction in the generation of 13-hydroxyoctadecadienoic acid, and a reduction in mitogen-activated protein kinase. This series of events has been proposed to reduce the proliferation of cancer cells, thereby reducing tumor growth. Additionally, via its free radical scavenging actions, melatonin limits DNA damage and, as a consequence, the likelihood of tumor initiation.

### Melatonin and Immune Function

Melatonin also has actions on the immune system that seem to be mediated by both membrane melatonin receptors and nuclear binding sites. Melatonin has direct effects on T and B lymphocytes and on monocytes, including the regulation of a number of cytokines. Melatonin influences a variety of immune parameters including both specific and nonspecific immunity. In general, the actions of melatonin at the level of the immune system are immunoenhancing; as a result, exogenously administered melatonin may be contraindicated in conditions where the immune system is already hyperfunctional, e.g., Crohn's disease.

### Melatonin as an Antioxidant

Melatonin's ability to directly scavenge a variety of free radicals as well as to promote the activities of several antioxidative enzymes has been well documented (Fig. 2). Although these actions have been



**Figure 2** An example of melatonin's free-radical scavenging activity. In this reaction, melatonin scavenges two highly toxic hydroxyl radicals ( $\bullet\text{OH}$ ), with the resulting product being cyclic 3-hydroxymelatonin, which is excreted in the urine.

most eloquently verified using pharmacological concentrations of melatonin, loss of the endogenous nocturnal rise in melatonin due to pinealectomy also exaggerates oxidative damage in many organs, suggesting a role for physiological concentrations of melatonin in restraining free radical damage to essential molecules and subcellular organelles. The antioxidative actions of melatonin may also relate to its ability to reduce electron leakage from the mitochondrial respiratory pathway.

## CONCLUDING REMARKS

Via its chief secretory product, melatonin, the physiological relevance of the pineal gland is much more widespread than initially envisaged. Indeed, there may not be an organ in the body or a subcellular organelle that escapes melatonin's direct or indirect influence. Melatonin's ubiquitous actions relate to the several means by which it can exert its effects. Thus, melatonin influences cellular processes through a variety of membrane receptors, possibly through cytosolic and nuclear binding sites and due to its ability to directly neutralize highly reactive free radicals and associated oxygen- and nitrogen-based reactants. More

extensive reviews should be consulted for detailed information on these diverse actions of melatonin.

## See Also the Following Articles

Circadian Rhythms: Hormonal Facets • Melatonin • Oxidative Stress and Aging • Pineal Gland, Evolution of • Pineal Tumors

## Further Reading

- Acuña-Castroviejo, D., Escames, G., Carazo, A., Leon, J., Khaldy, H., and Reiter, R. J. (2002). Melatonin, mitochondrial homeostasis and mitochondrial-related diseases. *Curr. Top. Med. Chem.* **2**, 133–152.
- Allegra, M., Reiter, R. J., Tan, D. X., Gentile, C., Tesoriere, L., and Livrea, M. A. (2003). The chemistry of melatonin's interaction with reactive species. *J. Pineal Res.* **34**, 1–10.
- Blask, D. E., Sauer, L. A., and Dauchy, R. T. (2002). Melatonin as a chronobiotic/anticancer agent: Cellular, biochemical, and molecular mechanisms of action and their implications for circadian-based cancer therapy. *Curr. Top. Med. Chem.* **2**, 113–132.
- Guerrero, J. M., and Reiter, R. J. (2002). Melatonin-immune system relationships. *Curr. Top. Med. Chem.* **2**, 167–180.
- Pevet, P. (2002). Melatonin: Pharmacological aspects. *Dialog Clin. Neurosci.* **4**, 57–72.
- Reiter, R. J. (1991). Pineal melatonin: Cell biology of its synthesis and of its physiological interactions. *Endocr. Rev.* **12**, 151–180.
- Reiter, R. J. (1993). The melatonin rhythm: Both a clock and a calendar. *Experientia* **49**, 654–664.
- Tan, D. X., Reiter, R. J., Manchester, L. C., Yan, M. T., El-Sawi, M., Sainz, R. M., Mayo, J. C., Kohen, R., Allegra, M., and Hardeland, R. (2002). Chemical and physical properties and potential mechanisms: Melatonin as a broad spectrum and free radical scavenger. *Curr. Top. Med. Chem.* **2**, 181–198.
- von Gall, C., Stehle, J. H., and Weaver, D. R. (2002). Mammalian melatonin receptors: Molecular biology and signal transduction. *Cell Tissue Res.* **309**, 151–162.
- Wehr, T. A. (2001). Photoperiodism in humans and other primates: Evidence and implications. *J. Biol. Rhythms* **16**, 348–364.

that do not possess or express reduced pineal tissue, such as some adult salamanders (class Amphibia; order Caudata), geckos (class Reptilia; order Squamata; family Gekkonidae), crocodilians (class Reptilia; order Crocodylia), and some species of owls (class Aves; order Strigiformes), its absence is a derived characteristic (i.e., they have lost their pineal glands in the course of their evolution). It is interesting to note that these groups contain primarily nocturnal species.

### Cellular Anatomy

The cellular anatomy of pineal organs among vertebrate groups shares one common feature: the major cell type, the pinealocyte, derives from photoreceptive neurons. Other cell types within pineal organs of diverse groups include neuroglia, endothelial cells, and, in some species, B lymphocytes. In nonmammalian vertebrates, the pinealocyte is characterized by ciliary outer segments with photoreceptor-like conformations; that is, there are nine pairs of microtubules radially arranged at the base of the cilium with no central pair of microtubules, forming the base of the pineal outer segment. The pineal outer segments of nonmammalian vertebrates are diverse in their cellular anatomy, ranging from cone-like photoreceptors in most fish to membranous whorls in birds and some reptiles. In those species in which it has been studied, these outer segments contain high concentrations of classical (e.g., cone opsins) and/or pineal-specific (e.g., pinopsin) opsin-based photopigment molecules. In addition, these cells contain high concentrations of flavin-based cryptochromes, which have been implicated as photoreceptive molecules in plants and insects. Interestingly, even though there is no evidence for photoreception in adult mammalian pineal glands, mammalian pinealocytes retain several of these photopigments. The function of these molecules is not known.

In most amniote vertebrates, pinealocytes project axons to pineal ganglion cells, which in turn project down the pineal stalk to diencephalic and mesencephalic sites within the brain. In addition, in some species, pinealocytes may directly innervate the brain, especially within epithalamic structures such as the habenula, and synapse on blood vessels residing within the pineal parenchyma. This latter motif is retained within several amniote reptile groups, especially lizards, but is lost among the turtles, birds, and mammals. In those groups, the pinealocytes appear to express an endocrine morphology, directly contacting blood vessels within the pineal organ, thereby achieving the designation “pineal gland” in birds and mammals.

Conversely, the pineal organ of amniote vertebrates receives little if any autonomic innervation, whereas the pineal glands of both birds and mammals are heavily innervated by sympathetic fibers arising from the superior cervical ganglion. The picture that emerges is that, in amniotes, the pineal organ is a sensory structure that conveys photic information to diencephalic and mesencephalic structures in the brain. This motif is converted to a neuroendocrine structure in amniotes over the course of evolution, under direct autonomic regulation in reptiles, birds, and mammals. Interestingly, many transitional patterns can be observed among the reptiles, suggesting that comparative analysis of reptilian pineal anatomy, physiology and molecular biology may reveal important clues to the evolution of the epiphysis cerebri.

### MELATONIN BIOSYNTHESIS

#### Melatonin Is a Phylogenetically Ancient Molecule

The indoleamine hormone melatonin is present in pineal and retinal photoreceptor cells in all vertebrate groups studied. It is therefore safe to conclude that melatonin itself is an ancient feature of vertebrate photoreceptors, including pinealocytes. In all vertebrate species studied, melatonin is produced at higher levels during the night than during the day. Furthermore, in most species, this rhythm continues in constant darkness (DD), such that melatonin levels are high during the animal’s “subjective night.” Thus, melatonin rhythmicity is a circadian rhythm in most vertebrates, the result of an endogenous circadian oscillator.

In fact, the hormone is also produced by several invertebrate species, higher plants and microbes, although the biosynthetic pathways for melatonin production in these groups appear to be different from those that produce melatonin in vertebrate photoreceptors. Indeed, to punctuate this distinction, melatonin is produced primarily during the day in the fruit fly *Drosophila melanogaster*. It is therefore not clear whether melatonin is an ancient feature of all photoreceptive tissues or rather arises as a derived feature in each of these groups independently. More comparative research at the molecular level will be required to determine these relationships.

#### Vertebrate Melatonin Biosynthesis

In all vertebrate pineal organs, tryptophan is taken up by pinealocytes and converted to 5-hydroxytryptophan

(5-HTP) by tryptophan hydroxylase (EC 1.14.16.4). Then, aromatic amino acid decarboxylase (EC 4.1.1.28) converts 5-HTP to serotonin (5-HT). During the night, 5-HT is converted to *N*-acetylserotonin (NAS) by arylalkylamine-*N*-acetyltransferase (AANAT; EC 2.3.1.87). NAS is the primary substrate for hydroxyindole-*O*-methyltransferase (EC 2.1.1.4), which converts NAS to melatonin.

The major biosynthetic enzyme responsible for melatonin biosynthesis is probably AANAT in that it represents the rate-limiting step in melatonin biosynthesis in many (but not all) species and in that this enzyme is typically regulated at the transcriptional, translational, and posttranslational levels in many (but not all) species. The deduced amino acid sequences of AANAT in mammals, birds, and teleost fish reveal a highly conserved enzyme that bears very little relationship to any other protein in sequence databases. The closest neighbor is the arylamine-*N*-acetyltransferases (aNAT), which are primarily responsible for detoxification of cyclic hydrocarbons and are, along with AANAT, members of the A/B acetyltransferase superfamily. Both aNAT and AANAT are found in all vertebrate taxa, suggesting that these two enzymes diverged before different vertebrate classes diverged from their ancestral stock. Interestingly, an AANAT has been cloned in *Drosophila melanogaster*, which shares characteristics with both aNAT and AANAT, corroborating the view that aNAT and AANAT are homologous enzymes that share a common evolutionary ancestor that predates the emergence of vertebrates. In several species of teleost fish, there are two genes that encode AANAT. In pike and zebrafish, AANAT1 is expressed primarily within the ocular retina, whereas AANAT2 is expressed primarily within the pineal organ. These enzymes are likely the result of gene duplication and divergence, a common phenomenon among teleosts. They are, furthermore, differentially regulated in these two tissues.

## PHOTORECEPTION

### Visually Evoked Responses

As stated above, the pinealocytes in all vertebrate groups contain opsin-based and possibly flavin-based cryptochrome photopigments. The presence of photopigments raises the possibility that these cells are photoreceptive and indeed they are in many species. Perhaps, the most extensively studied group among the vertebrates in terms of visually evoked electrophysiological responses is the squamate reptiles. In

these animals, pinealocytes and photoreceptors in the associated parietal eye respond by depolarization to photic stimulation. These responses are characterized by chromatic and achromatic responses. This differential response to different wavelengths of light (chromatic response) has been suggested as a mechanism to differentiate dawn from midday from dusk, each of which has different light qualities. Similar responses have been observed in several teleosts and amphibians.

Attempts to demonstrate visually evoked responses in birds and mammals have been difficult. Patch clamping of individual avian pinealocytes has revealed graded photic responses that are dependent on calcium influx.

### Inhibition of Melatonin Biosynthesis

Certainly, the best known photic response within the pineal organ is the inhibitory effect of light on melatonin biosynthesis. In all species studied, the biosynthesis of melatonin is restricted to the night and, if the pineal organ is illuminated during the night, melatonin biosynthesis is halted, and melatonin levels decline rapidly.

In nonmammalian vertebrates, including teleosts, reptiles, and birds, this response occurs both *in vivo* and *in vitro*, indicating that photopigments within the pinealocytes themselves mediate this process. The best studied of these is the chick pineal gland, in which light has essentially three effects on melatonin rhythms: (1) light acutely suppresses melatonin biosynthesis, (2) light phase-shifts the circadian rhythm of melatonin release, and (3) light increases the amplitude of the melatonin rhythm. The acute suppression of melatonin by light is believed to result from the activation of opsin-like photopigments, since vitamin A depletion abolishes this effect. Action spectral analysis of chick pinealocytes suggests an opsin-like photopigment with peak absorption of approximately 500 nm. Illumination ultimately down-regulates cyclic AMP (cAMP) and its downstream signal cascades. When cAMP declines, AANAT protein is ubiquitinated and degraded rapidly by proteosomal proteolysis. The mechanisms for the phase-shifting and amplitude effects of light are not known, although it is clear that neither vitamin A-based photopigments nor cAMP is necessary for the effect. It is likely that these effects involve the regulation of the circadian clock itself.

In mammals, the adult pineal gland is not directly photoreceptive. Photoreceptors are restricted to the ocular retinae in mammals. However, remarkably, the traditional rods and cones are only part, perhaps

a minor part, of the light response. Instead, photoreceptive retinal ganglion cells project directly to the hypothalamic suprachiasmatic nucleus (SCN), the central circadian clock of mammals. The SCN in turn projects to the pineal gland via a multisynaptic pathway that includes the sub-paraventricular zone of the hypothalamus, the intermediolateral cell column of the thoracic spinal cord, and pre- and postganglionic fibers of the sympathetic nervous system. This pathway is homologous to a similar innervation of the pineal gland in birds. In nocturnal rodents, sympathetic norepinephrine is released during the night under SCN control, which stimulates  $\beta$ -adrenergic receptors, which in turn stimulates cAMP production. Then, AANAT is stimulated at the transcriptional, translational, and posttranslational levels to increase melatonin biosynthesis. It is not clear whether this exact mechanism also underlies melatonin biosynthesis in other mammalian groups. In sheep, for example, significant  $\alpha_1$ -adrenergic activation is also required. When mammals are illuminated in the middle of the night, sympathetic stimulation is interrupted, and similar to the situation in the photoreceptive chick pineal gland, the decline in cAMP precipitates a proteosomal proteolytic degradation of AANAT. It is believed, but it has not been determined, that the phase-shifting effects of light on melatonin rhythms are also mediated via regulation of sympathetic activity by the SCN. It is interesting to note that pineal glands from neonatal rodents retain photoreceptive properties until they are innervated by sympathetic fibers, suggesting that primitive photoreceptive properties of mammalian pinealocytes are lost during development.

## CIRCADIAN OSCILLATION

### Nonmammalian Vertebrates

Every class of nonmammalian vertebrate studied contains species that have pineal glands that are capable of both photoreception and generation of circadian rhythms of melatonin both *in vivo* and *in vitro*. These include the agnathan lampreys, teleost fish, squamate reptiles, and birds. Therefore, circadian oscillation within pinealocytes is almost certainly a primitive characteristic. The several exceptions to this rule are very likely derived characteristics, in which the pineal gland has secondarily lost its oscillatory capability. These include the salmonid teleosts, such as salmon and trout, whose pineal organs produce melatonin rhythmically *in vivo*, but when their pineal organ is placed *in vitro*, they express constitutively high levels of the hormone in darkness. However, as stated

above, these pineal organs retain their photoreceptive capability, since illumination suppresses melatonin biosynthesis.

In agnathan lampreys, other species of teleost fish, reptiles, and birds, pineal glands placed *in vitro* express a circadian pattern of melatonin biosynthesis such that melatonin is released during the night if the gland is placed in a light:dark cycle and persists for at least four circadian cycles in DD. In most cases, this rhythm begins to damp after extended *in vitro* analysis. This damping is the result of both individual pinealocytes drifting out of phase with one another and the oscillators within each cell experiencing a reduction in amplitude.

Understanding of the mechanisms of circadian oscillation has undergone a revolution in the past 10 years with the discovery of mammalian "clock genes" and the subsequent cloning and isolation of orthologues of these genes among nonmammalian vertebrates. These genes include the so-called "positive elements" *BMAL1* and *Clock* and "negative elements" *Period 1*, *2*, and *3* and the cryptochromes. These genes are expressed at very high concentrations in the pineal organs of zebrafish and chicks.

The evolutionary relationships among these clock genes and their relatives in other taxa are just beginning to be studied. It is likely that a complete phylogeny of clock genes will soon be known. Interestingly, these clock genes are also expressed within the mammalian pineal gland, even though the gland does not appear to express endogenous circadian rhythmicity. It is likely that the roles of these clock genes have only just begun to be understood.

## PINEAL'S ROLE IN VERTEBRATE FUNCTION

### "Nonspecific" Effects

A growing body of evidence suggests that melatonin may affect tissues very broadly in a "nonspecific" fashion in that the molecule possesses potent antioxidant activity. Melatonin scavenges hydroxyl free radicals more efficaciously *in vitro* than do many "traditional" antioxidants, such as glutathione. In addition, administration of very high concentrations of the hormone to animals ameliorates many cellular disorders associated with free radical activity. As such, the hormone has been touted as an "anti-aging" therapeutic agent. If this activity holds true in future research, it is likely to be a common mechanism among all animals (and perhaps plants and algae as well, where no melatonin receptor has been identified). However, the evidence

is not strong enough to warrant such a statement at this stage.

### “Specific” Effects—Melatonin Receptors

The discovery and subsequent use of the melatonin agonist 2-[<sup>125</sup>I]iodomelatonin as a ligand for the study of high-affinity melatonin binding strongly pointed to specific, high-affinity melatonin receptors. Pharmacological analysis of this binding strongly indicated that melatonin receptors were G<sub>i</sub>-GTP-binding protein-associated receptors that were expressed broadly in the brains and bodies of many vertebrate species. The subsequent cloning of a melatonin receptor in the African clawed frog *Xenopus laevis* confirmed this view. Sequence analysis of this melatonin receptor indicated that it was a member of the seven-transmembrane domain G protein receptors and expression of the protein in eukaryotic expression vectors further confirmed this view. Melatonin inhibited the accumulation of cAMP in these cells.

Since the discovery of this first melatonin receptor, at least three related high-affinity melatonin receptors have been isolated, cloned, and characterized in a wide variety of species, ranging from zebrafish to humans. Three very closely related genes encode the Mel<sub>1A</sub>, Mel<sub>1B</sub>, and Mel<sub>1C</sub> receptor proteins in all nonmammalian vertebrates. Interestingly, no Mel<sub>1C</sub> receptor has been isolated from mammals. Although these receptor proteins appear to be nearly identical in their affinities and specificities, they are differentially expressed. The Mel<sub>1A</sub> receptor is expressed widely in endothelial and neuronal tissues in all species studied, whereas the Mel<sub>1B</sub> receptor is predominantly expressed by the retinae. The Mel<sub>1C</sub> receptor is typically expressed in the nervous system and in cells derived from the neural crest, including melanocytes of the African clawed frog, in which the receptor was first isolated. Each of these receptor subtypes, with the exception of the mammals, is expressed in each class studied. Therefore, the three receptor subtypes are phylogenetically older than the divergence of these vertebrate groups.

### Physiological and Behavioral Effects

As stated above, though the affinity and specificity of the known receptor subtypes are similar, the receptors are differentially expressed, at least in the brain. First, in most vertebrates studied, melatonin affects circadian behavior. In many species of fish, reptiles, and birds, rhythmic melatonin is required for circadian behavior. In mammals and many species of birds,

removal of endogenous melatonin by pinealectomy or pharmacological means does not abolish rhythms, although administration of melatonin entrains circadian patterns of activity. The sites of action for this effect are likely to be the hypothalamic SCN or homologous sites within the hypothalamus. However, this aspect of circadian physiology has not been studied in detail.

In nonmammalian vertebrates, melatonin receptor binding predominates in structures involved in vision and visual integration. Concomitantly, several studies have shown that a circadian clock regulates visual system function and that this circadian variation in visual function is at least in part regulated by melatonin. Typically, in diurnal reptiles and birds, visual sensitivity is higher during the night than during the day and visual response (amplitude) is higher during the day than during the night. These changes are similar to what one might expect in dark and light adaptation, respectively. However, these changes occur even if the animals are maintained in constant darkness, showing that a circadian clock regulates the visual system.

Whereas visual function in mammals is also regulated by melatonin, this effect is not as dramatic as it is in reptiles and birds. Instead, melatonin receptors predominate in the SCN and tuberal hypothalamus. Melatonin affects circadian behavior through its effects on the SCN, but in seasonally reproducing mammals, the duration of the melatonin cycle also synchronizes reproductive, metabolic, and pelage functions to the prevailing photoperiod. It is interesting to note that although birds depend on melatonin for circadian activity, the hormone has little effect on seasonal functions in birds. Conversely, whereas the circadian effect of melatonin is limited in mammals, melatonin plays a major role in photoperiodic time measurement. Thus, the photoperiodic role for melatonin is a derived characteristic in this group.

### SYNOPSIS

In summary, the epiphysis cerebri, pineal organ, or pineal gland, as it is variously called, is a phylogenetically ancient photoreceptive structure that has diverged dramatically over the evolutionary history of vertebrates. It is a primitive characteristic of all vertebrate classes that has been periodically lost or reduced in function in distinct groups, most of which are nocturnal. Its actions are predominantly through the release of the indoleamine melatonin, either as a neurotransmitter or neuromodulator or as a hormone released directly into the bloodstream. Melatonin is believed to

bear antioxidant characteristics and this function may be its most primitive, since this function is shared with invertebrate and plant species. Its specific actions vary among vertebrate groups, but regulation of circadian and visual function appears to be the most primitive, because it is shared by all groups. In contrast, melatonin's role as a regulator of seasonal/photoperiodic function is likely derived, since it predominates in mammals alone.

### See Also the Following Articles

Circadian Rhythms: Hormonal Facets • Melatonin • Pineal Gland • Pituitary Gland, Evolution of

### Further Reading

- Bailey, M. J., Chong, N., Xiong, J., and Cassone, V. M. (2002). Chickens' *Cry2*: Molecular analysis of an avian cryptochrome in retinal and pineal photoreceptors. *FEBS Lett.* **513**, 169–174.
- Begay, V., Falcon, J., Cahill, G. M., Klein, D. C., and Coon, S. L. (1998). Transcripts encoding two melatonin synthesis enzymes in the teleost pineal organ: Circadian regulation in pike and zebrafish, but not trout. *Endocrinology* **139**, 905–912.
- Butler, A. B., and Hodson, W. (1996). "Comparative Vertebrate Neuroanatomy: Evolution and Adaptation." Wiley-Liss, New York.
- Cahill, G. M. (2002). Clock mechanisms in zebrafish. *Cell Tissue Res.* **309**, 27–33.
- Cassone, V. M., and Natesan, A. K. (1997). Time and time again: The phylogeny of melatonin as a transducer of biological time. *J. Biol. Rhythms* **12**, 489–497.
- Cassone, V. M., Brooks, D. S., and Kelm, T. A. (1995). Comparative distribution of  $2[^{125}I]$ iodomelatonin binding in the brains of diurnal birds: Out-group analysis with turtles. *Brain Behav. Evol.* **45**, 241–256.
- Collin, J. P., Voisin, P., Falcon, J., Faure, J. P., Brisson, P., and Defayre, J. R. (1989). Pineal transducers in the course of evolution: Molecular organization, rhythmic metabolic activity and role. *Arch. Histol. Cytol.* **52**, 441–449.
- Gwinner, E., and Brandstätter, R. (2001). Complex bird clocks. *Proc. Trans. R. Soc. London* **356**, 1801–1810.
- Hastings, M. H., Vance, G., and Maywood, E. (1989). Some reflections on the phylogeny and function of the pineal. *Experientia* **45**, 903–909.
- Iigo, M., Hara, M., Ohtano-Kaneko, R., Hirata, K., Tabata, M., and Aida, K. (1997). Photic and circadian regulations of melatonin rhythms in fishes. *Biol. Sign.* **6**, 225–232.
- Jackman, W. R., Langeland, J. A., and Kimmel, C. B. (2000). *islet* reveals segmentation in the amphioxus hindbrain homolog. *Dev. Biol.* **220**, 16–26.
- Kawamura, S., Blow, N. S., and Yokoyama, S. (1999). Genetic analyses of visual pigments of the pigeon. *Genetics* **153**, 1839–1850.
- Klein, D. C., Coon, S. L., Roseboom, P. H., Weller, J. L., Bernard, M., Gastel, J. A., Zatz, M., Iuvone, P. M., Rodriguez, I. R., Begay, V., Falcon, J., Cahill, G. M., Cassone, V. M., and Baler, R. (1997). The melatonin rhythm generating enzyme: Molecular regulation of serotonin *N*-acetyltransferase in the pineal gland. *Rec. Prog. Horm. Res.* **52**, 307–358.
- Korf, H. W. (1994). The pineal organ as a component of the biological clock: Phylogenetic and ontogenetic considerations. *Ann. N. Y. Acad. Sci.* **719**, 13–42.
- Kusakabe, T., Kusakabe, R., Kawakami, I., Satou, Y., Satoh, N., and Tsuda, M. (2001). *Ci-opsin*, a vertebrate-type opsin gene, expressed in the larval ocellus of the ascidian *Ciona intestinalis*. *FEBS Lett.* **506**, 69–72.
- Mandado, M., Molist, P., Anadon, R., and Yanez, J. (2001). A DiI-tracing study of the neural connections of the pineal organ in two elasmobranchs (*Scyliorhinus canicula* and *Raja montagui*) suggests a pineal projection to the midbrain GnRH-immunoreactive nucleus. *Cell Tissue Res.* **303**, 391–401.
- Mano, H., Kojima, D., and Fukada, Y. (1999). Exo-rhodopsin: A novel rhodopsin expressed in the zebrafish pineal gland. *Mol. Brain Res.* **73**, 110–118.
- Okimoto, D. K., and Stetson, M. H. (1999). Presence of an intrapineal circadian oscillator in the teleostean family Poeciliidae. *Gen. Comp. Endocrinol.* **114**, 304–312.
- Oksche, A. (1984). Evolution of the pineal complex: Correlation of structure and function. *Ophthalmic Res.* **16**, 88–95.
- Yokoyama, S. (1996). Molecular evolution of retinal and non-retinal opsins. *Genes Cells* **1**, 787–794.

**Table I** Histological Distribution of 281 Pineal Region Tumors Gathered in a Multicentral European Study

Histological diagnosis	Number/% of total
Germ cell tumors	96/34%
Pineal parenchymal tumors	76/27%
Astrocytomas	52/19%
Meningiomas	20/7%
Ependymomas	13/5%
Oligodendrogliomas and mixed gliomas	7/2%
Primary melanomas	4/1.5%
Metastases	4/1.5%
Probable parenchymal tumors with insufficient clinical data or histological material	9/3%
Total	281/100%

Adapted from Fauchon *et al.* (2000).

or show intermediate differentiation and do not constitute a unified biological entity.

The revised World Health Organization (WHO) classification on central nervous system tumors divides them into well-differentiated pineocytomas, poorly differentiated pineoblastomas, and mixed pineocytomas–pineoblastomas with intermediate differentiation. A large French and English cooperative study carefully reviewed their results over 24 years; they observed that the presence of more than six mitoses and the presence of necrosis were significantly associated with a poorer outcome, whereas positive immunolabeling for neurofilaments was related to a better survival. This led them to propose a new prognostic classification with four histological grades (Table II). Germ cell tumors, histologically and biologically homologous to gonadal germ cell neoplasms, will characteristically present positive markers for  $\alpha$ -fetoprotein (AFP) and  $\beta$  human

chorionic gonadotropin ( $\beta$ -hCG), with more (teratomas) or less differentiation (germinomas), as well as intermediate degrees (yolk sac tumors) of differentiation. Very rarely, pineal region tumors may derive from meningotheial, mesenchymal, ependymal, and choroid plexus elements, giving rise to gangliogliomas, melanocytic neoplasms, atypical teratoid/rhabdoid tumors, meningiomas, cavernous angiomas, and hemangiopericytomas.

## CLINICAL PRESENTATION

The clinical presentation of pineal tumors varies depending on age at onset and histology. Over 90% of patients present with raised intracranial pressure, often with obstructive hydrocephalus; presenting symptoms are headache, nausea, vomiting, and blurred vision. Visual signs including diplopia, cranial nerve palsies, papilledema, and ptosis as well as Parinaud's syndrome (failure of upward gaze, pupillary dilation, and diminution of pupillary light reflex) are present in 50 to 70% of patients. Abnormalities of other cranial nerves, ataxia, diabetes insipidus, and hypopituitarism may also be diagnosed at presentation, reflecting compression on the brain, cerebellum, hypothalamus, and pituitary. Pineal tumors can interfere with puberty, due to pressure of the tumor on the hypothalamic centers that govern gonadotropin secretion, excessive melatonin secretion by pinealocyte tumors causing delayed puberty in adolescents, or a reduction of the potential anti-gonadotropic effect of melatonin, which, together with  $\beta$ -hCG secretion by destructive germ cell tumors, could explain precocious puberty in prepubertal children. Melatonin deficiency may produce sleeping disorders or behavioral problems.

**Table II** Parenchymal Pineal Tumor Classification

Tumor grade	Morphology	Number of mitoses plus neurofilament (NF) immunostaining (indicative of differentiation)	Prognosis
1	Pineocytoma	0 plus very positive NF	Good
2	Transitional, lobulated, or diffuse tumor	< 6 plus positive NF	↓
3	Lobulated or diffuse tumor	≥ 6 plus positive or negative NF or < 6 plus low expression of NF	
4	Pineoblastoma	Variable plus positive or negative NF	Bad

Adapted from Jouvét *et al.* (2000).

*Note.* This classification differentiates those of intermediate differentiation into two separate groups, with prognostic implications. Data were obtained from a total of 66 patients (11 patients with pineocytomas, ages 10–65 years, 16 patients with pineoblastomas, ages 1–36 years, and 39 patients with intermediate tumors, ages 5–64 years). Median followup was 3.5 years, with overall and event-free survival rates a 5 years, of 60 and 69%, respectively.

## Astrocytic Tumors

These lesions may affect persons of any age, from children to adults. Pilocytic astrocytomas usually present before the age of 20 years, with no sex bias, whereas other types are more frequent in adults.

## Parenchymal Tumors

Pineocytomas present more often in adults over the age of 25 years, do not show a sex bias, evolve slowly (the interval between onset of symptoms and surgery may last several years), and do not invade contiguous tissue or seed the cerebrospinal fluid (CSF). Reports of metastasizing pineocytomas most likely represent misclassifications as pineocytomas of mixed tumors of intermediate differentiation or as pineoblastomas in the current WHO system classification. With strictly defined tumor series, no metastases and a 5-year survival of 86% have been reported. Presenting manifestations are nonspecific and reflect compression of adjoining structures (tectal plate, aqueduct of Sylvius, cerebellum, brainstem, hypothalamic–pituitary circuits): increased intracranial pressure, neuro-ophthalmologic dysfunction, changes in mental status, dysfunction of the brainstem and/or cerebellum, and hypopituitarism with or without hyperprolactinemia. Rarely, intratumoral hemorrhage (pineal apoplexy) with subarachnoid extravasation may be encountered. Concurrent uveoretinitis in occasional patients with pineocytomas probably reflects the common photoreceptor activity of pineal and retinal cells.

Pineoblastoma, the least differentiated and most aggressive pineal parenchymal tumor, typically appears before the age of 20 years and most often in young children, but there are reports in middle age and later adulthood. There is a slight male bias. Clinical manifestations do not differ from those described for pineocytomas, except that they are of more rapid progression and shorter duration. The interval between initial symptoms and surgery may be less than 1 month. Median postsurgical survival varies from 24 to 30 months. They are both locally invasive and prone to disseminate through the CSF; they are often fatal, but may be controlled in some cases by a multimodal combination of aggressive surgery, radiotherapy, and chemotherapy. The association of a pineoblastoma in a child with familial bilateral retinoblastoma (due to a germ-line retinoblastoma gene mutation) is known as a trilateral retinoblastoma, with a median survival of only 6 months.

## Germ Cell Tumors

Approximately 90% of germ cell tumors appear in patients under the age of 20 years. Most arise around the third ventricle, most commonly in the pineal region followed by the suprasellar compartment; 5–10% of patients harbor both lesions. Globally they are more frequent in boys than in girls (2.5:1); however, pineal region germ cell tumors are more frequent in boys and suprasellar lesions are more frequent in girls. An increased risk of intracranial germ cell tumors has been associated with Klinefelter's syndrome, Down's syndrome, and neurofibromatosis type 1.

## Other Tumors

Pineal meningiomas, usually slow growing, have an insidious presentation, with Parinaud's syndrome and hearing impairment. Pineal metastases, most frequently of breast or lung origin, may occur, often with other brain metastases; symptoms and signs will reflect the extent of the disease.

## DIAGNOSIS

An appropriate tissue specimen for accurate histological diagnosis and determining tumor type is critical to optimize subsequent management.

Serum AFP (synthesized mainly by yolk sac tumors and also by teratomas) and  $\beta$ -hCG (in choriocarcinomas or germinomas) concentrations, if markedly elevated, are of diagnostic utility. It is also recommended that these markers be measured in the CSF for initial staging and, if positive, for follow-up. CSF cytological examination should be delayed for at least 2 weeks after surgery to increase the chance of reflecting true dissemination of viable tumor rather than tumor spillage from the time of operation. In the presence of clearly raised markers, histological verification may not always influence management and is not required by some authors, although others believe that even in patients with markedly elevated tumor markers indicative of germ cell tumors (i.e., hCG > 50 mIU/liter) outcome seems to be better if a histological diagnosis is obtained.

Biopsies may be obtained by classical surgical routes (posterior interhemispheric transcallosal, suboccipital transtentorial, and infratentorial–supracerebellar routes) or may be replaced by microsurgical techniques, with significantly reduced perioperative mortality rates (< 2%). Even though stereotactic sampling under endoscopic control has been tried, it should be stressed that the diagnosis of mixed or intermediate

tumors may be very difficult without extensive tissue sampling, so operative risk should be balanced with the risk of not obtaining an accurate histological diagnosis, with prognostic implications.

## IMAGING

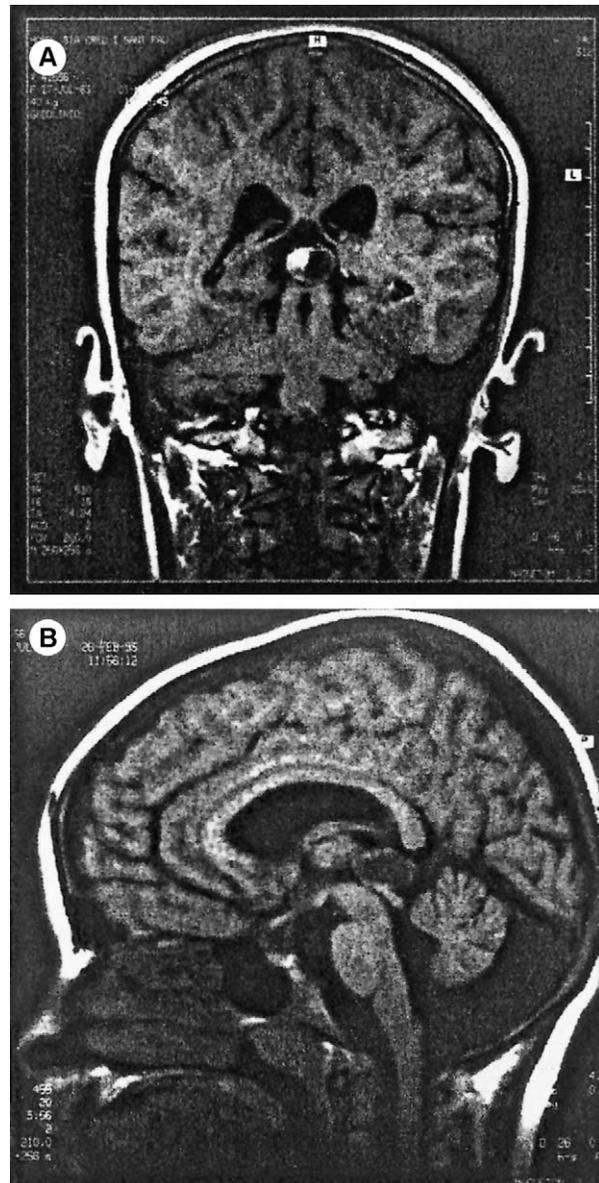
A neuroradiological examination, preferably by magnetic resonance imaging (MRI), will disclose the size and extent of the tumor as well as possible metastases, but cannot accurately identify the histological nature of the tumor, which relies on biopsy or serum/CSF tumor markers. In the more malignant tumors (pineoblastomas, germinomas, teratomas), the spine as well as the brain should be imaged, since spreading into the subarachnoid space and the spine is frequent.

Neuroimaging of astrocytic tumors can be highly variable (Figs. 1 and 2); MRI usually shows hypodensity on T1-weighted images and hyperintensity on T2-weighted images; gadolinium enhancement is uncommon, except when active tumor progression occurs.

When comparing parenchymal tumors, pineocytomas appear as noninvasive, solid masses in the posterior third ventricular region and tend to be smaller (< 3 cm in general), rounder, hypodense, homogenous masses with dispersed calcifications, particularly peripheral, that enhance heterogeneously or diffusely on computed tomography (CT) and MRI and present a lesser degree of hydrocephalus (Fig. 3). Macrocystic presentation is rare but small cysts may be present. T1-weighted images are hypointense, whereas T2-weighted images are hyperintense. Hemorrhage and necrosis are exceptional.

Pineoblastomas are larger, lobulated, homogenous tumors, are rarely calcified, and present with a greater degree of hydrocephalus and local invasion of contiguous brain or leptomeninges. They may exhibit distant subarachnoid and extracranial metastases, more frequently in young females. They are hyperdense and enhance homogeneously on CT, whereas on MRI they appear as hypo- to isointense on T1-weighted images and enhance diffusely or heterogeneously with contrast. Hemorrhage and necrosis are common.

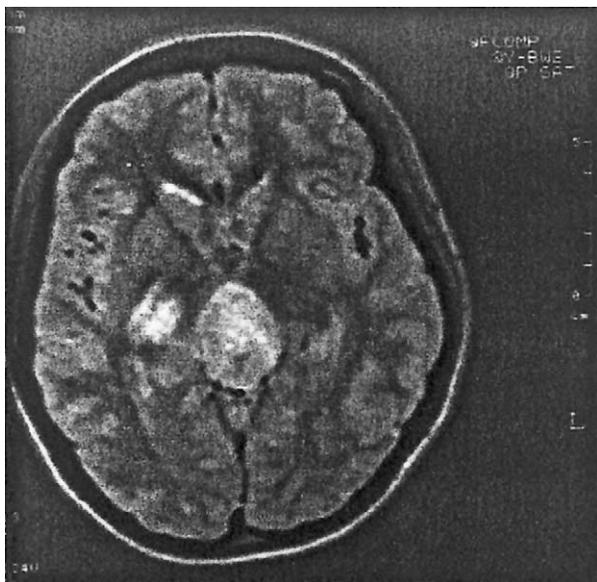
Germ cell tumors (except teratomas) appear as solid masses on MRI that are iso- or hyperdense, which enhance after contrast (Fig. 4); small nodular calcifications may be seen on CT scans. Teratomas tend to contain intratumoral cysts next to calcifications and display low-attenuation signals, typical of fat. Hemorrhages are common in choriocarcinomas and mixed neoplasms.



**Figure 1** T1-weighted gadolinium-enhanced coronal view (A) and midsagittal view (B) of an MRI of a partially cystic pineal astrocytomas in a 10-year-old girl. Additionally, there is an enlargement of the cisterna magna in the posterior fossa, which may be an arachnoidal cyst.

## PROGNOSIS

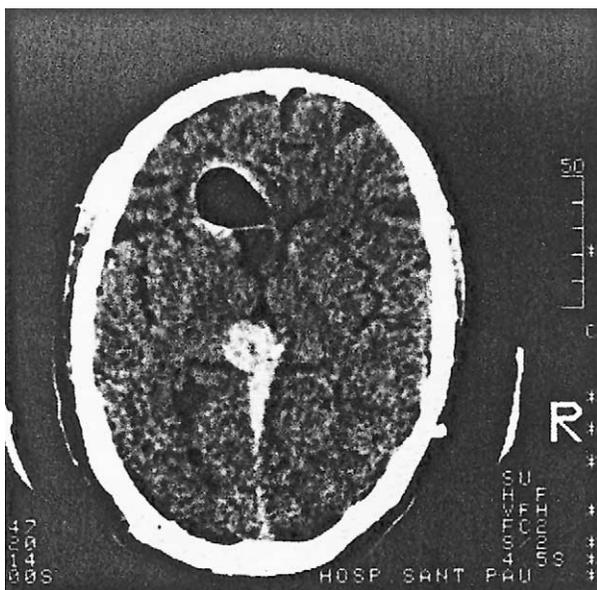
The prognosis of pineal tumors has improved over the years. Major indicators of prognosis are histology and neuroaxis staging. Accurate histological diagnosis is also important to optimize treatment. Thus, the benefit of a histological diagnosis is generally considered to outweigh the risk of biopsy, especially



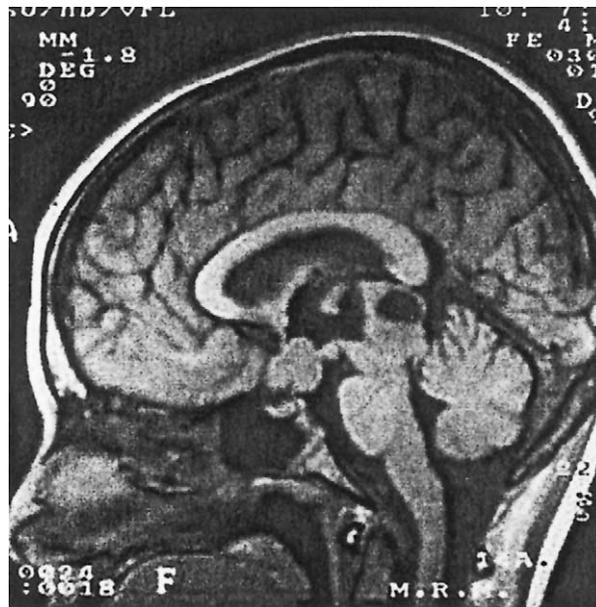
**Figure 2** Axial view of a DP-T2-weighted MRI of a 3.5 cm pineal solid astrocytoma in a 23-year-old woman.

since morbidity and mortality of this procedure are lower with experienced neurosurgical teams.

Mean survival after surgery of astrocytic tumors is approximately 6–8 years with marked individual variation. Malignant progression to glioblastoma, which



**Figure 3** Axial view with contrast enhancement of a CT scan of a pinealoma in a 55-year-old woman; calcifications were observed on nonenhanced images. Enlargement of the left ventricular frontal horn is also observed.



**Figure 4** Midsagittal T1-weighted MRI of a hypothalamic and pineal dysgerminoma (the latter partially cystic) in a 14-year-old boy. There is also marked tumoral enlargement of the quadrigeminal lamina next to the pineal.

tends to occur after a mean interval of 4–5 years, determines final outcome. Positive glucose consumption with positron emission tomography scans of glioblastoma clearly correlates with cellularity and reduced survival.

Whereas prognosis of pineocytomas is generally good (86% at 5 years), the extent of the disease at the time of diagnosis, as determined by the CSF examination and MRI of the neuroaxis, directly affects the survival of patients with pineoblastomas, and metastases to the CNS and vertebral column are the most common causes of death. Prognosis is even worse in sporadic or familial trilateral retinoblastoma (approximately 6 months). Intermediate or mixed parenchymal tumors (grade 2 and 3 in [Table II](#)) have an intermediate prognosis.

Prognostic factors were investigated in 76 parenchymal pineal tumors identified from 281 pineal region tumors gathered over 5 years from 12 different neurosurgical centers in Europe by two experienced neuropathologists; it is the largest series thus far published ([Table I](#)) and includes patients that were diagnosed and followed for 25 years. Eight patients were lost to follow-up between 1 and 78 months after diagnosis. The median follow-up for the whole patient population was 3.5 years and it exceeded 5 years in 25 of the 76 cases. There was a 1.5:1 (39:27)

female:male sex distribution (not previously observed in smaller series), with a median age of 32 years (range 1 to 65 years). Fifty-seven patients were adults and 19 patients were children with a median age of 2.5 years (range 9 months to 16 years). Positive cytology was seen only in grade 4 tumors (Table II) and positive spinal MRI was seen only in grade 3 and 4 tumors. Median diameter was 33.9 mm (range 10 to 74 mm). Sixty-three patients required a CSF shunt (83%), whereas 74/76 patients underwent surgery at presentation; the other 2 patients were diagnosed at autopsy or from a biopsy obtained at laminectomy for spinal cord compression during follow-up. Tumor removal was attempted in 19% of patients treated between 1972 and 1980, in 42% of patients treated between 1981 and 1990, and in 39% of patients treated between 1991 and 1997. Unifactorial and multifactorial analyses (including sex, age at treatment, tumor size, hydrocephalus, calcifications on imaging, and meningeal seeding) for overall and event-free survival were performed. A second analysis added treatment information (extent of surgery, radiation doses and techniques, and chemotherapy) and a third analysis included the 4-grade histological classification (see Table II). Younger patients were found to have more mitoses, more tumor necrosis, more high-grade tumors, and larger tumors. Chemotherapy and radiotherapy were administered significantly more often to patients with high-grade and large tumors, with high mitotic activity. The impact of radiotherapy on survival was not significant for low-grade tumors, but was significant ( $P < 0.02$ ) for high-grade tumors. Significant variables on survival or disease-free survival included histological grading (as seen in Table I), age  $< 20$  years, tumor diameter  $\geq 40$  mm, and chemotherapy (which had a negative effect), whereas diameter  $< 25$  mm had a positive effect. Multivariate analysis showed a major role of histological grade alone on disease-free survival, whereas both grade and tumor size had an effect on overall survival. Specifically, 5-year survival was 91% for grade 1, 74% for grade 2, 39% for grade 3, and 10% for grade 4, whereas event-free survival was 100% for grade 1, 96% for grade 2, 42% for grade 3, and 0% for grade 4. The prognostic importance of this classification, especially in the more heterogeneous mixed tumors (grades 2 and 3), would seem to indicate that they might benefit from different therapeutic approaches.

Pure germinomas, which are exquisitely radiosensitive, have an especially favorable prognosis, with a 5-year survival rate after radiotherapy alone of approximately 65–95%. The addition of chemotherapy may allow a comparable prognosis at a reduced

radiation dose. In other germ cell tumors, prognosis is worse, unless total surgical resection is possible (i.e., in the case of mature, noninvasive teratomas), followed by chemotherapy. In general, the greater extent of disease at diagnosis is associated with a poorer outcome, which may be precipitated by metastases to the brain or neuroaxis, metastases to the lung or bone, or abdominal contamination via ventriculoperitoneal shunts.

## TREATMENT

Advances in surgery, chemotherapy, and radiation have changed the manner in which pineal region tumors are diagnosed and treated. Open surgery may be replaced by stereotactic or endoscopic approaches to obtain a biopsy and are mandatory to obtain a definite histological diagnosis. Morbidity and cure rates have improved thanks to a greater understanding of the nature of the different tumors, more accurate neurosurgical experience, selective use of chemotherapy, and the introduction of modern irradiation techniques. However, given the rarity of pineal tumors, large prospective multicenter international studies would be desirable to define their optimal management.

Treatment of pineal tumors depends on histology obtained after surgery, which apart from the biopsy can resolve intracranial hypertension with a ventricular shunt (atrial or peritoneal). Partial debulking of the tumor should be performed if possible; total resection is rarely possible (Table III).

### Astrocytomas

Treatment for astroglial cell-derived malignant gliomas is local radiotherapy to the tumor (54 Gy), either conventional or stereotaxic, whereas surgery may be curative for the more benign pilocytic astrocytomas.

### Pineal Parenchymal Tumors

Pineocytomas require only local radiotherapy to the tumor (54 Gy).

In pinealoblastomas, a high probability of spinal seeding should lead to craniospinal radiotherapy, since they are radiosensitive (25 to 30 Gy on the neuroaxis with a pineal boost of 40 Gy aimed at more effective local disease control). However, routine craniospinal irradiation has been called into question and may not be necessary in patients with negative staging. New techniques of administering radiotherapy, such as highly localized radiosurgery, need to be evaluated, since radiosurgery may control local

Table III Schematic Therapeutic Guidelines for the Treatment of Pineal Region Tumors

Tumor type	Radiotherapy	Chemotherapy <sup>a</sup>	Surgery
<b>Glial origin</b>			
Juvenile pilocytic astrocytoma	No	No	Complete resection
Intermediate/diffuse/anaplastic/astrocytomas/glioma	Local	No	Debulking
Malignant glioblastoma	Local	No	Debulking
<b>Parenchymal tumors</b>			
Pineocytoma	Local	No	Biopsy
Intermediate or mixed tumor	Local ± craniospinal	Yes in more undifferentiated tumors	Biopsy
Pineoblastoma	Local + (?) routine craniospinal; age < 5 years: lower dose after chemotherapy	Yes (role on final outcome unclear)	Biopsy
<b>Germ cell tumors</b>			
Germinoma	Local + craniospinal (unless convinced of negative staging)	Yes (alone not curative)	Biopsy
Non germinatous tumors	Local + craniospinal	Yes (pre- or post surgery)	Resection as much as possible, without increased morbidity

Note. Initial surgery for histological biopsy and, if necessary, a ventricular shunt should always be performed, with the possible exception of germ cell tumors with diagnostically elevated tumor markers.

<sup>a</sup>Chemotherapy includes cisplatin, etoposide, and cyclophosphamide or ifosfamide.

progression but is associated with a high risk of marginal recurrence and distant metastases and is therefore not considered the treatment of choice for infiltrative but curable tumors. Furthermore, complications such as ataxic gait and gaze palsy were reported several months after radiosurgery. In comparison to conventional radiotherapy, radiosurgery has the advantage of reducing the radiation dose to surrounding normal brain while augmenting the radiobiological effect on the tumor volume; this reduced total brain dose may be especially important in prepubertal patients with pineoblastomas, in whom total brain irradiation is associated with neurocognitive dysfunction, endocrinopathy, vascular complications, and spinal growth impairment. In these infants, chemotherapy with cisplatin, etoposide, cyclophosphamide, and vincristine, which alone is not curative, may allow a lower dose of radiotherapy to have similar effects. In older children with pineoblastoma, craniospinal irradiation is followed by chemotherapy (even though its effect on final outcome has not been fully defined).

In mixed or intermediate pineal parenchymal cells (grade 2 or 3, Table II), apart from local radiotherapy, craniospinal irradiation and chemotherapy should be considered when there is an increased number of mitoses and less differentiation.

Chang and colleagues reported a series of 11 adults with pineoblastoma (ages 17–59 years), evaluated retrospectively, who were diagnosed between 1975

and 1992. All presented with hydrocephalus and the 5 with positively staged disease had focal or spinal progression after a median of 10 months and died 1 to 20 months after recurrence (median survival time from the date of surgery was 30 months). One patient refused treatment and died 6 months after diagnosis. The 5 patients with negative staging at diagnosis were alive without disease progression for a median of 26 months after follow-up. In view of these findings, the authors recommend initial staging to include examination of the CSF and MRI of the brain and spine; however, the benefits of chemotherapy after craniospinal irradiation also remained unclear in these adult pineoblastomas.

Cohen and colleagues reported results from a randomized clinical trial in children (up to the age of 19 years); pineoblastomas had a better prognosis than other primitive neuroectodermal tumors such as medulloblastomas (located in the posterior fossa) or central neuroblastomas (originating in the cortical regions of the brain) with a 3-year survival and progression-free survival rates of  $73 \pm 12\%$  and  $61 \pm 13\%$  for the 17 pineal tumors and  $57 \pm 8\%$  and  $45 \pm 8\%$  for the whole group of 55 supratentorial primitive neuroectodermal tumors ( $P < 0.03$ ). In this randomized study, all patients underwent craniospinal radiotherapy followed by different protocols of chemotherapy, which did not seem to influence outcome significantly; the authors recommend eight cycles of CCNU,

vincristine, and prednisone, since this combination was less toxic than other chemotherapeutic combinations. The presence of initial metastases had an ominous prognosis (with no progression-free survival in these patients) and younger age (< 2 years) was a worse univariate prognostic factor than age > 3 years ( $25 \pm 13\%$  versus  $53 \pm 9\%$ ,  $P < 0.02$ ).

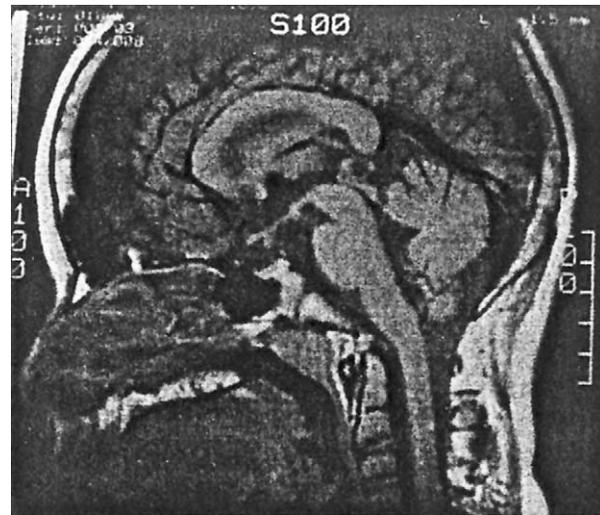
### Germ Cell Tumors

Surgery is not considered curative in germinomas, which are radiosensitive and should therefore receive local radiotherapy. Unless firmly confident of negative staging (by negative tumor markers—AFP and  $\beta$ -HCG—in blood and CSF and by negative MRI), craniospinal radiotherapy should be offered, given the high probability of spinal seeding. Germinomas are also highly chemosensitive, and excellent responses to postoperative cisplatin and cyclophosphamide have been reported; however, toxicity of chemotherapy, which alone is not curative, should be considered, since the effect of chemotherapy on final outcome is unclear. Survival is high in patients with localized germinomas, using chemotherapy, focal radiotherapy, or craniospinal irradiation; however, focal irradiation alone has a worse outcome. In metastatic germinomas, craniospinal irradiation is the treatment of choice (25–35 Gy to the spine and a local pineal boost of 40 Gy). Lower irradiation doses are being considered, especially if adjuvant chemotherapy is offered. Bifocal lesions in the pineal and hypothalamus should be considered localized germinomas rather than metastatic disease and receive irradiation to both locations.

Other germ cell tumors are less radiosensitive than germinomas, with a poor survival after radiotherapy alone (median survival of under 2 years), and require multimodal treatment. Aggressive tumor resection, which may improve survival, should not justify increased morbidity. Surgical resection after tumor reduction with initial chemotherapy with cisplatin, etoposide, and ifosfamide is a modern alternative. Tumor markers are useful for follow-up. Chemotherapy should be combined with radiotherapy (local up to 54 Gy or craniospinal up to 36 Gy), in which case long-term survival may reach 80%.

### Other Tumors

Surgery is the treatment of choice of pineal meningiomas and other localized pineal tumors if possible; alternatively, localized stereotactic radiosurgery may be offered with good long-term prognosis.



**Figure 5** T1-weighted midsagittal MRI of a pineal cyst; it is asymptomatic and creates no hydrocephalus.

### PINEAL CYSTS

Masses in the pineal region are most commonly non-neoplastic cysts incidentally discovered at autopsy or on a radiographic work-up for symptoms not reasonably attributed to the cyst (Fig. 5). Very rarely can they act as a mass lesion and produce signs of increased intracranial pressure, by compressing the aqueduct (obstructive hydrocephalus) or tectal plate (Parinaud's syndrome). On MRI, they appear as a 1 to 3 cm mass, they are equally as dense as, or slightly denser than, CSF in T-1 weighted image studies, and they brightly enhance in T-2 weighted images, reflecting their fluid nature. Evidence of hemorrhage and peripheral calcification may be found. If asymptomatic, pineal cysts do not generally require treatment; if such a cyst is large enough to increase intracranial pressure, resection may be necessary, with an excellent long-term outcome.

### Acknowledgments

The helpful suggestions of M. Fèvre-Montange and A. Jouvret are gratefully acknowledged.

### See Also the Following Articles

Adrenal Tumors, Molecular Pathogenesis • Circadian Rhythms: Hormonal Facets • Craniopharyngiomas • Melatonin • Pineal Gland • Pineal Gland, Evolution of • Pituitary Tumors, Clonality • Pituitary Tumors, Molecular Pathogenesis

## Further Reading

- Bailey, S., Skinner, R., Lucraft, H. H., Perry, R. H., Todd, N., and Pearson, A. D. (1996). Pineal tumours in the north of England 1968–93. *Arch. Dis. Child* **75**, 181–185.
- Chang, S. M., Lillis-Hearne, P. K., Larson, D. A., Wara, W. M., Bollen, A. W., and Prados, M. D. (1995). Pineoblastoma in adults. *Neurosurgery* **37**, 383–390.
- Cohen, B. H., Zeltzer, P. M., Boyett, J. M., Geyer, J. R., Allen, J. C., Finlay, J. L., McGuire-Cullen, P. L., Milstein, J. M., Rorke, L. B., Stanley, P., Stehens, J. A., Shurin, S. B., Wisoff, J., Stevens, K. R., and Albright, A. L. (1995). Prognostic factors and treatment results for supratentorial primitive neuroectodermal tumors in children using radiation and chemotherapy: A children's cancer group randomized trial. *J. Clin. Oncol.* **13**, 1687–1696.
- Fauchon, F., Jouvét, A., Paquis, P., Saint-Pierre, G., Mottolese, C., Ben Hassel, M., Chauveinc, L., Sichez, J. P., Philippon, J., Schlienger, M., and Bouffet, E. (2000). Parenchymal pineal tumors: A clinicopathological study of 76 cases. *Int. J. Radiat. Oncol. Biol. Phys.* **46**, 959–968.
- Jackacki, R. I., Zeltzer, P. M., Boyett, J. M., Albright, A. L., Allen, J. C., Geyer, J. R., Rorke, L. B., Stanley, P., Stevens, K. R., Wisoff, J., McGuire-Cullen, P. L., Milstein, J. M., Packer, R. J., and Finlay, J. L. (1995). Survival and prognostic factors following radiation and/or chemotherapy for primitive neuroectodermal tumors of the pineal region in infants and children: A report of the Children's Cancer Group. *J. Clin. Oncol.* **13**, 1377–1383.
- Jouvét, A., Saint-Pierre, G., Fouchon, F., Privat, K., Bouffet, E., Ruchoux, M. M., Chauveinc, L., and Fèvre-Montange, M. (2000). Pineal parenchymal tumors: A correlation of histological features with prognosis in 66 cases. *Brain Pathol.* **10**, 49–60.
- Mena, H., Nakazato, Y., and Scheithauer, B. W. (1997). Pineal parenchymal tumors. In "Pathology and Genetics: Tumors of the Nervous System" (P. Kleihues and W. K. Cavenee, eds.), pp. 115–121. International Agency for Research on Cancer, Lyon, France.
- Oi, S., Shibata, M., Tominaga, J., Honda, Y., Shinoda, M., Takei, F., Tsugane, R., Matsuzawa, K., and Sato, O. (2000). Efficacy of neuroendoscopic procedures in minimally invasive preferential management of pineal region tumors: A prospective study. *J. Neurosurg.* **93**, 245–253.
- Schild, S. E., Scheithauer, B. W., Schomberg, P. J., Hook, C. C., Kelly, P. J., Frick, K., Robinow, J. S., and Biskirk, S. J. (1993). Pineal parenchymal tumors: Clinical, pathologic and therapeutic aspects. *Cancer* **72**, 870–880.

tumors might make surgical resection difficult and traumatic and might explain poor surgical results in patients with macro- or invasive adenomas.

TSH-secreting pituitary adenomas constitutively secrete TSH and exhibit a defect in the negative regulation of TSH by thyroid hormones. High thyroid hormone concentration in the face of normal TSH level is explained, at least in part, by an increased bioactivity of neoplasm-derived TSH. A somatic mutation of TR $\beta$  gene, impairing negative regulation of  $\alpha$ -subunit and TSH $\beta$  genes, has been reported in a TSH-secreting pituitary adenoma, and it might be one mechanism for the defective regulation of TSH by triiodothyronine (T3) observed in these rare tumors.

Thyrotropin-releasing hormone (TRH) receptors are found on the majority of tumor cells. However, the lack of response of TSH and  $\alpha$ -subunit to TRH in the majority of patients with TSH-secreting pituitary adenomas suggested that the TRH receptors are non-functional on these tumors. However, partial or even normal response to TRH is present in other patients, suggesting etiological differences among TSH-secreting tumors. Most studies have revealed high concentrations of different subtypes of somatostatin receptors on TSH-secreting adenomas, explaining the *in vivo* response to somatostatin analogue administration. Somatostatin and its analogues (e.g., octreotide, lanreotide) reduce TSH secretion and TSH bioactivity immediately due to a change in glycoisomer distribution. Dopamine D2 receptors have been detected in TSH-secreting tumors, but the effect of long-term bromocriptine therapy is limited due to incomplete suppression of inappropriate TSH secretion. Finally, estrogen receptors have been found on TSH-secreting tumors and, at least in theory, could accelerate the growth of these tumors during a hyperestrogenic state such as pregnancy.

## CLINICAL PRESENTATION

Clinical manifestations of central hyperthyroidism secondary to a TSH-secreting pituitary adenoma are frequently mild and not different from those occurring in the other forms of hyperthyroidism such as autoimmune Graves' disease, multinodular goiter, and toxic nodular goiter. TSH-secreting pituitary adenomas occur at any age and without a female predominance in the majority of the series. Clinical features of hyperthyroidism are progressive in their installation, and the mean latency between onset of hyperthyroidism and correct diagnosis is 4 years. Mild, moderate, and severe thyrotoxicosis with atrial fibrillation and

cardiac failure are observed in such patients. The presence of a goiter is a rule, and about two-thirds of patients have thyroid nodules or multinodular goiters, probably due to sustained TSH stimulation of the thyroid cells over many years. The occurrence of differentiated thyroid carcinoma is very rare. In many patients, tumoral signs and symptoms due to compression of the surrounding nervous structures, such as visual field defects (80%) and headaches (20%), prevail over those due to hyperthyroidism. Unilateral exophthalmos due to orbital invasion by the pituitary tumor is rare. In mixed secreting tumors, clinical findings are dependent on the nature of the hormone cosecreted (e.g., acromegalic features, amenorrhea and/or galactorrhea). In macroadenomas, concomitant hyposecretory syndromes (e.g., impotence, reduced libido) are not rare. TSH-secreting pituitary adenomas have been reported in pregnant women, in patients with multiple endocrine neoplasia type I, and in atypical McCune–Albright syndrome.

## HORMONAL EVALUATION

TSH-secreting pituitary adenomas are now diagnosed with the introduction of ultrasensitive assay for TSH routinely performed as firstline thyroid function tests. Detectable or elevated TSH concentration with biochemical hyperthyroidism (increased free thyroid hormone concentrations) exclude primarily hyperthyroidism such as Graves' disease. Although TSH pulsatility is respected, the physiological circadian TSH variation is absent. Increased plasma  $\alpha$ -subunit level and increased  $\alpha$ -subunit/TSH molar ratio (i.e., greater than 1) are considered as diagnostic tools, but  $\alpha$ -subunit level is usually normal in patients with TSH-secreting pituitary microadenomas. Recent studies show a relationship between hypersecretion of other pituitary hormones and tumor volume; high prolactin level and/or elevated growth hormone and insulin-like growth factor-1 (IGF-1) concentrations are reported in macroadenomas. In many patients with TSH-secreting pituitary adenomas, dynamic tests reveal a decrease or lack of response of TSH during the TRH test, absence or inappropriate negative feedback of thyroid hormone on TSH secretion (T3 suppression test), and/or a marked decrease in TSH level after subcutaneous injections of somatostatin analogue octreotide.

## PITUITARY IMAGING

The presence of a pituitary tumor in a patient with inappropriate secretion of TSH, although strongly

suggestive, is not diagnostic of a TSH-secreting pituitary tumor given that pituitary incidentalomas have been found on magnetic resonance imaging (MRI) in up to 10% of normal individuals. In patients with TSH-secreting pituitary adenomas, there is no correlation between serum TSH levels and tumor size. On the other hand, the development of a TSH-secreting tumor is rare in patients with resistance to thyroid hormone syndrome. In the oldest series, many patients presented with macroadenomas with suprasellar and sphenoidal extension, whereas in more recent reports, MRI frequently revealed medially localized microadenomas. Several observations of patients with neoplastic-inappropriate secretion of TSH have been reported with normal MRI, and the diagnosis of TSH-secreting pituitary adenoma has been made with bilateral petrous sinus sampling with TRH stimulation test, <sup>111</sup>indium pentreotide scan, or <sup>11</sup>C methionin PET scan.

## DIFFERENTIAL DIAGNOSIS

Elevated levels of total thyroid hormones and inappropriately normal or elevated TSH levels might be related to increased levels of thyroxine-binding proteins, familial dysalbuminemia, treatment with drugs (amiodarone), or the presence of heterophilic antibodies or autoantibodies able to interfere, depending on the methodology used, with the measurement of TSH and thyroid hormones. In fact, the differential diagnosis between a TSH-secreting pituitary adenoma and a resistance to thyroid hormone syndrome might be difficult because age, sex, TSH, and free thyroid hormone concentrations are not different. The diagnosis of a

TSH-secreting pituitary adenoma is likely in the presence of an absence of family history, an elevated serum  $\alpha$ -subunit level with a high  $\alpha$ -subunit/TSH molar ratio, an absence of circadian TSH variation, a high level of sex hormone-binding globulin (SHBG), an absent or impaired TSH response to TRH administration, an absent TSH suppression after oral administration of T<sub>3</sub>, and a long-term response of thyroid parameters to somatostatin analogue injections. The diagnosis of central hyperthyroidism secondary to a TSH-secreting pituitary adenoma is confirmed by the finding of a pituitary lesion on MRI and by genetic analysis with the absence of mutation in the sequence of the TR $\beta$  gene (Table I). Finally, few cases of coexisting Graves' disease or resistance to thyroid hormone syndrome and TSH-secreting pituitary adenoma have been reported.

## TREATMENT

In patients with TSH-secreting pituitary adenomas, the goal of therapy is to restore euthyroidism in hyperthyroid patients and to eliminate the symptoms of mass effect in patients with large tumors. Early diagnosis and correct treatment of these rare tumors prevent complications, such as visual abnormalities due to compression of optic chiasma and hypopituitarism, and also might improve the cure rate. The success of treatment depends on the criteria used. An early test of cure might be an undetectable TSH concentration 7 days after surgery (with measured TSH level reflecting tumoral TSH) when the normal thyrotrophs are suppressed by the thyroid hormones, whereas normalization of dynamic tests (e.g., TRH

Table I Differential Diagnosis Between TSH-Secreting Pituitary Adenoma and Resistance to Thyroid Hormone Syndrome

	TSH-secreting adenomas	Resistance to thyroid hormone
Family history	Absent	Possible
Increased free thyroid hormone	+	+
Normal or high TSH	+	+
Circadian TSH secretion	Absent	Present
SHBG	Increased	Normal
$\alpha$ -subunit level	Increased	Normal
$\alpha$ -subunit/TSH molar ratio	>1	<1
Response to TRH test	Absent	Normal
Response to T <sub>3</sub> test	Absent	Normal
Hormonal cosecretion	+	-
MRI	Adenoma	Normal
Genetic analysis	-	+

test, T3 suppression test) might represent a later test of cure. In any case, long-term follow-up is necessary to detect relapse and recurrence.

Transphenoidal surgery is the firstline therapy in patients with TSH-secreting pituitary microadenomas. The occurrence of inappropriate secretion of antidiuretic hormone seems to be frequent after surgery of such tumors. Pituitary surgery alone leads to normalization of thyroid hormone secretion and disappearance of the pituitary mass in approximately 50% of patients, leads to normalization of thyroid parameters despite incomplete tissue removal in approximately 25% of patients, and is unsuccessful in nearly 30% of cases. Therefore, the results of surgery are rather disappointing in patients with TSH-secreting pituitary adenomas. However, an increase in the surgical rate has been reported in the more recent series, probably reflecting the early diagnosis of these tumors.

Conventional radiotherapy and stereotaxic radiosurgery are associated with low percentage remission of hyperthyroidism, underscoring the relative resistance of TSH-secreting pituitary adenomas to external radiotherapy. Also taking into consideration the long time period necessary for the full effect to be realized and the well-known late side effects (e.g., hypopituitarism, impaired cognitive function), radiotherapy is now prescribed less frequently in patients with central hyperthyroidism related to TSH-secreting pituitary adenomas.

Medical treatment is an alternative option to pituitary surgery in patients with TSH-secreting pituitary tumors. Dopamine agonists are an effective treatment in mixed TSH/prolactin-secreting pituitary adenomas, whereas success is limited with bromocriptine treatment in patients with pure TSH-secreting pituitary adenomas. The presence of somatostatin receptors on TSH-secreting pituitary adenomas has allowed consideration of treatment with somatostatin analogues in patients with TSH-dependent hyperthyroidism related to TSH-secreting pituitary tumors. Somatostatin analogue octreotide suppresses TSH secretion in more than 90% of TSH-secreting pituitary adenomas, normalizes thyroid hormone concentration in approximately 75% of patients, and decreases adenoma size in approximately 50% of cases with an amelioration of visual disturbances in patients with macroadenomas. However, the effects of somatostatin analogues are reversible, with the need for long-term administration, with possible tachyphylaxis requiring increasing doses of the drug to maintain good control of the disease in 10% of cases, and

with true resistance in few patients. This treatment is expensive and might be associated with side effects such as cholelithiasis and carbohydrate intolerance. In the oldest reports, long-term medical therapy with somatostatin analogues was indicated adjunctively in patients with TSH-secreting pituitary adenomas who failed to be cured after surgery or who were awaiting the full effect of radiotherapy. The use of somatostatin analogues as primary therapy was generally reserved for patients who refused surgery or who were poor surgical candidates. Recent studies have demonstrated that primary treatment with somatostatin analogues might be a reasonable option in patients with TSH-secreting pituitary tumors to preoperatively control the TSH-dependent hyperthyroidism and that long-term treatment is effective in patients with TSH-secreting macroadenomas or invasive tumors treated with surgery and/or pituitary radiotherapy as well as in previously untreated individuals. Finally, long-acting somatostatin analogues (e.g., Sandostatin LAR, Somatuline LP), injected intramuscularly every 10 to 28 days, are useful therapeutic tools to improve compliance of patients and to facilitate the medical treatment of TSH-secreting pituitary tumors in patients who need long-term somatostatin analogue therapy.

## CONCLUSION

TSH-secreting pituitary adenomas are a rare cause of hyperthyroidism. The main prognostic factors of these adenomas are size and invasiveness of the tumors, duration of symptoms, and intensity of hyperthyroidism. The early recognition of central hyperthyroidism due to these pituitary tumors is possible with the availability of sensitive TSH assays and the improvement in pituitary imaging. Therefore, the spectrum of the diagnosis and treatment of these rare pituitary tumors has changed during the past decade. Patients present with mild or moderate symptoms and signs of hyperthyroidism, hormonal evaluation shows increased free thyroid hormone concentrations with detectable serum TSH levels, and MRI leads to increase the percentage of patients with microadenomas. On the other hand, transphenoidal surgery remains the treatment of choice for patients with TSH-secreting pituitary microadenomas, whereas long-acting somatostatin analogue administration seems to be indicated as a primary treatment in patients with macroadenomas or invasive pituitary tumors.

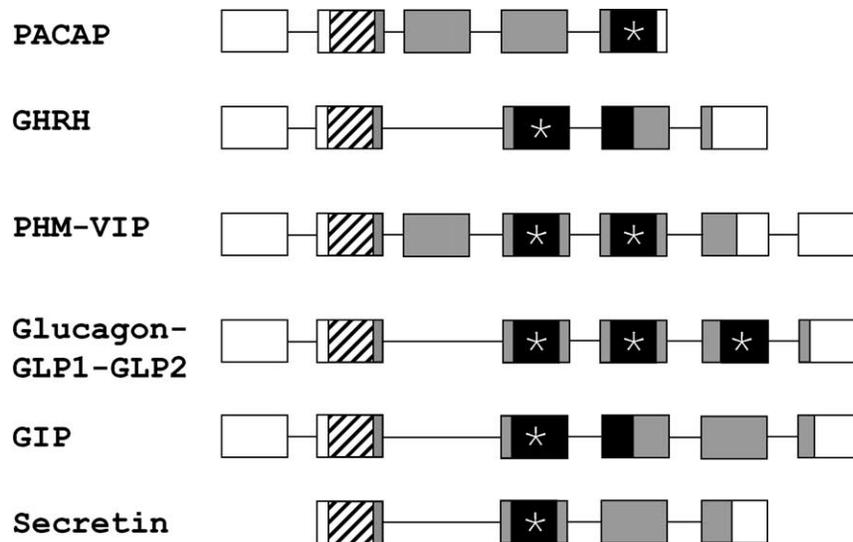
## See Also the Following Articles

Medullary Thyroid Carcinoma • Pituitary Region, Non-Functioning Tumors of • Pituitary Tumors, ACTH-Secreting • Pituitary Tumors, Clonality • Pituitary Tumors, Molecular Pathogenesis • Pituitary Tumors, Surgery • Thyroid Carcinoma • Toxic Adenoma • TSH (Thyroid-Stimulating Hormone; Thyrotropin)

## Further Reading

- Ando, S., Sarlis, N. J., Oldfield, E. H., and Yen, P. M. (2001). Somatic mutation of TR $\beta$  can cause a defect in negative regulation of TSH in a TSH-secreting pituitary tumor. *J. Clin. Endocrinol. Metab.* **86**, 5572–5576.
- Beck-Peccoz, P., Brucker-Davis, F., Persani, L., Smallridge, R. C., and Weintraub, B. D. (1996). Thyrotropin-secreting pituitary tumors. *Endocrine Rev.* **17**, 610–638.
- Brucker-Davis, F., Oldfield, E. H., Skarulis, M. C., Doppman, J. L., and Weintraub, B. D. (1999). Thyrotropin-secreting pituitary tumors: Diagnostic criteria, thyroid hormone sensitivity, and treatment outcome in 25 patients followed at the National Institutes of Health. *J. Clin. Endocrinol. Metab.* **84**, 476–486.
- Caron, P., Arlot, S., Bauters, C., Chanson, P., Kuhn, J. M., Pugeat, M., Marechaud, R., Teutsch, C., Vidal, E., and Sassano, P. (2001). Efficacy of the long-acting octreotide formulation (Octreotide-Lar) in patients with thyrotropin-secreting pituitary adenomas. *J. Clin. Endocrinol. Metab.* **86**, 2849–2853.
- Caron, P., Gerbaud, C., Pradayrol, L., Simonetta, C., and Bayard, F. (1996). Successful pregnancy in an infertile woman with a thyrotropin-secreting macroadenoma treated with somatostatin analog (octreotide). *J. Clin. Endocrinol. Metab.* **81**, 1164–1168.
- Chanson, P., Weintraub, B. D., and Harris, A. G. (1993). Treatment of TSH-secreting pituitary adenomas with octreotide: A follow-up of 52 patients. *Ann. Int. Med.* **119**, 236–240.
- Iglesia, P., and Diez, J. J. (1998). Long-term preoperative management of thyrotropin-secreting pituitary adenoma with octreotide. *J. Endocrinol. Invest.* **21**, 775–778.
- Kuhn, J. M., Arlot, S., Lefebvre, H., Caron, P., Cortet-Rudelli, C., Archambaud, F., Chanson, P., Tabarin, A., Goth, M. I., Blumberg, J., Catus, F., Ispas, S., and Beck-Peccoz, P. (2000). Evaluation of the treatment of thyrotropin-secreting pituitary adenomas with a slow release formulation of the somatostatin analog lanreotide. *J. Clin. Endocrinol. Metab.* **85**, 1487–1491.
- Mixson, A. J., Friedman, T. C., Katz, D. A., Feuerstein, I. M., Taubenberger, J. K., Colandrea, J. M., Doppman, J. L., Oldfield, E. H., and Weintraub, B. D. (1993). Thyrotropin-secreting pituitary carcinoma. *J. Clin. Endocrinol. Metab.* **76**, 529–533.
- Sanno, N., Teramoto, A., and Osamura, R. Y. (2001). Thyrotropin-secreting pituitary adenomas: Clinical and biological heterogeneity and current treatment. *J. Neuro-Oncol.* **54**, 179–186.
- Socin, H. V., Chanson, P., Delemer, B., Tabarin, A., Rohmer, V., Mockel, J., Stevenaert, A., and Beckers, A. (2003). The changing spectrum of TSH-secreting pituitary adenomas: Diagnosis and management in 43 patients. *Eur. J. Endocrinol.* **148**, 433–447.





**Figure 2** Organization of the genes encoding the PACAP/glucagon superfamily members. White boxes represent 5'- and 3'-untranslated regions; crosshatched boxes represent signal peptides; black boxes represent bioactive superfamily members, and shaded boxes represent N- and C-terminal peptides and intervening domains. Introns are depicted as lines, but are not drawn to scale. PACAP, pituitary adenylate cyclase-activating polypeptide; GHRH, growth hormone-releasing hormone; PHM, peptide histidine-methionine; VIP, vasoactive intestinal peptide; GLP1, glucagon-like peptide-1; GLP2, glucagon-like peptide-2; GIP, glucose-dependent insulinotropic polypeptide.

There are three types of PACAP receptors, each encoded by a separate gene. One of these receptors, the PAC<sub>1</sub> receptor, is specific for PACAP. The other two receptors, VPAC<sub>1</sub> and VPAC<sub>2</sub>, bind both PACAP and VIP. The most interesting of these three receptors is PAC<sub>1</sub> because it is not only specific for PACAP, but has nine variant forms. These variants have insertions in the third intracellular loop, deletions in the extracellular domain, or substitutions in the second and fourth transmembrane regions. The effect is that some of the variants couple to different intracellular signaling pathways that act through second messengers such as cyclic AMP (cAMP), inositol triphosphate, and calcium ions. In contrast, the receptors shared with VIP do not have variant forms; they induce postreceptor effects primarily through cAMP.

The target tissues for the action of PACAP are those expressing the receptors. The PAC<sub>1</sub> receptors are widely distributed in the brain. VPAC<sub>1</sub> and VPAC<sub>2</sub> receptors are in the brain with some concentration in the hippocampus and hypothalamus, but are far less abundant than PAC<sub>1</sub> receptors in the brain. In peripheral tissues, PAC<sub>1</sub> receptors are found in the eye, pituitary, adrenal medulla, pancreas, liver, ovary, lung, gut, and lymphoid tissue. The VPAC<sub>1</sub> receptors are in the eye, adrenal gland, pancreas, liver, testis, lung, gut, and lymphoid tissue. The VPAC<sub>2</sub> receptors are in similar

locations as well as in the pituitary and ovary. There is less overlap among receptor location than the list suggests as the receptors may be expressed only in specific cells within the organs and at specific times.

The functions of PACAP are diverse, but the physiological stimulus that releases PACAP from nerve endings or cells has not been elucidated. Regardless of the control of PACAP, this hormone has a number of functions. PACAP stimulates the release of other hormones; it releases insulin and glucagon from the pancreas, catecholamines from the adrenal medulla, glucocorticoids from the adrenal cortex in some species, and growth hormone from the pituitary gland. PACAP acts on blood vessels to cause vasorelaxation. In the developing nervous system, PACAP acts to alter proliferation and differentiation. PACAP's actions have been examined in mice in which either the PACAP gene or the receptor (PAC<sub>1</sub>) gene has been knocked out by targeted disruption of the gene. These studies show that PACAP plays a role in behavior and in lipid and carbohydrate metabolism, including the response to metabolic stress, such as an insulin challenge. One hypothesis to explain the many actions of PACAP is that it is used for responses to environmental or metabolic stress.

In regard to disease, PACAP is overexpressed in a number of tumors including gliomas, neuroblastomas,

and pheochromocytomas and in tumors of the pancreas, ovary, and pituitary. However, solid evidence is lacking as to symptoms associated with alterations in the human genes encoding PACAP or its receptors.

## VIP

Vasoactive intestinal peptide was named because it was isolated from the intestine, but later it was discovered that VIP is widely distributed in the central and peripheral nervous systems. VIP is closely related to PACAP in peptide structure, gene organization, distribution, and shared receptors. VIP is a 28-amino-acid peptide that is 70% identical at the amino acid level with PACAP<sub>1-27</sub> in humans (Fig. 1). Only one gene for VIP is present in humans and it encodes a large precursor containing both PHM and VIP (Fig. 2). The two hormones are on separate exons (exons 4 and 5). Although the PHM-VIP gene has seven exons, it is very close in structure to the PACAP gene in the first five exons. The final two exons in VIP encode only a nonfunctional peptide and the 3'-untranslated region.

Peripheral nerves with VIP are distributed to the intestine, lungs, pancreas, and adrenal. Also, VIP has been found in gonadal and immune cells. A specific receptor for VIP has not been reported, but two different receptors are shared with PACAP (VPAC<sub>1</sub> and VPAC<sub>2</sub>). Both receptors are seven-transmembrane G protein-coupled receptors. These two receptors are widespread, as discussed above for PACAP. A variety of functions are described for VIP. In the brain, VIP is described as part of a "visceral forebrain system" that can superimpose control over the brainstem center for cardiovascular, respiratory, and gastrointestinal functions. Like PACAP, VIP is a vasorelaxant. VIP can regulate other hormones; VIP releases prolactin and at times growth hormone (GH) and luteinizing hormone from the pituitary and it releases insulin or glucagon from the pancreas depending on the glucose level in the blood. VIP has some effect on release of the catecholamines from the adrenal and indeed is in the same nerve terminals as PACAP. VIP is expressed early in the embryonic brain and is reported to increase growth and proliferation. Although inhibition of VIP in embryos resulted in microcephaly, this experiment could not be repeated by other researchers.

A connection with disease has been observed in that pancreatic and bronchial tumors known as VIPomas can result in a watery diarrhea syndrome. The symptoms are probably related to VIP's action on the exocrine pancreas and intestine to alter secretions and absorption of water and ions in the gut. These tumors

may be derived from neural tissue innervating the pancreas, intestine, or bronchial tubes, as the highest levels of VIP are found in neuroblastomas. In addition, VIP has been shown in animal models to prevent lung injury, improve survival in acute respiratory disease syndrome, and protect cardiac and neural tissue.

## PHM

Human PHM is a 27-amino-acid peptide encoded by exon 4 of the same gene that encodes VIP (Fig. 2). Cells express and release PHM and VIP together, so PHM expression has the same wide distribution as VIP in the gastrointestinal tract as well as in the central and peripheral nervous systems.

PHM acts as a less potent version of VIP, exerting similar biological effects, including the release of prolactin, insulin, and glucagon. PHM possibly acts through its ability to weakly bind to G protein-coupled VPAC<sub>1</sub> and VPAC<sub>2</sub> receptors. A specific PHM receptor, for which there is evidence based on binding studies, has not yet been isolated and characterized in humans. There are no disorders or diseases that are unique to PHM, although an aberrant profile of expression and amounts of PHM and VIP are seen in some disease states. VIP is of more interest for therapy because of its increased potency compared to PHM.

## GHRH

GHRH is well established as a hormone made in the brains of mammals. In humans, GHRH is a 44-amino-acid peptide (Fig. 1) that is released from nerve cells in brain and then binds to pituitary cells to release GH. In humans, GHRH is encoded by a single gene, with a GHRH-coding region that extends across exon 3 and part of exon 4 (Fig. 2). GHRH has a long evolutionary history in which GHRH shared a gene with PACAP until the appearance of mammals.

The distribution of GHRH is primarily in the brain, although it is reported to be present in the gut, gonads, placenta, pancreas and immune cells. There are both long and short versions of the GHRH receptor, which is the usual seven-transmembrane G protein-coupled receptor. The target organs for GHRH are primarily the pituitary and brain, although low levels of receptors are reported for the gonads, placenta, and kidney. The function of GHRH is to increase the synthesis and release of GH from somatotrophs in the pituitary, but also to stimulate these somatotroph cells for proliferation, differentiation, and growth. Centrally, GHRH is implicated in

the control of food intake and enhancement of sleep. In the periphery, GHRH is reported to release insulin and to have a role in the gut, in reproduction, and in fetal growth.

Rare tumors that overexpress GHRH can occur in humans, especially in the pancreas, gastrointestinal tract, and lung. The symptoms in adults are typical of acromegaly with enlarged hands, feet, and jaw. Two such tumors were used for the first structural determination of human GHRH.

## GLUCAGON

The glucagon gene encodes a precursor with a number of peptides, three of which are bioactive (Fig. 2). Differential cleavage of this precursor results in glucagon production in the pancreas, but production of GLP-1 and GLP-2 in the intestine. In the pancreas, alpha cells use a proteolytic cell-specific enzyme, prohormone convertase, to produce glucagon (active), glicentin-related pancreatic peptide (inactive), and the major proglucagon fragment containing uncleaved GLP-1 and GLP-2 (inactive). In contrast, cells of the terminal small intestine and large intestine cleave proglucagon into glicentin (inactive), GLP-1 (active), and GLP-2 (active). Mature glucagon is a 29-amino-acid peptide, which includes an N-terminal histidine that is vital for biological activity.

The glucagon receptor is a G protein-coupled receptor that stimulates adenylate cyclase, resulting in increased intracellular cAMP and increased protein kinase A activity downstream. It has homology with the receptors for the other PACAP/glucagon superfamily members. Glucagon receptor mRNA is also present outside the liver in kidney, heart, stomach, pancreatic islets, spleen, thymus, adrenal, and skeletal muscle, although the functions in these tissues are not yet known.

Glucagon controls glucose metabolism by activating gluconeogenesis and glycogenolysis in the liver, effectively counteracting the action of insulin. Also, glucagon is synthesized in the CNS where it is thought to play a role in the regulation of peripheral glucose levels.

The glucagon pathway is both a contributor to the symptoms of diabetes and a candidate for the treatment of diabetes. Diabetes is a disorder involving reduced or absent insulin production and secretion. Glucagon increases to excess in proportion to a lack of insulin because transcription of the glucagon gene is normally inhibited by insulin. Excess glucagon results in hyperglycemia. In diabetes, glucagon favors glucose

production in the liver, indirectly compounding the effect of low insulin in ketone production. The glucagon receptor is one target of therapy for diabetic patients with excessive glucose production. Glucagon receptor antagonists represent an avenue to decrease liver glucose production and lower the blood glucose of diabetic patients.

## GLP-1

The full-length N-terminal-extended forms of GLP-1 (1-37 and 1-36 amide) were initially found to have little biological activity. Removal of the first six amino acids was necessary for full biological activity. Cells in the intestine produce the two biologically active forms (GLP-1<sub>7-36</sub> and GLP-1<sub>7-37</sub>), both of equal potency, but the major circulating form is GLP-1<sub>7-36</sub>. The enzyme dipeptidyl-peptidase IV cleaves the two N-terminal amino acids from the active peptide, rendering a peptide (GLP-1<sub>9-36</sub>) with little or no biological activity. The GLP-1 receptor is expressed in the gastrointestinal tract, pancreatic islets, kidney, lungs, heart, CNS, and possibly adipose tissue.

Major actions of GLP-1 are reported in the pancreas and the central nervous system. GLP-1 is released from intestinal cells in response to nutrients in the gut and then enhances the synthesis and release of insulin. At the same time, GLP-1 suppresses gastric emptying and directly or indirectly suppresses secretion of glucagon. GLP-1 helps to clear glucose following a glucose challenge. Another important action of GLP-1 is its ability to stimulate islet cell proliferation and growth as shown in mice and cell line model studies.

In the central nervous system, GLP-1 is reported to suppress food intake. Although it is not yet understood if the mechanism involved is direct or indirect, brain GLP-1 is not essential for physiological control of nutrient intake and body weight regulation *in vivo*. Another role for the GLP-1 is in the stress response. For example, mice with targeted disruption of the GLP-1 receptor gene have exaggerated levels of corticosterone in stress tests. The expression of GLP-1 receptors in extrapancreatic tissues suggests that GLP-1 may have other actions. In rats, intravenous injection of GLP-1 increases heart rate and blood pressure.

GLP-1 is a candidate in the treatment of diabetes. GLP-1 lowers blood glucose as well as appetite in diabetic patients, although the long-term therapeutic potential of GLP-1 in treating diabetes is still unknown. Analogues of GLP-1 resistant to degradation

in the body will likely meet the criteria for a therapeutic agent.

### GLP-2

GLP-2 is 33 amino acids long and is located following GLP-1 on their common gene (Fig. 2). Tissue-specific posttranslational processing results in the production of GLP-2 in the intestine and brain but not the pancreas. The active circulating form of GLP-2 is GLP-2<sub>1-33</sub>, whereas GLP-2<sub>3-33</sub> is inactive due to processing by dipeptidyl-peptidase IV cleavage. The GLP-2-specific receptor is a G protein-coupled receptor expressed in the stomach, small bowel, and colon, as well as the hypothalamus.

GLP-2, like GLP-1, is involved in nutrient assimilation and energy homeostasis. However, an increasing number of functions unique to GLP-2 are being uncovered. GLP-2 inhibits gastric secretion and affects intestinal permeability. Acute GLP-2 infusion in rats results in increased intestinal hexose transport. Larger intestinal surface area results from the growth-stimulating properties of GLP-2 in the intestine. GLP-2 stimulates the proliferation of crypt cells, resulting in villi of greater height. Apoptosis of enterocytes following intestinal injury appears to be inhibited by GLP-2, which has led to further study for a direct role for GLP-2 receptor activation in reducing certain cell death events. The bioactive properties of GLP-2 related to the protection and restoration of intestinal cells are being tested for their potential in treating human gastrointestinal diseases.

### GIP

GIP is a 42-amino-acid peptide hormone encoded by a single gene. GIP mRNA is expressed only in the gut, specifically in the K cells, the majority of which are located in the proximal duodenum. GIP is secreted from these cells on stimulation by nutrients in the gut. GIP, along with GLP-1 and GLP-2, is inactivated by the enzyme dipeptidyl-peptidase IV, which cleaves the two N-terminal amino acids from the peptide to produce inactive GIP<sub>3-42</sub>. Mice that lack the ability to produce dipeptidyl-peptidase IV have significantly increased levels of GIP. GIP has a specific 59 kDa G protein-coupled receptor. Unlike its ligand, GIP receptor mRNA is widely distributed in peripheral organs including the pancreas, gut, heart, brain, adipose tissue, and adrenal cortex.

GIP was initially discovered and named for its ability to inhibit gastric acid secretion. Later it was

found that at physiological levels, GIP stimulates insulin synthesis and release in rodents and humans. The function of GIP as an insulinotropic agent stimulating insulin synthesis and release depends on an increase in blood glucose levels, resulting in the name change to glucose-dependent insulinotropic polypeptide to reflect these findings. Also, GIP appears to compensate for lost GLP-1-regulated insulin release in studies where the GLP-1 receptor gene has been disrupted. GIP may have a more restricted role in glucose homeostasis than GLP-1, as GIP has been shown to regulate glucose absorption and glycemic excursions only following enteral glucose challenge.

### SECRETIN

Secretin has retained a basic 27 amino acid length. The structure of the secretin gene is different in many ways from the genes for other PACAP/glucagon superfamily members. Most notably, it contains only four exons, with exon 1 serving for both the transcription and translation start sites. These and other properties may reflect the long period of time since the secretin gene diverged from the other PACAP/glucagon superfamily members.

Secretin is synthesized in endocrine S cells of the small intestine and in beta cells of the developing pancreas. There are inconsistent reports of secretin mRNA expression in other tissues including brain, heart, lung, kidney, and testis. Secretin mRNA and secretin protein reach their highest levels in the developing intestine 2 days before birth, whereas they peak in the developing pancreas at day 19, but are not detectable in the adult pancreas. The synthesis and secretion of secretin in the fetus are likely under regulation by different mechanisms, as the stimuli from nutrients are not yet a factor compared to postnatal animals.

A secretin-specific receptor has been cloned and characterized in humans. There are multiple subtypes of this G protein-coupled receptor. Secretin receptor mRNA is expressed in the gut and related organs and in the spermatids of the testis, although it has not been detected in brain.

Many actions have been assigned to secretin since its discovery. The well-known functions of secretin at physiological doses are actions on gut, liver, and pancreas. Secretin, which is released in the small intestine in response to gastric acid and fat, acts on the pancreas to release water and a bicarbonate-rich fluid to neutralize the acidity. Also, secretin stimulates gastric pepsin secretion and inhibits gastric acid secretion and

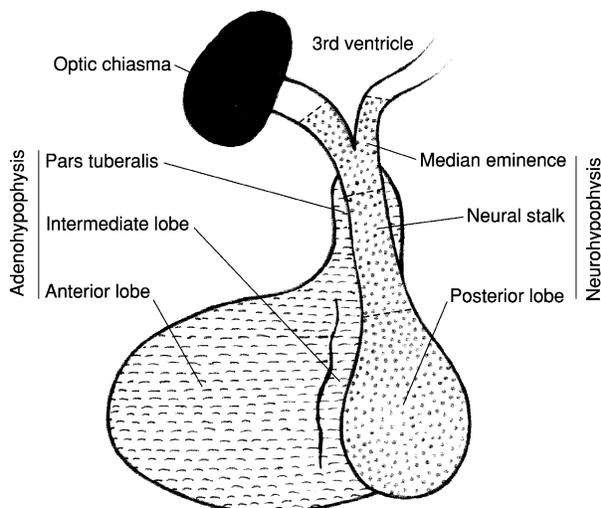
gut motility. Secretin has been shown to stimulate the release of somatostatin in the small intestine. In PC12 cells, secretin is able to stimulate tyrosine hydroxylase, an enzyme involved in catecholamine synthesis. Injection of physiological amounts of secretin in adults has coincided with the release of insulin and improved glucose tolerance. A specific disease has not been directly connected with a problem in secretin production.

### See Also the Following Articles

GI Hormone Development (Families and Phylogeny) • GI Hormones Outside the Gut: Central and Peripheral Nervous System • Glucagon and Glucagon-like Peptides • Peptide Neurotransmitters and Smooth Muscle in the Gut

### Further Reading

- Arimura, A. (1998). Perspectives on pituitary adenylate cyclase activating polypeptide (PACAP) in the neuroendocrine, endocrine, and nervous systems. *Jpn. J. Physiol.* **48**, 301–331.
- Drucker, D. J. (1999). The glucagon-like peptides. *Endocrinology* **142**, 521–527.
- Drucker, D. J. (2001). Glucagon-like peptide 2. *J. Clin. Endocrinol. Metab.* **86**, 1759–1764.
- Gray, S. L., Cummings, K. J., Jirik, F. R., and Sherwood, N. M. (2001). Targeted disruption of the pituitary adenylate cyclase-activating polypeptide gene results in early postnatal death associated with dysfunction of lipid and carbohydrate metabolism. *Mol. Endocrinol.* **15**, 1739–1747.
- Larsen, P. R., Kronenberg, H. M., Melmed, S., and Polonsky, K. S. (2003). “Williams Textbook of Endocrinology,” 9th ed. W. B. Saunders, Philadelphia, PA.
- Said, S. I. (2000). Protection by VIP and related peptides against cell death and tissue injury. *Ann. N. Y. Acad. Sci.* **921**, 264–274.
- Sherwood, N. M., Krueckl, S. L., and McRory, J. E. (2000). The origin and function of the pituitary adenylate cyclase-activating polypeptide (PACAP)/glucagon superfamily. *Endocr. Rev.* **21**, 619–670.
- Shioda, S. (2000). Pituitary adenylate cyclase-activating polypeptide (PACAP) and its receptors in the brain. *Acta Anat. Nippon.* **75**, 487–507.
- Strand, F. L. (1999). Gut and brain neuropeptides. I. Neuropeptide families: Gastrin/CCK; VIP/secretin/glucagon; PP/PYY/NPY; BN/GRP. In “Neuropeptides: Regulators of Physiological Processes,” pp. 383–429. Massachusetts Institute of Technology Press, Cambridge, MA.
- Vaudry, D., Gonzalez, B. J., Basille, M., Yon, L., Fournier, A., and Vaudry, H. (2000). Pituitary adenylate cyclase-activating polypeptide and its receptors: From structure to functions. *Pharmacol. Rev.* **52**, 269–324.



**Figure 1** Schematic of the divisions and subdivisions of the pituitary as seen in a midsagittal section of the gland.

pituitary. The optic chiasma lies directly above the sellar diaphragm ahead of the hypophysial stalk. Growth of a pituitary tumor may compress the chiasma and impair vision. The tuber cinereum of the hypothalamus also lies above the roof of the sella. Space-occupying lesions in the pituitary may compress and compromise the tuber cinereum and cause hypothalamo-hypophysial dysfunctions. The lateral walls of the sella are close to the cavernous sinuses containing the internal carotid arteries and several nerves including the oculomotor, trochlear, and abducens nerves and the first two branches of the trigeminal nerve. The sphenoid sinus, separated from the sella by a thin layer of bone, is inferior to the

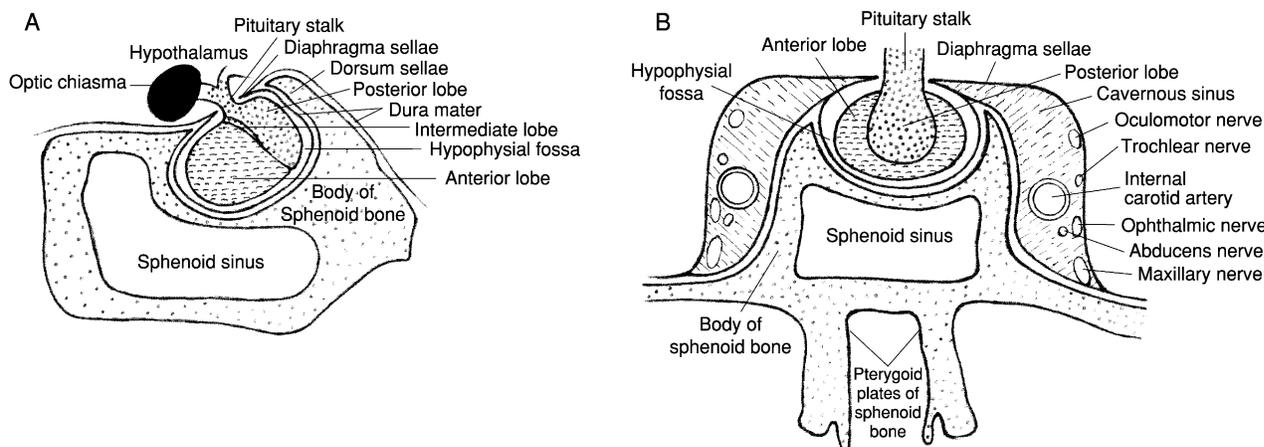
pituitary. In the case of a pituitary tumor, this bone may be resorbed and eroded, leading to tumor penetration into the sinus.

### Blood Supply

The blood supply of the pituitary derives from two groups of arteries: from above, the right and left superior hypophysial arteries; from below, the right and left inferior hypophysial arteries. Both groups of arteries arise from the internal carotid arteries. The superior hypophysial arteries supply blood to the median eminence and proximal portion of the pituitary stalk. Here these arteries break up into a primary capillary plexus. The capillaries of this plexus rejoin to form the long portal vessels that traverse the pituitary stalk and break up into a secondary capillary plexus in the anterior lobe of the pituitary in close relationship to the cells of anterior pituitary. This hypophysial portal vascular system is of the utmost importance in regulating hormone secretion of the anterior pituitary. The posterior pituitary receives its blood supply from the inferior hypophysial arteries. Some vessels from the posterior lobe penetrate into the anterior pituitary (short portal vessels).

Venous blood leaves the pituitary through dural channels and enters the cavernous sinuses, which drain to the inferior petrosal sinus and the internal jugular vein. There may be anastomoses between the petrosal sinuses, which can lead to confusing observations with petrosal sinus sampling.

Although most blood flow is from the hypothalamus to the pituitary, there is evidence that some blood may flow in the opposite direction, from the anterior pituitary to the hypothalamus.



**Figure 2** Schematic of the location and topography of the pituitary in a midsagittal section (A) and a coronal section (B) of the sphenoid bone.

**Innervation**

The anterior lobe is poorly innervated. A few nerve fibers reach the anterior pituitary along the blood vessels. These nerve fibers are not believed to have major importance in the control of hormone secretion. The posterior pituitary contains the axon terminals of the supraoptic and paraventricular neurons, synthesizing and releasing oxytocin and vasopressin. In addition, the intermediate pituitary and the posterior pituitary are innervated by hypothalamic dopaminergic neurons of the periventriculo-hypophysial and the tuberohypophysial dopaminergic system.

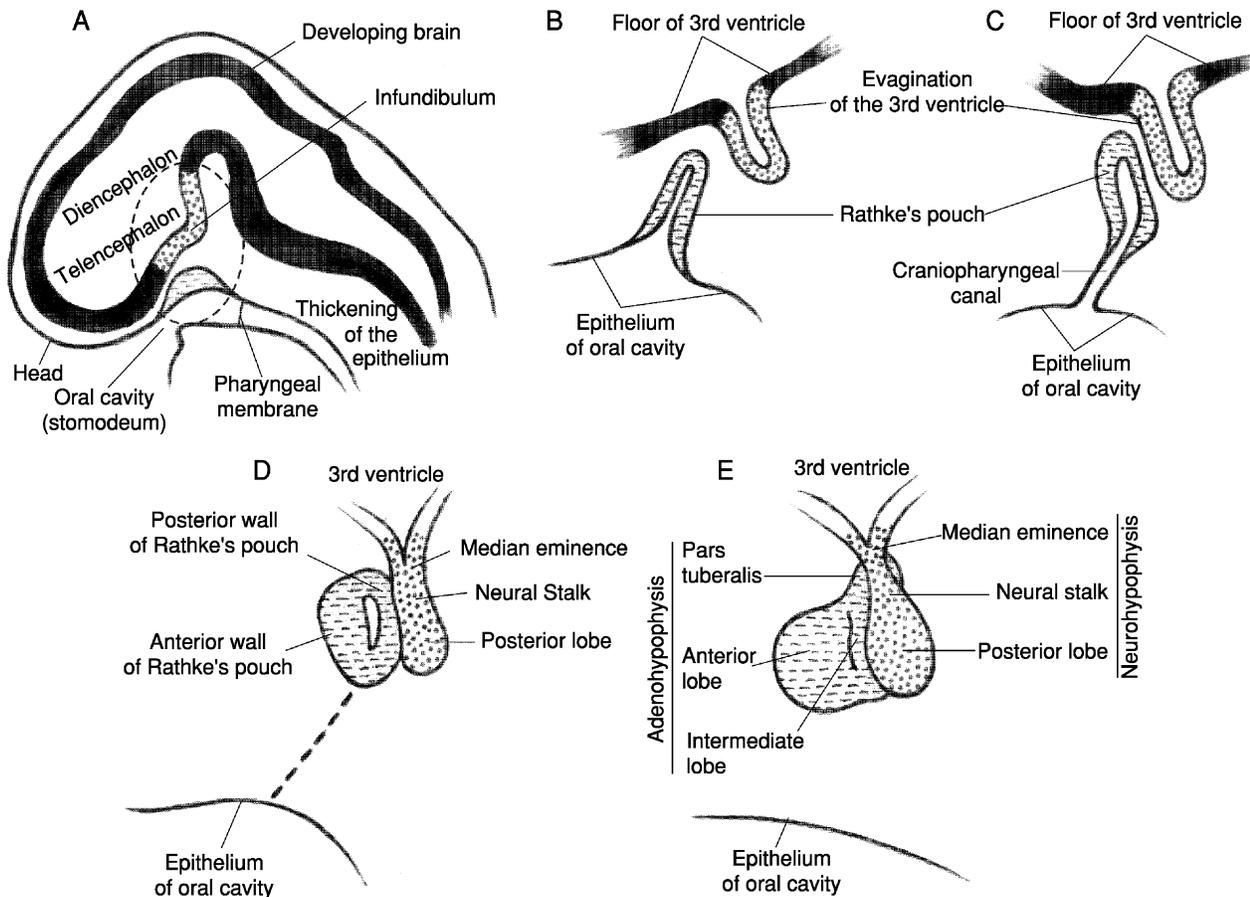
**EMBRYOLOGY**

The adenohypophysis originates from an evagination of the epithelium covering the vault of the stomodeum (primary oral cavity). The neurohypophysis develops as a process growing downward from the floor

of the diencephalon, i.e., as an evagination of the third ventricle of the brain (Fig. 3). During embryonic development, the adenohypophysial primordium becomes located anterior to the neural primordium.

**Development of the Adenohypophysis**

The epithelium of the stomodeum (primitive oral cavity) becomes thicker just ahead of the pharyngeal membrane (Fig. 3A). This flat primordium invaginates and penetrates the connective tissue in the direction of the diencephalon, forming a diverticulum, Rathke's pouch (Fig. 3B). Initially, Rathke's pouch is composed of a small, thin-walled vesicle in the roof of the primitive oral cavity. It subsequently expands in the direction of the evagination of the third ventricle and will be situated just ahead of the evagination (Fig. 3C) and then it adheres to it. Rathke's pouch is attached to the stomodeal vault by a stalk (cranio-pharyngeal canal), which regresses, is obliterated, and



**Figure 3** Schematics of the development of the pituitary as seen in midsagittal sections. The drawings show the various stages of development: (A) very early stage; (B, C) intermediate stage; (D) nearly developed stage; and (E) completely developed stage.

usually disappears (Fig. 3D). Parts of it, however, may persist in a more or less differentiated and sometimes functional condition. Nests of adenohypophysial tissue may be deposited along the route of the cranio-pharyngeal canal. Remnants of the pharyngeal hypophysis may be capable of hormone synthesis and may give rise to ectopic adenomas. Occasionally, cranio-pharyngiomas (tumors) develop in the pharynx or in the sphenoid bone of the skull, but most often they form in or above the sella turcica of the sphenoid bone at the base of the skull.

Just behind the pharyngeal membrane, the entodermal epithelium also forms a pouch, the pouch of Sessel. This structure is involved in the formation of the adenohypophysis in lower vertebrates. However, this involvement decreases as the evolutionary scale increases and disappears completely in primates and human. In human, it sometimes persists and may cause certain tumors.

#### **Development of the Anterior Lobe**

The anterior lobe develops from the anterior wall of Rathke's pouch. Cells of this wall proliferate actively and give rise to the anterior lobe of the pituitary gland (Fig. 3D). Proliferation occurs in such a way that a small basin is formed, separated into two compartments by a median cellular septum. The compartments disappear progressively, due to the growth of the wall. The median septum forms the pars medialis and the lateral portions form the lateral part of the anterior lobe. The extensive proliferation of the anterior wall of Rathke's pouch reduces the lumen to a narrow residual cleft (Fig. 3E). It is usually not recognizable in the adult gland and is represented by a zone of cysts.

Acidophilic cells are detectable in the anterior lobe of the human embryo at approximately the third month of gestation; basophilic cells can be detected somewhat later. Adrenocorticotrophic hormone (ACTH)- and growth hormone (GH)-synthesizing cells are identifiable by the end of the second month of gestation. This is followed by the production of glycoprotein hormones. Blood vessels grow into the anterior lobe and establish a direct neurovascular link between the anterior lobe and the hypothalamus at approximately the eighth week of gestation. The mammothroph cell type (prolactin-producing) appears late, at approximately the fifth month of gestation. Hormone-producing cells differentiate in the absence of a hypothalamic influence. In anencephaly, all cell types of the anterior lobe except corticotrophs develop and are capable, to some extent, of hormone synthesis and release.

#### **Development of the Intermediate Lobe**

The posterior wall of Rathke's pouch gives rise to the intermediate lobe of the adenohypophysis (Figs. 3D and 3E). In humans, cells of the posterior wall of Rathke's pouch do not proliferate; instead, they form the poorly defined intermediate lobe, which becomes an inconspicuous, discontinuous layer.

#### **Development of the Pars Tuberalis**

The pars tuberalis of the adenohypophysis develops from the anterior wall of Rathke's pouch. The cells of the median septum of the anterior wall of Rathke's pouch proliferate upward along the pituitary stalk, which becomes gradually encircled by the cellular expansion of the developing pars tuberalis (Fig. 3E).

#### **Development of the Neurohypophysis**

The neurohypophysis develops from the evagination of the wall of the third ventricle (Fig. 3). The floor of the third ventricle becomes depressed and produces the infundibulum (Fig. 3A). This depression penetrates progressively toward the adenohypophysial primordium. Its ventral end forms a diverticulum (Figs. 3B and 3C). The wall of the diverticulum thickens and its lumen gradually fills, forming the neural or posterior lobe, which is attached to the posterior wall of Rathke's pouch (Fig. 3D). It remains to be connected to the diencephalon part of the brain by the thin neural stalk. The ascending part of the infundibulum is the median eminence. The median eminence, the neural stalk, and the posterior lobe (or neural lobe or infundibular process) form the neurohypophysis (Fig. 3E). The neural lobe differentiates and specific neuroglial cells, called pituicytes, appear in it. The neurohypophysis is then colonized by axons coming from the hypothalamic paraventricular and supraoptic nuclei. In the human, neurosecretory material, characteristic of the posterior pituitary, is demonstrable at approximately the fifth month of gestation.

## **HISTOLOGY**

### **Adenohypophysis**

#### **Anterior Lobe (or Anterior Pituitary or Pars Distalis)**

The anterior lobe accounts for approximately 75–80% of the whole gland. It is highly vascular and contains various cell types. The lobe is largely enclosed by a dense collagenous capsule and is composed of glandular cells arranged in irregular cords or clumps, which are intimately related to an extensive system of dilated capillaries (sinusoids) of the blood

vascular system. Reticular fibers surround the cords of parenchymal cells and are also present close to the wall of the sinusoids.

The glandular cells were originally classified as chromophilic and chromophobic on the basis of their avidity or lack of affinity for the dyes used in routine staining of histological sections. The chromophilic cells were subdivided into acidophilic and basophilic cells on the basis of staining with combinations of an acidic dye and a basic dye. In the human pituitary, acidophilic cells are most numerous in the posterolateral portions of the anterior lobe. They are rounded, small cells with a well-developed Golgi complex and rod-shaped mitochondria. The basophilic cells are larger and less numerous.

The most meaningful method applied at both the light and electron microscopic levels for identification of the cells producing various hormones involves the use of immunocytochemical procedures. Antibodies to a specific hormone induced in another species are conjugated with horseradish peroxidase or with fluorescence dyes. These labeled antibodies are reacted with sections of the anterior lobe and the sites of the antigen in the tissue are localized by the histochemical method for peroxidase or by fluorescence microscopy.

The anterior pituitary produces six hormones: GH (or somatotropin), prolactin, thyrotropin (TSH), ACTH, and two gonadotropins—follicle-stimulating hormone (FSH), and luteinizing hormone (LH). The acidophilic cells secrete simple proteins (somatotropin and prolactin) and the basophils secrete glycoprotein hormones (TSH, FSH, LH, and the precursor of ACTH). It has become a common practice to use the terms of the hormone secreted (ACTH cell, TSH cell, FSH/LH cell) or the name of the target organ stimulated (corticotroph, thyrotroph, gonadotroph) to denote the cell type. In addition to these cells, some other cell types, such as folliculostellate cells and null cells, are also found in the pituitary.

#### *Somatotrophs*

Somatotrophs represent approximately 50% of the anterior lobe cells. They are found in groups along the sinusoids and are located mainly in the two lateral wings of the anterior lobe. The cells are usually medium-sized and contain numerous spherical, evenly electron-dense granules, the majority being 350 to 500 nm in diameter. They have a well-developed endoplasmic reticulum and secrete GH. Immunocytochemical observations strongly suggest that the somatotrophs are not a uniform mass of cells. It appears that they consist of several subpopulations, all of which express somatotropin, but are also capable

of producing other hormones, such as prolactin and TSH, in special circumstances. There is evidence that transdifferentiation of GH-/prolactin-synthesizing cells (somatomammotrophs) may contribute to the mass of prolactin-producing cells during development of prolactin cell hyperplasia in human pregnancy and further, that somatotrophs may transform into large thyrotroph cells in rats during experimental hypothyroidism.

#### *Mammotrophs*

Mammotrophs (or lactotrophs or prolactin cells) produce prolactin and constitute 10 to 25% of the anterior lobe cells. They are relatively small cells, are distributed individually in the interior of the cell cords, and are located throughout the entire anterior lobe. The cells are especially numerous at the posterolateral and posteromedial edges of the adenohypophysis. Two types of lactotrophs can be distinguished. The majority of the cells are small or medium in size and sparsely granulated. The secretory granules are spherical and evenly electron-dense, measuring 150 to 350 nm in diameter. The other, less frequently occurring type is densely granulated and the granules are larger (300 to 600 nm). The number of prolactin cells varies considerably. In pregnancy and lactation, lactotrophs are increased in number. During pregnancy, the cells undergo considerable hypertrophy: their Golgi complex enlarges and the endoplasmic reticulum becomes more extensive; multiple layers parallel to the cell membrane develop and the granules become larger (550 to 600 nm in diameter) and often irregular in outline. The mammotrophs are most active during lactation. After weaning has occurred, lysosomes play an important role in the elimination of excess secretory granules and the hypertrophied cellular organelles involved in the lactation period of active protein synthesis. Lysosomes fuse with the secretory granules to form autophagic vacuoles. The secretory granules are degraded by hydrolytic enzymes in these vacuoles. This procedure of disposal of secretory product that is no longer needed is called crinophagy. Excess cytomembranes and ribosomes are also enclosed in vacuoles and degraded by autophagy.

#### *Thyrotrophs*

Thyrotrophs, secreting TSH, are located mainly in the anteromedial part of the anterior lobe and constitute less than 10% of the lobe. The cells are medium or large in size and polygonal with long cytoplasmic processes. In electron micrographs, they are characterized by spherical secretory granules, are variably electron-dense, measure 100 to 200 nm in diameter,

and often line up along the cell membrane, with short stacks of rough endoplasmic reticulum membranes and a Golgi complex with flattened sacculi and several vesicles. In hypothyroidism, thyrotrophs increase in size and number and transform into so-called thyroidectomy cells or thyroid-deficiency cells. These cells are large and contain widely dilated endoplasmic reticulum membranes, conspicuous Golgi complexes, and a varying number of secretory granules.

#### **Gonadotrophs**

Gonadotrophs synthesize both gonadotropic hormones, FSH and LH. These cells are larger than other cells of the anterior lobe, constitute approximately 15 to 20% of the lobe, and are located throughout the anterior pituitary. They are near capillaries and often in close proximity to mammotrophs, suggesting the possibility of paracrine action between the two cell types. FSH and LH are located mostly in the cytoplasm of the same cells, but there are also gonadotrophs that show only FSH or LH positivity. In electron micrographs, the rough endoplasmic reticulum is prominent and forms slightly dilated stacks; the Golgi complexes are conspicuous. The secretory granules represent two populations: the diameter of one granule type is 150 to 250 nm and that of the other type is 350 to 450 nm. Following castration, the size and number of gonadotrophs increase and the cells show enlargement of the cytoplasm, proliferation, and dilation of the endoplasmic reticulum membranes.

#### **Corticotrophs**

Corticotrophs (or corticotropin- or ACTH-producing cells) are located mainly in the central part of the pituitary. Some cells are scattered in the lateral wings. They represent 10 to 15% of the anterior lobe cells. Corticotrophs are medium-sized or large oval cells and show ACTH,  $\beta$ -lipotropin, and  $\beta$ -endorphin immunoreactivity (proopiomelanocortin-derived peptides). In electron micrographs, they have well-developed rough endoplasmic reticulum membranes and are usually numerous, spherical, irregular secretory granules measuring 250 to 400 nm, showing varying electron density. The granules tend to be located adjacent to the cell membrane. In addition, the presence of type 1 filaments, which are bundles of filaments located mainly in the perinuclear area, are a characteristic feature of corticotrophs.

#### **Folliculostellate Cells**

Folliculostellate cells, called also follicular cells, are mostly agranular cells with branching processes

among the secretory cells. There are data indicating that they are derived from granulated cells after having been joined by junctional complexes around damaged and ruptured adenohypophysial cells and are capable of forming follicles. Folliculostellate cells produce several substances (interleukin G, follistatin, etc.). It is assumed that these cells play an important role in the paracrine mechanisms controlling pituitary functions.

#### **Null Cells**

Null cells are relatively small, chromophobic cells that do not contain known anterior pituitary hormones in their cytoplasm. The cells have all the cytoplasmic organelles necessary for secretion. They contain secretory granules that may contain hormone fragments, precursors, or biologically inactive substances. They may represent resting cells, precursors of various cell types, or an unknown cell type.

It should be mentioned that the majority of what were originally called chromophobic cells contain specific granules with hormone content. The cells classified as chromophobes by light microscopy are not a homogenous population. Some are evidently chromophils agranulated to the point at which their specific nature is not detectable. There seems to be a considerable degree of cytological specialization among the cells normally classified as chromophobes. It is probable that many of the apparent chromophobes have already been determined and are capable of differentiating into only one of the chromophil types.

#### **Intermediate Lobe**

There is considerable variation among species in the degree of development of the intermediate lobe. It is poorly developed in the human. It contains a few dilated cavities lined by a single layer of cuboidal or columnar epithelial cells and is filled with an amorphous proteinaceous material. Many cells lining the cystic cavities and cells between the cysts give a positive immunostaining for proopiomelanocortin-derived peptides. These cells are smaller than the corticotrophs in the anterior lobe.

#### **Pars Tuberalis**

The pars tuberalis is an upward extension of the anterior lobe and is attached to the neural stalk. It contains small groups of cells that produce mainly glycoprotein hormones, gonadotropins and thyrotropin. The functional significance of the pars tuberalis is not known.

## Neurohypophysis

The human neurohypophysis consists of approximately 10,000 axons and terminals of neurosecretory cells located in the hypothalamic supraoptic and paraventricular nuclei. The unmyelinated fibers of the neurosecretory cells descend, converge, and form the supraoptico- and paraventriculohypophysial tract or pathway. These fibers descend through the median eminence and neural (or infundibular) stalk into the neural lobe of the hypophysis and make up the bulk of the substance of the lobe.

### **Median Eminence and Infundibular Stalk**

The median eminence is made up of two layers: the external layer and the internal layer. Not the capillary loops but the hypophysiotrophic neurons whose axon terminals are in the external layer are producing the neurohormones. This layer is a key structure of the neurovascular contact between the hypothalamus and the anterior pituitary.

The internal layer of the median eminence and the infundibular stalk contain the descending unmyelinated axons of the supraoptico- and paraventriculohypophysial tract, terminating in the posterior pituitary, transporting the neurosecretory material from the hypothalamic neurons to the posterior lobe at a rate of 1 to 4 mm/h. In histological preparations stained with chrom-alum hematoxylin, deeply stained neurosecretory material is seen in aggregations of varying sizes throughout the infundibular stalk and neural lobe. These aggregates are called Herring bodies. In electron micrographs, they are found to be large aggregations of the small secretory granules. The axons of these neurosecretory neurons vary greatly in caliber and have numerous dilations along their length. Approximately 60% of all neurosecretory material resides in these dilations and approximately 30% resides in axon endings in the posterior pituitary.

### **Posterior Lobe (or Neural Lobe, or Infundibular Process)**

The posterior pituitary is made up of axons and axon endings of the neurosecretory neurons of the hypothalamic supraoptic and paraventricular nuclei synthesizing oxytocin and vasopressin (called also antidiuretic hormone) and also contains an intrinsic population of cells called pituicytes, which resemble astrocytes and do not appear to be secretory. Axons of the neurosecretory neurons terminate blindly in close relation to the basal lamina of a rich capillary plexus.

The neurosecretory material is stored in granules at the dilated blind endings of the axons and are released as needed. The neurosecretory granules have a diameter of 100–200 nm, are surrounded by a membrane, and are more numerous apposed to fenestrated blood capillaries. The neurosecretory material consists of oxytocin or vasopressin and a binding protein (neurophysin) specific for each hormone. The hormone-neurophysin complex is synthesized as a single, long peptide on ribosomes attached to the membranes of the rough endoplasmic reticulum. The endoplasmic reticulum is a cell organelle from where the synthesized material is passed onto the Golgi complex. As the granules pass down axons of the supraoptico- and paraventriculo-hypophysial tract, proteolysis of the precursor occurs, yielding the hormone and its specific neurophysin. Vasopressin and oxytocin are stored in the posterior pituitary and are released into the blood by impulses in the nerve fibers from the hypothalamus.

In addition to the axons from hypothalamic nerve cells, approximately 25–30% of the volume of the posterior pituitary consists of a distinctive type of glial cell, called a pituicyte. Pituicytes are highly branched cells with processes that form a three-dimensional network ensheathing the neurosecretory axons. Their cytoplasmic processes meander among groups of preterminal secretory axons and often intimately envelop their granule-filled terminal expansions. In the human, they are highly variable in size and shape. The cytoplasm of the pituicytes may contain lipid droplets and pigment. The processes of pituicytes are connected by gap junctions. Pituicytes are believed to have a trophic and supportive function and to maintain the appropriate ionic composition of the extracellular fluid compartment.

## See Also the Following Articles

Growth Hormone (GH) • Hypopituitarism • Hypothalamic-Pituitary Unit • Hypothalamus–Pituitary–Thyroid Axis • Pituitary Gland, Evolution of • Pituitary Tumors, ACTH-Secreting • Pituitary Tumors, Clonality • Pituitary Tumors, Molecular Pathogenesis • Pituitary Tumors, Surgery

## Further Reading

Bhatnagar, A. S. (ed.) (1983). "The Anterior Pituitary Gland." Raven Press, New York.  
 Fawcett, D. W. (1986). "Bloom and Fawcett: A Textbook of Histology." W. B. Saunders, Philadelphia, PA.

- Horvath, E., and Kovacs, K. (1994). Morphology of adenohypophysial cells and pituitary adenomas. In "The Pituitary Gland" (H. Imura, ed.), 2nd ed., pp. 29–62. Raven Press, New York.
- Ikeda, H., Suzuki, J., and Sasano, N. (1988). The development and morphogenesis of the human pituitary gland. *Anat. Embryol.* **178**, 327–336.
- Moore, K. L. (1982). "The Developing Human: Clinically Oriented Embryology," 3rd ed. W. B. Saunders, Philadelphia, PA.
- Williams, P. L., Warwick, R., Dyson, M., and Bannister, L. H. (1989). "Gray's Anatomy," 37th ed. Churchill Livingstone, London/New York.

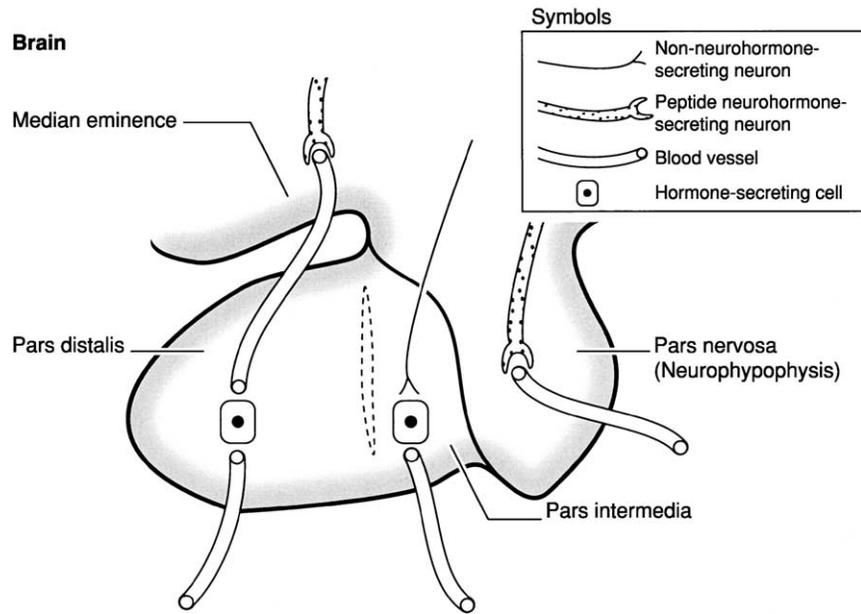


Figure 1 Diagrammatic representation of the relationship of the brain with different parts of the pituitary gland.

## PARS DISTALIS

### Evolution of the Brain–Pituitary Connection

The method of delivery of the brain neuropeptide hormones that regulate secretion of pars distalis hormones has evolved differently in different vertebrate groups. In the most primitive vertebrates, the Agnatha (hagfish and lampreys), the simplest relationship exists. The Agnatha have a broad, thin neurohypophysis in which the releasing hormone-bearing neurosecretory axons end. Applied to this is an equally broad and thin coextensive adenohypophysis (hagfish) or pars distalis (lampreys). Regulatory neuropeptides merely diffuse between the two.

In the bony fishes, the neurohypophysis is intimately interdigitated with the pars distalis. Secretory cells of the pars distalis are directly innervated or are near the endings of brain neurosecretory axons.

In all other vertebrates, the relationship is as illustrated in Fig. 1. The regulatory neurosecretory axons end on blood vessels in the adjacent brain in a region known as the median eminence. From this location, portal blood vessels extend into the pars distalis.

Dependence on diffusion in the simplest arrangement dictates that the pars distalis must be thin and may be broken up into follicles. Direct innervation and vascular relationships allow the pars distalis to be larger, more globular in shape, and more compact.

### Additional Lobes of the Pars Distalis

In several of the fish groups, there are lobes or lobules of glandular tissue in addition to the main structure of the pars distalis. In the selachians (e.g., sharks and rays), there is a so-called “ventral lobe,” which may remain attached to the pars distalis by a stalk and which is known to secrete gonadotropic and thyrotropic hormones.

The puzzle that these additional pars distalis structures present is that they appear to be outside of the brain’s control. What is the significance of a hormonally active pars distalis system that is not under brain control? This question and its evolutionary value need to be studied.

## INVERTEBRATE EVOLUTIONARY HOMOLOGUES OF THE PITUITARY OR HYPOTHALAMO–HYPOPHYSEAL SYSTEMS

All vertebrates have a complex pituitary gland system that is regulated by the brain. This must have required an extensive evolutionary history that preceded its appearance in the earliest vertebrates. The logical place to look for an anatomical precursor homologous system is in the protochordates, the organismal group most closely related to the vertebrates.

### The Cephalochordate Brain–Hatschek’s Pit System

The cephalochordates (amphioxi) present the best invertebrate homologous structure. It consists of two elements, a shallow pouch-like ingrowth from the mouth, Hatschek’s pit, and a lobe of the brain, the infundibulum. The infundibulum is a down-growth from the brain that extends around the notochord and makes near contact with Hatschek’s pit. The part of Hatschek’s pit that is so contacted contains a peptide that is immunoreactive with an antibody to vertebrate pituitary luteinizing hormone. The additional functional properties of this pituitary-like system require further study.

### The Neural Gland–Neural Ganglion System of Urochordates (Ascidians)

In the “anterior” region of ascidians, there is a concentration of nerve cells, the neural ganglion, sometimes referred to as the brain. It is in contact with a rounded, glandular-appearing structure, the neural gland. The neural gland opens to the mouth cavity through a short duct. The brain–pituitary homology of this system is strengthened by the presence of the vertebrate brain hormone gonadotropin-releasing hormone (GnRH) in the system. GnRH is found also in the network of nerves derived during development from the neural ganglion.

### The GnRH Cells of Hemichordates

The hemichordates are the least studied and the most primitive members of the chordate group. They are worm-like burrowing animals. Unlike the cephalochordates and urochordates, they have no organ structures reminiscent of the vertebrate pituitary system. However, they do have many integumentary nerve-like cells that contain numerous large granules in which there is GnRH. These mulberry cells are

oriented such that they appear to secrete GnRH externally. It has been suggested that in the hemichordates, the primitive function of GnRH is as a pheromone, a chemical released into the environment that affects other members of the same group. GnRH is believed to integrate in this way the simultaneous sexual ripening and shedding of gametes in whole populations of a species, thus ensuring success in fertilization.

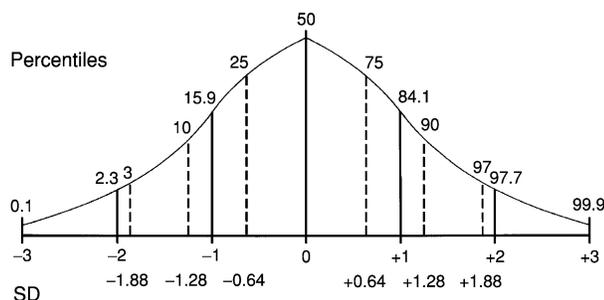
This illustrates the exciting possibility that the three protochordate groups provide a pattern of evolution of the vertebrate hypothalamo–hypophyseal system. That is, GnRH may have been at first a reproductive pheromone produced by the nervous system in response to environmental signals. It could have begun to have an internal function as well as pheromonal activity in the urochordates. Finally, GnRH from the nervous system could have assumed an endocrine role in cephalochordates by taking over control of an epithelial gonadotropin-secreting structure, such as Hatschek’s pit. Hatschek’s pit strongly resembles the vertebrate embryonic adenohypophyseal diverticulum of the mouth, not only in structure and position, but also in its content of gonadotropin. This is an attractive and satisfying evolutionary scenario. It only requires more study and more pertinent facts to either establish or discredit it.

### See Also the Following Articles

Gonadotropin-Releasing Hormone, Family of • Hypothalamic-Pituitary Unit • Hypothalamus–Pituitary–Thyroid Axis • Pineal Gland, Evolution of • Pituitary Gland Anatomy and Embryology

### Further Reading

- Gorbman, A. (1997). Amphioxus as a prevertebrate endocrine model. *In* “Advances in Comparative Endocrinology” (S. Kawashima and S. Kikuyama, eds.), pp. 227–228. Monduzzi, Bologna, Italy.
- Matsumoto, A., and Ishii, S. (1992). “Atlas of Endocrine Organs.” Springer Verlag, New York.
- Norris, D. O. (1997). “Vertebrate Endocrinology.” Academic Press, San Diego, CA.



**Figure 1** Relationship between heights expressed as percentiles and standard deviations of the mean.

scheduled approximately at the same hour (morning is recommended).

### Parental Height

Parental height is also essential to assess a child's stature; real measures are much better than self-reported heights. Target height can then be estimated by simply averaging the parents' height (adding 13 cm to the mother's stature for males or by subtracting these 13 cm from the father's height for females). The resulting "midparental height" represents the target height for the a given child within a  $\pm 10$  cm range; there is a 95% probability the child will reach that height.



**Figure 2** Correct handling of patient's head to obtain precise and repeatable measurements.

### Standard Deviation Score

When reporting a child stature simply as "below the 3rd centile" one does not provide any information on how far his or her stature deviates from normal; furthermore, in clinical studies one could not group patients as simply "below the  $x$ th centile" or directly compare different statures of patients of different heights. Therefore the best way of representing stature data is the worldwide adopted standard deviation score (SDS):

$$\text{SDS} = (x - \hat{Y}) / \text{SD}$$

where  $x$  is the child's height (in cm),  $\hat{Y}$  is the mean stature for age, and SD is the standard deviation of mean stature for age. With this simple method one can very easily appreciate changes in growth (either spontaneous or due to treatment). For example, when a child's stature changes from a score of  $-3.4$  to a score of  $-2.9$  in a 6 month period, this means that the child is gaining centimeters toward the normal range. With this method, patients with different ages and degrees of short stature can be grouped and their SDS treated statistically.

### Growth Velocity

The precise measurement of height is the first step of the assessment of growth pattern, which is an important parameter. In fact, the current stature could simply reflect a slowed growth during intrauterine life or early years that has not yet been recovered. A better source of information on growth is provided by assessing growth rate: changes in height plotted versus time. Considering the intrinsic error in measurements and the expected changes in growth for a given age, the minimum interval between two subsequent measures should be 3 months; however, the minimal length of time to extrapolate yearly growth velocity should be at least 6 months (the ideal would be an entire year of observation with four measurements at 3-month intervals). Height velocity, expressed in cm/year, can be estimated by plotting longitudinal measurements of stature on the appropriate growth chart. It varies with age and sex: in the first 6 months of life a boy can grow 15 cm (i.e., 30 cm/year); this gradually slows down to about 9 cm/year at 2 years, and to 5 cm/year of mean velocity at 10 years. Within the next 2–3 years the pubertal spurt re-accelerates to a maximal growth velocity of 12–14 cm/year, this however almost closes the epiphyseal cartilages, leaving only a few centimeters to be gained until growth is completed.

According to the mathematical model of growth elaborated by J. Karlberg, one can distinguish three different phases of growth: infancy, childhood, and puberty. They are integrated to accomplish a non-linear pattern; they also reflect different underlying biological mechanisms: during infancy, as well as during fetal life, nutritional and metabolic factors play the major role, while growth hormone (GH) comes into play mostly during the second phase; both GH and sex hormones interact during pubertal growth. Estrogens are particularly involved: even in males estradiol is the most active steroid acting on hypothalamic centers responsible for growth hormone secretory peaks that are particularly elevated in pubertal subjects.

## DIAGNOSTIC APPROACH TO SHORT CHILDREN

Once stature has been accurately determined and found to be pathological, and growth rate has also been recorded as slower than expected for age and sex, the diagnostic approach implies further steps. Physical examination could have already shown possible dysmorphic features such as those of Down's and Turner's syndromes, Noonan syndrome, etc.; hypothyroidism and rickets also exhibit typical body proportions. Maternal medical history is very important as far as ascertaining exposure to alcohol, smoking, or other toxins during pregnancy.

### Bone Age

Bone age reflects the biological maturation of the child. Different techniques exist to assess skeletal maturation; the most widely used is an X-ray of the left hand and wrist. The Greulich and Pyle method consists of matching the overall appearance of the bones to the standards shown on the atlas and consequently to assign the closest age. The Tanner–Whitehouse technique (known as TW2) is relatively more complex: it assigns a score to each of 20 bones of the wrist so that all contribute to the final calculation of bone age. Bone age is delayed in most growth disorders and in itself it has no diagnostic value, although it can be used together with chronological age and height to predict final height more accurately than by simple midparental stature. Equations that incorporate height velocity and the occurrence of menarche (for pubertal girls) can also be used for this purpose. Children with delayed bone maturation and short stature have a larger chance of recovering from growth defects (either spontaneously or following

therapy) than those in whom bone age and chronological age are similar.

## Growth Charts

The standard reference charts on which a child's height should be plotted vary from country to country; in Britain and other European countries, the Tanner and Whitehouse charts are commonly used, while in the United States the NCHS (National Center for Health Statistics) tables are more common. Although some of these charts are based on data collected 30 years ago, they are still appropriate. When considering children of different ethnic or racial origin, standards charts for these groups should be consulted whenever possible. Distance charts (i.e., stature plotted against age) can be either cross-sectional or longitudinal: the former are built from data collected in large samples of children of an entire age range at the same time, whereas the latter are based on measurements collected from smaller samples of the same children that have been followed for several years until complete bone maturation. If the purpose is simply to screen a population or to search for differences between two different populations the cross-sectional charts are suitable. On the contrary, if one wants to follow the growth pattern of a single patient, the longitudinal type charts are more appropriate.

It must be taken into account that growth velocity might be in the normal range in two of the most common causes of short stature: genetic short stature (also called familial short stature) and in constitutional delay of growth. Therefore, not only could stature per se be insufficient for a diagnosis, but in some cases even growth rate may not be enough for the complex assessment of growth failure.

## Clinical Evaluation

Children growing at a rate lower than the 5th centile for their age should undergo a careful overall medical investigation taking into consideration all possible causes starting from malnutrition (in turn secondary to malabsorption, celiac disease, etc.), lung and heart diseases, renal failure, etc. (Table I) reports the most frequently observed causes of growth failure; it should be stressed that the endocrine diseases represent only a small percentage of the affected patients. The endocrine assessment should consider, for instance, hypothyroidism, Cushing's disease, and GH deficiency in all its variants. Hypothyroidism can easily be ruled out by measuring FT4 and TSH plasma levels, while

**Table I** General Causes of Short Stature

---

**Primary growth defects** (intrinsic growth plate defects)

- Chondrodysplasias
- Chromosomal short stature
  - Turner's syndrome
  - Down's syndrome
  - 18 q deletions
- Genetic short stature
  - Familial short stature
- Intrauterine growth retardation
  - Intrinsic fetal abnormalities
  - Maternal abnormalities
  - Placental disorders

**Secondary growth defects** (low growth velocity)

- Severe malnutrition
- Renal diseases
- Chronic liver diseases
- Hematologic disorders
- Diabetes mellitus
- Chronic pulmonary diseases
- Cardiac diseases
- Gastrointestinal diseases (inflammatory and celiac disease)
- Malignancies and/or chemotherapy and/or radiotherapy
- Corticosteroid treatment
- Immunological diseases (AIDS)
- Vitamin D deficiency or resistance
- Endocrine diseases (GH deficiency and related disorders, Cushing's syndrome, hypothyroidism)

**Delayed growth**

- Constitutional delay of growth and puberty
- Malnutrition of moderate degree
- Mild underlying medical disease

**Final short stature due to early accelerated growth**

- Precocious puberty
- Virilizing forms of congenital adrenal hyperplasia

---

Cushing's disease can be suggested on the basis of the three most common symptoms reported in affected children: typical body features, weight gain, and short stature; the diagnosis can be confirmed by hormonal tests and imaging procedures.

## GH DEFICIENCY

### Clinical Aspects

GH deficiency (see (Table II) for classification) can be attributed to either congenital or acquired forms. All the forms have in common the most relevant clinical sign: growth defect. In both in the United States and Europe the prevalence appears to be 1:4000 children of school age. A severe form of congenital GH

**Table II** GH Deficiency and Related Disorders

---

Hypothalamic-pituitary abnormalities

- Genetic
  - Abnormalities determining multiple hormone deficiencies
  - Molecular defects of GH-RH
  - Molecular defects of the GH-RH receptor
  - Isolated deficiency (type I A, type I B, type II, type III)
  - Bio-inactive growth hormone
- Congenital malformations
- Trauma
- Inflammatory disorders
- Tumors involving hypothalamus or pituitary and radiation
- Psychosocial dwarfism
- GH neurosecretory dysfunction
- Prader-Willi syndrome

Growth hormone insensitivity

- Abnormalities of GH receptor and postreceptor defects
- Primary defects of IGF-1 biosynthesis
- Genetic insensitivity to IGF-1

---

deficiency can be suspected at birth by the association of small birth size and micropenis; in patients with multiple pituitary deficiencies (in particular the association of GH and ACTH deficiency), postnatal spontaneous hypoglycemia might be a relevant symptom. In congenital forms the growth rate remains low and, within the first year, the patient's length can already be 3-4 standard deviations below the expected mean values.

In the acquired forms (which represent about 25% of all the patients with GH deficiency), both stature and growth rate can be normal until the appearance of the underlying disease, then they fall to lower percentile lines and other clinical features such as obesity, delayed puberty, and visual field defects may appear.

In the complete (congenital) form, the triad of short stature, bone age retardation, and absent growth hormone response to provocative test is of diagnostic relevance. Low levels of insulin-like growth factor-1 (IGF-I), the main peripheral effector of growth promoting action of GH, generally accompany this triad.

### Causes of GH Deficiency

#### **Hypothalamic-Pituitary Dysfunctions**

Because many hypothalamic dysfunctions disturb pituitary function, it is frequently impossible to distinguish between the two locations of the defect. Several genetic causes such as errors of Pit1 gene and Prop-1 mutations (the transcription factors responsible for the control of the genes of GH and other pituitary hormones) have been reported (see Table II). GHRH

receptor gene defects located on chromosome 7 may induce GH unresponsiveness, and, finally, mutations of the GH-1 gene (the one encoding for growth hormone) have been described.

### **Congenital Malformations of the Hypothalamic-Pituitary Region**

The septo-optical displasia is one example in which anatomical abnormalities of this region are accompanied by GH, ACTH, and other pituitary deficiencies. This is due to mutations in the HESX1 gene.

### **Trauma of the Brain and/or the Hypothalamus**

Trauma of the brain and/or the hypothalamus may cause multiple pituitary insufficiencies; difficult delivery or hypoxemia in the perinatal period are also part of this condition.

### **Inflammation**

Fungal, viral, and bacterial infections as well as granulomatous diseases involving this area may lead to pituitary insufficiency.

### **Tumors**

In addition to tumors of this brain area (meningiomas, gliomas, ependymomas, etc.), craniopharyngiomas are rather frequent in children. Radiation therapy, used on some of these tumors or in hematologic disorders, may determine GH deficiency.

### **Psychosocial Dwarfism**

Psychosocial dwarfism is the consequence of a severe form of emotional deprivation frequently ascribed to a familiar or to social inadequate conditions in which GH secretion is low and behavioral abnormalities of eating habits are also common. The reversibility to a condition of normal GH secretion in these children may confirm the diagnosis.

### **GH Neurosecretory Dysfunction**

In affected patients, GH responses to provocative tests are normal but IGF-I is low; the diagnostic clue is the greatly decreased spontaneous peaks of GH secretion observed during a 24 hour continuous sampling.

## **GH RESISTANCE SYNDROMES**

In GH resistance syndromes, the common feature is represented by lack of GH action in spite of elevated plasma levels of the hormone; IGF-I is generally low in these forms and the growth defect is particularly severe. Zvi Laron first described such a syndrome in 1965, years before the somatomedin hypothesis was

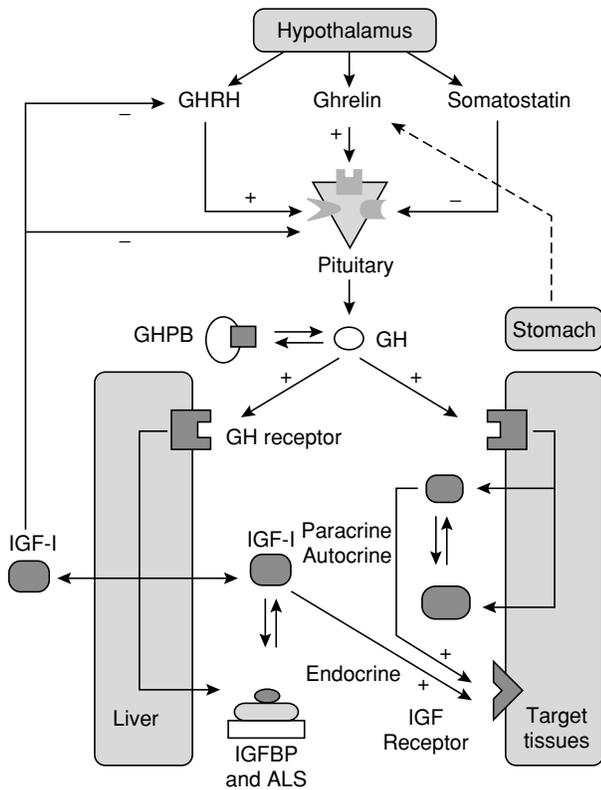
presented. It is now accepted that mutations in the extracellular domain of the GH receptor are present in patients with these clinical features. About 200 patients in the Mediterranean area and 20 in Ecuador have been reported thus far.

In GH resistance (or insensitivity), IGF-I is also unresponsive to induction test with hGH; there are studies under way showing positive growth acceleration in these patients during treatment with human recombinant IGF-I.

## **A SHORT REVIEW OF GH AND IGF-I PHYSIOLOGY**

GH, a polypeptide hormone of 191 amino acids with molecular weight of about 22,000, is produced by specific cells of the anterior pituitary and is released in bursts lasting only minutes, followed by long periods during which plasma levels are almost undetectable. The most frequently observed peaks are those following meals and those in the first part of slow-wave sleep. The physiological basis for this pulsatile secretory pattern is found in the hypothalamic nuclei (in particular the ventromedial nucleus) where the two main hypothalamic factors controlling GH are produced: growth hormone-releasing hormone (GHRH, positive regulator) is mainly responsible for the secretory peaks, but its activity is evident only when the other factor, somatotrophic-releasing inhibiting factor (SRIF, negative regulator), is inhibited. These two hormones, together with other hypothalamic factors such as GH-releasing peptides and neurosecreted amines interplay with each other: GH released, in every minute, is then the algebraical result of these opposite forces (Fig. 3). Feedback mechanisms also play a role both at hypothalamic and at pituitary level controlling plasma GH levels and IGF-I values. Moreover, ultrashort feedback mechanisms are also exerted by the levels of the two neuroregulatory hormones GHRH and SRIF, which influence each other, acting back on hypothalamic neurosecretory amines.

GH in itself has little effect on bone and cartilage development, but its growth-promoting activity is mediated by substances that are induced by the hormone: GH-free serum has mitogenic properties in cell cultures and insulin-like activity in adipose tissue and promotes the incorporation of amino acids and sulfate into the cartilage. In 1972, W. H. Daughaday proposed the somatomedins hypothesis to explain these effects. It is now recognized that the most important of these substances is IGF-I, produced in the liver and kidney as well as in the growing cartilages. What is



**Figure 3** Regulation of growth hormone secretion and interplay among the main influences.

actually measured in the plasma is mostly the IGF-I coming from the liver and kidney, although IGF-I mediating the growth effect of GH is mainly the amount locally generated in chondrocytes under stimulation by GH. However, plasma IGF-I levels provide a reliable indicator of the effective amount of GH produced, since it is found to be very low in almost all the conditions of endocrine-mediated growth defect. When measured in a given patient after a short cycle (2 days) of exogenous hGH administration (induction test) it also provides relevant information about sensitivity to GH.

In conclusion, IGF-I reflects the integrated concentrations of the previously secreted GH (i.e., in the preceding 24–36 hours); at variance with the pulsatility of GH, their levels are quite stable during the day. Although IGF-I levels may be low even in conditions different from GH deficiency (hypothyroidism, renal diseases, etc.), they reflect GH spontaneous secretion better than the provocative tests. At present, however, these tests remain the gold standard for the diagnosis of GH deficiency.

## ASSESSMENT OF GH SECRETION

From the previously discussed considerations, and due to the pulsatile nature of GH secretion, it emerges that baseline plasma GH levels are very frequently found undetectable if single random samples are taken, even in normal subjects. Therefore, to confirm a diagnosis of GH deficiency, subsequent samples must be obtained in standard conditions and appropriate stimulation tests should be performed. Some of the following standardized tests are considered useful for initial screening purpose: (1) a post-prandial (3–4 hours after a protein-rich meal) or (2) a post-exercise sample (30 min after a strenuous exercise) that shows GH levels above the cutoff value of 10 ng/ml could rule out GH deficiency and avoid further testing. These procedures, however, are not very common, and consequently the most widely used tests have been: (1) arginine infusion (0.5 g/kg by slow i.v. infusion within 30 min) and (2) insulin-induced hypoglycemia (0.1 U/kg as i.v. bolus). These tests are not devoid of side effects and should be performed by trained personnel in appropriately equipped centers.

The main defect of these two tests is their high false positive rate: from 10 to 20% of normal children fail to respond to the test substances and would then be wrongly diagnosed as GH deficient. Moreover, approximately 30% of normal children will show discordant response to the two tests. The basis for interpreting this discrepancy is to be found in the interplay between the two neurohormones GHRH and SRIF: a GH response to the stimulations mentioned above would be possible only when the stimulus arrives during a decreasing phase of SRIF (and the reciprocal GHRH increase).

One of the main problems common to the various provocative tests resides in the interpretation of the results: different countries have adopted different cutoff levels to separate normal from pathological responses. Cutoff values ranging from 5 to 10 ng/ml of GH at the peak are widely accepted. It must also be considered that the commercial kits used for the assay of GH in plasma with monoclonal antibodies have reached a level of sensitivity much higher than that obtained at the time when the set points of these test were established. Some authors have proposed a provocative test based on the administration of arginine (0.5 g/kg i.v.; supposed to act by shutting off the endogenous SRIF tone) together with an intravenous injection of synthetic GHRH (1  $\mu$ g/kg). This double test is a very potent GH secretory stimulus, and the cutoff limit for normal subjects has been set to 20 ng/ml. The main advantage of this test is its greater

capability of discriminating normal from pathological responses (GH deficiency) due to the minimal number of false positives.

In conclusion, it is recommended that each Pediatric Endocrinology Center should adopt two or three provocative tests, and use always the same tests, after having established its own limits of normal responses. Baseline IGF-I levels are also of great help: GH-deficient patients generally have very low levels (i.e., below  $-2$  SDS of IGF-I). However, the provocative tests for GH secretion retain some limits: they are not physiological, they are poorly reproducible, and the normal range of responses is arbitrary.

These reasons emphasize the recommendation that the poorly growing child should undergo a complete examination, including psychological, medical, and hormonal aspects. GH deficiency remains a clinical diagnosis, corroborated by auxological data concerning the growth pattern, by medical history, and by biochemical tests; it is not the simple observation of hormonal levels below a set value.

## IMAGING

Once this clinical diagnosis is obtained a magnetic resonance image of the pituitary and hypothalamic area is mandatory. It can show morphological abnormalities: pituitary stalk agenesis or interruption, pituitary hypoplasia, or other lesions. In acquired forms it will show the cystic or solid lesion damaging pituitary/hypothalamic function.

## TREATMENT

The starting dose of hGH to be used in children with GH deficiency is in the range of 0.6 to 0.7 mg/kg body weight per day. It must be administered in the evening at bedtime, in order to mimic the nocturnal spontaneous GH secretory peaks. Daily subcutaneous injections are recommended, although different devices for administration (such as needle-free percutaneous jet injector systems and long acting intramuscular preparations requiring only fortnightly injections) have been used. Pen devices (either with almost painless needles or the new needle-free devices) have almost completely replaced the old use of syringes; they are accurate and easily by the children themselves. The treatment efficacy is generally monitored by regular assessment of growth rate.

The growth-promoting effects observed with the treatment are a function of the administered doses: increased doses lead to accelerated growth rates. Moreover, the younger the patient age at the beginning of treatment, the better the final results that will be obtained. The administered doses must be regularly re-set according to the effects on weight/height obtained.

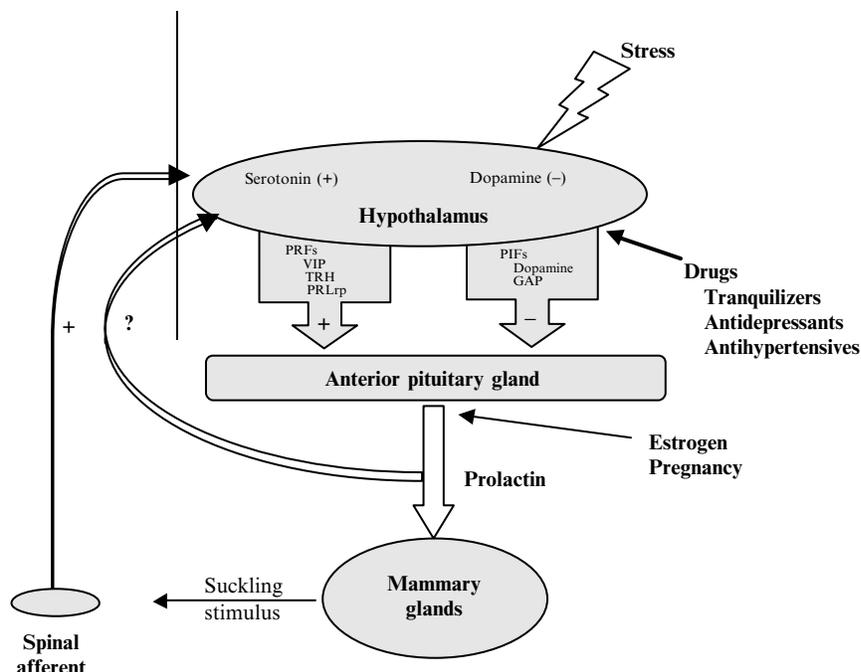
Until about 1990, GH therapy was discontinued after growth rate was observed to be greatly decreased and cartilages closed, but now that metabolic effects of GH are better understood the treatment regimen has gradually changed to "adult" doses (such as 0.04–0.06 mcg/kg/week; given daily or on alternate days). Reports confirming preliminary positive effects on bone mineral density, muscle strength, exercise performance, cardiovascular function, and body composition are encouraging this view.

## See Also the Following Articles

Gigantism: Excess of Growth Hormone • Growth Hormone (GH) • Growth Hormone Deficiency, Genetic • Growth, Normal Patterns and Constitutional Delay • Intrauterine Growth Retardation • *SHOX* Disorders • Postnatal Normal Growth and Its Endocrine Regulation • Puberty: Physical Activity and Growth • Short Stature and Chromosomal Abnormalities • Skeletal Development During Childhood and Adolescence • Turner Syndrome

## Further Reading

- American Academy of Pediatrics. (1997). Considerations related to the use of recombinant human growth hormone in children. *Pediatrics* **99**, 122–129.
- Drugs and Therapeutics Committee of the Lawson Wilkins Pediatric Endocrinology Society. (1995). Guidelines for the use of growth hormone in children with short stature. *J. Pediatr.* **127**, 857–867.
- GH Research Society. (2000). Consensus guidelines for the diagnosis and treatment of growth hormone (GH) deficiency in childhood and adolescence: Summary statement of the GH Research Society. *J. Clin. Endocrinol. Metab.* **85**, 3990–3993.
- Reiter, E. O., and Rosenfeld, R. G. (2003). Normal and aberrant growth. In "Williams Textbook of Endocrinology" (H. M. Kronenberg, S. Melmed, P. R. Larsen, and K. S. Polonsky, eds.), 10th ed., pp. 1003–1114. Saunders, Philadelphia.
- Rosenfeld, R. G., Albertsson-Wikland, K., Cassorla, F., Frasier, S. D., Hasegawa, Y., Hintz, R. L., Lafranchi, S., Lippe, B., Loriaux, L., and Melmed, S. (1995). Diagnostic controversy: the diagnosis of childhood growth hormone deficiency revisited. *J. Clin. Endocrinol. Metab.* **80**, 1532–1540.



**Figure 1** Relationship between hypothalamic–pituitary–mammary gland axis and control of prolactin secretion. The primary target organ of prolactin is the breast. Suckling activates neural pathways to the hypothalamus. An appropriate increase in prolactin-releasing factors (PRFs) or a decrease in prolactin inhibitory factors (PIFs) affects prolactin release. Within the hypothalamus, serotonergic pathways are stimulatory and dopaminergic pathways are inhibitory to prolactin release. Prolactin release is stimulated by a number of PRFs including vasoactive intestinal peptide (VIP), thyrotropin-releasing hormone (TRH), and prolactin-releasing peptide (PRLrp). Prolactin is inhibited by PIFs, predominantly dopamine. GAP, gonadotropin-associated peptide.

addition, suckling activates neural afferent pathways that go through the spinal cord to the hypothalamus, where increases in PRFs or decreases in PIFs influence prolactin release during this time period.

Prolactin is released in a sleep-related circadian rhythm characterized by pulses superimposed on a continuous basal secretion in both men and women. Approximately 350  $\mu\text{g}$  of prolactin is secreted daily. The highest plasma concentrations occur during nocturnal sleep, peaking between 3 AM and 5 AM and falling rapidly during the first hour of waking, with the lowest concentrations occurring by late morning. There is a midday surge of prolactin and cortisol secretion in response to a high-protein and/or high-fat meal. (Carbohydrates do not affect prolactin levels.)

Estradiol augments the nocturnal prolactin rise, making serum prolactin concentrations highest during the reproductive years in women. With the onset of puberty in girls, serum prolactin levels rise significantly, reaching a peak at midcycle and remaining elevated during the luteal phase. The prolactin pulses are amplified during the high-estrogen phases of the menstrual cycle, resulting in increased serum prolactin levels.

## ETIOLOGY OF HYPERPROLACTINEMIA

Various physical and/or emotional stress-inducing stimuli, such as pregnancy, suckling, prolonged breast manipulation, coitus, physical exercise, and surgery, may induce hyperprolactinemia. Stress can inhibit the release of hypothalamic prolactin inhibitory factors, which increases the secretion of prolactin and galactorrhea. Tactile stimulation of the breast and nipple increases prolactin levels; said stimulation is responsible for the rise in serum prolactin seen 20–30 min after breastfeeding is initiated.

## PREGNANCY AND PROLACTIN

Maternal serum prolactin levels rise in a linear fashion beginning in the first trimester, reaching concentrations 10 times higher than those in a nonpregnant woman at term, as high concentrations of estrogen induce hyperplasia and hypertrophy of the pituitary lactotrophs, resulting in enlargement of the pituitary to twice its normal size. Similarly, in the fetus, prolactin synthesis, storage, and secretion can be

demonstrated after the 12th week of gestation. Pituitary lactotrophs can be first detected at 18 weeks gestation with a sharp increase in fetal serum levels from 22 weeks gestation to term. By the end of the neonate's first day, prolactin levels progressively decline to normal. Not only is prolactin important in intrauterine fetal tissue differentiation and organ development, but experimental evidence has also demonstrated that fetal prolactin may participate in kidney osmoregulation, lung maturation, and the regulation of salt and water balance.

During parturition, prolactin secretion follows a multiphasic pattern, not demonstrable in patients undergoing elective cesarean delivery. The level of prolactin rises throughout pregnancy and precipitously falls during active labor, reaching a nadir approximately 2 h prior to delivery. A surge of prolactin release then occurs immediately before and after delivery, achieving a peak within 2 h postpartum. Prolactin also exerts mammogenic and lactogenic effects on the mammary gland.

## SIGNS AND SYMPTOMS OF HYPERPROLACTINEMIA

Irregular menses and galactorrhea are the most frequently observed abnormalities associated with hyperprolactinemia. Women may also present with decreased libido and infertility, whereas men may display decreased libido, infertility, gynecomastia, or impotence.

Galactorrhea, defined as a watery, milky breast discharge containing neither pus nor blood, can be confirmed by the presence of milk, which is ascertained by examining a sample of breast secretion for the microscopic presence of fat droplets on a glass slide. Oligomenorrhea and amenorrhea frequently occur with hyperprolactinemia as well. Dopamine and norepinephrine are neurotransmitters that influence the synthesis and secretion of gonadotropin-releasing hormone (GnRH), luteinizing hormone (LH), and follicle-stimulating hormone (FSH). Any abnormalities in their synthesis or secretion will result in menstrual irregularities. Primary or secondary amenorrhea results from hyperprolactinemia directly inhibiting GnRH secretion, consequently inhibiting the release of LH and FSH. This negative feedback can be overcome by exogenous administration of GnRH, which produces abnormally high levels of LH and FSH. Hyperprolactinemia may also alter the frequency or amplitude of GnRH pulses through changes in dopamine release, which also produce further menstrual abnormalities. Only one-third

of women with amenorrhea demonstrate clinical galactorrhea despite hyperprolactinemia.

## FACTORS AFFECTING PROLACTIN SECRETION

### Drugs

The most common causes of hyperprolactinemia are medications that alter the inhibitory actions of dopamine. These agents include neuroleptics, which block dopamine receptors, the tricyclic antidepressants and monoamine oxidase inhibitors, which facilitate several possible stimulatory pathways, serotonin reuptake inhibitors, opiates, and cocaine, and the antihypertensive medications including  $\alpha$ -methyldopa, reserpine, and verapamil. Protease inhibitors have been shown to cause hyperprolactinemia, although the mechanism is unknown.

### Hypothyroidism

Primary hypothyroidism, which exists in approximately 3 to 5% of women with galactorrhea and hyperprolactinemia, is characterized by low serum levels of thyroxine (T4) and decreased negative feedback on the hypothalamic-pituitary axis. Decreased levels of triiodothyronine (T3) and T4 increase the sensitivity of pituitary cells, which easily respond to normal or slightly elevated levels of TRH. Therefore, when thyrotrophs and lactotrophs are stimulated by TRH, there is an exaggerated response, resulting in increased levels of both thyroid-stimulating hormone (TSH) and prolactin (PRL) secretion. Along with the diminished negative feedback seen in primary hypothyroidism, patients may also experience an alteration in the positive feedback loop between T4 and dopamine, in which decreased dopamine secretion will also lead to elevated levels of serum TSH and PRL.

Women with galactorrhea should undergo measurement of both serum PRL and TSH as the initial diagnostic tests to distinguish primary hypothyroidism from a TSH-secreting pituitary adenoma. Primary hypothyroidism, such as Hashimoto's thyroiditis, will exhibit elevated serum TSH, whereas a TSH-secreting pituitary adenoma will also exhibit elevated serum T3 levels and symptoms of hyperthyroidism, along with elevated serum TSH levels.

### Renal Disease

Individuals with acute or chronic renal failure will develop hyperprolactinemia as a result of delayed

clearance of PRL. They also have a reduced PRL response to TRH stimulation resulting from pituitary dysfunction. When such individuals take medications known to alter the hypothalamic regulation of prolactin, such as methyl dopa or metoclopramide, prolactin levels may rise to over 2000 ng/ml. Treatment other than that indicated for renal failure is rarely necessary.

## DISORDERS OF PROLACTIN SECRETION

### Craniopharyngiomas

Two hundred fifty new cases of craniopharyngiomas are diagnosed annually in the United States. These tumors, which are the nonpituitary tumors most commonly associated with hyperprolactinemia, arise from epithelial remnants of Rathke's pouch that can be found distributed along the pituitary stalk from the pars distalis to the floor of the third ventricle. Fifty-five percent of these tumors are cystic, whereas 15% are solid; however, 30% of patients express both solid and cystic components. Craniopharyngiomas are typically diagnosed during the second and third decades of life, but calcifications within these tumors can be noted during childhood by plain skull X-rays and computerized tomography (CT) scans.

A craniopharyngioma damages the hypothalamus or extends into the sella turcica, interfering with the transport of hypothalamic hormones and neurotransmitters, resulting in pituitary dysfunction. The levels of growth hormone (GH), TSH, adrenocorticotrophic hormone (ACTH), and antidiuretic hormone are all diminished. In addition, a deficiency in gonadotropins (FSH, LH) produces amenorrhea. The degree of impairment depends on the extent of involvement with either the hypothalamus or the pituitary stalk. Expansion of the craniopharyngioma into the optic chiasm produces local compression, demanding surgical resection or decompression. Survival rates of 80% have been reported for surgical resection followed by radiotherapy and many patients will require subsequent hormone replacement therapy.

### Prolactin-Secreting Adenoma

Prolactin-secreting adenomas are the most common pituitary tumor. Microadenomas, which are adenomas of less than 1 cm diameter confined to the sella turcica, constitute the majority of prolactin-secreting adenomas. In contrast, prolactin-secreting macroadenomas, which are adenomas of more than 1 cm diameter extending beyond the sella turcica, are very rare.

Galactorrhea and hyperprolactinemia are common findings in women with pituitary prolactinomas. Ten percent of women with galactorrhea and radiologic evidence of a pituitary tumor have normal serum PRL levels and approximately 50% of women with hyperprolactinemia will demonstrate radiographic changes of the sella turcica compatible with an adenoma. Most women have normal basal levels of FSH and LH and produce a normal or elevated response following administration of GnRH.

Numerous retrospective and prospective studies have demonstrated that hyperprolactinemia with or without a microadenoma nearly always follows a benign clinical course without treatment, with spontaneous remissions occurring in a substantial portion of women. However, data from several series of patients with microadenomas observed over long periods of time without treatment demonstrated that the risk of progression from a microadenoma to a macroadenoma was 7%. In another series of patients carefully observed for up to 15 years, Jeffcoate and co-workers found that approximately one-third of patients who were either untreated or treated with dopamine agonists intermittently had a return to normal levels of prolactin off all therapy. Patients who experienced a pregnancy during this interval had a higher rate of remission than patients who did not (35% versus 14%, respectively).

### Multiple Endocrine Neoplasia Type I

Prolactinomas occur in approximately 20% of patients with multiple endocrine neoplasia type 1 (MEN1). The MEN1 gene, localized to chromosome 11q13, is a constitutive tumor suppressor gene in that an inactivating mutation results in tumor development. Prolactinomas occurring in patients with MEN1 may be more aggressive than sporadic prolactinomas.

### Prolactinomas and Pregnancy

Numerous studies suggest that pregnancy exerts beneficial effects on women with functional hyperprolactinemia or prolactin-secreting microadenomas and increases the chance that hyperprolactinemia will spontaneously regress. Following parturition, mean prolactin levels decrease in approximately half of all women with hyperprolactinemia, eventually normalizing in approximately one-third of these patients. There is also an increased frequency of empty sella syndrome in multiparous women compared with nulliparous women due to autoinfarction of lactotrophic adenomas in pregnancy. Breastfeeding does

not stimulate tumor growth; hence, postpartum women with prolactinomas can safely breastfeed their infants. Hyperestrogenism associated with oral contraceptives or estrogen replacement therapy does not cause significant growth of macroadenomas.

### Lactotroph Hyperplasia

Women with elevated prolactin levels without clinical evidence of a prolactin adenoma are given the clinical diagnosis of functional hyperprolactinemia, which may arise from decreased dopamine inhibition resulting in hyperplasia of lactotrophs. This diagnosis can be made only at the time of surgical exploration of the pituitary gland. Lactotroph hyperplasia has also been reported to be due to pituitary enlargement with suprasellar extension.

### Empty Sella Syndrome

Empty sella syndrome is characterized by herniation of the subarachnoid membrane into the sella turcica through a defect in the pituitary sella diaphragm, which allows the cerebrospinal fluid (CSF) that fills the subarachnoid space to protrude into the sella turcica whenever intracranial pressure increases. This in turn compresses the pituitary gland and remodels the contour of the sella.

Primary empty sella syndrome results from a congenital defect in the pituitary sella diaphragm. This defect may coexist with pituitary adenomas that secrete PRL, GH, or ACTH or reflect alterations in the circulatory dynamics of the CSF. Secondary empty sella syndrome may occur after either radiation therapy or surgical intervention in the sellar region.

### Diagnosis

On radiographic examination, the empty sella appears symmetrically enlarged, with or without the erosion of bone. Definitive diagnosis can be made by magnetic resonance imaging (MRI) or CT scan utilizing metrizamide, a contrast medium injected into the CSF through a lumbar puncture, which fills the intrasellar defect, thus confirming the diagnosis.

### Endocrine Dysfunction

Although endocrine abnormalities are uncommon, pituitary dysfunction, panhypopituitarism, and pituitary adenoma may coexist with empty sella syndrome in some patients. Galactorrhea with normal or moderately elevated levels of serum PRL may also be present. The elevation in PRL may reflect arachnoid herniation

that compresses the pituitary stalk, interfering with the transport of dopamine.

### Acromegaly

Galactorrhea and elevated serum PRL may exist in women with acromegaly. *In vitro* studies with GH-secreting pituitary adenomas demonstrate that these tumors may secrete both PRL and GH. The galactorrhea that occurs in women with acromegaly is thought to derive from increased secretion of PRL rather than the lactogenic properties of growth hormone.

### Sheehan Syndrome/Autoimmune Lymphocytic Hypophysitis

Sheehan syndrome results from infarction of the anterior lobe of the pituitary gland as a direct result of a major hemorrhage in pregnancy that reduces the blood supply to the gland. Women with this syndrome are unable to lactate postpartum and do not show serum responses of GH, cortisol, or PRL to insulin-induced hypoglycemia. There is also blunted LH, FSH, PRL, and TSH response to TRH and GnRH in patients with complete infarction of the anterior pituitary. Autoimmune lymphocytic hypophysitis involves selective loss of PRL-producing cells and occasionally other anterior pituitary hormones.

## DIAGNOSTIC EVALUATION

In evaluating a woman who presents with galactorrhea or amenorrhea, a detailed history and physical examination must be obtained to exclude the use of drugs that have prolactin-stimulatory properties; in addition, a panel of screening tests including blood chemistries, thyroid function tests, prolactin levels, and a pregnancy test must be performed to exclude all physiologic causes of hyperprolactinemia other than hypothalamic-pituitary disease.

A radiologic evaluation of the hypothalamic-pituitary region should then ensue to exclude a mass lesion such as a macro- or microadenoma. CT scans provide information on the bony structure of the sella turcica but provide less information on soft tissue lesions, such as extrasellar extension and empty sella turcica. MRI, on the other hand, is more sensitive and expensive than CT, providing better contrast resolution. MRI is a more effective method for detecting an empty sella, optic chiasm, optic nerves, cavernous sinuses, and carotid arteries.

## TREATMENT

Treatment options include observation with careful follow-up, medical therapy, surgery, and irradiation.

### Observation

Periodic observation can be utilized for patients with galactorrhea, regular ovulatory cycles, and normal or slightly elevated serum prolactin levels (less than 40 mg/ml). Treatment with a dopamine agonist or estrogen is unnecessary in these patients. If anovulation with unopposed estrogen secretion occurs, cyclic progestins or oral contraceptives should be administered to induce regular uterine bleeding.

Patients with documented radiologic evidence of microadenomas should be monitored closely for interval growth. In 93% of patients with microadenomas, there is no demonstrable enlargement over 4- to 6-year observation periods. Hence, serial prolactin levels with follow-up scans are sufficient. If prolactin levels rise significantly, repeat scanning is indicated. A microadenoma documented to have undergone interval growth requires therapy as the tumor may be one of the 7% that will grow to become a macroadenoma.

### Medical Therapy

#### **Dopamine Agonists**

Bromocriptine, a dopamine agonist, binds to dopamine receptors and inhibits pituitary prolactin secretion. Administration of bromocriptine also induces a cyclic and physiologic estrogen secretion and is useful in treating symptomatic galactorrhea and inducing ovulation. Doses of 2.5 to 5 mg/day are sufficient to suppress prolactin levels and restore ovulatory function; however, higher doses, up to 10 mg/day, may be necessary. Discontinuation of therapy results in a return of hyperprolactinemia, galactorrhea, and amenorrhea in 70% of women. After 1 year of bromocriptine treatment, 10% of women experience complete regression of their microadenomas. In these patients, therapy should be discontinued for 6 weeks every year to determine whether prolactin levels have normalized and regular menses have resumed.

#### **Side Effects of Dopamine Agonist Therapy**

Nausea, vomiting, fatigue, headaches, dizziness, and syncope have been observed as side effects in 50–70% of women following oral administration of bromocriptine. Most of these side effects are mild and transient if administered at bedtime and/or with food. Orthostatic hypotension may also occur and can be

avoided if the initial dosage is taken at bedtime. Approximately 10% of patients experience severe side effects requiring discontinuation of bromocriptine. Avoidance of these side effects can be attained by intravaginal administration of bromocriptine, which allows usage of a lower dosage, once-daily treatment, and slower absorption, all of which minimize the side effects. Bromocriptine does not interfere with sperm function, as women using vaginal bromocriptine have become pregnant and delivered normal infants. Bromocriptine is also available as a long-acting injectable form (depot bromocriptine), which is as effective as the oral form and has similar side effects.

#### **Ovulation Induction**

Ovulation following administration of bromocriptine can occur in almost all women with hyperprolactinemia. Approximately 50% will ovulate after receiving 5 mg/day of bromocriptine, whereas 50% will require higher doses. In women with pituitary adenomas, the mean length of treatment necessary to induce ovulation is 16 weeks, whereas patients without adenomas usually respond within 10 weeks.

#### **Treatment during Pregnancy**

Once pregnancy is confirmed, bromocriptine therapy should be discontinued unless a macroadenoma is present, for which treatment may be maintained throughout pregnancy or reinstated for tumor enlargement. Microadenomas have not been demonstrated to grow substantially during pregnancy. However, frequent visual field examinations are encouraged at 20, 28, and 38 weeks gestation, with follow-up MRI or CT scan for suspected headaches or abnormal examinations. If there is evidence of suprasellar extension, bromocriptine treatment should be instituted or increased and maintained for the remainder of pregnancy. Bromocriptine has not been shown to carry an increased risk of fetal malformations and its use can be discontinued postpartum to allow breastfeeding.

#### **Treatment of Macroadenomas**

During pregnancy, women with macroadenomas that fail to demonstrate regression to bromocriptine therapy may be candidates for adenomectomy, as no increased surgical morbidity has been reported. Bromocriptine has been shown to produce regression of macroadenomas within 3 to 6 months of initiating therapy. Repeat MRI or CT scans should be repeated 6 months after initiation of therapy and bromocriptine should be continued as long as the adenoma continues to regress. Prolactin values cannot be used

to gauge the response to treatment as levels may not correspond with tumor regression. In addition, surgery should be considered in patients who do not demonstrate a response to bromocriptine therapy. Cabergoline, a drug that binds with high affinity to pituitary dopamine receptors, and Pergolide, a drug approved by the U.S. Food and Drug Administration for the treatment of Parkinson's disease, are additional medications comparable to bromocriptine in efficacy and tolerance and both demonstrated reductions in prolactin levels and tumor size. Pergolide is longer-acting and more potent than bromocriptine and is better tolerated in some patients, whereas Cabergoline can be given once weekly and may have fewer side effects than bromocriptine.

### Surgical Therapy

Transsphenoidal surgery (via sublabial incision) is the surgical procedure used for microadenomas and most macroadenomas that allows microsurgical exploration of the sella turcica, permitting removal of the pituitary tumor, while preserving the functional capacity of the remaining gland. Surgical success rates are highly dependent on the experience and expertise of the surgeon, as well as the size of the tumor. Cure rates of 50 to 80% have been reported in patients with microadenomas, using resumption of menses, euprolactinemia, and cessation of galactorrhea as endpoints. In patients with macroadenomas, long-term cure rates between 10 and 30% have been reported, suggesting that surgery should be performed before the adenomas reach diameters greater than 1 cm. Reported complications include transient or chronic diabetes insipidus, hemorrhage, meningitis, cerebrospinal fluid leak, and panhypopituitarism. The reported mortality and morbidity rates from transsphenoidal surgery for microadenomas are 0.3 and 0.4%, respectively.

Given the favorable results of bromocriptine therapy, surgical excision of macroadenomas should be performed only if complete or partial failure of medical therapy or poor patient compliance occurs. In addition, maximum reduction of tumor size should be obtained with bromocriptine therapy before

attempted surgical excision. Postsurgical bromocriptine therapy should be continued to prevent rapid regrowth of the adenoma.

### Radiation Therapy

Due to the excellent therapeutic responses to medical therapy and transsphenoidal surgery, radiotherapy is not considered the primary mode of treatment for pituitary adenomas. A dose of 4500 rad of cobalt may be used to arrest progression of tumor growth. Regular menses usually do not return after treatment and galactorrhea may not be corrected for up to 1 year. Secondary hypothalamic or pituitary damage is also prone to occur with radiation therapy as adenoma tissue is often more resistant to treatment than surrounding structures. Proton-beam and heavy-particle irradiation methods may be used as alternatives to cobalt irradiation and have been shown to be associated with increased visual-field defects or oculomotor palsies. Consequently, this form of therapy should be reserved for patients who are unresponsive to medical and surgical management.

### See Also the Following Articles

Acromegaly, Clinical Features of • Craniopharyngiomas • Pregnancy Endocrinology • Prolactin, Evolution of • Prolactin (PRL) • Prolactinoma, Clinical Manifestations • Prolactinoma, Diagnosis • Prolactinoma, Pathogenesis • Prolactinoma, Therapy

### Further Reading

- Luciano, A. A. (1999). Clinical presentation of hyperprolactinemia. *J. Reprod. Med.* **44**, 1085–1090.
- Molitch M. E. (2001). Neuroendocrinology—Disorders of prolactin secretion. *Endocrinol. Metab. Clin.* **30**.
- Shoupe, D., and Mishell, D. R. (1997). In "Infertility, Contraception, and Reproductive Endocrinology" (R. A. Lobo, *et al.*, eds.), 4th ed. Blackwell Science, Malden, MA.
- Yen, S. S. C., and Jaffe, R. B. (1999). "Reproductive Endocrinology, Physiology, Pathophysiology, and Clinical Management," 4th ed. W. B. Saunders, Philadelphia, PA.

somatotroph or corticotroph adenomas. Only 6% are completely negative for all pituitary hormones.

## CRANIOPHARYNGIOMAS

Craniopharyngioma is a frequent if not the most frequent sellar and suprasellar tumor of childhood and adolescence (up to 9 or 10%). Most patients with such a tumor are children and young adults, but craniopharyngiomas have also been described in newborns and elderly patients. The tumor is usually suprasellar but may extend to the sella. It probably derives from remnants of the Rathke's pouch. The tumor may sometimes achieve huge proportions (up to 12 cm; giant cystic craniopharyngioma) and grow well beyond the suprasellar region. Rarely, tumors may be confined in the chiasm or optic nerve. Craniopharyngiomas are the most heterogeneous tumors of the sellar region. They are mostly cystic or cystic solid, filled with a lipid and cholesterol-rich viscous fluid. Magnetic resonance imaging (MRI) is superior to computed tomography (CT) in determining the full extent of the tumor. The pituitary is generally pushed toward the bottom of the sella. Calcifications are frequent (75% of the cases in childhood and 35% in adulthood). Areas of the tumor may degenerate, leading to an intense inflammatory and foreign body giant cell reaction.

Despite being histologically benign, craniopharyngiomas are rather biologically aggressive tumors, often invading or infiltrating the surrounding neurological structures, such as the hypothalamus and the optic nerve. Malignant transformation has been described.

The clinical presentation of craniopharyngiomas is directly related to their location and growth behavior. If they compress the optic pathways and infiltrate the hypothalamus and the third ventricle, they can cause visual disturbances, dysfunction of the hypothalamo-pituitary axis, and hydrocephalus.

The most frequent endocrine signs are partial anterior pituitary deficiency (mostly short stature due to growth hormone deficiency), hyperprolactinemia due to compression of the pituitary stalk and/or infiltration of the hypothalamus interfering with the dopamine secretion that normally suppresses pituitary prolactin secretion, and diabetes insipidus. Most unexpected, the syndrome of inappropriate secretion of antidiuretic hormone has also been reported in rare cases.

The optimal treatment for craniopharyngiomas is controversial. For some authors, radical surgical excision, if feasible, constitutes the only chance at a cure, whereas for others the complications of radical

surgery outweigh the benefits and they propose only partial resection followed by radiotherapy. The growing consensus seems to be that the goal should not be to aim for radical resection at all costs in every patient but to reserve this procedure only for tumors for which it can be achieved without damaging vital structures. It has been shown that subtotal excision followed by adjuvant radiotherapy produces excellent results, with 10-year recurrence-free survival rates of 80–90%.

Alternative therapies to surgery have also been developed. One is the stereotactically guided instillation of  $\beta$ -emitting isotopes directly into the craniopharyngioma cysts. Another option is stereotactic radiosurgery with a gamma-knife.

## MENINGIOMAS

Meningiomas are tumors of arachnoid and meningotheelial cells. They represent approximately 25% of intracranial tumors in women and 13% in men. The reason for this gender difference may be related to the expression of sex steroid receptors by these tumors. Indeed, they contain estrogen receptors and may increase in size during pregnancy or even during the menstrual cycle. Twenty percent of meningiomas arise in the sellar and parasellar regions. Sella involvement by a meningioma is usually due to the extension of a suprasellar meningioma that invades the pituitary. Intrasellar meningiomas mimic nonfunctioning pituitary adenomas. It is therefore sometimes very difficult to distinguish a suprasellar meningioma with intrasellar extension from a silent pituitary adenoma with suprasellar extension. They can both present with visual disturbances, partial or complete hypopituitarism, hyperprolactinemia, or a combination of these signs. Deterioration of cognitive function, confusion, and memory loss may also occur in 20% of patients. Interestingly, a few patients have had concomitantly pituitary adenoma and sellar meningioma and one patient developed an intrasellar meningioma 8 years after radiotherapy treatment for a pituitary prolactinoma. MRI is superior to CT in delineating the lesion, but unless a dural origin of the tumor can be clearly demonstrated, radiographic distinction from pituitary adenoma is difficult.

The treatment of choice is surgical resection. However, meningiomas are highly vascular and intraoperative bleeding can be a complication in approximately one-third of cases. This complication can be prevented by selective preoperative endovascular embolization. Stereotactic radiosurgery has been proposed as an excellent alternative therapy.

## GANGLIOCYTOMAS

Gangliocytomas are rare hypothalamic or intrasellar tumors composed of mature neurons without a glial component. If a glial component is present, the tumors are called gangliogliomas. Gangliocytomas are either pure gangliocytomas or mixed adenoma–gangliocytomas since the latter have a second component that is indistinguishable from pituitary adenoma. This adenomatous component may be mixed or form a separate, discrete nodule.

Interestingly, the ganglionic component that resembles hypothalamic neurons has been shown to produce and secrete various hypothalamic-releasing hormones, such as growth hormone-releasing hormone, gonadotropin-releasing hormone (GnRH), and corticotropin-releasing hormone, but also other hormones such as gastrin and vasopressin, and, rarely, pituitary hormones. Most of the mixed tumors are associated with endocrine syndromes, such as acromegaly, Cushing's disease, and other endocrine disturbances (hyperprolactinemia), whereas most of the pure gangliocytomas are endocrinologically silent. The mixed adenoma–gangliocytomas could be due to the growth of the adenoma secondary to stimulation by the hypothalamic-releasing hormones secreted by the gangliocytoma.

## GRANULAR CELL TUMORS

These are benign tumors of the neurohypophysis, roughly evenly distributed between the infundibulum and the posterior lobe. They are mostly asymptomatic since they only rarely become large enough to produce symptoms. There is a female predominance and most lesions present in the fourth or fifth decade. The most common presentation is visual disturbances and/or hypopituitarism. Despite their location in the neurohypophysis and the infundibulum, diabetes insipidus is very rare. Radiological sellar changes are only observed in approximately 50% of cases. Treatment is surgical, with postoperative radiotherapy for incompletely resected lesions.

## GLIOMAS

Glioma of the optic nerve occurs predominantly in children and adolescents. It is often associated with neurofibromatosis type 1 (NF-1), and bilateral optic pathway gliomas are even diagnostic for NF-1. Loss of vision is obviously the main clinical manifestation, but depending on the tumor's extension, glioma may cause hydrocephalus and endocrinopathy, such as

hypopituitarism, diabetes insipidus, or precocious puberty.

In children, histologically these tumors are usually of a low-grade glioma called pilocytic astrocytoma. Optic gliomas in adults are usually anaplastic astrocytomas, which are highly infiltrative and aggressive lesions. The prognosis is very poor with a few exceptions. Gliomas involving only the pituitary gland are very rare. They can be mistaken as nonfunctioning pituitary adenomas by causing hypopituitarism often associated with hyperprolactinemia due to the pituitary stalk compression.

Treatment of gliomas in children is a matter of debate between the indication for surgery and the role of radiotherapy and chemotherapy. In most cases, there is a need for lifelong hormonal substitution due to the persistent endocrinopathy.

## CHORDOMAS

Chordomas are rare, slowly growing, locally aggressive tumors arising in the midline, almost always involving the clivus area. Chordomas of the clivus are infiltrative and present with headache, visual symptoms, intracranial hypertension, and/or anterior pituitary insufficiency. Radiological evaluation should include both CT to determine the extent of bone involvement and MRI to delineate accurately the soft tissue infiltration. Treatment is predominantly surgical with or without radiotherapy. It seems that patients treated with surgery and radiotherapy do better than those treated with either approach alone. It is fortunate that chordomas rarely occur in children since they are very aggressive in patients younger than 5 years of age.

## GERM CELL TUMORS

These tumors are classified as germinomas, teratomas, embryonal carcinomas, yolk sac tumors, or choriocarcinomas depending on their stage of embryonic development. Germinomas and teratomas are the most frequent. These tumors occur mostly in people younger than 20 years of age, with a peak age of diagnosis of 10–12 years, and males seem to be more affected than females. They commonly arise in the pineal region, followed by the suprasellar region. Rarely, they can arise from the pituitary fossa or even be limited to it. Interestingly, intracranial germ cell tumors have been reported in association with Klinefelter's syndrome. If the tumor contains syncytiotrophoblastic cells, it can secrete  $\beta$ -human chorionic

gonadotropin ( $\beta$ -hCG) as well as  $\alpha$ -fetoprotein in the blood and cerebrospinal fluid (CSF). Measurements of these markers are important for follow-up since monitoring their levels in the patient's blood and CSF will allow early detection of recurrence. Angiotensin converting enzyme is also sometimes secreted by germinomas.

Mass effect or precocious puberty are the most common presenting symptoms of the suprasellar germ cell tumors. Hypopituitarism, diabetes insipidus, visual disturbances, hydrocephalus, intracranial hypertension as well as symptoms due to hypothalamic compression, such as bulimia, anorexia, psychosis, or seizures, can all be observed.

Interestingly, tiny germinomas in the pituitary stalk have been shown to be responsible for what was first considered to be "idiopathic" central diabetes insipidus. Precocious puberty may be caused by the tumoral secretion of  $\beta$ -hCG, which directly stimulates the Leydig cells of the testes to produce androgens, resulting in sexual precocity. Association with a pituitary adenoma has been described.

Since the histologic type is a major determinant of therapy and prognosis, a biopsy of suspected germ cell tumors should be performed before therapy is initiated. Mature teratomas are usually benign and can be cured by surgical resection alone. Pure germinomas are highly radiosensitive, and radiotherapy achieves excellent results. However, the other types of germ cell tumors are more aggressive and often need to be treated with a combination of surgery, radiotherapy, and chemotherapy. For these tumors, survival rates are approximately half of those for pure germinomas.

## HYPOTHALAMIC HAMARTOMAS

Hamartomas are benign hyperplastic malformations composed of a fibrosis glial matrix with mature ganglion cells that resemble hypothalamic neurons. The lesions sometimes contain myelinated nerve fibers. These tumors are slow growing and rarely invasive. Approximately 90% of patients develop precocious puberty of central origin. There is often a neurodevelopmental delay, with low IQ and seizures. There is also a tendency for patients to become obese in late childhood and adolescence. Some tumors have been shown to express GnRH, which may explain the association with precocious puberty. It has also been proposed that hamartomas may either mechanically stimulate the median eminence to secrete GnRH or interrupt interneuronal pathways that tonically inhibit the GnRH-secreting neurons in the median eminence,

allowing these neurons to be disinhibited and to secrete GnRH.

The treatment of choice of the precocious puberty is the blockade of GnRH action using GnRH analogs. It is not necessary to attempt a total neurosurgical removal of the lesions because the hamartomas in most instances do not appear to progress. Surgery should therefore be restricted to patients who have signs of increased intracranial pressure or neurologic deterioration from progressive growth of the hamartoma.

## METASTASES

Metastasis to the pituitary gland is not uncommon, occurring in 1–5% of cancer patients. Metastases more frequently affect neurohypophysis than adenohypophysis. Any tumor can metastasize to the pituitary gland, but the most frequent are carcinomas of the breast and lungs, which represent 50 and 20% of cases, respectively. Among other carcinomas spreading to the pituitary gland are those of the stomach, kidney, thyroid, prostate, ovary, and thymus and melanomas and hematologic malignancies. In patients with disseminated cancer, pituitary metastases are often asymptomatic and are found incidentally at autopsy. The pituitary metastasis may be the first sign of cancer or no primary tumor may be identified. A metastasis can grow rapidly and the presenting symptoms include visual field defects due to chiasma compression, ophthalmoplegia, and ptosis secondary to infiltration into the cavernous sinus. It can rapidly lead to hypopituitarism. The involvement of the infundibulum or the posterior pituitary can lead to diabetes insipidus. No radiological characteristics allow one to distinguish a metastasis from a pituitary adenoma. Interestingly, metastases have also been shown to spread into a previously existing pituitary adenoma, leading to a rapid increase in size of the adenoma.

## HEMATOLOGICAL PATHOLOGIES

Leukemias and lymphomas can infiltrate the pituitary region. True intraparenchymal deposits are very rare, but periglandular infiltration is observed in approximately 50% of cases of adult lymphoblastic leukemia. Endocrine functions do not appear to be affected except in patients who present with a syndrome of inappropriate secretion of antidiuretic hormone. Rarely, pituitary infiltration may produce a major enlargement of the gland, causing visual disturbances.

Plasmacytoma can also involve the sella and mimic nonfunctioning pituitary adenoma. Langerhans cell

histiocytosis (LCH), formerly known as histiocytosis X, is also a rare cause of lesions in the sellar region. The lesions usually affect the hypothalamus or the pituitary stalk, but they may also arise in the sella turcica, mimicking a pituitary adenoma. Several formerly distinct pathologies are currently classified under the denomination of LCH: eosinophilic granuloma, Hand-Schüller-Christian disease, and Letterer-Sieve disease. LCH is a proliferation and dissemination of abnormal histiocytic cells of the Langerhans system, mixed with eosinophils, foamy macrophages, and lymphocytes. Langerhans cells express S-100 protein and the CD1a antigen. They also contain peculiar intracytoplasmic organelles, called Birbeck granules, which are pathognomonic.

LCH occurs mostly in children, and besides the lesions in bones, skin, and lungs, diabetes insipidus and growth failure are the most frequent manifestations. LCH can also produce isolated hypothalamic involvement. For lesions causing mass effect, the treatment is surgery with or without radiotherapy.

Erdheim-Chester disease is a rare non-LCH affecting multiple organ systems. Among patients with central nervous system involvement, the most frequent manifestations are diabetes insipidus, cerebellar syndromes, and orbital lesions (exophthalmia). When the disease involves the hypothalamus, the pituitary stalk, or the pituitary, pituitary insufficiency may also be observed in addition to diabetes insipidus. Erdheim-Chester disease usually appears in adulthood and has histological characteristics different from those of LCH since there are no Birbeck granules and the histiocytes do not express CD1a and only occasionally express S-100, but they do express CD68.

Rosai-Dorfman disease is another rare histoproliferative disease that may have central nervous system manifestations, which are commonly dura based, resembling meningiomas. One case of this disease presenting as a pituitary tumor has been reported. In this disease, the proliferative histiocytes exhibit cytoplasmic staining against S-100.

## SARCOIDOSIS

Sarcoidosis involves the central nervous system in approximately 10% of cases. In this situation, hypothalamic and/or sellar localizations are most frequent. In most cases, central nervous system involvement is associated with the presence of the disease in other organs (e.g., the lungs and lymph nodes), which greatly facilitates the diagnosis. However, isolated pituitary involvement is possible, making the etiological

diagnosis of pituitary insufficiency and/or pituitary mass difficult.

Diabetes insipidus occurs in 30–40% of patients, whereas endocrine dysfunction of the anterior pituitary is variable, with hyperprolactinemia present in approximately half of cases. Sarcoidosis of the sellar region may be complicated by aseptic meningitis and cranial nerve palsies. Measurement of blood angiotensin converting enzyme concentration may help the diagnosis. If hypothalamo-pituitary involvement is the sole manifestation of sarcoidosis, it may be necessary to perform a pituitary biopsy to make an accurate diagnosis and determine the best treatment. Glucocorticoids may have good efficacy on the symptoms due to the mass effect, but they often do not eliminate the lesions or fully restore the function of the disturbed endocrine axes.

## TUBERCULOSIS

Intracranial tuberculomas account for 0.15–4% of space-occupying lesions in developed countries, but tuberculosis involving the sella and parasellar regions is very rare, with only approximately 40 cases reported in the literature. Pituitary tuberculomas present with signs of a mass lesion, headache, visual disturbances, hypopituitarism, and, in some cases, diabetes insipidus. On imaging, tuberculosis lesions can appear as an intrasellar-enhancing mass or as a parasellar mass. Tuberculosis can also present with thickening and enhancement of pituitary stalk or the meninges, with the latter suggesting an inflammatory process. The lesions are quite aggressive toward the endocrine functions of the pituitary. Not all patients with tuberculosis of the pituitary region have signs of active tuberculosis elsewhere, making the diagnosis of tuberculosis difficult. It may require a surgical biopsy of the lesion. The surgical specimen shows an inflammatory lesion containing granulomas with epithelioid cells and giant cells, and occasionally central caseating necrosis, but without acid-fast bacilli on staining or after culture. However, polymerase chain reaction examination of the CSF or of the biopsy for *Mycobacterium tuberculosis* DNA may be very useful in the diagnosis. The treatment is obviously antituberculosis chemotherapy.

## WEGENER'S GRANULOMATOSIS

Wegener's granulomatosis can sometimes involve the central nervous system, leading to cerebrovascular lesions and/or meningitis. Rarely, the posterior and anterior pituitary may be involved, leading to diabetes insipidus, hyperprolactinemia, and/or anterior

pituitary insufficiency. Usually, symptoms caused by pituitary involvement occur after or concomitantly with the general symptoms of the disease. Rarely is pituitary dysfunction the sole or first expression of the disease.

### INTRASELLAR ABSCESS

Pituitary abscess is a rare but potentially life-threatening pathology caused either by hematogenous seeding of the gland or by direct extension of an adjacent infection. The abscesses can be due to a broad spectrum of bacterial microorganisms or fungi. A pre-existing tumor is evident in some cases. Infectious processes can present with symptoms undistinguishable from those of patients with pituitary tumors. The most common clinical features are headache, visual disturbances, and pituitary insufficiency. Surprisingly, evidence of a serious infection is not common. An exact diagnosis may often require surgical biopsy. Broad-spectrum antibiotic therapy is usually used at the beginning of treatment and subsequently narrowed or stopped according to the results of the cultures.

### LYMPHOCYTIC HYPOPHYSITIS

Lymphocytic hypophysitis occurs most often in women during the latter half of pregnancy or the postpartum period as well as in 20–30% of patients with coexisting autoimmune diseases, such as thyroiditis, adrenalitis, or autoimmune hypoparathyroidism. Anti-pituitary antibodies have been detected in a minority of patients. Histologically, it is an extensive inflammatory infiltration of glands consisting of B and T lymphocytes and plasma cells. Often, it presents as a mass lesion of the sella turcica, simulating a pituitary adenoma. Headache is often the main symptom; there is also anterior pituitary insufficiency, which is often only partial. Visual field defects may also occur. Diabetes insipidus is rare except when there is infiltration of the infundibulum. Lymphocytic hypophysitis may also result in isolated hormone deficiencies, especially ACTH. Conservative management may lead to resolution. Transsphenoidal surgery, however, is both diagnostic and therapeutic and should be performed in cases with progressive compression. Although glucocorticoid therapy may result in a decrease in the size of the pituitary mass through a reduction of the inflammatory process in a few patients, these cases are difficult to distinguish from those of spontaneous shrinkage.

## CYSTS

### Rathke's Cleft Cyst

Rathke's cleft cysts originate from the remnants of Rathke's pouch, which may fail to obliterate completely, leaving cystic remnants at the interface between the anterior and posterior lobes of the pituitary. This lesion is quite common; it is found in 13–33% of autopsy pituitaries. Most of these cysts are intrasellar with or without suprasellar extension. The cyst lumen is filled with mucus, cell debris, and cholesterol crystals. Most of these cysts are asymptomatic and are found at autopsy. When symptomatic, they often have a suprasellar extension and present clinical symptoms due to the mass effect: headache, visual field defects, hyperprolactinemia, partial anterior pituitary insufficiency, and, rarely, diabetes insipidus. Rupture of a cyst may produce aseptic meningitis and Tolosa–Hunt syndrome, which is a painful ophthalmoplegia. A progressive enlargement of the cyst can lead to an empty sella. These cysts may mimic craniopharyngiomas in radiologic appearance. Surgical decompression with drainage is the treatment of choice when they become symptomatic. Except for pituitary insufficiency, most deficits may resolve at least partially, and recurrence is unusual.

### Dermoid and Epidermoid Cysts

Symptoms of these cysts arising in the sellar and parasellar regions produce a mass effect resulting more often in visual field defect than anterior pituitary insufficiency. Rupture of the cyst may cause chemical meningitis secondary to spillage of keratinous debris. The transformation into a squamous cell carcinoma in the cyst is very rare. The unusual presentation includes subarachnoidal hemorrhage and stroke. These cysts may coexist with arachnoid cysts. Treatment is surgical.

### See Also the Following Articles

Acromegaly, Clinical Features of • Craniopharyngiomas • Diabetes Insipidus, Nephrogenic • Gonadotropin-Secreting Tumors • Hamartoma, Pituitary • Hypopituitarism • Pineal Tumors • Pituitary Adenomas, TSH-Secreting • Pituitary Tumors, Molecular Pathogenesis

### Further Reading

Albrecht, S., Bilbao, J. M., and Kovacs, K. (2002). Nonpituitary tumors of the sellar region. *In* "The Pituitary" (S. Melmed, ed.), pp. 592–609. Blackwell, Malden, MA.

- De Vile, C. J., Grant, D. B., Hayward, R. D., and Stanhopez, R. (1996). Growth and endocrine sequelae of craniopharyngioma. *Arch. Dis. Child.* **75**, 108–114.
- Freda, P. U. (2002). Tuberculosis of the pituitary and sellar region. *Pituitary* **5**, 147–148.
- Gsponer, J., De Tribolet, N., Déruaz, J. P., Janzer, R., Uské, A., Mirimanoff, R. O., Reymond, M. J., Rey, F., Temler, E., Gaillard, R. C., and Gomez, F. (1999). Diagnosis, treatment, and outcome of pituitary tumors and other abnormal intrasellar masses. *Medicine* **78**, 236–269.
- Hansen, I., Petrossians, P., Thiry, A., Flandroy, P., Gaillard, R. C., Kovacs, K., Claes, F., Stevenaert, A., Piguët, P., and Beckers, A. (2001). Clinical case seminar: Extensive inflammatory pseudotumor of the pituitary. *J. Clin. Endocrinol. Metab.* **86**, 4603–4610.
- Isaacs, R. S., and Donald, P. J. (1995). Sphenoid and sellar tumors. *Otolaryngol. Clin. North. Am.* **28**, 1191–1229.
- Kretschmar, C. S. (1997). Germ cell tumors of the brain in children: A review of current literature and new advances in therapy. *Cancer Invest.* **15**, 187–198.
- McCormick, P. C., Post, K. D., Kandji, A. D., and Hays, A. P. (1989). Metastatic carcinoma to the pituitary gland. *Br. J. Neurosurg.* **3**, 71–80.
- Nozaki, K., Nagata, I., Yoshida, K., and Kituchi, H. (1997). Intrasellar meningioma: Case report and review of the literature. *Surg. Neurol.* **47**, 447–454.
- Randall, R. V., Laws, E. R., Jr., and Abboud, C. F. (1984). Clinical presentations of craniopharyngioma: A brief review of 300 cases. In “The Hypothalamus” (J. R. Givens, ed.). Year Book, Chicago.
- Schaller, B., Kirsch, E., Tolnay, M., and Mindermann, T. (1998). Symptomatic granular cell tumor of the pituitary gland: Casereport and review of the literature. *Neurosurgery* **42**, 166–171.
- Scheithauer, B. W. (1985). Pathology of the pituitary and sellar region: Exclusive of pituitary adenoma. *Pathol. Annu.* **20**, 67–155.
- Thodou, E., Asa, S. L., Kontogeorgos, G., Kovacs, K., Horvath, E., and Ezzat, S. (1995). Clinical case seminar: Lymphocytic hypophysitis: Clinicopathological findings. *J. Clin. Endocrinol. Metab.* **80**, 2302–2311.
- Vates, G. E., Berger, M. S., and Wilson, C. B. (2001). Diagnosis and management of pituitary abscess: A review of twenty-four cases. *J. Neurosurg.* **95**, 233–241.

laboratory findings resembling those of Cushing's syndrome and are called "pseudoCushing's states."

## ETIOLOGY OF CUSHING'S SYNDROME

Cushing's syndrome recognizes two main etiologies: ACTH-dependent and ACTH-independent. In the former, glucocorticoid hypersecretion is due to continuous stimulation of the adrenal by ACTH produced by pituitary or extrapituitary tumors (i.e., ectopic ACTH secretion). Conversely, in ACTH-independent Cushing's syndrome, the adrenal produces cortisol autonomously due to benign or malignant neoplastic transformation (i.e., adenoma or carcinoma); stimulation by aberrant receptors specific to catecholamines, serotonin, or peptides such as gastrointestinal inhibitory peptide (GIP), glucagon, thyrotropin-releasing hormone, leptin, interleukin-1, luteotropin, and vasopressin; or genetic causes (e.g., primary pigmented nodular adrenal hyperplasia [PPNAD], possibly part of Carney's complex). Cases due to aberrant receptors generally present as macronodular adrenal hyperplasia, whereas the adrenal is not enlarged in PPNAD. Cushing's syndrome due to administration of exogenous glucocorticoids also is included with ACTH-independent forms.

## DIAGNOSIS OF CUSHING'S SYNDROME

The diagnosis of Cushing's syndrome is usually made on the basis of clinical and laboratory data rather than on the radiographic visualization of the pituitary tumor. Traditionally, diagnostic milestones for the screening of this disorder are considered the detection of elevated 24-h urinary-free cortisol (UFC) levels (indicating increased glucocorticoid production), absent cortisol suppression by dexamethasone (the set point for ACTH suppressibility by glucocorticoids is altered, thus larger doses of exogenous glucocorticoids are needed to inhibit pituitary-adrenal secretion), and altered cortisol circadian rhythm with elevated late-evening cortisol levels. Although most patients with Cushing's syndrome do have increased UFC levels and fail to suppress serum cortisol values below 1.8  $\mu\text{g}/\text{dl}$  after an overnight low-dose (1 mg orally) dexamethasone suppression test, false-positive results may be found, resulting in suboptimal diagnostic specificity. Similarly, increased midnight serum cortisol ( $>7.5 \mu\text{g}/\text{dl}$ ) reportedly has an excellent sensitivity for the diagnosis of Cushing's syndrome, but

its specificity remains uncertain when applied to patients with pseudoCushing's states. Furthermore, the test is burdened by costs of inpatient admission. As an alternative, measurement of late-night cortisol in saliva can be easily performed at home and yields good diagnostic accuracy at far lesser expense. However, each of these tests fails to reliably identify all patients with Cushing's syndrome, so further testing may be required. Second-line tests are stimulation with desmopressin, a vasopressin analogue that elevates plasma ACTH and cortisol levels in the majority of patients with Cushing's disease but not in normal individuals or patients with pseudoCushing's, and/or combined dexamethasone-corticotropin-releasing hormone (CRH) stimulation. In the latter test, CRH stimulation is performed after 2 days of 2 mg dexamethasone daily (orally), and attainment of plasma cortisol levels greater than 1.4  $\mu\text{g}/\text{dl}$  favors the existence of Cushing's syndrome. However, the complexity and cost of these procedures call for specialized endocrine centers. Indeed, this and the subsequent diagnostic steps are best performed by endocrinologists experienced with the disease, as recommended by a recent consensus conference.

## DIAGNOSIS OF CUSHING'S DISEASE

### Noninvasive Testing

Once the diagnosis of Cushing's syndrome has been established, plasma ACTH levels should be measured to ascertain the site driving excess glucocorticoid secretion. ACTH levels within or above the normal range ( $>10 \text{ pg}/\text{ml}$ ) reasonably rule out primary adrenal disease and indicate ACTH-dependent Cushing's syndrome. At this stage, a pituitary source of ACTH secretion can be expected in 85% of cases, with the remaining 15% presenting an extrapituitary etiology, mostly neuroendocrine tumors (e.g., bronchial and midgut carcinoid tumors and small cell lung carcinomas).

Baseline ACTH levels may also provide a clue as to the origin of ACTH secretion given that ectopic ACTH production is often characterized by extremely elevated plasma ACTH levels. However, there is considerable overlap with ACTH levels measured in Cushing's disease, and other tests need to be performed. These tests rely on the fact that the tumoral corticotrope, although partly autonomous, does still retain some features of normal corticotropes (i.e., sensitivity to glucocorticoid feedback and CRH stimulation). In contrast, ACTH-producing cells outside the pituitary by and large do not possess these

features and, therefore, are not likely to respond to glucocorticoid or CRH administration. However, it should be mentioned that, as for the diagnosis of Cushing's syndrome, none of these tests is 100% accurate and false-positive or false-negative results may be observed.

Historically, inhibition with high doses of dexamethasone (8 mg overnight or 2 mg every 6 h for 48 h) was the first test used for the differential diagnosis of ACTH-dependent Cushing's syndrome. A 50 or 80% suppression of endogenous cortisol secretion (evaluated in serum or 24-h urine) will identify 70 to 90% of patients with a pituitary adenoma but may also yield a false-positive response in up to 30% of patients with ectopic ACTH secretion. Given this low diagnostic accuracy, this test is best employed to confirm a high suspicion of Cushing's disease using strict criteria for cortisol suppression (e.g., 80 or 90% suppression from baseline). Thus, only a marked decrease in cortisol should be taken as indicative of Cushing's disease.

Over the past two decades, stimulation with CRH has been accepted as the most accurate noninvasive tool to discriminate pituitary from ectopic sources of ACTH secretion. In the majority of patients with Cushing's disease, intravenous administration of 1  $\mu$ g/kg body weight or 100  $\mu$ g ovine CRH causes a marked rise in plasma ACTH and cortisol, whereas the hormonal response is modest or absent in patients with ectopic ACTH syndrome. The test is well tolerated, with mild, short-lived facial flushing and tachycardia as the most common side effects and with more severe reactions, such as hypotension and chest pain, occurring only rarely. Criteria consistent with Cushing's disease are a rise in plasma ACTH of at least 50% over baseline and a 20 to 50% maximal cortisol increase. These cutoffs will identify 85% of patients with Cushing's disease and yield only a small number of false positives.

### Invasive Testing

When the diagnosis of Cushing's disease cannot be firmly established by noninvasive procedures, inferior petrosal sinus sampling (IPSS) has to be performed, preferably in a highly specialized center. This technique involves sampling both the right and left inferior petrosal sinuses, which drain the respective hemipituitaries, and calculating the ratio between petrosal and peripheral plasma ACTH concentrations obtained before and after stimulation with CRH. The rationale for this procedure is the presence of higher ACTH

levels in the veins draining the pituitary if the gland harbors an adenoma, whereas petrosal ACTH levels will not differ from peripheral concentrations if the tumor is elsewhere in the body. A baseline center/periphery ratio greater than 2 and/or greater than 3 after CRH stimulation is indicative of a pituitary source of ACTH hypersecretion ("gold standard criteria"). However, the occasional patient with Cushing's disease may fail to meet these criteria, possibly due to anomalous venous drainage. Thus, the diagnosis of ACTH-secreting pituitary adenoma cannot be completely excluded if no center/periphery gradient is detected. We usually perform IPSS if CRH and dexamethasone testing yield contradictory results and pituitary imaging fails to clearly visualize an adenoma. IPSS may also provide useful information for identification of the site of the adenoma within the pituitary, but its predictive value falls short of two-thirds and so may be misleading in more than 30% of cases.

### IMAGING

ACTH-secreting tumors are visualized by magnetic resonance imaging (MRI) with gadolinium enhancement in slightly over 50% of cases due to their exceedingly small size (mostly 3–5 mm). The detection rate of MRI seems only slightly higher than that of computed tomography (CT) but may increase with the advent of dynamic procedures. It is worth emphasizing that pituitary imaging may also visualize other small lesions (e.g., incidentalomas) that may be erroneously interpreted as ACTH-secreting adenomas. Thus, only lesions greater than 5 to 6 mm in patients where the diagnosis of Cushing's disease is established by testing can be interpreted as adenomas.

### SPECIAL FEATURES

#### ACTH-Secreting Pituitary Macroadenomas

The vast majority of ACTH-secreting pituitary adenomas are less than 10 mm in diameter and, hence, are microadenomas. In about 10% of patients, Cushing's disease is due to a pituitary macroadenoma (i.e., a tumor >1 cm in diameter). In addition to features of glucocorticoid excess, approximately one-third of these patients present neurological symptoms and signs such as frontal and occipital headache and visual field defects due to compression of the optic chiasm. Patients with macroadenomas often show high baseline plasma ACTH immunoreactivity, albeit with reduced bioactivity, and may fail to respond to CRH or suppress after

high-dose dexamethasone. A peculiar variant of macroadenoma is a pituitary ACTH-secreting tumor developing in patients adrenalectomized for Cushing's disease (i.e., Nelson's syndrome).

### ACTH-Secreting Pituitary Carcinomas

Pituitary neoplasms are classified into three groups according to their behavior—benign adenomas, invasive adenomas, and carcinomas—with benign adenomas representing the majority. Invasive adenomas account for 10 to 30% of all pituitary tumors and can infiltrate the dura mater, cranial bone, or sphenoid sinus. These tumors are only locally invasive and do not metastasize. Very rarely (fewer than 40 cases have been described), Cushing's disease originates from an ACTH-secreting pituitary carcinoma. Because of the uncertainties in identifying a carcinoma at pathological examination, the diagnosis of malignancy is assigned to neoplasms that have metastasized to the central nervous system or outside the cerebrospinal axis via hematogenous or lymphatic spread, mostly the liver (70%) or bones and lungs (less frequently).

### THERAPY

The aims of treatment are removal of the pituitary adenoma, correction of hypercortisolism, and preservation of pituitary function. Pituitary exploration by transsphenoidal surgery is the treatment of choice. The majority of microadenomas can be resected by selective adenomectomy, resulting in approximately 80% cure rates. When no adenoma can be found at surgery, hemi- or total hypophysectomy can be performed, although even these procedures are not necessarily curative. Likewise, debulking of pituitary macroadenomas is successful in only one-third of cases and may be followed by tumor regrowth. Remission of hypercortisolism is defined by the detection of low to undetectable serum cortisol levels (e.g.,  $<1.8 \mu\text{g/dl}$ ), normalization of UFC, and normal cortisol inhibition after low-dose dexamethasone. Other parameters, such as baseline ACTH levels and response to CRH or desmopressin, are more useful as predictors of relapse.

Following successful removal of the ACTH-secreting pituitary tumor, the hypothalamic-pituitary-adrenal axis is transiently suppressed because normal ACTH-secreting cells have been inhibited by long-standing

hypercortisolism. Thus, patients will typically require glucocorticoid replacement therapy for 5 to 18 months.

If the patient is not cured by surgery, a second transsphenoidal exploration, radiotherapy, medical treatment, or bilateral adrenalectomy may be chosen as an alternative. Repeat pituitary surgery may be performed, but the likelihood of cure is not greater than 50%. Pituitary radiotherapy can be used in both adult and pediatric patients but takes months or years to be effective and, therefore, requires adjunctive medical therapy. Cure rates vary from 40 to 70%. Unfortunately, a variable degree of hypopituitarism occurs in nearly 50% of treated patients, although this proportion could be smaller with the new forms of stereotactic radiotherapy (e.g., radiosurgery). Moreover, a number of other complications, such as optic nerve damage and neoplastic transformation of adjacent tissues, should not be underestimated.

Medical therapy is not considered the primary therapy for Cushing's disease but is used in conjunction with radiotherapy or preoperatively to improve the patient's clinical condition before surgical treatment. Drug therapy targeted at pituitary ACTH secretion using compounds such as dopamine agonists (e.g., bromocriptine, cabergoline), serotonin antagonists (e.g., cyproheptadine, ritanserine, ketanserin),  $\gamma$ -aminobutyric acid agonists (e.g., valproic acid), and somatostatin analogues (e.g., octreotide) is efficacious only in anedoctal cases. Conversely, medical therapy aimed at the adrenal is more effective and includes several inhibitors of adrenal steroidogenesis. Ketoconazole, an antifungal drug that also inhibits adrenal cortisol synthesis, has been shown to effectively suppress cortisol levels in approximately 70% of patients with Cushing's disease. The major side effects include alterations in hepatic function, gastrointestinal discomfort, and (at high doses) impairment of testicular function. Liver function tests should be monitored during therapy. If ketoconazole is not effective or well tolerated, other adrenal enzyme inhibitors such as metyrapone, aminoglutethimide, and *o,p'*DDD (mitotane) can be used. A combination of various drugs at lower doses can minimize side effects. Etomidate, an imidazole derivative related to ketoconazole, has been used intravenously to rapidly lower cortisol levels during emergencies or in patients who cannot take medications by mouth. All of these compounds strongly suppress adrenal cortisol synthesis, and steroid replacement therapy may be required. Mifepristone (RU-486), a glucocorticoid receptor

antagonist, has been used with some efficacy in a small number of patients with Cushing's disease, especially in glucocorticoid-related psychoses. However, monitoring its efficacy and counteracting overtreatment remain unsolved problems that hamper its widespread use.

Bilateral adrenalectomy was used extensively prior to the advent of transsphenoidal pituitary surgery, although it was burdened by high peri- and postoperative risks. Nowadays, the development of laparoscopic approaches allows this operation to be performed with far greater ease and safety. Removal of both adrenals might need to be performed in patients who do not achieve disease control by pituitary surgery and/or radiotherapy. Patients require lifelong replacement therapy with glucocorticoids and mineralocorticoids and may develop an ACTH-secreting pituitary macroadenoma (i.e., Nelson's syndrome). The risk of Nelson's syndrome (~15%) may be reduced by associating adrenalectomy with pituitary radiotherapy.

### See Also the Following Articles

ACTH,  $\alpha$ -MSH, and POMC, Evolution of • Gonadotropin-Secreting Tumors • Hypercortisolism and Cushing's Syndrome • ACTH (Adrenocorticotrophic Hormone) • Pituitary Adenomas, TSH-Secreting • Pituitary Gland Anatomy and Embryology • Pituitary Tumors, Clonality • Pituitary Tumors, Molecular Pathogenesis • Pituitary Tumors, Surgery

### Further Reading

- Blevins, L. S., Jr., Christy, J. H., Khajavi, M., and Tindall, G. T. (1998). Outcomes of therapy for Cushing's disease due to adrenocorticotropin-secreting pituitary macroadenomas. *J. Clin. Endocrinol. Metab.* **83**, 63–67.
- Freda, P. U., and Wardlaw, S. L. (1999). Diagnosis and treatment of pituitary tumors. *J. Clin. Endocrinol. Metab.* **84**, 3859–3866.
- Invitti, C., Pecori Giraldi, F., De Martin, M., and Cavagnini, F., for the Study Group on the Hypothalamo–Pituitary–Adrenal Axis of the Italian Society of Endocrinology. (1999). Diagnosis and management of Cushing's syndrome: Results of an Italian Multicentre Study. *J. Clin. Endocrinol. Metab.* **84**, 440–448.
- Kaltsas, G. A., and Grossman, A. B. (1998). Malignant pituitary tumors. *Pituitary* **1**, 69–81.
- Lüdecke, D. K., Flitsch, J., Knappe, U. J., and Saeger, W. (2001). Cushing's disease: A surgical view. *J. NeuroOncol.* **54**, 151–166.
- Moro, M., Putignano, P., Losa, M., Invitti, C., Maraschini, C., and Cavagnini, F. (2000). The desmopressin test in the differential diagnosis between Cushing's disease and pseudo-Cushing states. *J. Clin. Endocrinol. Metab.* **85**, 3569–3574.
- Morris, D., and Grossman, A. B. (2002). The medical management of Cushing's syndrome. *Ann. N.Y. Acad. Sci.* **970**, 119–133.
- Newell-Price, J., Trainer, P., Besser, M., and Grossman, A. B. (1998). The diagnosis and differential diagnosis of Cushing's syndrome and pseudo-Cushing's states. *Endocrine Rev.* **19**, 647–672.
- Oldfield, E. H., Doppman, J. L., Nieman, L. K., Chrousos, G. P., Miller, D. L., Katz, D. A., Cutler, G. B., Jr., and Loriaux, D. L. (1991). Petrosal sinus sampling with and without corticotropin-releasing hormone for the differential diagnosis of Cushing's syndrome. *N. Engl. J. Med.* **325**, 897–905.
- Pecori Giraldi, F., Moro, M., and Cavagnini, F., for the Study Group on the Hypothalamo–Pituitary–Adrenal Axis of the Italian Society of Endocrinology (2003). Gender-related differences in the presentation and course of Cushing's disease. *J. Clin. Endocrinol. Metab.* **88**, 1554–1558.

Table 1 Primary Site of Initiation for Pituitary Tumors

Hypothalamus	Pituitary
<p>Evidence in favor:</p> <ul style="list-style-type: none"> <li>• “Feedback” tumors in longstanding untreated end-organ failure</li> <li>• GHRH transgenic mice develop multiple adenomas on background of somatotroph hyperplasia</li> <li>• Dopamine D2 receptor knockout mice develop lactotroph hyperplasia and multiple discrete adenomas</li> <li>• GHRH can induce GH cell proliferation and <i>c-fos</i> gene <i>in vitro</i></li> <li>• Cushing’s disease with corticotroph “hyperplasia” may be “cured” by surgery</li> </ul> <p>Evidence against:</p> <ul style="list-style-type: none"> <li>• No multiple tumors</li> <li>• Normal tissue surrounds tumor</li> <li>• Low recurrence rate</li> <li>• How to account for nonfunctional tumors</li> </ul>	<p>Evidence in favor:</p> <ul style="list-style-type: none"> <li>• Anatomically and functionally discrete, <i>not</i> surrounded by hyperplasia</li> <li>• Histologically homogenous</li> <li>• “Mixed” GH/Prl tumors derived from pluripotential progenitor cell</li> <li>• Surgery for small adenomas followed by long-term “cure”</li> <li>• Genetically monoclonal</li> </ul> <p>Evidence against:</p> <ul style="list-style-type: none"> <li>• In MEN1 hyperplasia occurs in parathyroids and islets (? pituitary)</li> <li>• Hyperplasia and multiple adenomas coexist in Carney complex somatotrophinomas</li> <li>• May be multiclonal/oligoclonal from outset</li> <li>• Occasionally polyclonal</li> </ul>

## METHODS FOR ESTABLISHING CLONALITY

### X-Chromosome Inactivation

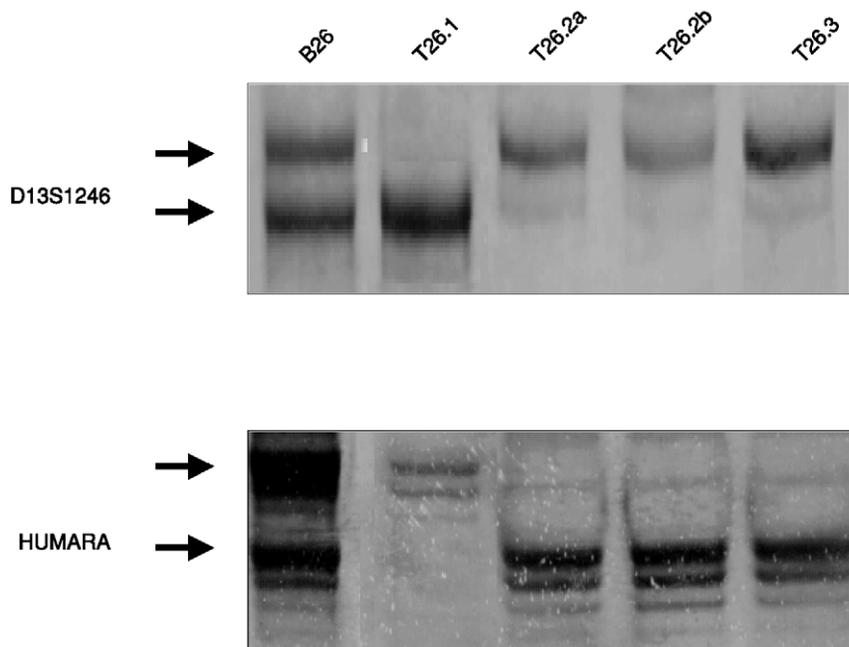
This is the “gold standard” method for assessment of clonality. In any tissue fragment that contains several cell generations, there will be an approximately equal complement of cells with paternal and maternal X chromosomes inactivated. Fortunately, there are several genes on the X chromosome that are “polymorphic”; that is, each allele can be distinguished by the presence or absence of a restriction enzyme site or the number of tandem repeats. The number of CAG repeats in the human androgen receptor gene (HUMARA) on each X chromosome is determined by polymerase chain reaction (PCR) amplification with primers spanning the microsatellite repeat sequence, after the template DNA has first been digested with the methylation-sensitive enzyme *HpaII*. Although restriction fragment length polymorphism (RFLP)–Southern blotting avoids any possibility of PCR artifacts, the latter method becomes the only practical one when sample size is small, since it can be performed with small amounts of DNA extracted from sections of archival paraffin-embedded tissue.

There are two limitations to X-chromosome inactivation analysis:

1. It can only be performed on samples from women
2. A reasonable amount of starting DNA is required, so very small samples cannot be analyzed.

### Autosomal Allelic Deletion (Loss of Heterozygosity)

The demonstration that a tissue or tumor contains only one of two heterozygous alleles is also evidence that the sample is monoclonal; this method of analysis does not depend on the X chromosome and so can be applied to samples from men. However, it does depend on the degree of constitutive heterozygosity of DNA markers or microsatellite polymorphisms at a given locus. The key advantage is that these polymorphisms are frequent and evenly spaced across the genome, so this is generally more informative than either RFLP or X-inactivation analysis. Where there have been direct comparisons in women, autosomal loss of heterozygosity (LOH)-determined clonality is confirmed by X-chromosomal clonality (see Fig. 1 for comparison). In general, therefore, PCR-based techniques for clonality analysis are preferred when the source of DNA is limited. It has been argued that because of the extreme sensitivity of PCR, any contamination of tumor DNA with “normal” constitutive DNA (such as that derived from entrapped leukocytes, blood vessels, or fibrous



**Figure 1** Clonality determined by LOH and X-chromosome inactivation analyses. Analysis of LOH at the microsatellite locus D13S1246 in a nonfunctioning tumor reveals uniform loss of either the upper or lower allele, indicating monoclonal tumors (top). A similar interpretation is revealed by the loss of the lower allele at the HUMARA locus in T26.1 and the upper allele in the recurrences T26.2 and T26.3. T26.2a and T26.2b are separate fragments from the same tumor sample (the first recurrence). B, blood; T, tumor; the number shown indicates patient number; .1 denotes first tumor; .2 denotes second tumor.

tissue) may obscure the ability to detect LOH or loss of one X chromosome allele but in practice this is not an issue.

## CLONALITY OF PITUITARY TUMORS

It is just over 10 years since the first reports appeared on the clonality of sporadic adenomas assessed by X-chromosome inactivation. These indicated that all pituitary tumor subtypes were monoclonal. Subsequently, autosomal allelic deletion analysis by RFLP–Southern blotting from whole tissue or by PCR microsatellite analysis of archival sections has confirmed monoclonality in much larger numbers of all the major subtypes of adenoma. Although monoclonality is the rule, this concept has been challenged.

Of particular interest with respect to the potential influence of the hypothalamus in the initiation of pituitary tumors are the corticotrophinomas. Levy uses this subtype to build an argument for physiological multiple monoclonality within anterior pituitary corticotrophs. If his argument is correct, it might be expected that a proportion of corticotroph adenomas may be polyclonal by X-inactivation and LOH analysis and this is indeed the case.

However, both micro- and macro-corticotrophinomas whose tumor fragments are macroscopically homogenous generally show a monoclonal pattern, though there are some exceptions (see below). However, predictably, a pituitary biopsy that showed multifocal hyperplastic corticotroph nodules from a patient with ectopic CRF production was polyclonal. Unpublished data by the authors adds to this debate in that some microcorticotrophinomas were found to be polyclonal and others to be monoclonal. However, in patients with Cushing's disease apparently cured by removal of nonadenomatous hyperplastic or normal pituitary tissue, this has, without exception, been polyclonal. The proportion of unambiguous adenomas that turn out to be polyclonal remains to be determined but if it were on the order of 30%, this would provide substantial support for the multiple monoclonality concept.

## DOES MONOCLONALITY EQUATE WITH "TUMOR"?

A review by Levy challenges the assumption that monoclonality is synonymous with neoplasia in the pituitary. The grounds put forward for this are as follows:

1. The rare cyclic adrenocorticotropic hormone (ACTH) secretory activity by corticotroph adenomas with return to normality at intervals.

2. The expression of multiple hormones (both pituitary and hypothalamic) by subsets of cells within the adenoma.

3. During normal pituitary development, there is nonrandom dispersion of cell types such that there are clusters of particular cell types in discrete regions of the pituitary. These might have arisen by repetitive and controlled division from a single progenitor cell, thus giving rise to “physiological” zonal clonality within a “normal” gland.

4. There are precedents for microscopic monoclonality, in the absence of neoplasia, in other tissues. Smooth muscle cells in atherosclerotic plaques exhibit a monoclonal pattern and the normal aorta appears to be a mosaic of small overlapping monoclonal “tiles.” The same appears to hold true of the bladder epithelium and uterus. Regenerating liver nodules are frequently monoclonal but not neoplastic, though presumably they have the potential to become so as hepatocellular carcinoma frequently develops in certain types of cirrhotic livers.

Levy proposes that an expansion of a “normal” monoclonal population of cells could occur in response to a growth stimulus from outside or within the pituitary to produce a histologically discrete adenoma that is truly neoplastic with respect to unrestrained cell division. If there were several normal monoclonal clusters, all with different genetic and functional characteristics, that responded to the same stimulus, this could produce an oligoclonal tumor, which by conventional clonality analysis would appear polyclonal. So “monoclonality” of discrete areas might be normal but this could only be tested by carefully examining a normal pituitary gland in multiple discrete areas, which remains to be carried out. Furthermore, if Levy’s idea of oligoclonal tumors is true, it may be possible to detect the different clones by examining multiple biopsies from different regions of the same tumor.

Undoubtedly, in the majority of pituitary adenomas monoclonality indicates neoplasia. The most persuasive evidence for this is data from recurrent tumors, a significant proportion of which show additional genetic changes in recurrent/regrown tumors. However, data from recurrent/regrown tumors, and also from several examples of pituitary carcinomas with metastases, provide evidence in support of a multiple monoclonal nature of the primary/initial tumor.

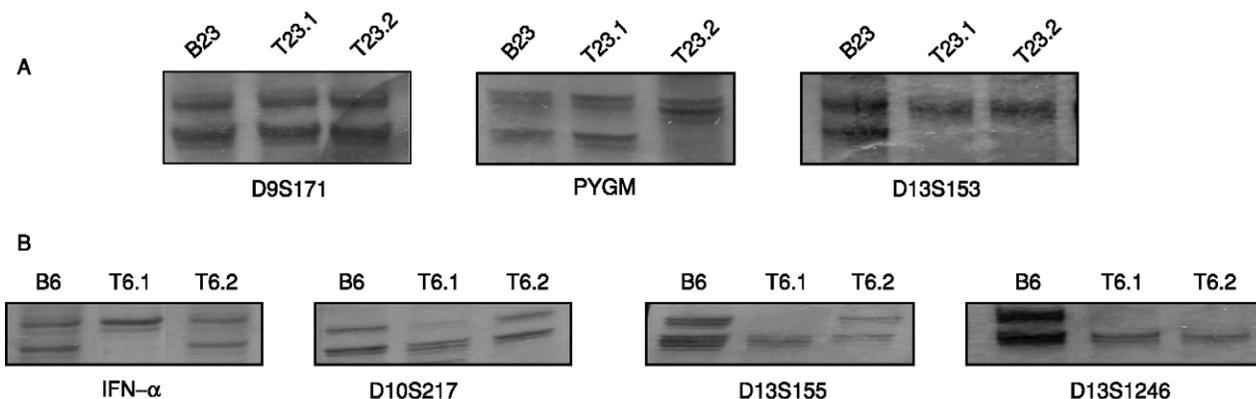
## MULTICLONALITY OF PITUITARY TUMORS

### Evidence from Recurrent Tumors

No study has previously examined initial and recurrent/regrowth pituitary tumors from the same individual for clonality or allelic deletions. Therefore, paraffin-embedded samples from patients who had two or more pituitary operations separated by at least 1 year were obtained. All clinical subtypes were represented. Most initial and recurrent tumors were invasive, meaning that tumor was present in one or both cavernous sinuses or invading into the sphenoid sinus.

Tumors from 33 patients showed allelic loss in the first and subsequent tumors at one or more loci. On the basis of LOH pattern in first and subsequent tumors, two common LOH patterns were seen; in pattern A, LOH observed in the original tumor was identical (preserved) to that in the patient’s subsequent tumor(s) and in some cases this was accompanied by additional losses (Fig. 2A); in pattern B, LOH was observed in the first sample but both alleles were retained (loss to retention) in a subsequent sample(s) (Fig. 2B). In some cases, this was accompanied by additional losses. From the patterns of LOH observed, it is suggested that LOH pattern A is consistent with a single monoclonal origin, since within each individual patient, the loss pattern seen in the first tumor is preserved in subsequent samples. An example is shown in subject 23, where loss at the marker *D13S153* is seen in the first (23.1) and subsequent tumors (23.2) with an additional loss at the marker *PYGM* in the recurrence (23.2) consistent with a progressive accumulation of losses with time (Fig. 2A). This single monoclonal interpretation was seen with all the tumor subtypes.

In samples showing retention of heterozygosity in the second or subsequent samples (loss pattern B, Fig. 2), this could most likely be explained by the second sample being derived from a distinctly separate clone, although both are monoclonal tumors. For example, in patient 6 (Fig. 2B), the tumor showed loss at four of the microsatellite markers, whereas in the second sample (6.2), retention of heterozygosity at three of these markers was observed, making it highly unlikely that this sample is clonally related to the first. The possibility that retention of heterozygosity is due to contamination by normal tissue is excluded since LOH is still found in tumor 6.2 at the marker *D13S1246*, indicating monoclonality. Again this interpretation applied to all tumor subtypes. The above interpretation of clonality by X-inactivation



**Figure 2** Examples of LOH in recurrent pituitary tumors representing a common plus additional deletion in the regrown nonfunctional tumor from a 30-year-old woman (A, Patient No. 23, loss pattern A) or a “loss-to-retention” pattern in a 29-year-old man with a nonfunctional tumor (B, Patient No. 6, loss pattern B). In this example, markers interferon- $\alpha$  (9P), D10S217, and D13S155, which are lost in the first sample (T6.1), are retained in the recurrence (T6.2), which is still monoclonal as evidenced by LOH at locus D13S1246. Thus, the second sample is of a different clonal origin than the first. B is the matched blood DNA sample.

analysis in the female patients with informative HUMARA or PGK-1 alleles was confirmed.

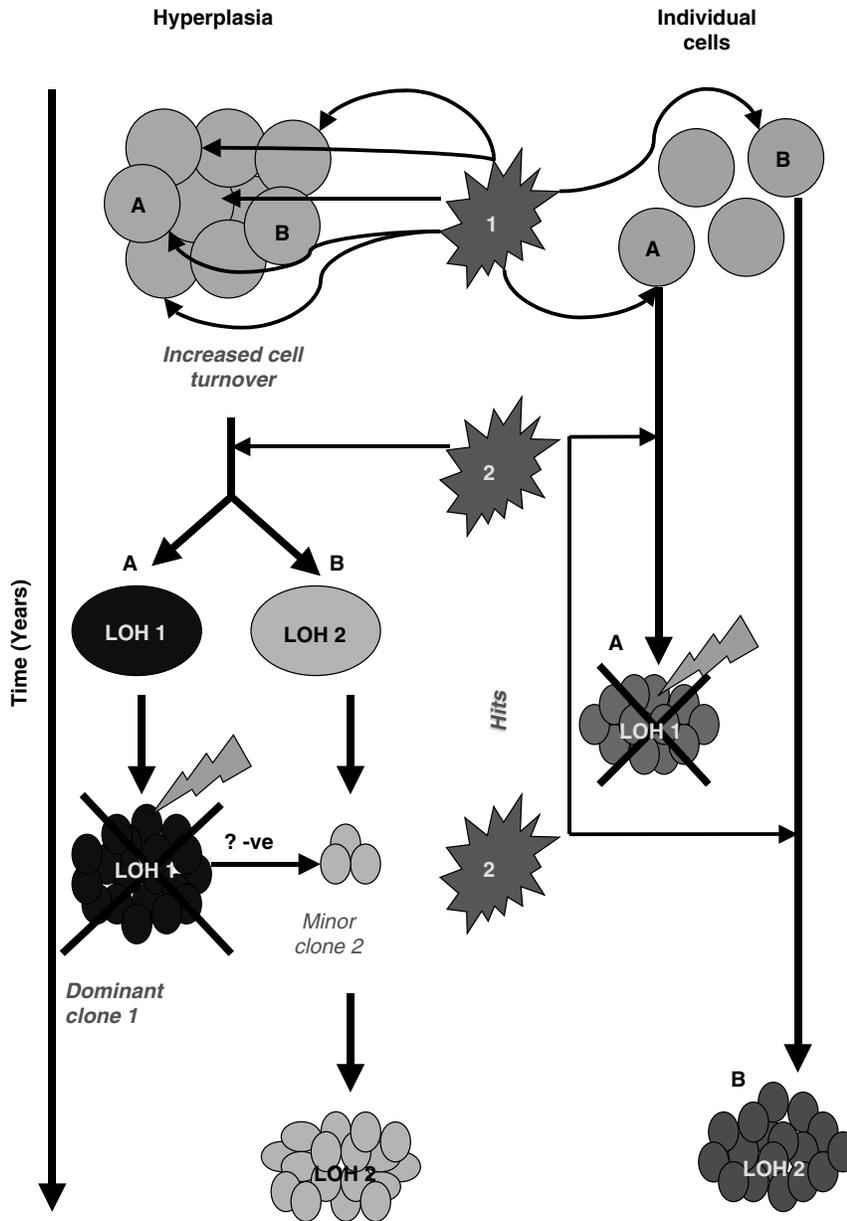
Thus, up to 60% of recurrent/regrown pituitary tumors appear to be clonally distinct from the first tumor. Perhaps even differing clonality of recurrent tumors is the rule rather than the exception.

### Evidence from Individual Cases with Metastases

Further evidence for more than one clone in pituitary tumors is available from rare case reports of carcinomas with metastases. In a woman with recurrent Cushing’s syndrome, the primary pituitary tumor, its recurrence, and a cervical lymph node metastasis were examined. Morphologically, all samples were identical with both ACTH and growth hormone immunopositivity. With the autosomal LOH analysis, the primary and recurrent tumors showed loss of the same allele at four loci, whereas the metastasis showed retention of heterozygosity at two of these loci. This pattern is consistent with a different clonal origin of the metastasis (as in the example in Fig. 2B).

The authors have also examined another case of a nonfunctioning tumor in a man who, 5 years after his initial pituitary surgery, developed dural metastases. All three metastases/seedlings were from the same clone but this was a different clone from the original pituitary tumor. Immunohistochemically, the original tumor and the dural metastases were identical.

How should the aforementioned data be interpreted with respect to tumor initiation/progression? Given that a significant proportion of recurrent/regrown tumors have different clonal origins, are two or more clones present from the very beginning of the tumor? Or does the second independent clone arise later in the progression of the pathogenic process? Figure 3 is an attempt to represent two possible options diagrammatically in relation to hypothetical genetic damage (“hits”). Scenario 1 is on the left-hand side of Fig. 3 and envisages that the first hit results in hyperplasia of several cells of a given lineage. This hit may be exogenous or endogenous to the pituitary. The authors favor an exogenous source, such as hypothalamic stimulation, since this is more likely to be cell subtype specific. If the stimulus to hyperplasia was from within the pituitary, e.g., growth factor overexpression, this might be expected to target cells of several lineages simultaneously. Hyperplasia, with its associated increased rate of cell turnover, then predisposes these cells to a second hit that occurs in one or more cells (A and B), causing different genetic changes (LOH) in each. The changes in cell A produce more rapid cell division than in cell B such that clone A becomes dominant, although clone B grows slowly. At surgery, the dominant clone is removed and this is monoclonal by LOH analysis. Depending on the completeness of surgical removal, a recurrence may derive either from the remnants of clone A, with or without additional genetic damage (LOH), or from clone B. It is also conceivable that clone A might produce factors that inhibit the expansion of clone



**Figure 3** Two possible scenarios to explain the differing clonal origins of recurrent/regrown tumors. Time, in years from initiation, is indicated, proceeding from the top to the bottom of the figure. “Hits” denote events leading to genetic damage and ultimately LOH, without defining what these might be. A large X represents surgical removal and elimination of the relevant clone of the tumor. LOH denotes loss of heterozygosity, with 1 and 2 being different.

B, since growth-inhibiting as well as growth-stimulating factors may be produced from pituitary tumors. Scenario 2 is on the right-hand side of Fig. 3 and is not predicated by preexisting cellular hyperplasia but assumes that the first hit targets two individual normal cells of the same subtype contemporaneously. The second hit then affects only cell A (not cell B at this

time), enabling clone A to develop while cell B does not expand into even a minor clone but retains its predisposition to expand if it sustains its second hit at a time before or after removal of clone A (the initial tumor). The authors believe this scenario to be less likely than the first in most cases for the following reasons:

1. Statistically, the first hit would be expected to randomly target any cell of any lineage, not cells of the same lineage, unless this hit was cell subtype specific. If this were the case, clone tumor A and clone tumor B would be expected to be of different subtypes. Clinically, this is rare since most recurrences/regrowths are of the same cell subtype as the original tumor, with very few rare exceptions, e.g., silent corticotrophinomas, which are originally nonfunctional (though by definition ACTH immunopositive) and then change phenotype to become highly functional).

2. If cell B sustained genetic damage and did not expand until some time (months to years) later, there would be a greater chance that it would be eliminated through apoptosis, which in fact may be enhanced because of the genetic change and which does occur in the pituitary albeit at a slow rate.

The results from the two cases with metastases are perhaps more difficult to interpret. In the first case, the lymph node metastasis (albeit in the territory of jugular vein drainage) must presumably have arisen by hematogenous spread. Did this occur at the time of primary or second surgery to the pituitary? The answer cannot be determined, and the interpretation is compatible with either contemporaneous or later independent clonal development.

The second case is perhaps easier to interpret in that it is suspected that the dural deposits arose as a result of seedlings shedding at the time of first surgery into the cerebrospinal fluid (CSF) (the patient did have a transient spontaneously closing CSF leak after surgery). If this is the case, then there must have been at least two clones present at the time of the first surgery, though clearly one was dominant given that the initial tumor was monoclonal (or it was multiclonal and the fragment that was analyzed was from only one of the clones).

## EVIDENCE OF DIFFERING CLONAL ORIGIN FROM OTHER TUMOR TYPES

One crucial question to be answered in LOH and clonal analysis studies relates to that of sampling bias. How does one know that the extracted DNA, especially from a small biopsy represented on a slide, is representative of the whole tumor? Intratumoral DNA heterogeneity has been found in prostate, breast, and colorectal carcinomas by LOH analysis. How are these data to be interpreted? One distinct possibility is that these solid tumors are composed of several distinct clones of cells, some of which may have greater malignant potential than others. This might explain why in

the aforementioned cancers the prognostic significance of DNA aberrations has been inconclusive. One of these studies formally assessed clonal composition by X-chromosome inactivation and showed that separate areas within a given tumor showed an identical X-inactivation pattern, although LOH analysis was variable within a given tumor. The authors would interpret this as a single clonal origin for the tumor but different subclones sustaining accumulating genetic damage as cells within the tumor progress at different rates. What has not yet been determined is whether similar genetic heterogeneity applies in larger pituitary tumors, although studies on patients with metastases suggest that this is a possibility.

Is it a surprise that many pituitary tumors might be multiclonal? In the sense that these are generally benign adenomas rather than carcinomas, perhaps it is, but there are several examples in other tumors, some benign and some malignant, of different clonal derivations. In multinodular goiters, coexisting individual nodules may be either polyclonal or monoclonal and if monoclonal, the clones may be different. This clearly indicates that whatever the goitrogen(s), it is capable of targeting several cells of the same lineage, possibly contemporaneously. Another study indicated that solitary follicular thyroid adenomas or carcinomas are monoclonal although nodules from multinodular goiters are largely polyclonal.

Although solitary parathyroid adenomas are monoclonal, monoclonal parathyroid tissue is common in hyperplastic glands from patients with uremic hyperparathyroidism as well as in patients with primary parathyroid hyperplasia. Histological categories of nodular versus diffuse hyperplasia did not predict clonal status. Thus, the chronic stimulation of uremia induces hyperplasia followed by monoclonal expansion from this pool of cells in the parathyroid gland. So why could the same thing not happen in the pituitary? In contrast, a polyclonal pattern of X-chromosome inactivation was observed in medullary carcinomas of the thyroid from both MEN2 patients and sporadic cases. This is a rather unexpected finding given earlier studies that were more suggestive of a single clone of origin of MTC with distinctly progressive subclones. In benign pancreatic endocrine tumors, polyclonality is the rule and progresses to monoclonality as a single, more aggressive clone develops the potential for invasiveness and metastatic spread. In this example, too, the clonality pattern could not be predicted on the basis of histochemistry, proliferation index, or growth pattern. Somewhat surprisingly, sporadic gastrinomas occurring at multiple sites in the same patient appear to have the

same clonal origin, whereas in MEN1 patients multiple gastrinomas may be of independent clonal origin.

The presence of precursor populations of cells with different clonal composition appears quite commonly in other organs prior to the development of invasive cancer (e.g., cervix, bladder cancers), although when these cancers are multifocal they appear to have the same clonal origin. Similarly, in prostatic intraepithelial neoplasia and multifocal prostate cancer, the distribution of LOH indicates areas of similar and differing clonality.

The aforementioned evidence indicates several things about cancer clonality:

1. morphology cannot predict clonality;
2. clonality within a given tumor may be multiple or single;
3. multiple tumors arising on the background of hyperplasia may be of identical or differing clonality;
4. multiple "sporadic" tumors within an organ may be of differing clonal origin.

One should therefore not be surprised that pituitary tumors may consist of an admixture of clones and that their recurrences/regrowths may be from any one of such clones. Thus, although the early available evidence indicated that pituitary tumors appear largely monoclonal, it is simplistic to assume that this is inevitable and that these cannot be multiclonal at the outset. These observations would be entirely compatible with an initiating stimulus resulting in cell subtype-specific hyperplasia, which itself gives rise to several distinct clones with variable potential to

develop into tumors. Such stimuli might include hypothalamic tropic factors, intrapituitary growth factors, or pituitary-specific oncogenes.

### See Also the Following Articles

Hamartoma, Pituitary • Hypopituitarism • Hypopituitarism, Hormonal Therapy for • Pituitary Adenomas, TSH-Secreting • Pituitary Region, Non-Functioning Tumors of • Pituitary Tumors, ACTH-Secreting • Pituitary Tumors, Molecular Pathogenesis • Pituitary Tumors, Surgery • Prolactinoma, Diagnosis • Prolactinoma, Pathogenesis

### Further Reading

- Asa, S. L., and Ezzat, S. (1998). The cytogenesis and pathogenesis of pituitary adenomas. *Endocr. Rev.* **19**, 798–827.
- Clayton, R. N., and Farrell, W. E. (2001). Clonality of pituitary tumors: More complicated than initially envisaged? *Brain Pathol.* **11**, 313–327.
- Clayton, R. N., Pfeifer, M., Atkinson, A. B., Belchetz, P., Wass, J. A. H., Kyrodinou, E., Vanderpump, M., Simpson, D., Bicknell, J., and Farrell, W. E. (2000). Different patterns of allelic loss (see of heterozygosity) in recurrent human pituitary tumors provides evidence for multiclonal origins. *Clin. Cancer Res.* **6**, 3973–3982.
- Dahia, P. L. M., and Grossman, A. B. (1999). The molecular pathogenesis of corticotroph tumors. *Endocr. Rev.* **20**, 136–155.
- Derwhal, M., and Studer, H. (2002). Hyperplasia versus adenoma in endocrine tissues: Are they different? *Trends Endocrinol. Metab.* **13**, 23–28.
- Farrell, W. E., and Clayton, R. N. (2000). Molecular pathogenesis of pituitary tumors. *Front. Neuroendocrinol.* **21**, 174–198.
- Fearon, E. R., and Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. *Cell* **61**, 759–767.
- Levy, A. (2000). Is monoclonality in pituitary adenomas synonymous with neoplasia? *Clin. Endocrinol.* **52**, 393–397.

suggesting that continuous stimulation by thyrotropin-releasing hormone (TRH) may lead to thyrotroph adenoma. TRH has also been shown to be expressed in the pituitary and by the different types of pituitary adenomas.

TRH signaling appears to be intact in pituitary adenomas as evidenced by intact binding and normal release of thyroid stimulating hormone (TSH) and prolactin (PRL) in response to TRH exposure. TRH receptor expression and structure are grossly unaltered even in thyrotroph adenomas. TRH mRNA is alternatively spliced in some pituitary tumors with deletion of exon 3 resulting in a truncated product that does not bind TRH. The relatively higher levels of the truncated forms compared to the full-length form of the TRH receptor in lactotroph adenomas may explain some of the paradoxical *in vivo* responses to TRH administration.

### **Growth Factors and Their Receptors**

#### *Transforming Growth Factor- $\alpha$*

Transforming growth factor- $\alpha$  (TGF- $\alpha$ ) is expressed as a membrane-anchored protein by human adenohypophysial cells and tumors. TGF- $\alpha$  may alter pituitary production of GH, PRL, and TSH as well as cell proliferation. Estrogen stimulation has been implicated in pituitary tumorigenesis and TGF- $\alpha$  appears to mediate some estrogenic effects. Targeted overexpression of TGF- $\alpha$  to the pituitary results in lactotroph adenomas, providing compelling evidence for the significance of this growth factor in pituitary tumorigenesis.

#### *Epidermal Growth Factor and Its Receptor*

Epidermal growth factor (EGF) is detectable by immunohistochemistry in most adenohypophysial cells and its mRNA is expressed with marked variation in all types of functional and nonfunctional adenomas. EGF potently stimulates PRL and adrenocorticotrophic hormone secretion with variable effects on rat pituitary cell proliferation. The selective expression and specific effects of EGF suggest that the pituitary is an important target site for this growth factor's action.

The common receptor of EGF and TGF- $\alpha$ , EGF receptor (EGF-R), is a 170 kDa plasma membrane tyrosine kinase product of the proto-oncogene *v-erbB*. EGF-R is overexpressed in several types of human cancers and in most instances this overexpression is accompanied by TGF- $\alpha$  expression; expression of this receptor appears to correlate with tumor aggressiveness. EGF-R is expressed by human pituitary adenomas with the highest levels in recurrent somatotroph adenomas and suggesting a selective mechanism for the EGF/EGF-R family in the growth of aggressive

somatotroph tumors. The importance of the EGF-R in the somatotroph is further supported by the selective loss of somatotrophs in transgenic mice overexpressing a dominant-negative EGF-R mutant lacking the intracellular protein kinase domain (EGFR-tr). These findings point to EGF-R as an integral component in the differentiation and proliferation of somatotrophs.

#### *Fibroblast Growth Factors and Their Receptors*

Basic fibroblast growth factor (bFGF, also known as FGF-2) is one of an expanding family of at least 23 members of ligands with variable mitogenic, angiogenic, and hormone regulatory effects. FGF-2 immunoreactivity was described originally in non-hormone-producing bovine pituitary folliculo-stellate cells and has since been shown to regulate GH, PRL, and TSH secretion by the rodent pituitary. In the human pituitary, in contrast, FGF-2 is produced by adenohypophysial cells that constitute pituitary adenomas. Pituitary-derived FGF-2 has been shown to stimulate replication of PRL-secreting cells but also may inhibit DNA synthesis in pituitary adenoma cells, suggesting that some forms of the growth factor or its receptor may act as growth inhibitors. Elevated concentrations of FGF-2-like immunoreactivity have been documented in patients with multiple endocrine neoplasia type 1 (MEN1) and in patients with sporadic pituitary adenomas. The FGF-related *hst* has been found in transforming DNA of human PRL-secreting tumors and facilitates lactotroph proliferation. Transgenic mice expressing FGF-2 under the control of the GH and the  $\alpha$ -subunit promoters develop hyperplasia of several adenohypophysial cell types but not frank adenomatous transformation. FGFs or homologous family members may, therefore, play an important role in pituitary tumorigenesis.

There are four mammalian FGF receptor (FGFR) genes encoding a complex family of transmembrane receptor tyrosine kinases. Each prototypic receptor is composed of three immunoglobulin-like extracellular domains (two of which are involved in ligand binding), a single transmembrane domain, a split tyrosine kinase, and a COOH-terminal tail with multiple autophosphorylation sites. Multiple forms of cell-bound or secreted receptors are produced by the same gene. Tissue-specific alternative splicing, variable polyadenylation sites, and alternative initiation of translation result in truncated receptor forms. Structural alterations of FGFRs may play a role in human tumorigenesis. Antisense targeted interruption of FGFR1 reduces malignant melanoma cell proliferation and differentiation. The normal pituitary expresses mRNAs for FGFR1, 2, and 3. An interesting finding was the

presence of novel truncated mRNA for a kinase-containing variant of FGFR4 with an alternative initiation site in pituitary adenomas. This tumor-derived kinase encodes a polypeptide receptor that lacks a signal peptide and the first two extracellular immunoglobulin-like domains. *ptd*-FGFR4 has a distinctive cytoplasmic localization and is constitutively phosphorylated. Transfected cells display transforming features *in vitro* in soft agar and develop tumors *in vivo* in nude mice. More importantly, targeted expression of *ptd*-FGFR4 to pituitary lactotrophs, but not wild-type FGFR4, results in hyperprolactinemia and pituitary adenomas that morphologically mimic primary human pituitary adenomas. These findings suggest that dysregulated FGFR4 function plays an important role in pituitary tumorigenesis and that FGFR4 is a candidate tumor-specific kinase.

## Loss of Function

### p53

This nuclear protein is critical in the control of G<sub>1</sub>/S phase progression by regulating the gene expression of p21, the first of the group of cyclin-dependent kinase inhibitors to be identified. Following DNA damage, p53 can induce growth arrest during the G<sub>1</sub> or G<sub>2</sub> phase of the cell cycle or by stimulating apoptosis, thereby protecting the normal cell from replicating damaged DNA. Conversely, p53 mutations render the cell genome more vulnerable to mutations that represent the rate-limiting steps in tumor progression. Indeed, inactivating mutations of p53 have proven to be among the most commonly encountered gene alterations in human malignancies. However, studies have failed to identify p53 mutations at putative "hotspots" in the different types of human pituitary adenomas.

### The Retinoblastoma Gene

The retinoblastoma (Rb) gene is another member of the family of tumor suppressor genes that has been implicated in several neoplasms including retinoblastoma and osteosarcoma. Mice heterozygous for an Rb mutation develop pituitary tumors of intermediate lobe corticotroph differentiation. No such mutations have been identified in human pituitary adenomas. Instead, there is loss of heterozygosity (LOH) at sites telomeric and centromeric to the Rb locus in some aggressive pituitary adenomas. These data argue for an independent tumor suppressor gene (TSG) on 13q, which is closely linked with but distinct from Rb. In addition, methylation of the Rb promoter region has been described in association with diminished detectable Rb protein expression.

### Cyclins, Cyclin-Dependent Kinases, and Cyclin-Dependent Kinase Inhibitors

A series of cyclins and cyclin-dependent kinases (cdks) plays a central role in the regulation of cell cycle progression. The cyclin D (cdk4) and E (cdk2) complexes are catalytically active during the late G<sub>1</sub> phase and are implicated in the regulation of G<sub>1</sub>/S progression. The Rb protein is one of the putative substrates of the cdks. Rb phosphorylation abrogates the ability of these proteins to inhibit transactivation of transcription factors important in cell cycle control. In turn, cdk activity is modulated by cdk inhibitors. These include p27<sup>kip1</sup>, p57<sup>kip2</sup>, p16<sup>ink4A</sup>, p15<sup>ink4B</sup>, p18<sup>ink4c</sup>, and p19<sup>ink4D</sup>.

Prompted by the potentially pivotal role of the Rb protein in regulating pituitary cells and the negative findings involving the *Rb* gene itself, factors governing Rb phosphorylation became obvious candidates. Specifically, cdk inhibitors have received extensive attention. As with the Rb itself, p16 has not been found to be mutated in pituitary adenomas, but is frequently silenced by means of extensive methylation in human pituitary adenomas. These findings suggest an alternative mechanism in modulating Rb-related protein control of the pituitary tumor cell cycle.

Mice lacking p27<sup>kip1</sup> have an increased propensity for the development of multiorgan neoplasia including pituitary tumors. As with the *Rb* gene, however, mutations of the p27<sup>kip1</sup> gene do not appear to play a role in human pituitary tumorigenesis.

### MEN1

According to the two-hit model of tumorigenesis, both copies of a gene situated on opposite alleles must be inactivated, such as by deletion, rearrangement, or silencing through methylation, to confer a selective growth advantage to a precursor cell, which may subsequently proliferate in a clonal neoplastic fashion. Such genes that require homozygous inactivation of both copies are termed TSGs, anti-oncogenes, or recessive oncogenes. Examples of TSGs include Rb and p53. In the affected tumors, allelic loss of the gene in question is invariably noted.

After a decade of search, the MEN1 gene was identified by positional cloning with germ-line mutations characterized in kindreds with familial MEN1. The gene product was coined menin to describe the 610-amino-acid protein product. Targeted gene inactivation of menin results in prolactinomas characterized by somatic loss of the wild-type allele with remarkable similarity to humans with MEN1. As predicted, mutations in sporadic pituitary tumors have been described.

One copy of the MEN1 gene was found to be deleted in 10% of sporadic tumors examined. Missense mutations have been described throughout the gene of the remaining allele. Thus, as predicted by Knudson's two-hit tumor suppressor gene hypothesis, mutational inactivation of one copy of MEN1 coupled with deletion of the second allele strongly indicates that this gene is involved in the pathogenesis of hereditary (familial) pituitary tumors and in a subset of sporadic pituitary tumors. Up to 20% of sporadic pituitary adenomas exhibit LOH at the 11q13 locus even in the absence of menin mutations, suggesting additional TSG at this locus. The characterization of normal menin function will likely prove to be of evolving significance in pituitary tumorigenesis. Thus far, menin has been shown to modulate a number of growth-regulatory pathways including that of the inhibitory transforming growth factor- $\beta$  (TGF- $\beta$ ) as a result of interaction with the Smad3 transcription factor.

## Loss of Inhibitory Hormones

### Dopamine

The role of diminished hypothalamic inhibition was first suggested based on the observation of neovascularization in lactotroph adenomas. It was speculated that neovascularization would allow lactotrophs to escape from tonic dopaminergic inhibition. Dopamine signal transduction is mediated through D1 receptors that stimulate adenylyl cyclase activity and D2 receptors (D2R) that inhibit this enzyme. The family of dopamine receptors is complex in biochemical, physiological, and pharmacological diversity. Nevertheless, it appears that the predominant anterior pituitary dopamine receptor is the D2R. Targeted elimination of D2R activity in D2R knockout mice results in lactotroph hyperplasia and, subsequently, lactotroph adenoma formation in female D2R-deficient mice. Interestingly, these lesions are monohormonal PRL-immunoreactive neoplasms that display a characteristic juxtannuclear Golgi pattern of PRL staining and loss of the reticulin fiber network. Several of these adenomas have been noted to be much larger than normal glands, with marked suprasellar extension and invasion of brain but no gross evidence of distant metastases.

Although some pituitary tumors have been shown to be responsive to dopamine suppression, the dopaminergic resistance that is found in some tumors implicates diminished D2R activity as a putative factor in pituitary tumorigenesis. Thus far, however, investigation of the D2R gene has revealed it to be structurally intact in human lactotroph adenomas.

### Prolactin

The autoregulation of PRL secretion in the rat has been demonstrated at the levels of both the hypothalamus and the pituitary. In transfected somatolactotrophs, PRL exerts a strong and specific inhibition of PRL gene transcription. These effects appear to be mediated by intermediate and long forms of the PRL receptor. Deletions of the PRL promoter indicated that the autoregulatory effect of PRL requires the same regulatory domains that have been described for the other PRL gene regulators. These studies support an extra-short loop of PRL transcriptional autoregulation. In the normal human pituitary, PRL and luteinizing hormone cells exhibit the highest levels of PRL receptor (PRL-R) mRNA expression, whereas PRL-producing pituitary adenomas have significantly higher levels of PRL-R mRNA than other types of adenomas. Dopamine agonist treatment decreases the levels of PRL-R mRNA in PRL adenomas.

### Somatostatin

GH secretion is under contrasting influence from hypothalamic stimuli including GHRH, which stimulates GH secretion, and somatostatin (SS), which inhibits GH secretion. Specific receptors for SS (SSTRs) are expressed on somatotroph adenomas. Earlier studies suggested a relationship between the density of SS receptors on GH tumors and the secretory response to this analogue both *in vitro* and *in vivo*. Binding sites for SS, however, have been identified by autoradiography in tumors resistant to the GH-lowering effects of octreotide. These findings are consistent with differential adenylyl cyclase coupling by the five known subtypes of SSTRs and their heterogeneous expression in pituitary adenomas. Expression of SS itself in large invasive GH tumors appears to be diminished compared to that in the normal pituitary. These findings suggest multiple paracrine, autocrine, and endocrine mechanisms for SS-mediated control of somatotroph function and proliferation.

### Thyroid Hormones

The development of pituitary thyrotroph adenomas in patients with prolonged primary hypothyroidism has provided further evidence supporting the hypothalamic role in pituitary tumorigenesis. Thyroid hormones mediate their actions via nuclear thyroid hormone receptors (TRs) that bind to specific regulatory hormone-response elements. There are two major classes of TRs, designated  $\alpha$  and  $\beta$ , each of which undergoes alternative splicing to generate  $\alpha 1$  and  $\alpha 2$  and  $\beta 1$  and  $\beta 2$  isoforms. With the exception of the  $\beta 2$  form, which is predominantly expressed in the

hypothalamic–pituitary system, these receptor isoforms are ubiquitously expressed. Of interest, in the pituitary, the  $\beta 1$  and  $\beta 2$  isoforms appear to be expressed to a lesser extent in adenomas than in the normal gland. Screening TR- $\alpha$  mRNA identified three novel missense mutations, two in the common TR- $\alpha$  region and another that was  $\alpha 2$  specific. TR- $\beta$ -response elements failed to show any differences from published sequences.

Mice with targeted disruption of the entire TR- $\beta$  locus exhibit elevated thyroid hormone levels as a result of abnormal central regulation of thyrotropin but do not develop pituitary tumors. Thus, the putative differential hormone regulatory and mitogenic effects of the different THR isoforms in the pituitary remain unclear.

### TGF- $\beta$

TGF- $\beta$  has been implicated in the regulation of normal and neoplastic cell function. TGF- $\beta$  regulates the expression of various proteins, including the cell cycle inhibitory protein p27Kip1 (p27). TGF- $\beta$  1/2/3 isoforms and the TGF- $\beta$  receptor are expressed in normal and adenomatous pituitaries. Dispersed pituitary adenomas cells show a biphasic response to TGF- $\beta$  with changes in follicle-stimulating hormone (FSH) secretion. The TGF- $\beta$  family, however, is represented in at least three different forms in the pituitary. Inhibins and activins consist of two homo- or heterodimeric polypeptide subunits derived from a common precursor; inhibin A ( $\alpha$ - $\beta$ A) and inhibin B ( $\alpha$ - $\beta$ B) selectively inhibit the release of FSH from pituitary gonadotroph cells, whereas activin ( $\beta$ A- $\beta$ B), activin A ( $\beta$ A- $\beta$ A), and activin B ( $\beta$ B- $\beta$ B) stimulate its release. Inhibin subunits are expressed by pituitary gonadotroph adenomas and activin is known to stimulate

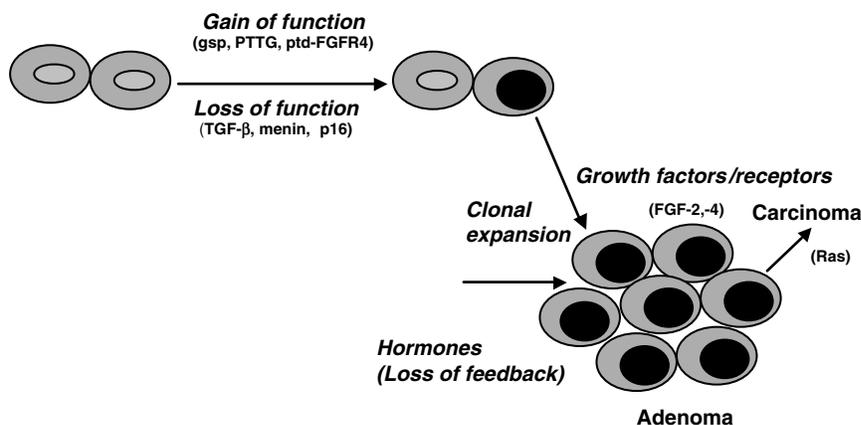
hormone secretion by these tumors. Activin, however, inhibits cell proliferation in a variety of cell types including pituitary adenomas. These actions are mediated through type I and type II serine/threonine kinase receptors. The predominant form is type I (ActRIB or Alk4). Truncated forms of the type I receptors expressed by gonadotroph adenomas interfere with wild-type receptor function and block the anti-proliferative effect of activin. These findings suggest an important role for the activin/follistatin/activin receptor balance in modulating pituitary tumor cell replication.

### Gonadal Steroids

The development of pituitary gonadotroph adenomas in patients with prolonged primary hypogonadism suggests that the lack of hormone negative feedback may facilitate pituitary tumor development. Again, however, the role of GnRH stimulation cannot be easily distinguished from that of gonadal hormone inhibition in the development of these adenomas.

## CONCLUSIONS

Accumulating evidence suggests that the control of pituitary cell multiplication rests on a precarious balance between growth factor stimulation and inhibition with dominant contributions from the FGF/FGFR and TGF- $\beta$  systems (Fig. 1). The events leading to pituitary tumorigenesis in the vast majority of tumors remain to be validated directly from the human situation in genetic mouse models. Differences between pituitary and other endocrine and nonendocrine neoplasms will likely emerge to explain the remarkably varied spectrum of biologic behavior in terms of tissue



**Figure 1** Proposed model of pathophysiology of pituitary tumorigenesis. An integrated approach incorporating the dominant gain and loss of function cascade events is summarized.

invasiveness and metastatic potential, which is unique to pituitary adenomas.

### Acknowledgment

This work was supported in part by grants from the Canadian Institutes of Health Research (MT-14404).

### See Also the Following Articles

Adrenal Tumors, Molecular Pathogenesis • EGF and Related Growth Factors • Fibroblast Growth Factor (FGF) • Genetic Testing for Pituitary Disease • Pineal Tumors • Pituitary Region, Non-Functioning Tumors of • Pituitary Tumors, ACTH-Secreting • Pituitary Tumors, Clonality • Pituitary Tumors, Surgery • Transforming Growth Factor (TGF) Alpha • TSH Receptor (Thyrotropin Receptor)

### Further Reading

- Asa, S. L., and Ezzat, S. (2002). Pathogenesis of pituitary tumors. *Nat. Rev. (Cancer)* **2**, 836–849.
- Crabtree, J. S., *et al.* (2001). A mouse model of multiple endocrine neoplasia, type 1, develops multiple endocrine tumors. *Proc. Natl. Acad. Sci. USA* **98**, 1118–1123.
- Ezzat, S., Zheng, L., Zhu, X.-F., Wu, G. E., and Asa, S. L. (2002). Targeted expression of a pituitary tumor-derived FGF receptor 4 isoform recapitulates pituitary tumorigenesis. *J. Clin. Invest.* **109**, 69–78.
- Kelly, M. A., *et al.* (1997). Pituitary lactotroph hyperplasia and chronic hyperprolactinemia in dopamine D2 receptor-deficient mice. *Neuron* **19**, 103–113.
- Landis, C. A., Masters, S. B., Spada, A., Pace, A. M., Bourne, H. R., and Vallar, L. (1989). GTPase inhibiting mutations activate the  $\alpha$  chain of  $G_s$  and stimulate adenylyl cyclase in human pituitary tumours. *Nature* **340**, 692–696.
- Nakayama, K., *et al.* (1996). Mice lacking p27<sup>Kip1</sup> display increased body size, multiple organ hyperplasia, retinal dysplasia, and pituitary tumors. *Cell* **85**, 707–720.

endocrinologically, have aided intraoperative orientation, and have improved intraoperative visualization.

### Imaging

The major advance in surgical planning over the past 50 years has been the introduction of sophisticated radiographic imaging techniques. Computerized tomography was introduced in the early 1980s, followed by magnetic resonance imaging (MRI) shortly thereafter. Current MRI scans will delineate all but the smallest microadenomas and demonstrate their proximity to surrounding structures. MRI scanners have been incorporated into the operating room environment and can provide near-real-time MRI for both intraoperative orientation and determination of the adequacy of resection. Whether this expensive technology leads to improvements in patient care and outcome remains to be shown.

### Surgical Techniques

Advances in surgical techniques have aided both intraoperative orientation and visualization. Navigational devices that will correlate intraoperative localization with preoperative imaging have become available and may offer some benefits when the traditional anatomical landmarks have been destroyed, as in cases of recurrent tumors. Improved visualization with a wider field of view and less tissue dissection is offered by the use of the endoscope, though traditional operative microscopy offers better optics, depth of field, and binocular vision.

## SURGICAL RESULTS

### Secretory Tumors

#### **Cushing's Disease**

Transsphenoidal surgery remains the primary mode of treatment for Cushing's disease, with cure rates approaching 90% in experienced hands. After surgical cure, patients will demonstrate profound hypoadrenalism for some months postoperatively and require cortisol replacement during this time. Tumors may recur at a rate of 7–10% at 5 years; reoperation for tumor recurrence is less successful, with remission rates of approximately 50%. Here, adjuvant therapy, including radiosurgery, adrenalectomy, or medical therapy with ketoconazole, may play a role. Since MRI is able to detect microadenomas in this disease in at best 60–70% of cases, accurate endocrine diagnosis, including inferior petrosal sinus catheterization, is necessary.

#### **Acromegaly**

Patients with acromegaly tend to present with larger, more invasive tumors and here transsphenoidal surgery is less successful. The remission rate is 50–60% in macroadenomas, whereas the remission rate may approach 90% in the less common microadenomas. Newer medical treatments, including the use of somatostatin analogues, are useful in controlling growth hormone levels in those patients not surgically cured and aggressive multimodality therapy (e.g., surgery, somatostatin analogues, radiosurgery) is necessary to control the increased mortality risk associated with the disease. The role of medical pretreatment, or primary medical therapy for acromegaly, remains unclear, but has been suggested for those patients who are not optimal surgical candidates or who have a minimal chance of surgical cure.

#### **Prolactinomas**

Women who present with amenorrhea–galactorrhea, mild elevations in prolactin level, and microadenomas on MRI scanning are usually best treated with dopamine agonist therapy. Surgery, though effective in approximately 90% of patients with microadenomas, is reserved for those patients who are unable to tolerate or who fail to respond to medical treatment. With newer, better tolerated dopamine agonists (cabergoline), patients only rarely require surgical treatment. Men with prolactinomas tend to present with either impotence or visual abnormalities and their tumors are more likely to be large and invasive. Dopamine agonist therapy remains effective in these cases, with surgery reserved for those cases that fail to respond. Even patients with bitemporal field defects will often respond dramatically after a few weeks of treatment with a dopamine agonist. It must be remembered that mild elevations in prolactin (less than  $\approx 200$  ng/ml) may have causes other than prolactin-secreting tumors (e.g., pregnancy, medication effects). A large tumor with a mild increase in prolactin level may represent the so-called “stalk effect” from compression of the pituitary stalk and normal gland, resulting in blockade of intrinsic dopamine inhibition and an elevated prolactin level. This tumor is unlikely to be a prolactinoma and may well require surgical removal. Dopamine agonist therapy in these cases may well normalize the prolactin level, but will do nothing to shrink the tumor.

### Nonfunctioning Tumors

The term “non-functioning” may well be a misnomer, as many of these tumors secrete inactive hormone fragments, especially the  $\alpha$ -subunit. Because of their

endocrine inactivity, they become large and invasive before being detected. They usually present with signs of mass effect, either with visual abnormalities, especially the classic bitemporal field defect of chiasm compression, or with endocrine abnormalities from destruction of the remaining gland. Surgery is required to decompress the optic apparatus and approximately 70% of patients will regain some degree of peripheral vision. Decompression of the remaining gland results in improvements in hormone function in approximately 50% of these patients. A small amount of residual tumor may be followed with serial MRIs, whereas an extensive amount of residual should be treated with radiotherapy. An incidentally found nonfunctioning tumor (the “incidentaloma”) has been reported in up to 10% of MRI studies. Microadenomas (<1 cm) can usually be followed with serial MRI; macroadenomas (>1 cm) often merit resection, though some would follow for a time with MRI to determine the rate of growth before recommending surgery.

## COMPLICATIONS

Transsphenoidal surgery in experienced hands carries a low risk of serious complications. New hormone insufficiency may occur in 5 to 10% of cases, although this may be considerably greater in cases of Cushing’s disease, in which extensive exploration of the gland is sometimes required to find a tiny microadenoma. The incidence of permanent diabetes insipidus is approximately 1–2%. Serious complications are unlikely. Carotid artery injury occurs at a rate of approximately 0.1%, whereas permanent postoperative visual decline occurs at a rate of approximately 0.5%. Postoperative CSF rhinorrhea is seen in approximately 1–2% of cases and may lead to meningitis if

untreated. Approximately 5% of patients report persistent sinus disease postoperatively.

## CONCLUSION

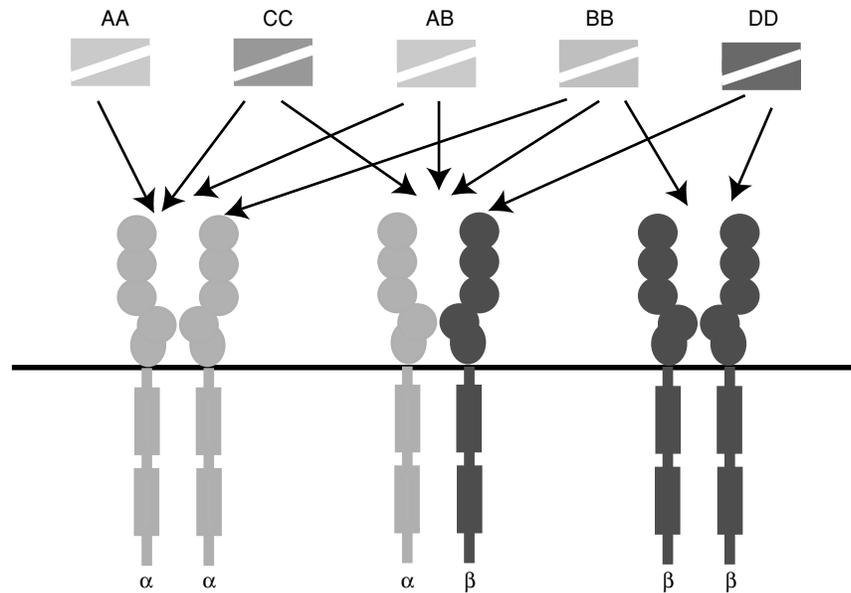
Transsphenoidal surgery remains the therapy of choice for nonfunctioning adenomas, especially in cases of optic chiasm compression or hormonal insufficiency. Secretory tumors, except for prolactinomas, require surgical therapy, although the treatment of acromegaly is in flux.

## See Also the Following Articles

Acromegaly, Clinical Features of • Parathyroid Surgery • Pituitary Tumors, ACTH-Secreting • Pituitary Tumors, Clonality • Pituitary Tumors, Molecular Pathogenesis • Prolactinoma, Clinical Manifestations • Prolactinoma, Pathogenesis • Prolactinoma, Therapy

## Further Reading

- Ciccarelli, E., Giusti, M., Miola, C., *et al.* (1989). Effectiveness and tolerability of long-term treatment with cabergoline, a new long-lasting ergoline derivative, in hyperprolactinemic patients. *J. Clin. Endocrinol. Metab.* **69**, 725–731.
- Ciric, I., Ragin, A., Baumgartner, C., and Pierce, D. (1997). Complications of transsphenoidal surgery: Results of national survey, review of the literature, and personal experience. *Neurosurgery* **40**, 225–237.
- Cushing, H. (1932). The basophil adenomas of the pituitary body and their clinical manifestations (pituitary basophilism). *Bull. Johns Hopkins Hosp.* **50**, 137–195.
- Landolt, A. M., and Schiller, Z. (1996). Surgical technique: Transsphenoidal approach. In “Pituitary Adenomas” (A. M. Landolt, M. L. Vance, and P. L. Reilly, eds.), pp. 315–332. Churchill Livingstone, New York.
- Melmed, S., Casanueva, F. F., Cavagnini, F., *et al.* (2002). Consensus: Guidelines for acromegaly management. *J. Clin. Endocrinol. Metab.* **87**, 4054–4058.



**Figure 1** Binding of the five isoforms of PDGF to the three dimeric receptor complexes. The antiparallel arrangement of subunits in PDGFs and the key features of the receptors (i.e., five extracellular Ig-like domains and the split tyrosine kinase domain) are schematically illustrated, together with binding specificities.

### Ligand–Receptor Interactions

The signaling ligand–receptor complex is composed of one PDGF dimer and a dimeric receptor complex, and it is stabilized by ligand–receptor and receptor–receptor interactions (Fig. 2). Both subunits of the ligand dimer contribute to the two binding epitopes. Amino acid residues in all three loops participate in receptor interactions. The receptor regions involved in ligand interactions are incompletely defined but appear to predominantly involve Ig domains 2 and 3. In addition to being kept together by the divalent ligand, the receptor dimer is stabilized by homophilic Ig domain 4-mediated receptor–receptor interactions.

## PDGF RECEPTOR SIGNALING

### PDGF Receptor Activation

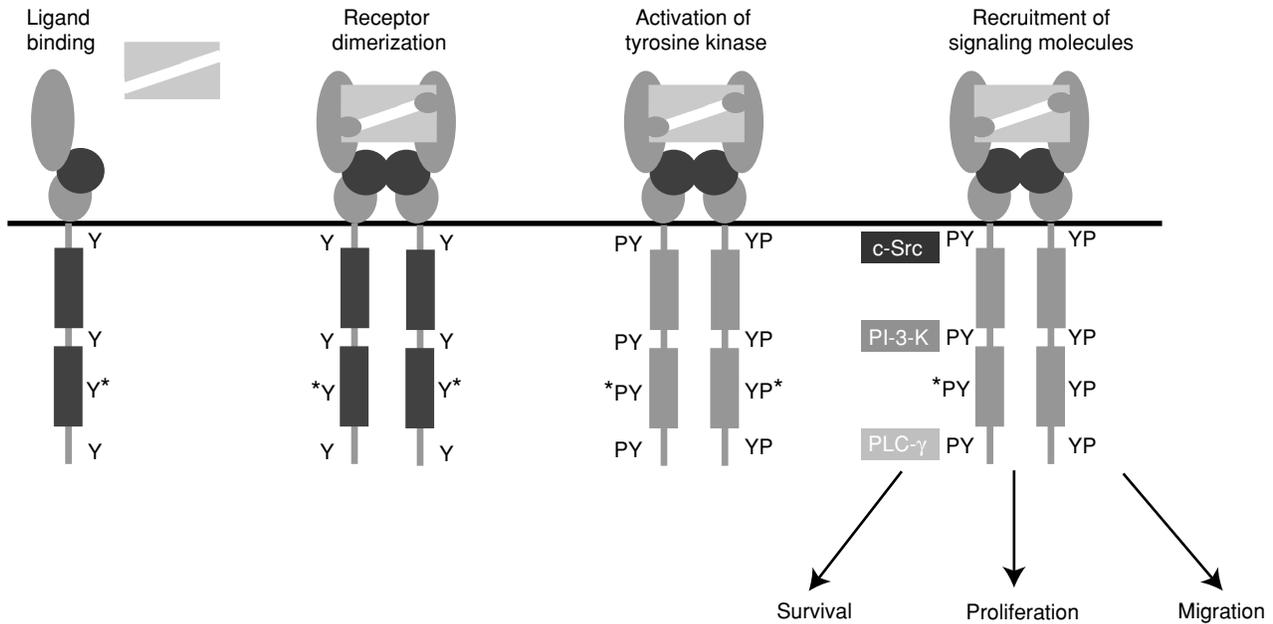
In the absence of ligands, PDGF receptors occur as monomeric forms with low basal tyrosine kinase activity, caused by a conformation unfavorable for ATP binding. Ligand binding and subsequent receptor dimerization lead to tyrosine kinase activity by stimulating transphosphorylation of the regulatory tyrosine in the activation loop of the kinase domain (Fig. 2). This is followed by phosphorylation of multiple tyrosine residues in the kinase insert, the juxtamembrane region, and the carboxyl-terminal tail region (Fig. 2).

### Intermediate Signaling

Phosphorylated tyrosine residues of PDGF receptors act as docking sites for downstream signaling molecules that, through SH2 domains, bind to PDGF receptors in a site-specific manner (Fig. 2). Signaling proteins that bind to PDGF receptors include signal transduction proteins, docking proteins, and transcription factors.

Well-characterized signal transduction proteins that bind directly to phosphorylated PDGF receptors include the tyrosine kinase c-Src, the protein tyrosine phosphatase SHP-2, phospholipase C- $\gamma$  (PLC- $\gamma$ ), the Ras activator and exchange factor Grb2/Sos1, and the p85/p110 class I phosphoinositide-3-kinase (PI-3-kinase). Activation of signaling downstream of these enzymes, upon SH2 domain-mediated receptor binding, occurs through different mechanisms, including induction of conformational changes (c-Src), tyrosine phosphorylation (PLC- $\gamma$ ), and translocation (Grb2/Sos1 and PI-3-kinase).

Shc, Nck, Crk, and Grb7 are docking proteins that associate with PDGF receptors. Through SH3 domains or other protein–protein interaction domains, these proteins act as scaffolds for the creation of multiprotein signaling complexes. The composition of these complexes and their role in receptor signaling remain to be fully elucidated.



**Figure 2** Schematic illustration of PDGF receptor activation. Ligand binding to Ig-like domains 2 and 3 of the receptors leads to formation of a signaling complex composed of one PDGF molecule and two receptor molecules. The complex is stabilized by ligand–receptor and Ig domain 4-mediated receptor–receptor interactions. Receptor dimerization is associated with phosphorylation of a regulatory tyrosine (asterisks) in the kinase domain and tyrosines outside the kinase domain that act as binding sites for signaling proteins.

Members of the STAT family of transcription factors also bind directly to PDGF receptors. After receptor binding, these factors undergo tyrosine phosphorylation, dimerization, and translocation to the nucleus, where they act as transcriptional regulators.

Ligand-induced receptor activation thus leads to an array of biochemical responses, including the production of various second messengers and concomitant activation of multiple signal transduction pathways. Ultimately, these events are translated into changes in gene expression and cellular responses.

### Cellular Responses

PDGFs were originally purified and characterized based on their ability to induce a proliferative response. Subsequent studies have revealed additional cellular responses to PDGF receptor activation, including the ability of PDGF to induce migration toward increasing PDGF concentrations. PDGF receptor activation also confers reduced sensitivity to apoptotic stimuli and thus acts as a survival signal. *In vitro* cellular responses to PDGF, where the physiological counterpart is less understood, include the formation of specific actin microfilament structures (e.g., circular ruffles) and contraction of three-dimensional collagen gels.

Numerous efforts have been made to link individual signaling pathways with different cellular responses. Although this approach initially appeared to be promising, accumulating evidence indicates that extensive cross-talk between different biochemical pathways takes place, which can occur in a cell type-specific manner. Thus, induction of a particular cellular response is dependent on the balance within a network of signaling components rather than activation of a specific biochemical pathway.

### DEVELOPMENTAL AND PHYSIOLOGICAL ROLES OF PDGF

Genome sequencing projects have allowed systematic characterization of the PDGF system in model organisms. In *Caenorhabditis elegans*, there are four VEGF receptor/PDGF receptor homologs and one putative ligand. In contrast, the *Drosophila melanogaster* genome encodes only one receptor gene but three ligand genes. Interestingly, many cell types, e.g., glial cells, cells of the gonads, and cells of the gut, which in vertebrate development involve PDGF receptors, appear to be controlled by the VEGF receptor/PDGF receptor in *D. melanogaster*.

### PDGF $\alpha$ -Receptor-Dependent Developmental Processes

PDGF  $\alpha$ -receptor knockout mice die between E8 and E16 with multiple mesenchymal defects, including abnormal somite patterning, cleft palate, and skeletal malformations. PDGF A-chain knockouts survive until perinatal age but show defects of the development of alveoli, intestinal villi, hair follicles, and the testis, all of which can be explained by defects in epithelial–mesenchymal interactions. Also, hypomyelination, caused by defects in oligodendrocyte development, occurs in PDGF A-chain  $-/-$  mice. The milder phenotype of the A-chain knock-out mice, as compared to the  $\alpha$ -receptor  $-/-$  mice, is probably due to partial rescue of PDGF  $\alpha$ -receptor signaling, in these mice, by other PDGF isoforms.

### PDGF $\beta$ -Receptor-Dependent Developmental Processes

The best characterized phenotypes of PDGF-B and PDGF  $\beta$ -receptor knockout mice are defects in kidney development and aberrations in vascular development. The kidney defects consist of incomplete development of capillary tufts in the glomeruli. The mesangial cells, which normally surround the glomerular capillaries, are primarily affected by receptor depletion. The vascular defects, involving dilated, ectopic, and unstable capillaries, are caused by a reduction in pericyte coverage. The defects occur with varying severity in different organ, suggesting tissue-specific PDGF dependence for pericyte recruitment.

### Physiological Functions of PDGF

Wound healing involves several steps that are stimulated by PDGF, including recruitment and proliferation of fibroblasts and smooth muscle cells and the production of extracellular matrix components. Topical application of PDGF, or PDGF gene transfer, improves wound healing in animal models. Beneficial effects of recombinant PDGF-BB have also been demonstrated in clinical trials.

PDGFs may also act as autocrine negative regulators of platelet aggregation and thus exert a negative feedback function in thrombosis formation. Additional, less characterized functions of PDGF in the vascular system include the control of vessel tonus, which can involve either relaxation or constriction of blood vessels.

Finally, a novel function of PDGF receptor signaling in the control of tissue homeostasis was revealed

by animal studies showing that local application of PDGF-BB in loose connective tissue increases interstitial fluid pressure (IFP). Liquid transport and convection-driven transport of macromolecules between vessels and extracellular compartments are partially controlled by the pressure difference between capillaries and the interstitial fluid. The effects of PDGF receptor on IFP thus imply a function of PDGF receptor signaling in the control of transvascular flow of water and macromolecules.

### Isoform-Specific Functions of PDGF Receptors

The major differences between the phenotypes of PDGF  $\alpha$ - and  $\beta$ -receptor knockout mice can largely be explained by differences in the expression patterns of the two proteins. However, intrinsic differences in signaling between the two receptors also occur, as exemplified by isoform-specific differences in the ability to stimulate chemotaxis. *In vivo* support for this notion was also provided by elegant experiments in which two complementary mice strains were generated, in which the intracellular domains of the two PDGF receptors were exchanged. In this setting, the intracellular  $\alpha$ -receptor domain failed to fully mediate the vascular development normally executed by the  $\beta$ -receptor.

## PDGF AND DISEASE

### PDGF Antagonists

As discussed later, PDGF receptor signaling is implied as a causative factor in numerous pathological processes. This has stimulated development of various types of PDGF antagonists that act by interference with ligand binding, receptor dimerization, or tyrosine kinase activity. Prototypic examples of the first type are inhibitory antibodies binding to ligand or receptor. Interference of receptor dimerization has been achieved with soluble Ig domain 4 or antibodies targeting this domain. Finally, low-molecular-weight kinase inhibitors with varying degrees of specificity for PDGF receptors have been developed. STI571/Glivec/Gleevec/Imatinib is the best characterized of this class of compounds. In addition to PDGF receptors, STI571 also inhibits the Abl-kinase and the c-Kit/CSF-1 receptor.

### PDGF and Cancer

The discovery in the early 1980s that the transforming protein of the simian sarcoma virus was a homologue

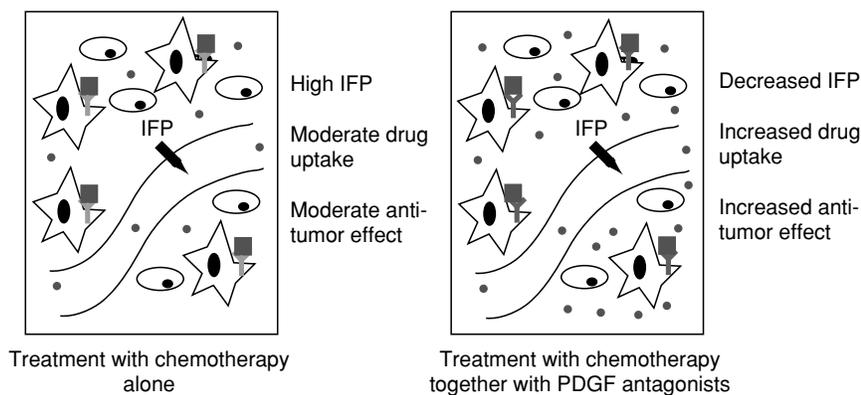
of the PDGF B-chain fuelled studies on the association between PDGF receptor signaling. The association between cancer and PDGF initially focused on PDGF autocrine stimulation of tumor cell growth. PDGF receptor signaling that contributes to the recruitment and function of tumor stroma and to tumor angiogenesis is receiving increased interest.

Different genetic mechanisms for the activation of PDGF receptor signaling, and subsequent autocrine growth stimulation, have been described. In subsets of glioblastoma multiforme, PDGF  $\alpha$ -receptors are amplified. Activating point mutations in the same receptor occur in a fraction of gastrointestinal stromal tumors. Translocations leading to novel ligand-independent fusion receptors, involving the  $\alpha$ - or  $\beta$ -receptor in idiopathic hypereosinophilic syndrome and chronic myelomonocytic leukemia, respectively, have also been described. Finally, genetic activation of the PDGF B-chain through fusion to the highly transcribed collagen gene is a characteristic feature of dermatofibrosarcoma protuberans. Successful use of PDGF antagonists for treatment of all these diseases, with the exception of glioblastoma multiforme, has been reported. There are also other solid tumors, such as soft tissue sarcomas and most glioblastomas, for which the basis for the aberrant coexpression of PDGF and PDGF receptors is unknown.

Regarding tumor angiogenesis, it is clear that a large majority of solid tumors express PDGF receptors on pericytes. Studies using animal cancer models also indicate that the extent of tumor vessel pericyte

coverage is dependent on paracrine stimulation of PDGF receptors on pericytes by tumor cells or endothelial cells. These findings have potential clinical implications since studies on antiangiogenic agents suggest that depletion of pericytes increases the sensitivity of tumor capillaries to antiendothelial agents such as VEGF antagonists. This led to the notion, which has been experimentally confirmed, that cotreatment with pericyte-targeting PDGF antagonists and antiendothelial agents will exert a more potent antiangiogenic and therapeutic effect.

The tumor stroma, mostly composed of cancer-associated fibroblasts and extracellular matrix, is emerging as a tumor compartment with important functions in tumor growth and tumor physiology. In contrast to most normal connective tissue, tumor stroma is characterized by expression of ligand-activated PDGF receptors. Studies in animal models with PDGF receptor expression restricted to tumor stroma have shown that inhibition of PDGF receptor expression is associated with an increased tumor uptake of systemically delivered standard chemotherapeutic agents (Fig. 3). Interestingly, this increased tumor drug uptake, which is not paralleled by increased uptake in other organs, is associated with an enhanced antitumor effect of standard chemotherapy such as taxol or 5-fluorouracil. The mechanism(s) underlying this effect is not fully understood, but available data indicate a role for the PDGF antagonist-induced reduction in tumor IFP as a causally related effect of PDGF receptor inhibition (Fig. 3).



**Figure 3** PDGF receptor inhibition in tumor stroma cells of experimental tumors is associated with increased tumor drug uptake and enhanced therapeutic response. In tumor models with expression of ligand-activated PDGF receptors on fibroblasts (stars), cotreatment with PDGF antagonists results in decreased tumor interstitial fluid pressure (IFP), an associated increase in tumor drug uptake, and enhanced antitumor effects of chemotherapy.

## PDGF in Atherosclerosis and Restenosis

Smooth muscle cell proliferation and migration are hallmarks of atherosclerosis as well as vessel obstruction following percutaneous transluminal angioplasty, bypass grafting, or transplantation. Up-regulation of PDGF receptors has been demonstrated in association with these pathological processes. Since PDGF is known to stimulate proliferation and chemotaxis of smooth muscle cells, PDGF antagonists have been used in a number of animal models; the results are encouraging.

Moderate inhibitory effects on fibrous cap formation in apolipoprotein E-deficient mice and on atherogenesis in cholesterol-fed rabbits occur after treatment with PDGF antagonists. However, studies with extended follow-up indicate that treatment with PDGF blockers in these models delays rather than prevents lesion formation.

Also, in rabbit, swine, and primate models of restenosis, beneficial effects on intima formation have been demonstrated after PDGF receptor inhibition. These findings have been confirmed in a swine model of stent-induced restenosis. These studies thus provide clear evidence that PDGF receptor stimulation of smooth muscle cells contributes to lesion formation in these models. Caution is warranted since there is also evidence from studies with longer follow-up that the effects are transient, with lesions reoccurring after withdrawal of treatment.

## PDGF and Fibrotic Disease

Evidence suggests a causal role of PDGF receptor signaling in proliferative glomerulonephritis. PDGF receptor up-regulation is a histological feature of the disease. Also, forced expression of PDGF by gene transfer, or infusion of PDGF, induces kidney fibrosis in animal models. Finally, using PDGF receptor kinase inhibitors or PDGF neutralizing agents in the Thy-1 model of nephritis, a few studies have provided encouraging results, demonstrating a reduction in mesangial cell proliferation and matrix accumulation.

Lung and liver fibrosis are additional examples of diseases for which several observations indicate a role of PDGF receptor signaling in the fibroproliferative aspects of the disease; up-regulation of PDGF receptors occurs in lesions, forced overexpression of PDGF stimulates disease development, and animal model studies with PDGF inhibitors demonstrate a partial inhibition of the disease process.

## PERSPECTIVES

There are a few key issues for future studies on the mechanism of action of PDGF, its role in disease, and its potential as a drug target. More detailed knowledge about the ligand–receptor interactions should assist in understanding ligand–receptor specificity and potentially help in designing isoform-specific antagonists. The determination of the structure of a PDGF ligand–receptor is therefore highly warranted. In the context of signaling mechanisms, the foundation concerning mediators of receptor signaling has been laid. Development of the current models to include spatial and temporal information about pathway activation is likely to increase understanding of how combinations of biochemical signals ultimately lead to changes in gene expression and to different cellular responses.

The role of PDGF in development and physiological processes has been greatly enhanced by studies of knockout phenotypes but has also been hampered by the embryonal or perinatal lethality associated with these genotypes. It can be predicted that analyses of tissue-specific and conditional knockout mice will provide additional information.

Recent clinical results with PDGF antagonists in autocrine malignancies have established PDGF receptors as “druggable” targets for disease management. These studies should stimulate further experimental and clinical use of PDGF antagonists, alone or in combination, in which PDGF receptors on pericytes and cancer-associated fibroblasts act as main targets. Improved methods for monitoring PDGF receptor activation and expression in tissue sections are highly warranted for optimal exploitation of the PDGF system as a cancer drug target. Also, it is likely that the availability of multiple isoform-specific antagonists will aid in reducing side effects and in allowing second-line strategies in case of molecular resistance. Finally, the transfer of the findings in animal models of vasculoproliferative diseases and fibrosis models from experimental to clinical settings is much awaited.

## See Also the Following Articles

Angiogenesis • Atherosclerosis • Bone Remodeling, Dynamics of • EGF and Related Growth Factors • Fibroblast Growth Factor (FGF) • Hepatocyte Growth Factor • Insulin-like Growth Factors • Peptide Hormones and Growth Factors: Cellular Signaling Mechanisms

## Further Reading

- Apperley, J. F., Gardembas, M., Melo, J. V., Russell-Jones, R., Bain, B. J., Baxter, E. J., Chase, A., Chessells, J. M., Colombat, M., Dearden, C. E., Dimitrijevic, S., Mahon, F. X., Marin, D., Nikolova, Z., Olavarria, E., Silberman, S., Schultheis, B., Cross, N. C., and Goldman, J. M. (2002). Response to imatinib mesylate in patients with chronic myeloproliferative diseases with rearrangements of the platelet-derived growth factor receptor beta. *N. Engl. J. Med.* **347**, 481–487.
- Bergers, G., Song, S., Meyer-Morse, N., Bergsland, E., and Hanahan, D. (2003). Benefits of targeting both pericytes and endothelial cells in the tumor vasculature with kinase inhibitors. *J. Clin. Invest.* **111**, 1287–1295.
- Betsholtz, C., Karlsson, L., and Lindahl, P. (2001). Developmental roles of platelet-derived growth factors. *Bioessays* **23**, 494–507.
- Duchek, P., Somogyi, K., Jekely, G., Beccari, S., and Rorth, P. (2001). Guidance of cell migration by the *Drosophila* PDGF/VEGF receptor. *Cell* **107**, 17–26.
- Enge, M., Bjarnegard, M., Gerhardt, H., Gustafsson, E., Kalen, M., Asker, N., Hammes, H. P., Shani, M., Fassler, R., and Betsholtz, C. (2002). Endothelium-specific platelet-derived growth factor-B ablation mimics diabetic retinopathy. *EMBO J.* **21**, 4307–4316.
- Fambrough, D., McClure, K., Kazlauskas, A., and Lander, E. S. (1999). Diverse signaling pathways activated by growth factor receptors induce broadly overlapping, rather than independent, sets of genes. *Cell* **97**, 727–741.
- Jones, S. M., and Kazlauskas, A. (2000). Connecting signaling and cell cycle progression in growth factor-stimulated cells. *Oncogene* **19**, 5558–5567.
- Klinghoffer, R. A., Muetting-Nelsen, P. F., Faerman, A., Shani, M., and Soriano, P. (2001). The two PDGF receptors maintain conserved signaling in vivo despite divergent embryological functions. *Mol. Cell* **7**, 343–354.
- Kozaki, K., Kaminski, W. E., Tang, J., Hollenbach, S., Lindahl, P., Sullivan, C., Yu, J. C., Abe, K., Martin, P. J., Ross, R., Betsholtz, C., Giese, N. A., and Raines, E. W. (2002). Blockade of platelet-derived growth factor or its receptors transiently delays but does not prevent fibrous cap formation in ApoE null mice. *Am. J. Pathol.* **161**, 1395–1407.
- Li, X., Ponten, A., Aase, K., Karlsson, L., Abramsson, A., Uutela, M., Backstrom, G., Hellstrom, M., Bostrom, H., Li, H., Soriano, P., Betsholtz, C., Heldin, C. H., Alitalo, K., Ostman, A., and Eriksson, U. (2000). PDGF-C is a new protease-activated ligand for the PDGF alpha-receptor. *Nat. Cell Biol.* **2**, 302–309.
- Pietras, K., Rubin, K., Sjoblom, T., Buchdunger, E., Sjoquist, M., Heldin, C. H., and Ostman, A. (2002). Inhibition of PDGF receptor signaling in tumor stroma enhances antitumor effect of chemotherapy. *Cancer Res.* **62**, 5476–5484.
- Pietras, K., Sjoblom, T., Rubin, K., Heldin, C. H., and Ostman, A. (2003). PDGF receptors as cancer drug targets. *Cancer Cell* **3**, 439–443.

granulosa cells. More granulosa cells with FSH receptors mean further increases in granulosa cell aromatase activity and further amplification of estrogen production, culminating in an LH surge and ovulation.

The polycystic ovary is characterized by increased intraovarian androgen concentrations. The increased androgen content interrupts follicle development through inhibition of FSH receptor recruitment and aromatase activity. Estrogen production is impeded and follicular growth stops before maturity and ovulation can be achieved. A pattern of hyperandrogenism and anovulation is thus established. *In vitro* studies have demonstrated that granulosa cells from PCOS ovaries function normally if adequate FSH levels are supplied and virtually all PCOS patients can be induced to ovulate, if enough exogenous FSH is supplied. The problem thus is one of inadequate FSH action at the level of the follicle and probably represents a deficiency of FSH or a blockade of FSH action by intraovarian regulators.

## DIAGNOSIS

Many diagnostic criteria have been developed, but no one item has been shown to be indispensable. Indeed, there is still substantial discussion as to what elements are crucial for the diagnosis. Evidence of peripheral hyperandrogenism (hirsutism), irregular menses, and chronic anovulation in the absence of other causes such as hypothyroidism and hyperprolactinemia constitutes the classic clinical presentation. Some, however, have argued that the appearance of multiple small cysts within the cortex of the ovary (representing incomplete maturation of follicles) is enough to make the diagnosis, even in the presence of normal menstrual function. Most authorities, however, advocate that evidence of menstrual dysfunction and hyperandrogenism should be present. Obesity is common, but in some studies as many as 50% of women with PCOS are not obese.

Biochemical markers include elevated serum LH levels, an elevated LH:FSH ratio (at least 2:1), elevated circulating androgens (total and free testosterone, androstenedione, dehydroepiandrosterone sulfate), decreased sex hormone-binding globulin, and evidence of hyperinsulinemia.

## TREATMENT

The initial treatment of Stein and Leventhal, that of bilateral ovarian wedge resection, was the only available treatment for polycystic ovary syndrome until the

1960s. That procedure is theorized to produce a sudden decrease in the concentration of intraovarian androgens through the removal of ovarian stromal tissue. The lower levels of intraovarian androgens release the follicle from interference with FSH action and follicle development and thus ovulation can occur. Due to concerns about postoperative adhesion formation and subsequent impairment of fertility, coupled with the introduction of effective medical treatments, the procedure is rarely performed anymore. Instead, treatment varies considerably depending on the individual patient's desires and most troubling symptoms.

For many years, PCOS was regarded by many as a "nuisance disease" with inconvenient symptoms, but presenting no major health consequences. As a result, treatment was directed at managing the most bothersome symptoms, such as irregular menstrual bleeding, hirsutism, and infertility. In women whose only concern was their irregular menstrual cycles, progestins were supplied to make up their deficiency of the hormone progesterone in the anovulatory state. In those who wished pregnancy, ovulation was induced with medications such as clomiphene citrate or injections of FSH. Those presenting with hirsutism and other signs of peripheral hyperandrogenism, such as acne, were given oral contraceptives and agents that blocked androgen action at the level of the hair follicle. As the relationship between hyperinsulinemia and PCOS began to be noted, a view of PCOS as a systemic disease with long-term health consequences, including increased cardiovascular risk and development of diabetes, emerged. It has become common for women with PCOS to routinely undergo screening for glucose intolerance and insulin resistance. Where these conditions are present, treatment with insulin-sensitizing agents, such as metformin, is often instituted.

## NEW DEVELOPMENTS—THE INSULIN CONNECTION

For some time, it has been observed that women with polycystic ovary syndrome have a higher incidence of long-term health problems than do normally menstruating women. The incidence of non-insulin-dependent diabetes mellitus is increased as is cardiovascular disease. Increasing attention has been given to the observation of insulin resistance in a large proportion of women with PCOS (up to 80% by some estimates). The elevated levels of insulin in these women that are necessary to maintain normal serum glucose levels cause direct stimulation of androgen production in the ovary, thus leading to the hyperandrogenism

characteristic of polycystic ovary syndrome. As insulin promotes fat storage and inhibits fat release from adipose tissue, these elevated insulin levels may play a role in promoting obesity, which, in turn, promotes further insulin resistance.

Findings of primary defects in the insulin receptor may explain much of the familial clustering of polycystic ovaries. Observed abnormalities of the insulin receptor in PCOS include excessive autophosphorylation of serine residues rather than threonine residues as is the case with normal insulin receptor function. Excessive serine autophosphorylation inhibits insulin receptor action by interfering with the translocation of intracellular glucose transporters to the cell surface, thus decreasing glucose uptake in those cells. Administration of medications designed to increase insulin sensitivity, such as metformin, have proven very useful in improving the hormonal and metabolic abnormalities of the syndrome and are rapidly becoming a primary treatment modality for PCOS.

### See Also the Following Articles

Hyperandrogenism, Functional • Hyperandrogenism, Hyperinsulinemic • Hypothalamic Anovulation, Functional • Infertility, Overview • Menstrual Cycle: An Integrative View • Ovarian Androgen-Producing Tumors • Ovarian

Hyperstimulation Syndrome • Polycystic Ovary Syndrome: Implications for Cardiovascular, Endometrial, and Breast Disease

### Further Reading

- Almahbobi, G., Anderiesz, C., Hutchinson, P., *et al.* (1996). Functional integrity of granulosa cells from polycystic ovaries. *Clin. Endocrinol.* **44**, 571–580.
- Barbieri, R. L., Makris, A., Randall, R. W., *et al.* (1986). Insulin stimulated androgen accumulation in incubations of ovarian stroma obtained from women with hyperandrogenism. *J. Clin. Endocrinol. Metab.* **62**, 904.
- Dunaif, A. (1997). Insulin resistance and the polycystic ovary syndrome: Mechanism and implications for pathogenesis. *Endocr. Rev.* **18**, 774–800.
- Nestler, J. E., Clore, J. N., and Blackard, W. G. (1989). The central role of obesity (hyperinsulinemia) in the pathogenesis of the polycystic ovary syndrome. *Am. J. Obstet. Gynecol.* **161**, 1095–1089.
- Nestler, J. E., Jakubowicz, D. J., Evans, W. S., and Pasquali, R. (1998). Effects of metformin on spontaneous and clomiphene-induced ovulation in the polycystic ovary syndrome. *N. Engl. J. Med.* **338**, 1876–1880.
- Stein, I. F., and Leventhal, M. L. (1935). Amenorrhea associated with bilateral polycystic ovaries. *Am. J. Obstet. Gynecol.* **29**, 181–191.
- Yen, S. S. C. (1999). Polycystic ovary syndrome (hyperandrogenic chronic anovulation). *In* “Reproductive Endocrinology: Physiology, Pathophysiology, and Clinical Management” (S. S. C. Yen, R. B. Jaffe, and R. L. Barbieri, eds.), 4th ed., pp. 436–476. W. B. Saunders, Philadelphia, PA.

to healthy controls. The results of a retrospective cohort follow-up study showed an increased prevalence of hypertension in women with PCOS. Another study of similar design documented an increased prevalence of coronary artery disease. However, the number of patients in both studies was very small (33 and 28 patients with PCOS, respectively).

Pierpoint *et al.* evaluated cardiovascular mortality in women with PCOS. A total of 786 patients diagnosed in the United Kingdom between 1930 and 1979 were traced from hospital records. The standardized mortality ratio was then calculated based on 59 deaths. The study found a higher mortality from diabetes mellitus but no increased average mortality from circulatory disease. However, the patients were included only on the basis of histopathology and surgical records. These criteria cannot be considered sufficient for the diagnosis of PCOS. Moreover, it can be assumed that a successfully performed procedure (wedge ovarian resection) can diminish some of the endocrine and metabolic disturbances and thus also alter the later risks.

In conclusion, there is no doubt that women with PCOS cluster many of the proven risk factors for cardiovascular disease. The syndrome is characterized by an increased prevalence of hyperinsulinemia, dyslipidemia, impaired glucose tolerance, and type 2 DM. In small groups of patients, atherosclerotic changes were found more often. Although evidence of an increased prevalence of coronary artery disease is low, on the basis of available data, women with PCOS should be considered a risk group in the population.

## ENDOMETRIAL CANCER

An increased risk of endometrial cancer is often presented as one of the health consequences of PCOS. Epidemiological studies have repeatedly found an increased prevalence of obesity, irregular menstrual cycles, amenorrhea, infertility, or hirsutism in patients with endometrial cancer. In a study evaluating 110 case-control pairs ages 45 years or less, current weight, amenorrhea, and infertility were associated with an increased risk. However, a large case-control Danish study of 237 cases and 538 controls under 50 years of age was not able to confirm BMI or amenorrhea as risk factors after careful adjustment for other confounders.

Evidence regarding the increased risk of endometrial cancer in PCOS patients is still weak. Published papers are mostly case series and include only a small number of cases (8–12). After careful evaluation, it is

found that most patients are obese or even extremely obese. One of the main arguments is the study by Coulam *et al.* in 1983, which evaluated the prevalence of neoplasia in 1270 women with chronic anovulation syndrome. An increased risk was found only for endometrial cancer; however, the relative risk reached only 3.1 (95% CI 1.1–7.3). It must be emphasized that this was not a case-control study, but the results were compared to the expected numbers based on an age-specific database from a certain region. Moreover, the calculation was made on the basis of only 5 observed and 1.59 expected cases. No prospective study has been published yet.

Unopposed estrogens due to long-term anovulation are presented as a principal mechanism of development of endometrial cancer in PCOS patients. However, it does not correspond to the finding of a thin endometrium in the majority of patients or the low incidence of dysfunctional uterine bleeding, which should be the result of endometrial hyperproliferation. It is also improbable that hyperandrogenemia would play an important role. A dose-dependent anti-estrogenic effect of androgens on the endometrium has been found in animals. It is thus likely that other mechanisms are involved. It has been demonstrated *in vitro*, on long-term cultures of human endometrial cells, that insulin, insulin-like growth factor-I (IGF-I), and IGF-binding proteins influence endometrial differentiation. Moreover, a greater number of binding sites for IGF-I were identified in cell cultures of endometrial cancer. If hyperinsulinemia is the main mechanism increasing the risk of endometrial cancer in PCOS patients, primarily obese women and women with insulin resistance would be at risk.

PCOS women cluster many of the risk factors for endometrial cancer. Although evidence is weak, it can be assumed that those patients will have an increased risk. However, it is not certain whether this risk applies only to certain subgroups of patients (obese women, patients with insulin resistance) or whether PCOS presents an independent risk. Hyperinsulinemia could play a significant role.

## BREAST CANCER

In comparison to the normal population, women with PCOS have a lower number of full-term pregnancies and a higher age at first pregnancy. Both of these parameters are known risk factors for the development of breast cancer. In addition to the consequences of infertility, other mechanisms may play a role. A common finding in obese and lean women with PCOS is hyperinsulinemia. Some experimental data

prove the important role of insulin and insulin-like growth factor in the development and proliferation of breast cancer.

Only a small number of authors have evaluated the risk of breast cancer in women with PCOS. Unfortunately, none of the published studies have fulfilled the criteria of level of evidence I or II. In addition, their inadequacy lies either in the small number of identified cases or in the insufficient diagnosis of PCOS based on self-reports by the patients.

In 1983, Coulam *et al.* evaluated the medical records of 1270 patients, in whom chronic anovulation was diagnosed between 1935 and 1980 at the Mayo Clinic. An increased risk of breast cancer was found only in postmenopausal women [relative risk (RR) 3.6; 95% confidence interval (CI) 1.2–8.3], whereas the total relative risk was not increased (RR 1.5; 95% CI 0.8–2.6). However, these results were based on only 12 identified cases. Furthermore, the observed cases were compared to the expected numbers using age-specific incidence rates for 1967 in Connecticut. Seven years later, Gammon and Thompson published the results of a case-control study analyzing the association between infertility and breast cancer in 4730 cases and 4688 controls. The authors did not find an increased risk in infertile women [odds ratio (OR) 1.01; 95% CI 0.89–1.15] or in the subgroup of infertile women with ovarian pathology (OR 0.75; 95% CI 0.48–1.24), after controlling for age, age at first delivery, and parity. In addition, 1 year later, on the basis of the same results, the authors reported a breast cancer age-adjusted odds ratio of 0.52 (95% CI 0.32–0.87) in women with PCOS. The main drawback of the study is the identification of PCOS on the basis of a self-reported history of physician-diagnosed polycystic ovaries. This also corresponds to the unexpected low prevalence of PCOS between cases and controls (0.49% versus 0.94%). Another prospective study of a cohort of 34,835 women followed for over a 6-year period was also based on self-reports of PCOS cases. In this series, no increased risk of breast cancer was documented (RR 1.3; 95% CI 0.7–2). Contributing to the discussion are the results of the above-mentioned study published by Pierpoint *et al.*, which focused on mortality in women with PCOS. No significant increase in standardized mortality ratio (SMR) for breast cancer was found (SMR 1.48; 95% CI 0.79–2.54).

If the known risk factors for breast cancer, such as the number of full-term pregnancies and age at first pregnancy, are evaluated separately, on the basis of available data PCOS should not be considered an independent significant risk factor for breast cancer.

## Acknowledgment

This study was supported by Grant No. NH/6558-3 of the Internal Grant Agency of the Ministry of Health of the Czech Republic.

## See Also the Following Articles

Cardiovascular Disease: Impact of Sex Steroid Replacement • Endometriosis • Hyperandrogenism, Functional • Hyperandrogenism, Hyperinsulinemic • Polycystic Ovary Syndrome (PCOS)

## Further Reading

- Balen, A. (2001). Polycystic ovary syndrome and cancer. *Hum. Reprod. Update* **7**, 522–525.
- Chang, R. J., Heindel, J. J., and Dunaif, A. (2002). “Polycystic Ovary Syndrome.” Dekker, Basel, Switzerland.
- Cibula, D., Cífková, R., Fanta, M., Poledne, R., Živný, J., and Skibová, J. (2000). Increased risk of non-insulin-dependent diabetes mellitus, arterial hypertension and coronary artery disease in perimenopausal women with a history of the polycystic ovary syndrome. *Hum. Reprod.* **15**, 785–789.
- Coulam, C. B., Annegers, J. F., and Kranz, J. S. (1983). Chronic anovulation syndrome and associated neoplasia. *Obstet. Gynecol.* **61**, 403–407.
- Dahlgren, E., Johansson, S., Lindstedt, G., Knutsson, F., Odén, A., Janson, P. O., Mattson, L., Crona, N., and Lundberg, P. (1992). Women with polycystic ovary syndrome resected in 1956 to 1965: A long-term follow-up focusing on natural history and circulating hormones. *Fertil. Steril.* **57**, 505–513.
- Parslov, M., Lidegaard, O., Klinton, S., Pedersen, B., Jonsson, L., Eriksen, P. S., and Ottesen, B. (2000). Risk factors among young women with endometrial cancer: A Danish case-control study. *Am. J. Obstet. Gynecol.* **182**, 23–29.
- Pierpoint, T., McKeigue, P. M., Isaacs, A. J., Wild, S. H., and Jacobs, H. S. (1998). Mortality of women with polycystic ovary syndrome at long-term follow-up. *J. Clin. Epidemiol.* **51**, 581–586.
- Talbott, E., Clerici, A., Berga, S. L., Kuller, L., Guzick, D., Detre, K., Daniels, T., and Engberg, R. A. (1998). Adverse lipid and coronary heart disease risk profiles in young women with polycystic ovary syndrome: Results of a case-control study. *J. Clin. Epidemiol.* **51**, 415–422.
- Wild, R. A. (2002). Long term health consequences of PCOS. *Hum. Reprod. Update* **8**, 231–241.

Table I Mean or Median of Adult Height (cm) in Various Countries

Study (year)	Males		Females	
	Mean or median	97th percentile ( $\pm 1.8$ SD)	Mean or median	97th percentile ( $\pm 1.8$ SD)
The Netherlands (1985)	182.0	194.5	168.3	179.8
Germany (1992)	179.9	192.5	167.0	179.0
Denmark (1982)	179.4	190.4	166.0	176.0
Switzerland (1976)	179.1	192.4	165.5	178.2
Czechia (1993)	178.3	191.7	165.0	176.8
United States (1977)	176.8	187.6 <sup>a</sup>	163.7	173.6 <sup>a</sup>
United Kingdom (1995)	176.4	190.5 <sup>b</sup>	163.6	176.0 <sup>b</sup>
Spain (1988)	175.6	186.9	161.2	172.0
France (1979)	175.0	187.0	163.0	174.5
Turkey (1978)	173.5	186.0	160.0	171.0
Mexico (1975)	172.8	186.3	160.6	174.5
Argentina (1987)	172.8	185.6	160.7	172.2
Korea (1979)	170.2	180.0	157.6	166.5

Source. Reprinted with permission from Drop, S. L. S., *et al.* (1998). Copyright 1998, The Endocrine Society.

<sup>a</sup>95th percentile.

<sup>b</sup>98th percentile.

dysmorphic proportions, such as in Marfan's syndrome, Beals' syndrome, and homocystinuria. (7) Alterations of genes involved in the regulation of the cell

cycle, proliferation, and growth, such as in Bannayan-Riley-Ruvalcaba syndrome, Cowden's disease, and Sotos' syndrome, can also lead to overgrowth.

Table II Classification of Tall Stature or Postnatal Nonendocrine Overgrowth

Normal variants
Familial (genetic) tall stature
Tall normal girl
Tall boy
Familial (genetic) rapid maturation
Overnutrition (obesity)
Chromosomal abnormalities
Trisomy X (47,XXX females)
Klinefelter's syndrome XXY
XYY syndrome
Fragile X chromosome
Syndromes and others
Marfan's syndrome
Beals' syndrome (CCA)
Homocystinuria
Beckwith-Wiedemann syndrome
Somatic overgrowth (H19 methylation)
Simpson-Golabi-Behmel syndrome
Sotos' syndrome
Weaver's syndrome
Bannayan-Riley-Ruvalcaba syndrome
Partington's syndrome

## NORMAL VARIANTS

### Familial or Genetic Tall Stature

This includes tall, otherwise normal children, who mature at the normal rate and attain puberty and a tall adult height at the usual age.

#### Tall Normal Girls

There is usually no difficulty in establishing the diagnosis of a tall, normal girl, by the normal history and physical findings, the normal pubertal development, the lack of dysmorphic features, and the family history of tall stature.

One or both of the parents and other members of the family are often tall. Genetic tall stature is believed to be determined by multiple genes, but the manner whereby these genes result in tall stature has not been determined. Patients with tall stature are normal at birth; an increased rate of linear growth is usually evident during the first 2 years and tall stature is obvious by the age of 3 or 4 years. Subsequently, growth continues at a rate slightly greater than normal and the child remains between 2 and 4 SD above the mean for height for age. Timing of puberty is normal, as is the time of attainment of the tall adult

Table III Molecular Basis and Etiopathogenesis of Overgrowth (Established or Suggested)

Disorder	Cause	Molecular basis
Extra growth genes		
Klinefelter's syndrome (47,XXY)	Extra growth gene SHOX	Extra X
Trisomy X (47,XXX)	Extra growth gene SHOX	Extra X
Males 47,XYY	Extra Y-specific growth control gene	Extra Y
Excessive GH secretion (pituitary tumors)		
Sporadic		
Gigantism/acromegaly	Mutations in $G_{\text{sa}}$ protein gene	20q12–q13.2
	LOH (allelic loss) of 11q.13 (no mutations of MEN1 gene)	11q13
	Overexpression of PTTG	
McCune-Albright syndrome	Mutations in $G_{\text{sa}}$ protein gene	20q12–q13.2
Familial		
MEN1	MEN1 gene mutations and LOH of 11q.13	11q13
Acromegaly/gigantism	LOH (allelic loss) of 11q.13 (no mutations of MEN1 gene)	11q13
Extra growth factors		
IGF-II		
Beckwith-Wiedemann syndrome	Overexpression of IGF-II	11p15.5
H19 silencing	Overexpression of IGF-II	11p15.5
Simpson-Golabi-Behmel syndrome	Modulation of IGF-II	Mutations
	Deficiency of glypican 3	GPC-3 gene
		Xq26
IGF-I–insulin		
Obesity	Hyperinsulinism—free IGF-I	
Lipodystrophy	Hyperinsulinism	
Infant of diabetic mother	Hyperinsulinism	
Infant giants with neonatal hypoglycemia	Hyperinsulinism	Various
Extra growth factors—receptors		
Partington's syndrome	Extra FGFR3 gene (?)	Duplication, 4p16.3
Deficiency of factors needed to arrest growth		
Aromatase deficiency	Estrogen deficiency	Mutations in Cyp19 gene, 15q21.1
Estrogen receptor deficiency	Estrogen deficiency	Mutations of estrogen receptor gene, 6q25.1
Hypogonadism	Estrogen deficiency (secondary)	Primary—various
Deficiency of factors needed to prevent elongation of bones		
Marfan's syndrome	Mutation of fibrillin gene	FBN1, 15q21.1
Beals' syndrome (CCA)	Mutation of fibrillin gene	FBN2, chromosome 5
Homocystinuria	Abnormal collagen	
	Mutations of CBS gene	21q21
Alterations of genes—involved in regulation of cell cycle, growth, and tumor suppression		
Bannayan-Riley-Ruvalcaba syndrome	PTEN gene mutations	10q23
Cowden's syndrome	PTEN gene mutations	10q23
Sotos' syndrome	NSD1 gene mutations	5q35
Weaver's syndrome	NSD1 gene mutations	5q35

*Note.* CBS, cystathionine  $\beta$ -synthase; CCA congenital contractural arachnodactyly; Cyp19, cytochrome P450 19-aromatase; FBN1 or 2, fibrillin gene 1 or 2; FGFR3, fibroblast growth factor receptor 3; GPC-3, glypican 3;  $G_{\text{sa}}$ , guanine nucleotide-binding protein 1 stimulatory  $\alpha$ -chain; IGF-I or -II, insulin-like growth factor-I or -II; LOH, loss of heterozygosity; MEN1, multiple endocrine neoplasia type 1; PTTG, pituitary tumor-transforming gene; NSD1, nuclear receptor-binding SET [where SET denotes SU(VAR)3-9,E(Z)trithorax] domain protein 1; PTEN, phosphatase and tensin homologue in chromosome 10; SHOX, short stature homeobox-containing gene.

height. The prediction of adult height is based on the bone age and height at the time.

Tall stature has become more acceptable in women than it was previously and consultation to curtail adult height has become much less frequent than in the past. Estrogens in large amounts, 3 to 10 times the replacement dose, have been proven effective in curtailing adult height. Bromocriptine does not appear to be effective. Octreotide, the long-acting somatostatin analogue (SMS-201-995), may be an alternative, but experience with this drug is limited.

The estrogen preparations most frequently used have been conjugated estrogens (7.5 to 10 mg) and ethinyl estradiol (0.1 to 0.3 mg) daily, without interruption. A progesterone-like compound should be given in cycles for 10 days every month, to induce menstrual periods and prevent metrorrhagia. Estrogen treatment increases GH secretion, but sharply decreases the rate of growth, partly as a result of a decrease in IGF-I levels. In addition, the bone age accelerates at an average of 1.8 years per year of treatment, with a range from 1.0 to 2.5 years. Treatment is usually begun after the age of 11 years and after the spontaneous onset of puberty.

The authors of this article have offered treatment to girls with a predicted adult height in excess of 178 or 180 cm. Treatment is also given to girls who are taller than 178 cm, when first seen, with a growth potential of another 2.5 cm or more, regardless whether they are pre- or postmenarcheal.

Duration of treatment has ranged from 8 months to 3 years.

The height reduction depends on the bone age and potential growth at the onset of treatment and may range from 2.5 cm in a girl with a bone age of 14 years to 14.0 cm in a girl with a bone age of 10½ or 11 years.

Grave side effects are quite rare. The main concern is the potential risk of thrombosis. In a survey of 904 cases, there was only one thromboembolic episode in a girl on treatment, after a crushing foot injury. Two other cases of thrombosis of the femoral vein have been reported: one occurred after a febrile viral disease, requiring bed rest for 4 days, and another occurred after arthroscopy of the knee with immobilization.

There is no evidence for an increased risk of endometrial, cervical, ovarian, or breast cancer with contraceptives in young women. There is no evidence that fertility is affected.

Because of the potential risk of thrombosis, it is not a treatment to be promoted, but it should be offered, when justified, because of the great benefit for some girls. This treatment is applicable and could

be helpful to other patients with overgrowth disorders (i.e., Marfan's syndrome, Sotos' syndrome, trisomy X, androgen resistance). It is contraindicated in patients with homocystinuria.

### **Tall Normal Boys**

Tallness in boys is better accepted by society and by the individual than it is in girls and rarely is medical attention sought. The pattern of growth and development of tall normal boys is similar to that mentioned for tall normal girls, as is the family history.

There is usually no difficulty in establishing the diagnosis of a tall normal boy by the normal history and physical findings, the normal pubertal development, the lack of dysmorphic features, and the family history of tall stature. A bone age and a prediction of adult height should be obtained.

Treatment with long-acting testosterone esters can reduce the adult height. Experience with such treatment is limited and the need to treat normal tall boys is rare. Treatment has been given to tall normal boys with a predicted adult height usually in excess of 198 cm. The preparations most frequently used have been testosterone enanthate or cypionate at a dose of 200 mg every 2 weeks for young adolescents and 500 mg every 2 weeks for older adolescents. The growth velocity increases, particularly in the first 6 months, but the bone age advances faster than the height, at an average of 1.8 years per year of chronological age, and as a result, adult height is reduced. The reduction in height depends on the bone age at the onset of treatment and ranges from  $3.0 \pm 2.29$  cm in those with a bone age higher than 15 years at the beginning of treatment to  $8.0 \pm 5.4$  cm in boys with a bone age of 12 to 14 years. It could be more, however. In a 12-year-old boy with a predicted height of 203 cm and a bone age of 9½ years, treated with testosterone enanthate (200 mg every 2 weeks) for 1½ years, the bone age advanced 6 years (4 years per year of chronological age) and the reduction of height was 21.5 cm. The mechanism of action is related to the anabolic and androgenic effects of testosterone, particularly in hastening bone maturation. This latter effect is probably the result of estrogen derived from testosterone. The side effects include weight gain, partly as a result of sodium retention at the initiation of treatment and partly from the protein anabolic effect. Gynecomastia occurs in some patients. During treatment, the testes do not grow in the youngest boys and decrease in volume in the older boys, from volumes of 10–12 to 8 cc, on average, at the end of treatment. There is oligo- and azoospermia. After treatment, the testicular volumes return to normal in the course of 1 to 2 years. Testicular function

is reversible by 1 year in the majority of patients, but recovery is slow in some patients, lasting 3 to 5 years. Aggravation of acne is a side effect reported in 30 to 60% of boys. Studies of subjects 10 years after termination of treatment showed normal testicular function. This treatment could also be offered to patients with other overgrowth conditions, such as Klinefelter's syndrome (XXY), XYY males, eunuchoidism (for virilization as well as for curtailment of height), and Marfan's syndrome, if they so desire.

Bromocriptine decreases the rate of growth somewhat, but does not advance the bone age and it is uncertain whether adult height will be reduced significantly.

Octreotide has also been used for treatment of tall boys and appears to be effective. The concerns are the cost, the need for two to three injections a day, and the risk of gallstone formation in 27% and bile sludge without stones in 22% of the patients receiving octreotide for 12 months or longer.

### Familial Rapid Maturation

These are normal children who may be tall in childhood, who have an advanced bone age corresponding to their height age, and who attain pubertal development and cessation of growth earlier than the usual age. The adult height is within the normal range. Family history usually reveals a similar growth pattern in one of the parents or in other members of the family. There is no need for treatment.

### OVERNUTRITION (OBESITY)

Obese children tend to be taller and grow faster than their lean peers during childhood and early adolescence. They do have advanced skeletal maturation, a blunted pubertal growth spurt, when compared with lean subjects, and normal adult heights. Longitudinal growth data on children who develop obesity during childhood showed a distinct tendency for height gain to accelerate coincident with or after the onset of excessive weight gain. Many of the tall obese children follow normal percentile lines for height, but some have excessive growth rates during infancy or childhood, reaching heights of 3 or 4 SDS, and other possibilities associated with abnormal excessive growth need to be considered. The diagnosis is based on clinical grounds and on the exclusion of other possible causes of overgrowth.

The tall stature and increased growth velocity associated with obesity are probably mediated by insulin and IGF-I. Growth hormone secretion is low.

Studies of patients with moderate and severe obesity and controls clearly showed an increase in serum insulin in males and females, a decrease in insulin-like growth factor-binding protein-1, and an increase in the free IGF-I. Thus, the increased growth velocity of obese children is most likely the result of the effect of hyperinsulinemia and increased free IGF-I on the insulin and IGF-I receptors.

The advisable treatment is to place the patient on a moderate calorie-restriction diet, until he or she reaches an ideal body weight, and then liberalize the caloric intake to resume normal growth velocity and weight gain.

## CHROMOSOMAL ABNORMALITIES

### Trisomy X (47,XXX Females)

Trisomy X should be included in the differential diagnosis of tall girls. Patients with trisomy X tend to be tall. Forty-two percent of the girls with trisomy X between the ages of 6 and 13 years are over the 90th percentile for height. Adolescent and adult 47,XXX females are tall, with heights generally at or above the 90th percentile (171 cm (67.32 in.)). They have long legs. Most of the patients have a normal phenotype, affected children are not recognized, and the condition is underdiagnosed. Trisomy X is the most frequent X chromosomal abnormality, occurring in 1 of 1000 newborn females; it is two to three times more frequent than Turner's syndrome. Some patients may have minimal dysmorphic features (clinodactyly, syndactyly). Most of the patients have normal sexual development, menarche, and ovarian function, but some may have secondary amenorrhea and premature ovarian failure. Primary amenorrhea and ovarian dysgenesis have been reported. The IQ may range from 55 to 115 and approximately one-half to two-thirds of patients may have an IQ that is slightly below normal.

The cause for the tendency to tall stature is probably the extra X chromosome. Many observations have clearly indicated that Xp deletions, particularly the lack of the tip of Xp, leads to short stature and that additional X chromosomes can compensate for that loss. In 1997, Rao *et al.* isolated a homeobox-containing gene from PAR1 that they named SHOX (for short stature homeobox-containing gene). The evidence suggests that deletions of the tip of Xp with the absence of SHOX could result in short stature and that additional inactivated X chromosomes, with active growth genes (SHOX) escaping inactivation, could cause tall stature, such as in Klinefelter's syndrome and trisomy X.

The presence of long legs in trisomy X as well as in Klinefelter's syndrome suggests that the characteristic is related to the extra X chromosome and to the extra growth gene *SHOX*, with the effect on bone growth. Large-scale deletions of *SHOX* in several families, and a nonsense mutation in one family, were reported in patients with dyschondrosteosis (Leri-Weill syndrome). This is an autosomal-dominant form of mesomelic dysplasia with deformity of the forearm (Madelung deformity), metaphysical changes, severe short stature, short arms, and short legs.

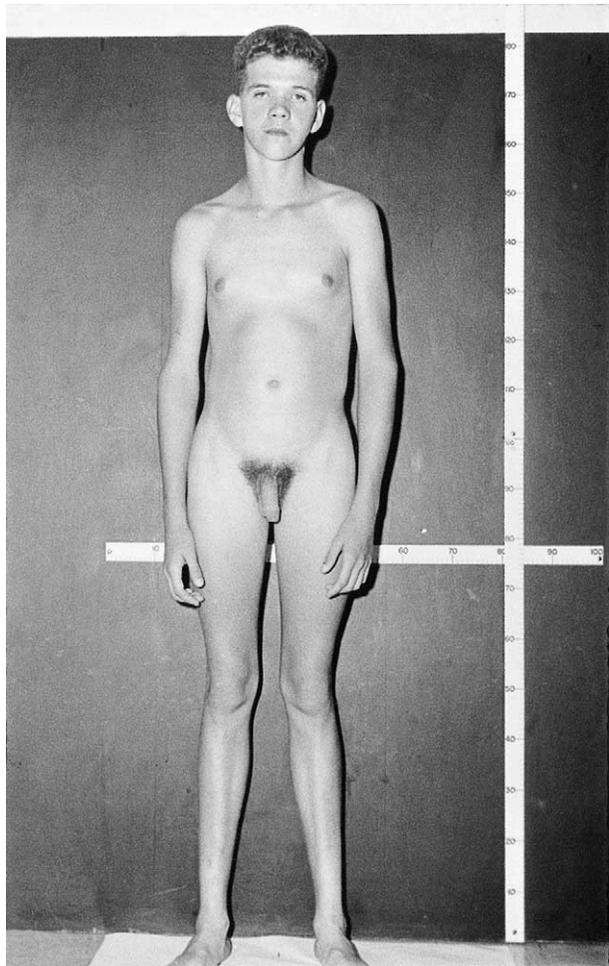
The possibility of 47,XXX should be suspected in girls of normal or tall stature with secondary amenorrhea, premature ovarian failure, or learning difficulties. A karyotype would seem advisable in the evaluation of tall girls.

Many of the patients are normal and need no treatment. For patients with excessive tall stature, treatment similar to that described for tall girls could be provided.

### Klinefelter's Syndrome XXY

This syndrome was first described by Klinefelter and co-workers in 1942 with the characteristic features, which become evident at adolescence, of gynecomastia, tallness, variable degree of eunuchoidism, imperfect virilization, small firm testes with hyalinization of the seminiferous tubules and aggregation or clumping of the Leydig cells, small penis, tendency to dull mentality, and increased urinary excretion of gonadotropins (Fig. 1). In 1959, a 47,XXY sex chromosome constitution in patients with this disorder was first reported. All the variants have in common the presence of at least two X chromosomes and a Y chromosome, except for rare patients who have only a 46,XX complement (XX males).

Patients are on average 10 cm (3.9 in.) taller than XY males. Tall stature is usually present before puberty, as is the disproportionate leg length, which suggests that this feature is not related to androgen deficiency and may be determined by the extra X chromosome, as discussed for trisomy X. The arm span is not increased and is equal to or less than the height. Affected children may be immature, excessively shy, anxious, and aggressive and may engage in antisocial acts, such as fire setting, theft, and cruelty to animals. There is a higher than average incidence of problems with speech development, with learning, and with social adjustment. The full-scale IQ is usually normal. Severe retardation is uncommon. Gynecomastia usually persists, contrary to adolescent



**Figure 1** A tall boy, age 15 years, 10 months, with Klinefelter's syndrome: karyotype XXY, small firm testes, gynecomastia, and long legs. Height is 185 cm (6 ft  $\frac{3}{4}$  in.).

gynecomastia that usually subsides after 2 to 3 years. Breast cancer occurs in approximately 4% of the patients, with an incidence 18 times higher than in normal males. Germ cell tumors, particularly in the mediastinum, that secrete human chorionic gonadotropin and cause luteinizing hormone-releasing hormone-independent sexual precocity, occur in 1.5 per 1000 patients.

Chromosome surveys at birth show an incidence of approximately 1 in 500 to 1 in 1000. As discussed in relation to trisomy X, the active genes (escaping inactivation) in the inactivated X chromosome may be responsible for the abnormalities found in Klinefelter's syndrome and, particularly, the extra growth gene, *SHOX*, may be responsible for the tall stature. Patients with Klinefelter's syndrome with an isochromosome Xq (47,XiXqY), lacking the extra short arm

of the second X chromosome and consequently the extra SHOX gene, show all the typical manifestations of Klinefelter's syndrome except tall stature.

Less than 10% of the estimated number of affected fetuses are detected prenatally and 75% of patients go through life undetected. Diagnosis of the syndrome in patients after puberty is not difficult, by the findings of the typical phenotype of tall stature, incomplete virilization, small firm testes, and elevated serum gonadotropin levels. The diagnosis can be confirmed by the karyotype.

With the onset of puberty, testosterone synthesis is impaired and gonadotropin levels are elevated. Many of the patients have low testosterone levels during adolescence and as a result have incomplete virilization, small penis, poor muscular development, and eunuchoid proportions. Osteoporosis occurs in 25% of the patients. To prevent the physical and psychological complications of hypogonadism, testosterone treatment is recommended. Treatment could be started, if needed, with 50 mg of testosterone enanthate (or cypionate) in oil, intramuscularly (im), monthly when the bone age is 12 to 13 years, increasing to 100 mg monthly when the bone age is 14 years, and increasing to 200 mg every 3 or 4 weeks when full virilization is desired or when growth is ending. At the same time, this could prevent excessive adult height. The adult replacement dose is 300 mg every 3 weeks or 200 mg im every 2 weeks.

Psychologically disturbing or persisting gynecostasia should be corrected by reduction mammoplasty. Early intervention for learning and behavioral disorders may be beneficial.

### 47,XYY Males

47,XYY males are known to have excessive height. The first report of an XYY male was made in 1961, when an essentially normal, fertile male of average intelligence was studied because he had a daughter with Down's syndrome. Subsequently, an increased prevalence of the XYY karyotype was found among tall, mentally retarded incarcerated males and created a stereotype of affected individuals as having deviant behavior, marked by physical aggressiveness and violence. This impression is not applicable to all patients. The only consistent physical feature of the syndrome is excessive height, with 50% of patients being above the 90th percentile of normal. Forty to 50% have learning difficulties. Adults with this syndrome are tall and have nodular cystic acne, large deciduous and permanent teeth, and some neurological abnormalities, such as

intention tremor, incoordination, and, frequently, radio-ulnar synostosis.

There is no evidence of any endocrine abnormality. The tall stature is probably determined by the additional Y chromosome. A region that has a great influence on growth and on tooth size was mapped by different investigators to the most proximal portion of the long arm of the Y chromosome, close to the centromere, and is called growth control in the Y (GCY) or Y-specific growth gene. GCY may be the gene responsible for the difference in height and tooth size between males and females. The observations suggest that tall stature in 47,XYY males is related to extra growth genes in the additional Y chromosome: SHOX and GCY or Y-specific growth gene.

As with trisomy X and Klinefelter's syndrome, XYY males go unrecognized, because of the lack of marked phenotypic changes. To detect such patients, a karyotype needs to be obtained in any patient with tall stature whose cause is unknown. The disorder should be suspected in tall males with nodular cystic acne, who exhibit antisocial behavior.

No treatment for the tall stature is needed, unless patients are affected psychologically and would so desire treatment. In that case, treatment will be the same as for tall normal boys.

### Fragile X Syndrome

In a report of seven young fragile X-positive girls, the two most common and most important findings were the overgrowth, present from birth onward, and behavioral features, including severe attention problems and extreme shyness and anxiety. The fragile X syndrome is the most common form of inherited mental retardation with a prevalence estimated to be 1 in 1250 males and 1 in 2500 females. In 1991, the gene, which when disrupted results in the fragile X site at Xq27.3, was isolated and characterized. The DNA segment shows a peculiar stretch of trinucleotide repeats (cytidine, guanosine, guanosine) in the fragile X mental retardation gene 1 (FMR-1) that increases the size of the specific DNA fragment of the X chromosome at Xq27.3.

In males, the most typical features are large ears and macro-orchidism, with approximately half of the patients having an ear length above the 90th percentile for normal (Fig. 2). Testicular volumes after puberty in normal males range from 10 to 25 ml and those in fragile X-positive patients range from 25 to 70 ml. Overgrowth and macrocephaly are seen in females. Females function in the borderline to mildly



**Figure 2** A 20-year-old mentally retarded young man with fragile X syndrome, prominent forehead and chin, and large ears. Height is 183 cm. Photograph courtesy of Dr. A. Sommer, Department of Genetics, Children's Hospital, Columbus, Ohio.

retarded range. Intelligence assessment showed an IQ of less than 70 in 25% and of less than 85 in 53% of the girls with a positive fragile X chromosome, by cytogenetic studies.

The possibility of fragile X syndrome should be suspected in males and females with mental retardation and the phenotypic changes and the diagnosis should be confirmed by DNA testing. No treatment is known except for supportive measures for the psychological and behavioral disorders. No treatment is usually needed for the overgrowth.

## SYNDROMES AND OTHER DISORDERS

### Marfan's Syndrome

Marfan's syndrome is an inherited disorder of connective tissue affecting the skeletal system, with elongation of the tubular bones (dolichomorphism), the cardiovascular system, and the ocular system. It is estimated to affect 1 in 20,000 people. The natural history of the syndrome, the autosomal-dominant inheritance, and the variability of the manifestations were pointed out by McKusick in 1955.

The clinical manifestations relating to the skeleton include tall stature, long and thin arms and legs (dolichostenomelia) (Fig. 3), arachnodactyly (spidery fingers), pectus excavatum (hollow chest) or



**Figure 3** A boy, age 14 years, 2 months, with Marfan's syndrome, showing dolichomorphism, slender extremities, and long limbs; arm span is 10 cm more than height. He has cardiovascular changes with dilation of the aortic root.

carinatum (pigeon breast), narrow facies with narrow palate, and scoliosis and kyphosis in 60 to 100% of the patients. Joint laxity and inguinal, femoral, and diaphragmatic hernias are other consequences of the abnormal connective tissue. The ocular manifestations include upward subluxation of the lenses, as a result of the defect of the suspensory ligament, increased axial globe length with myopia, and retinal detachment. The most life-threatening complications are those of the cardiovascular system, with dilation of the ascending aorta with or without a dissecting aneurysm and less commonly of the thoracic or abdominal aorta or pulmonary artery. As a consequence of the dilation of the aorta, there is secondary aortic regurgitation. Mitral valve prolapse is very common also. Without treatment for the cardiovascular complications, particularly the aortic dilation, the mean age of death is in the midforties, from aortic dissection and rupture.

Patients with Marfan's syndrome are mentally normal, but neuropsychologic impairment, including learning disability and attention deficit disorder, occurs in approximately 40% of patients.

This is an inherited disorder, transmitted in an autosomal-dominant fashion. Approximately 85% of patients have a positive family history and 15% of patients have a sporadic presentation. The basic defect in Marfan's syndrome has been traced to a defective fibrillin gene (FBN1) mapped to chromosome 15 (15q21.1). Fibrillin is a connective tissue protein found in microfibrils, a constituent of elastic tissue and abundant in tissues affected in Marfan's syndrome, including the aorta, the suspensory ligament of the lens, and the periosteum.

For the clinical diagnosis, the revised criteria of 1996 (Ghent Nosology) are followed. These criteria are based on four major diagnostic findings: (1) a positive family history and involvement of (2) the skeletal, (3) the ocular, and (4) the cardiovascular systems.

The diagnosis of Marfan's syndrome can be made in an index case, when there is a major involvement in two different organ systems and a minor involvement in a third system or when there is a mutation in the FBN1 gene, a major involvement of one system, and a minor involvement of a second system.

For a relative of an index case, the diagnosis can be made when there is a major criterion provided by family history, one major criterion in an organ system, and involvement of a second organ system.

Female and male patients with Marfan's syndrome may attain an excessively tall height. Treatment to curtail final height may be indicated, particularly for females, for the same reasons that one treats tall normal girls. In addition, since scoliosis and kyphosis might develop in 60 to 100% of patients with Marfan's syndrome, the arrest of growth may be beneficial. Estrogen treatment is effective. For boys, the indication for treatment would also be excessive height and scoliosis. The treatment is injection of testosterone esters, as indicated for tall normal boys. It is crucial that attention be paid to the possible development, prevention, and treatment of life-threatening cardiovascular abnormalities. Propranolol, which reduces the pounding of the ventricular ejection on the ascending aorta, is routinely used in patients who are beginning to have dilation of the aortic root. A composite graft operation to replace the ascending aorta and aortic valve has been highly successful.

### Beals' Syndrome (Congenital Contractural Arachnodactyly)

Although there were previous reports consistent with this disorder, Beals and Hecht delineated this syndrome in 1971 and it is known as Beals' syndrome. It is also known as congenital contractural arachnodactyly (CCA; joint contractures and arachnodactyly).

The skeletal features are similar to those for Marfan's syndrome, with long slender limbs (dolichostenomelia) and arachnodactyly. There is camptodactyly of the fingers. The difference between Beals' syndrome and Marfan's syndrome is that there are joint contractures, rather than looseness of the joints, and the eye and aorta are not affected. The helixes of the ears are folded and the ears have a wrinkled appearance. Kyphosis, scoliosis, or kyphoscoliosis occurs in 50% of the patients who are more severely affected.

This is an inherited disorder segregating in an autosomal-dominant fashion. It has been demonstrated that CCA is caused by a mutation in a second fibrillin gene (FBN2), which is mapped to chromosome 5q23-q31.

Diagnosis is based on clinical grounds. Although rare, at least 33 pedigrees with CCA have been described.

### Homocystinuria

Homocystinuria is an inherited inborn error of metabolism of methionine, due to a deficiency of the enzyme cystathionine  $\beta$ -synthase (CBS), originally reported by Carson *et al.* and Gerritsen and Waisman in 1963. CBS converts homocysteine to cystathionine, a reaction requiring pyridoxal phosphate (vitamin B6) as a cofactor. The prevalence is estimated at 1 in 200,000 live births. In Ireland, it is more frequent, occurring in 1 in 40,000 live births. Approximately 40% of patients respond to high doses of vitamin B6 (pyridoxine) and usually have milder clinical manifestations: a delay in the onset or a lower rate of complications than those who do not respond to vitamin B6 therapy.

Cystathionine  $\beta$ -synthase deficiency is inherited as a recessive trait, but there is considerable genetic heterogeneity in known patients. The gene for CBS has been mapped to the q21 region of chromosome 21.

The clinical manifestations are similar to those for Marfan's syndrome, but there are some differences. Major involvement relates to four different organ systems: the ocular system, the skeletal system, the vascular system, and the central nervous system

(CNS). Ectopia lentis or subluxation of the lens downward, contrary to the upward dislocation in Marfan's syndrome, is the most consistent finding. The skeletal abnormalities impart a phenotype that is similar to that in Marfan's syndrome, with dolichostenomelia (elongated and thin arms and legs), arachnodactyly, and tall stature (Fig. 4,) often with eunuchoid proportions. Among the most consistent skeletal abnormalities is osteoporosis. Scoliosis and kyphosis occur frequently. Other abnormalities include genu valgum and pectus carinatum or excavatum. The most frequent central nervous system manifestation is mental retardation, which may occur in as many as 50% of patients, with an IQ ranging from 30 to 75. Thromboembolic episodes involving the large and small vessels are life-threatening, particularly those in the brain; they are common and may occur at any age. The sulfhydryl groups of homocysteine interfere with collagen cross-linking and cause collagen abnormalities. In view of the similarities of many of the clinical features of homocystinuria and Marfan's syndrome, it is likely that many of the manifestations are related to qualitative changes in fibrillin. Sulfhydryl groups may also contribute to the disruption of the vascular endothelium and as a result contribute to thrombosis.

The most consistent biochemical finding has been homocystinuria. There is an increase in serum homocysteine and a reduced concentration or absence

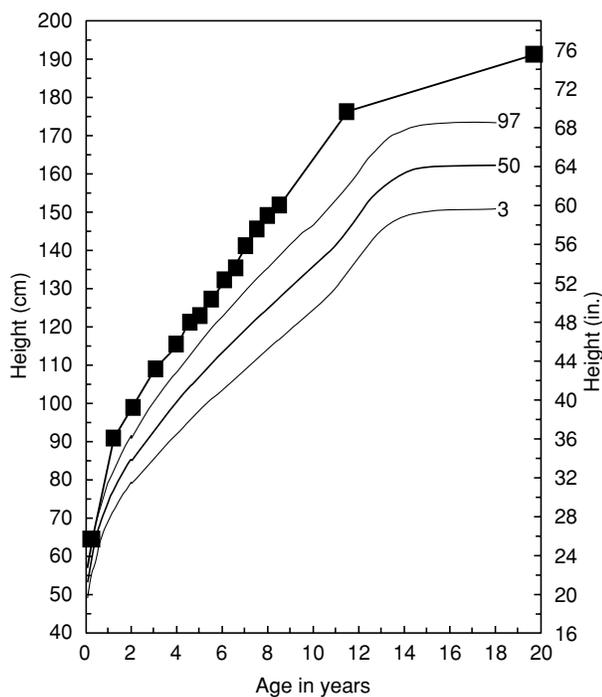


Figure 4 Growth chart of a female patient with homocystinuria.

of cysteine and cystathionine. In addition, serum methionine is increased in most patients.

Some patients may respond to 25 mg of vitamin B6 daily, but usually higher doses are needed. Treatment with high doses of vitamin B6, 200 to 1000 mg per 24 h, causes marked improvement in patients responsive to this therapy. Some patients may not respond because of folate depletion. Thus, folic acid at 1 to 5 mg per 24h should be added to the regimen. In patients who are unresponsive to vitamin B6, restriction of methionine intake, in conjunction with supplementation of cysteine, is recommended. Surgery should be avoided in these patients whenever possible. Should surgery be necessary, hydration pre- and postoperatively may lessen the risk of thrombosis.

### Beckwith-Wiedemann Syndrome

This syndrome was initially described independently by Beckwith and Wiedemann in 1963. It is associated with prenatal and postnatal overgrowth, with the most characteristic features occurring at birth: omphalocele or umbilical defects, macroglossia, and gigantism (Fig. 5). Thirty to 50% of affected children may have severe, persistent hypoglycemia beginning in the first days of life from hyperinsulinism, as a result of pancreatic islet cell hyperplasia. The hypoglycemia usually subsides by 4 months of age. There is visceromegaly with enlargement of the liver, kidneys, pancreas, and occasionally the heart; renal medullary dysplasia; fetal adrenal cortical cytomegaly; and interstitial cell hypoplasia of the gonads. Hemihypertrophy is present in 12.5% of cases.

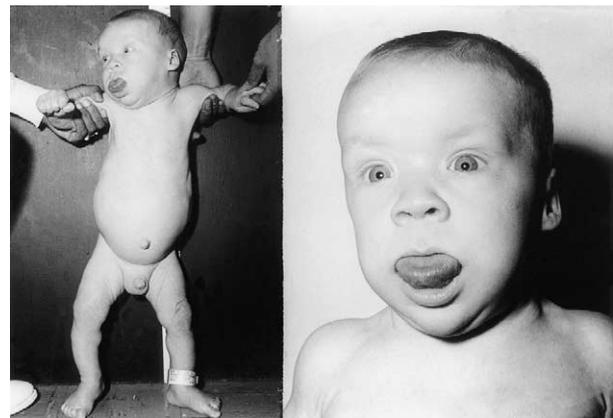


Figure 5 A 4½ month old infant with Beckwith-Wiedemann syndrome, showing macroglossia, umbilical hernia, prominent calf muscles, and overgrowth. Length is 67.3 cm (26½ in.) and weight is 7.7 kg (both >75th percentile).

There is an increased incidence of malignant tumors (7.4 to 10%), the most frequent being Wilms' tumors and adrenal cortical carcinoma. Other tumors could be nephroblastoma, hepatoblastoma, and rhabdomyosarcoma.

Children with Beckwith-Wiedemann syndrome (BWS) are large at birth. Growth velocity is usually above the 90th percentile until 4 to 6 years of age and is normal thereafter. Patients reach an average height of 2.5 SDS at or after puberty with weights between the 75th and 95th percentiles.

Several hundred cases have been reported. The prevalence is not known, but is estimated to be 1 in 14,000 persons. Approximately 85% of cases are sporadic and 15% are inherited, suggesting autosomal-dominant inheritance with incomplete penetrance.

BWS is a complex and heterogeneous genetic disorder resulting from alterations of the expression of imprinted genes, involved in growth and cell cycle control, in the 11p15 chromosomal region. The 11p15 region harbors a cluster of imprinted genes. IGF-II, insulin (Ins), and LIT1 (loss of intestine 1) normally show paternal expression, whereas H19, p57KIP-2, and KvLQT1, except for some tissue variation, are predominantly maternally expressed. The loss of imprinting of the maternal IGF-II gene, with resulting biallelic expression, is one of the most common molecular defects found in patients with Beckwith-Wiedemann syndrome without chromosomal abnormalities. In other patients, loss of maternal IGF-II imprinting is associated with complete suppression of maternal H19 expression. Overexpression of IGF-II can also result from paternal uniparental disomy of chromosome 11 or from duplications of the paternal 11p15 region associated with trisomy of 11p. The contribution or role of the loss of function of maternally expressed genes that act as growth or tumor suppressors is difficult to assess. Alterations in any of these suppressors could cause BWS and the phenotypic spectrum might depend on which maternally expressed gene is mutated. Loss of imprinting of the maternal LIT1 gene with biallelic expression of LIT1 is the other most frequent abnormality in BWS, accounting for 40 to 50% of patients, and is not linked to loss of imprinting of IGF-II.

The diagnosis is based on the clinical manifestations that have been described. Some of the phenotypic changes are somewhat similar to those of infants of diabetic mothers. Patients with Beckwith-Wiedemann syndrome may be difficult to distinguish from patients with Simpson-Golabi-Behmel syndrome (SGBS). Patients with the latter syndrome have some features similar to Beckwith-Wiedemann

syndrome. Distinguishing characteristics of SGBS are a cleft lip and palate, cardiac defects, polydactyly, vertebral and rib anomalies, hypoplastic or absent index fingernails, and X-linked recessive transmission.

Detection and treatment of hypoglycemia are most important for survival and to prevent neurological damage. The excessive rate of growth for the first few years and the tall stature during childhood and adolescence require no treatment, unless an excessive adult height is predicted, particularly in females. Regular follow-up for possible tumor development is needed and routine ultrasonography of the kidneys is mandatory (every 3 months for the first 6 years is recommended), because Wilms' tumor and adrenocortical carcinoma are the most frequent neoplasms.

### Somatic Overgrowth (H19 Methylation and Silencing)

As discussed in relation to Beckwith-Wiedemann syndrome, the H19 gene locus is in the chromosome 11p15 region and may function as a tumor suppressor. H19 is closely related to IGF-II and is paternally imprinted and maternally expressed in most tissues. Biallelic expression of IGF-II by disruption of maternal IGF-II imprinting has been shown in mice, as a result of a maternally inherited targeted H19 gene. Also, biallelic expression of IGF-II, associated with methylation and silencing of the maternal normally expressed H19 gene, has been found in BWS and Wilms' tumors.

Morison *et al.* examined H19 methylation and IGF-II expression in children with overgrowth, without diagnostic features of BWS or abnormalities suggesting any particular syndrome.

In three of the six children with somatic overgrowth, without features of BWS, the overexpression of IGF-II was attributable to abnormal methylation and silencing of H19. This observation has important implications for the evaluation of children with overgrowth without other manifestations suggestive of a syndrome.

### Simpson-Golabi-Behmel Syndrome

This syndrome was originally described by Simpson in 1975 and by Golabi and Behmel, independently, in 1984.

This syndrome is an X-linked recessive disorder characterized by prenatal and postnatal overgrowth, unusual facial appearance (described in the past as bulldog syndrome), and digital and other anomalies

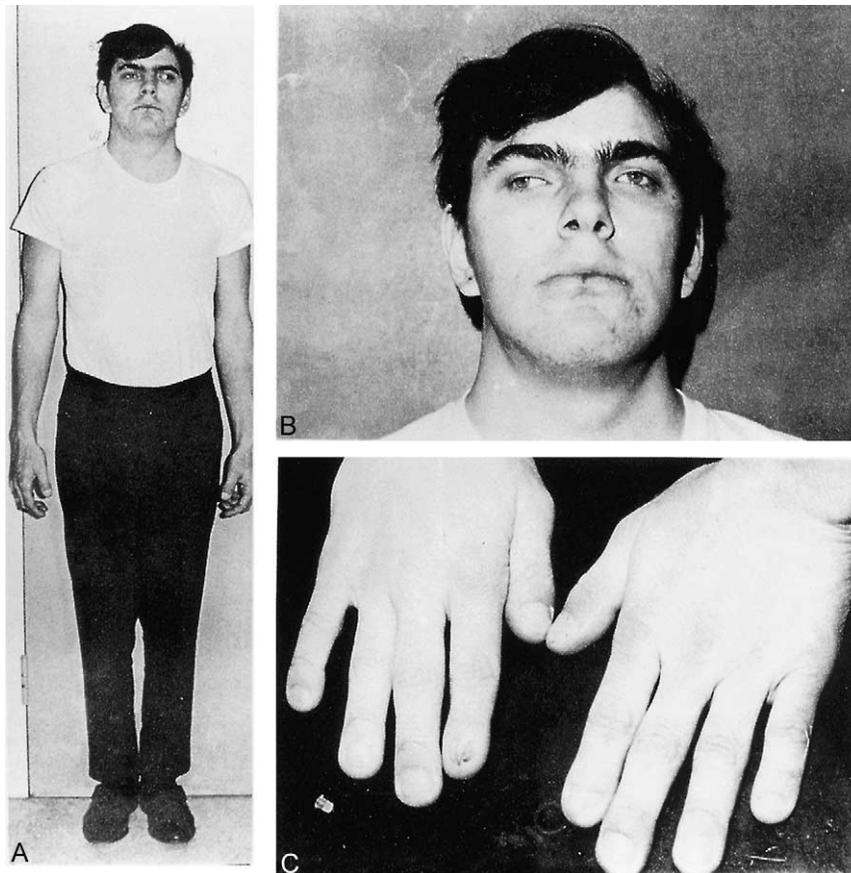
(Fig. 6). Female carriers sometimes can have some facial changes. Affected male patients may attain adult heights of 192 to 210 cm (6 ft 4 in. to 6 ft 11 in.).

Patients have macrosomia, macroglossia, visceromegaly, omphalocele, renal dysplasia, earlobe creases, neonatal hypoglycemia as a result of islet cell hyperplasia, and a risk of embryonal tumors, including Wilms' tumor, neuroblastoma, and hepatocellular carcinoma during early childhood, features similar to those in BWS.

The most prominent features of the disorder consist of overgrowth, characteristic facial changes (large protruding jaw, widened nasal bridge, upturned nasal tip, broad nose, wide mouth, large tongue, thick lips), high arched or cleft palate, large head, hypoplastic or absent index fingernails, and inguinal hernia. The bone age is advanced. Intelligence is usually normal or only mildly retarded in some cases. Hypotonia is common.

The gene for this syndrome has been identified by Pilia *et al.* This syndrome results from different microdeletions of the glypican-3 (GPC-3) gene in Xq26. Mutations or deletions were identified subsequently, suggesting that SGBS is caused by a nonfunctional GPC-3 protein. GPC-3 membrane-bound protein interacts with and binds IGF-II and may be important in the modulation of IGF-II. Theoretically, glypican-3 could complex with and sequester IGF-II or facilitate degradation of IGF-II through a further interaction with the IGF-II receptor. The loss of function of glypican-3 and decreased IGF-II binding would lead to an increase in the level of IGF-II that could activate the IGF-I receptor.

Patients with SGBS have many overlapping features with patients affected with Beckwith-Wiedemann syndrome and in some cases molecular studies will be needed to assign a proper diagnosis. SGBS does not overlap genetically but does overlap clinically with



**Figure 6** An 18-year-old with Simpson-Golabi-Behmel syndrome, illustrating the (A) broad stocky appearance, (B) unusual facial features, and (C) malformed nails. Reprinted from Simpson, J. L., *et al.* (1975). *Malformation Syndromes of the Birth Defects: Original Article Series* 11(2), 22, with permission from the March of Dimes Birth Defects Foundation.

BWS, which may be related to overexpression of IGF-II.

### Sotos' Syndrome

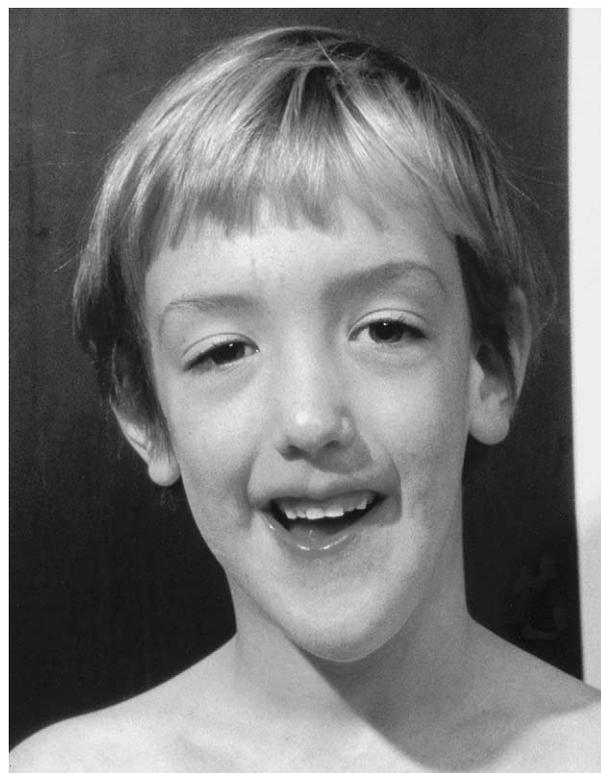
This syndrome was described in 1964. The major diagnostic features are excessive growth, large dolichocephalic head, distinctive facial configuration, advanced bone age, and a nonprogressive neurological disorder with mental retardation. More than 300 cases have been reported and many more are known. The prevalence is not known, but it is probably one of the most frequent overgrowth syndromes after BWS and Marfan's syndrome. Males and females are affected equally. It occurs in all ethnic groups and has been detected throughout the world.

The main clinical finding is prenatal and postnatal overgrowth. The growth velocity is particularly excessive in the first 3 to 4 years of life and subsequently proceeds at the normal rate, but in the high percentiles. The mean height increases from 2.2 SDS at birth to 2.8 SDS at 1 year of age and to mean values of 3 SDS during childhood. The weight is usually appropriate for the height and the bone age is advanced by 2 to 4 years over chronological age, during childhood. Adult height usually exceeds the 50th percentile of normal. Some individuals may reach excessive adult heights: males of 193 to 203 cm (6 ft 4 in. to 6 ft 8 in.) and females up to 188 cm (6 ft 2 in.) are known.

The craniofacial configuration is most characteristic, with a prominent forehead and receding frontoparietal hairline in 96% of the cases, dolichocephalic large head, hypertelorism, down-slanting of the palpebral features, high narrow palate, prominent palatine ridges, and pointed chin (Fig. 7). Premature eruption of teeth occurs in 60 to 80% of patients. CNS manifestations are frequent. Delay in the attainment of milestones of development, walking and in particular talking, is almost always present and clumsiness is frequent (60 to 80%), as are hypotonia and lax joints. Mental deficiency is present in 80 to 85% of the patients, with an average IQ of 72 and a range from 40 to borderline mildly retarded. Fifteen to 20% may be normal mentally and may have IQs of up to 129. Seizures may occur in 30% of the cases. Mildly enlarged ventricles and increased subarachnoid spaces may be present in some patients.

No endocrine abnormalities have been found to explain the rapid growth.

A genetic cause was suspected for a long time. Several families with members affected in two or three generations have been reported, suggesting autosomal-dominant inheritance. Most of the cases are



**Figure 7** A girl, age 4 years, 8 months, with Sotos' syndrome. Height was 124 cm (4 ft 7/8 in.; height age 7 years, 7 months) and weight was 35.2 kg (weight age 7 years, 9 months). Bone age was 4 years, 10 months. Her father (198 cm, or 6 ft 6 in.) and her half-sister from a different mother are also affected. Note the characteristic features: large head, frontal bossing, hypertelorism, down-slanting of palpebral fissures, pointed chin, and large ears.

sporadic, but could be due to new mutations. Deletions or point mutations of a single gene, nuclear receptor-binding SET (where SET denotes SU(VAR)3-9, E(Z)trithorax) domain protein 1 (NSD1), located at chromosome 5q35 were identified in 65-75% of the sporadic cases, indicating that haploinsufficiency of NSD1 is the major cause of Sotos' syndrome. Other genetic abnormalities are possible. The finding that all the NSD1 mutations identified were either heterozygous or hemizygous is consistent with an autosomal-dominant condition in the majority of cases of Sotos' syndrome.

The human NSD1 gene consists of 23 exons, encodes a protein of 2696 amino acids, and is expressed in the human fetal and adult brain, which could explain the findings of large brain and mental deficiency. It is also expressed in skeletal muscle, kidney, spleen, and thymus. The NSD1 protein interacts with the

ligand-binding domain of nuclear hormone receptors and may act as a corepressor or coactivator. The finding that haploinsufficiency of NSD1 induces overgrowth in Sotos' syndrome indicates that NSD1 acts as a corepressor of genes that promote growth.

There is no biochemical marker for the disease. The diagnosis is based on clinical grounds. The most characteristic manifestations are excessive growth and the craniofacial configuration. Without these signs, the diagnosis cannot be made. It is anticipated that future DNA studies of patients would permit a more certain diagnosis and delineation of the spectrum of the condition. The diagnostic tests are carried out mainly to exclude other possibilities, such as fragile X, Klinefelter's, and excessive growth hormone secretion. Magnetic resonance imaging of the brain, showing mildly enlarged ventricles and increased subarachnoid spaces in some patients, is helpful, but not diagnostic.

The management of the mental retardation is no different than for any child with a mental deficiency. The excessive height is usually not a handicap for males. Girls with a predicted ultimate height in excess of 178 cm (5 ft 10 in.) may benefit from treatment with high doses of estrogen to curtail linear growth, as for tall normal girls. Social and behavioral problems during childhood and immaturity in adulthood may improve with psychological counseling. Other important concerns are the possibility of tumor development (~3.0%) and the risk of transmission. Because the evidence suggests that this is an autosomal-dominant disorder affecting males and females, the affected individual has a 50% risk of having affected children. Affected individuals are fertile. There is no evidence that life span is shortened.

### Weaver's Syndrome

Weaver's syndrome was described in 1974. It is characterized by excessive growth prenatally or postnatally, unusual facies, advanced skeletal maturation, and camptodactyly. At least 37 cases have been reported: 21 males and 16 females.

Some males have attained an adult height of 194.2 cm (6 ft 4½ in.), a weight of 102.2 kg (225 lbs), and a head circumference of 61 cm (24 in.) and some females have attained an adult height of 176.3 cm (5 ft 9 in.), a weight of 87.6 kg (193 lbs), and a head circumference of 59.5 cm (23.43 in.).

The excessive growth is present at birth or has its onset during infancy. Eighty percent of patients are developmentally delayed or have mental retardation. Mental retardation can range from mild to severe.

Many of the features resemble those found in Sotos' syndrome. The craniofacial characteristics, although somewhat similar, are in some points different. Patients with Weaver's syndrome have hypertelorism, large ears, depressed nasal bridge, and down-slanting palpebral features, but the skull, in many cases, is not dolichocephalic, the occiput is often flat, patients do not have a prominent chin, and the face is broad. Another distinctive finding is that in many cases there is hypertonia rather than hypotonia. No consistent endocrinological abnormality has been found. The advanced skeletal maturation is striking, with the carpal age much higher than the maturation of the hand.

Most of the cases have been sporadic. Parent to child transmission has been reported five times, suggesting autosomal-dominant inheritance. NSD1 mutations have been described in three of seven patients with Weaver's syndrome, suggesting that Sotos' and Weaver's syndromes are allelic. Additional observations would be helpful to confirm these findings.

### Bannayan-Riley-Ruvalcaba Syndrome and Cowden's Disease

Bannayan-Riley-Ruvalcaba syndrome (BRRS) is an autosomal-dominant disorder characterized by macrocephaly, multiple hamartomas (lipomas, hemangiomas, lymphangiomas, intestinal polyps) and other tumors (seminoma, germinoma), macrosomia at birth, pigmented spots in the glans and shaft of the penis (Fig. 8), and pseudo-papilledema. Originally, it was described as three different syndromes: Bannayan-Zonana syndrome, Riley-Smith syndrome, and Ruvalcaba-Myhre-Smith syndrome. As more cases were described, it became evident that patients with the different syndromes shared similar manifestations and the unification of the three disorders into the Bannayan-Riley-Ruvalcaba syndrome was suggested.

The gene for BRRS has been localized to chromosome 10q23, overlapping the region for Cowden's disease. Marsh *et al.*, in 1997, demonstrated germline mutations in the PTEN gene (for phosphatase and tensin homologue deleted on chromosome 10) in patients with Bannayan-Zonana syndrome. Some of the mutations in the PTEN gene found in Bannayan-Zonana syndrome were the same as those previously reported in Cowden's disease. Because of the considerable phenotypic overlap between BRRS and Cowden's disease and because of the demonstration that in their patient the PTEN gene was deleted on chromosome 10, by an interstitial deletion of



**Figure 8** A 15-year-old with Bannayan-Riley-Ruvalcaba syndrome. Note pigmented spots on the glans and shaft of the penis. Reprinted from Ruvalcaba, R. H. A., Myhre, S., and Smith, D. W. (1980). *Clinical Genetics* 18, 413–416, with permission. Copyright 1980, Munksgaard International Publishers Ltd., Copenhagen, Denmark.

10q23.2–q24.1, Arch *et al.*, in 1997, suggested that BRRS and Cowden's disease are allelic disorders. In 1998, Zori *et al.* reported a mother with Cowden's syndrome whose son had BRRS; both were heterozygous for the same mutation of the PTEN gene.

Cowden's disease has been described mainly in adults. It is an autosomal-dominant condition characterized by macrocephaly, multiple hamartomatous lesions, especially of the skin and mucous membranes, verrucous skin lesions of the face and limbs, and cobblestone-like papules of the gingiva and buccal mucosa, but also involving hamartomas and neoplasms of internal organs, most commonly in the thyroid, breast, and ovary.

Progressive macrocephaly, scrotal tongue, and mild to moderate mental retardation are important signs of Cowden's syndrome in young children. Trichilemmomas in the nasolabial folds and palmar and plantar hyperkeratotic pits usually become evident later in childhood. They are often accompanied by the appearance of subcutaneous lipomas and cutaneous hemangiomas.

PTEN has been shown to play a large role in human malignancy. Somatic PTEN deletions and mutations have been observed in sporadic breast, brain, prostate, and kidney cancer cell lines and in several primary tumors, such as endometrial carcinomas, malignant melanoma, and thyroid tumors. The spectrum and genotype–phenotype analyses in Cowden's disease and Bannayan-Riley-Ruvalcaba syndrome were studied by Marsh *et al.* Some patients with BRRS and Cowden's

disease have similar manifestations. The difference in the manifestations in the two disorders may be related to different mutations, which needs to be further clarified by genotype–phenotype correlations in the future.

Because of a variety of hamartomas and tumors that could occur in patients with BRRS and Cowden's disease, long-term follow-up of these patients is needed. The treatment depends on the type of hamartoma or tumor.

### Partington's Syndrome

In 1997, Partington *et al.* reported three families with 11 members, male and female, showing a new overgrowth syndrome. The syndrome is characterized by generalized overgrowth, mild to moderate mental handicap, and duplication of 4p16.3. The soft tissues are involved, with thickening and coarsening of the face. The hair on the head is abundant and the eyebrows are bushy. Bony overgrowth leads primarily to a heavy body frame with a large head circumference, prominent supraorbital ridges, and a square jaw, but a rather small, upturned nose (Fig. 9). The hands and feet are large. The degree of mental handicap varies from moderate to mild with no specific behavioral characteristics. Considerable variation is seen in the manifestations of this syndrome between various members of the same family.

The height was well above average. Although physical overgrowth was detectable in childhood, it became most obvious in late adolescence and early adult life. Of the 10 subjects with overgrowth, all of whom were over the age of 15 years, all the heights were at or above the 75th percentile and 5 of them were above the 90th percentile. Some subjects experienced late growth. One patient between the age of 18 and 28½ years grew 15 cm (5.9 in.) and another patient between the age of 17½ years and 28 years grew 26 cm (10.2 in.), growth patterns that are quite abnormal. The patients are large people with large body frames.

The 4p16 region contains many genes, some of them responsible for well-known disorders such as Huntington's disease, achondroplasia, thanatophoric dysplasia, and hypochondroplasia. The last three result from mutations in the fibroblast growth factor receptor 3 (FGFR3) gene. In addition, deletions of this region result in Wolf-Hirschhorn syndrome and Pitt-Rogers-Danks syndrome.

The three families reported by Partington *et al.* had different translocations: translocation of 4p16.3 and 1q44 in family 1, translocation of 4p16.2 and 8p23.1 in family 2, and translocation of 4p16.3 and 21q22.3



**Figure 9** The two women, shown standing, are sisters and both have a translocation of 4p16.2 and 8p23.1: t(4;8) (16.2;p23.1). The two patients seated on the left (male, age 18 years, and female, age 16 years), affected with overgrowth disorder, are the children of the woman on the left. The two patients seated on the right are the children of the woman on the right. One is affected with overgrowth (female, age 15 years) and the other has Pitt-Rogers-Danks (PRD) syndrome (male, age 13 years). The children with overgrowth (two were tested) showed an unbalanced translocation with three doses of 4p16.3 (duplication of 4p16.3 and monosomy of 8p23.1). The patient with PRD syndrome showed one dose of 4p16.3 (deletion of 4p16.3 and duplication of 8p23.1). Reprinted from Partington, M. W., *et al.* (1997). *Journal of Medical Genetics* **34**, 719–728, with permission. Copyright 1997, BMJ Publishing Group, London, UK.

in family 3. One of the parents of the patients carried the balanced translocation. As a consequence, the progeny may receive one, two, or three copies of 4p16.3. All four subjects with overgrowth who were studied had an unbalanced translocation with three copies of 4p16.3. Partington *et al.* suggested that FGFR3 could be a candidate gene for growth abnormalities, with three doses leading to the overgrowth syndrome and one dose leading to growth failure (Pitt-Rogers-Danks syndrome).

The suggestion that additional doses of FGFR3 are responsible for the overgrowth has been questioned, because evidence suggests that FGFR3 is a negative regulator of bone growth. Known mutations of FGFR3 are activating mutations and lead to short-limb skeletal dysplasias (achondroplasia, thanatophoric dysplasia, and hypochondroplasia). Whether the overgrowth syndrome described by Partington *et al.*

is the result of overdosage of the FGFR3 gene, inactivating mutations of the gene, or other genes in 4p16.3 remains to be determined by further studies.

## Acknowledgements

This research was supported by S. Robert Davis Funds and The John W. Champion Center.

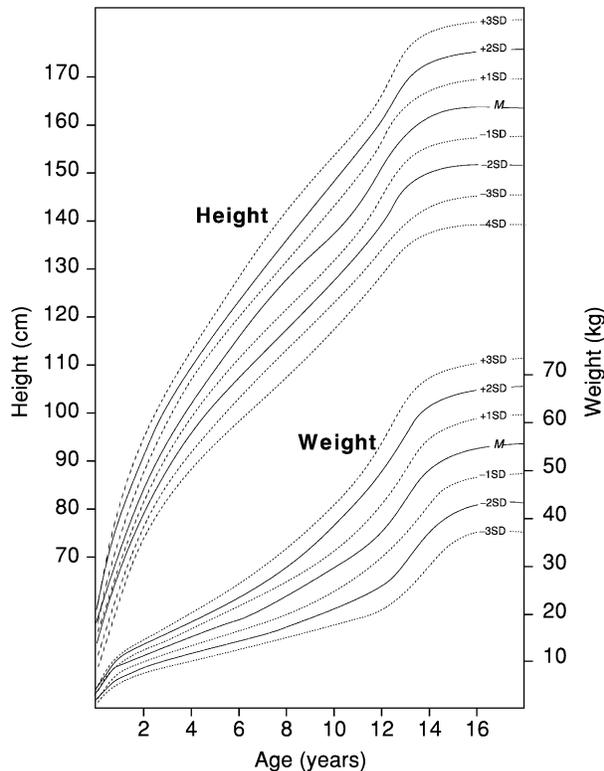
## See Also the Following Articles

Beckwith-Wiedemann Syndrome (BWS) • Body Proportions • Gigantism: Excess of Growth Hormone • Growth, Normal Patterns and Constitutional Delay • Gynecomastia • Klinefelter's Syndrome • Postnatal Normal Growth and Its Endocrine Regulation

## Further Reading

- Cohen, M. M., Jr. (1999). Overgrowth syndromes: An update. *Adv. Pediatr.* **46**, 441–491.
- Cole, T. R. P. (2001). Sotos syndrome. In "Management of Genetic Syndromes" (S. B. Cassidy and J. E. Allanson, eds.), pp. 389–404. Wiley-Liss, New York.
- de Paepe, A., Devereux, R. B., Dietz, H. C., Henekam, R. C. M., and Pyeritz, R. E. (1996). Revised diagnostic criteria for the Marfan syndrome. *Am. J. Med. Genet.* **62**, 417–426.
- Douglas, J., Hanks, S., Temple, I. K., Davies, S., Murray, A., Upadhyaya, M., Tomkins, S., Hughes, H. E., Cole, T. R., and Rahman, N. (2003). NSD1 mutations are the major cause of Sotos syndrome and occur in some cases of Weaver syndrome but are rare in other overgrowth phenotypes. *Am. J. Hum. Genet.* **72**, 132–143.
- Drop, S. L. S., DeWaal, W. J., and De Munick Keizer-Schrama, S. M. P. F. (1998). Sex steroid treatment of constitutionally tall stature. *Endocr. Rev.* **19**, (5) 540–558.
- Hoglund, P., Kurotaki, N., Kytola, S., Miyake, N., Somer, M., and Matsumoto, N. (2003). Familial Sotos syndrome is caused by a novel 1 bp deletion of the NSD1 gene. *J. Med. Genet.* **40**, 51–54.
- Kurotaki, N., Imaizumi, K., Harada, N., *et al.* (2002). Haploinsufficiency of NSD1 causes Sotos syndrome. *Nat. Genet.* **30**, 365–366.
- McKusick, V. A. (2000). Online Mendelian Inheritance in Man (OMIM). Center for Medical Genetics, Johns Hopkins University, Baltimore, MD, and National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD. Available at <http://www.ncbi.nlm.nih.gov/omim/>
- Nagai, T., Matsumoto, N., Kurotaki, N., Harada, N., Niikawa, N., Ogata, T., Imaizumi, K., Kurosawa, K., Kondoh, T., Ohashi, H., Tsukahara, M., Makita, Y., Sugimoto, T., Sonoda, T., Yokoyama, T., Uetake, K., Sakazume, S., Fukushima, Y., and Naritomi, K. (2003). Sotos syndrome and haploinsufficiency of NSD1: Clinical features of intragenic mutations and submicroscopic deletions. *J. Med. Genet.* **40**, 285–289.
- Pilia, G., Hughes-Benzie, R. M., MacKenzie, A., Baybayan, P., Chen, E. Y., Huber, R., *et al.* (1996). Mutations in GPC3, a glypican gene, cause the Simpson-Golabi-Behmel overgrowth syndrome. *Nat. Genet.* **12**, 241–247.
- Rao, E., Weiss, B., Fukami, M., Rump, A., Niesler, B., Mertz, A., *et al.* (1997). Pseudoautosomal deletions encompassing a novel

- homeobox gene cause growth failure in idiopathic short stature and Turner syndrome. *Nat. Genet.* **16**, 54–63.
- Rio, M., Clech, L., Amiel, J., Faivre, L., Lyonnet, S., Le Merrer, M., Odent, S., Lacombe, D., Eder, P., Brauner, R., Raoul, O., Gosset, P., Prieur, M., Vekemans, M., Munnich, A., Colleaux, L., and Cormier-Daire, V. (2003). Spectrum of NSD1 mutations in Sotos and Weaver syndromes. *J. Med. Genet.* **40**, 436–440.
- Salo, P., Kaariainen, H., Page, D. C., and de la Chapelle, A. (1995). Deletion mapping of stature determinants on the long arm of the Y-chromosome. *Hum. Genet.* **95**, 283–286.
- Sotos, J., and Argente, J. (2000). Hipercrecimiento (I) and (II). In “Tratado de Endocrinología Pediátrica y de La Adolescencia, Segunda Edición” (J. Argente, A. Carrascosa, R. Gracia, and F. Rodriguez, eds.), pp. 433–541. Doyma SL, Barcelona, Spain.
- Sotos, J. F. (1996/1997). Overgrowth disorders—Sections I through VI. *Clin. Pediatr.* **35**, 515–529; **35**, 577–590; **35**, 635–648; **36**, 37–49; **33**, 89–103; **36**, 155–170.
- Weksberg, R., and Shuman, C. (2001). Beckwith-Wiedemann syndrome. In “Management of Genetic Syndromes” (S. B. Cassidy and J. E. Allanson, eds.), pp. 49–69. Wiley-Liss, New York.



**Figure 1** Reference curves (in girls): height and weight for age are presented as the mean ( $M$ )  $\pm$  2 SD (standard deviations).

until cessation of growth. They have been analyzed in a normal population and standardized as a reference. For practical purpose, sequential X-rays of the left-hand wrist were selected and presented in an atlas by Greulich and Pyle, which is the internationally accepted reference for bone age assessment. Ideally, chronological and bone ages progress in parallel with little difference. Although the concept of bone age reflecting skeletal maturation, and possibly body maturation, has been a valuable tool in assessing abnormal growth, a number of caveats must be considered when applied to pathological conditions.

### Prediction of Adult Height

The ultimate and often important information is the predicted adult height. Methods have been developed based on bone age, height, and eventually chronological age for such a prediction. In addition, in some genetic diseases, such as Turner's syndrome or achondroplasia, disease-specific growth curves have been developed, which may also be helpful in the evaluation of therapeutic interventions. In normally growing children, because height remains canalized along a given

centile or SD value, adult height can be fairly well predicted by extrapolating within the adult height distribution; it can also be calculated as the mean parental height with the addition or subtraction of 6.5 cm for boys and girls, respectively. This value is only approximate. In addition, the parental genetic background can be accounted for by calculating "the parental target height." These predicted values, because they are derived from statistical calculations, remain approximate, but quite useful.

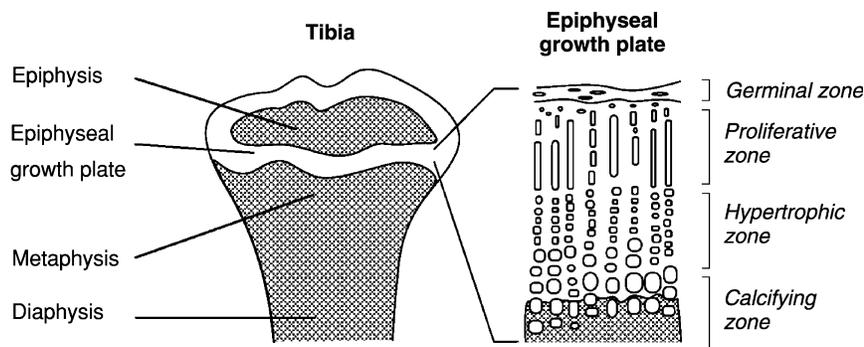
### Catch-Up Growth

The term "catch-up growth" was introduced to describe the phase of rapid growth that allows a child to accelerate growth until he has fully recovered his genetic pace of growth after a prolonged illness with inhibition of growth. The child would return to his original pre-retardation growth curve. It is a compensation phenomenon for the potential loss of growth. This is a reliable indicator of recovery and suppression of the growth-retarding agent. It is observed when growth retardation is secondary to malnutrition or failure to thrive or when replacement therapy is given in appropriate dosage in conditions such as hypothyroidism, complete GH deficiency, or delayed puberty. The mechanisms regulating catch-up growth are still unknown. One hypothesis suggests a cellular phenomenon in which the cell, which eventually becomes the chondrocyte, has a program and a stabilizing mechanism to recognize where it is in that program. Hormones would act as permissive agents to adjust at the program level.

## ENDOCRINE REGULATION OF GROWTH

### The Epiphyseal Growth Plate and Skeletal Growth

Body height depends on the growth of long bones and the spine. It originates from independent ossification centers: the growth plate cartilage separates the epiphysis (distal centers) and the metaphysis of long bones (Fig. 2). This plate is a distinct organ. It is the site of ossification forming the endochondral bone during growth. It is therefore the main target of hormones controlling skeletal growth. Normal growth is only possible if the growth plate is genetically normal and able to respond to endogenous hormonal stimulation. Cessation of growth is due to fusion of epiphyseal and metaphyseal bones with disappearance of the cartilage growth plate.



**Figure 2** Cartilage growth plate, such as that located at the upper end of the tibial bone. The microscopic structure shows the maturation of the chondrocytes from the germinal zone down to the calcifying zone, where bone formation is initiated.

The mechanisms that affect stature are complex: (1) some are intrinsic to the growth plate with local intracellular and intercellular factors controlling chondrocyte multiplication and maturation and (2) others are extrinsic, such as the hormones that promote these changes by acting directly at the cellular level or by interfering with locally produced growth factors. This endocrine control of growth also depends on local receptors specific to each hormone. A state of resistance to a given hormone would occur if its receptor is knocked out by a mutation of the gene encoding it. It would lead to disease, providing a model for the understanding of endocrine control of skeletal growth. This was typically observed in patients resistant to estrogens with inactivating mutations in the estrogen's receptor.

### The GH–Insulin-like Growth Factor-I Axis

By secreting GH, the pituitary gland plays the central role in the regulation of growth (Fig. 3). It is controlled by the hypothalamus, with which it has a close anatomical connection. As a matter of fact, GH secretion results from an interplay of two main hypothalamic factors: a positive factor, GH-releasing hormone (GHRH), and a negative factor, somatostatin- or somatotropin-release inhibiting factor. GHRH is required for GH stimulation as it controls the episodic release of GH. Somatostatin affects the timing and amplitude of GH secretion. The final pattern and amplitude are defined by this reciprocal interaction. A newly identified secretagogue is ghrelin, a GHRH-like peptide produced largely by the stomach, allowing a tight link with nutrition.

Stimulation of GH secretion occurs during sleep, fasting, hypoglycemia, and stress. GH production decreases after birth and remains unchanged during childhood. Its production is enhanced by sex steroids,

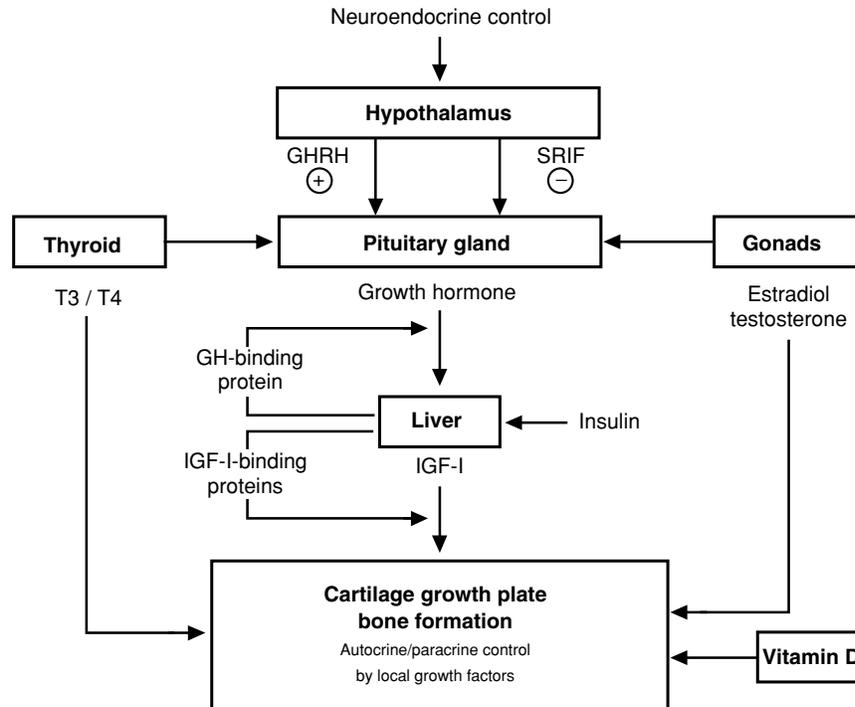
essentially estrogens, during puberty in both sexes. Importantly, GH acts through a specific receptor in target cells, principally in the liver, where most of the circulating insulin-like growth factor (IGF-I) is synthesized, but also in the hypothalamus, where IGF-I controls the GHRH–somatostatin system, and in the early differentiating chondrocyte in the growth plate. Mutations in the GH receptor cause a severe dwarfism, which was described by Laron, with high GH levels but low IGF-I levels and including the full clinical picture of GH deficiency.

A critical finding in the early 1950s showed that the action of GH was mediated by a factor that was later named IGF-I. Plasma IGF-I values are stable, in contrast with pulsatile GH concentrations. It is therefore accepted that IGF-I levels reflect the status of GH secretion in normal individuals. Variations are related to age and puberty. As a consequence of these findings, GH deficiency can be treated with recombinant GH, allowing the stimulation of IGF-I production and bone growth.

The predominant action of GH is mediated by liver-secreted IGF-I, which is delivered to the growth plate. However, there is an additional mechanism with local production of IGF-I that is also controlled by GH, promoting the proliferation of chondrocytes. The crucial role of IGF-I at that level is supported by the severe growth retardation induced by deletions in the gene coding for the GH receptor that abolish the production of IGF-I by the liver and dramatically reduce the circulating level of IGF-I.

### Thyroid Hormones

Hypothyroidism (decreased thyroid hormone activity) is a major treatable cause of linear growth and bone maturation retardation in children. Congenital forms of hypothyroidism can be prevented by neonatal



**Figure 3** The endocrine control of growth with the central role of the growth hormone–IGF-I axis. Hormones also have a direct effect on the growth plate.

screening, but hypothyroidism may also occur at any postnatal age. The active thyroid hormone is T3 (triiodothyronine), which is produced by conversion from thyroxine or directly by the thyroid gland. This hormone is necessary for appropriate chondrocyte differentiation in the growth plate. It acts directly through specific receptors and also regulates the local production of IGF-I and GH responsiveness. In addition to its direct effects on the growth plate, T3 controls the secretion of GH by the pituitary: in hypothyroidism, decreased GH secretion contributes to growth retardation. It is corrected by thyroid hormone replacement therapy. An opposite model of T3 action is observed in children with hyperthyroidism, as they show accelerated growth and bone maturation, leading to premature growth plate closure.

### Pubertal Growth and Sex Steroids

Until puberty, there is a slow but continuous bone growth essentially dependent on the activity of the GH–IGF-I axis and the thyroid hormones. At puberty, estrogens and androgens are secreted in both sexes. The major contribution to growth stimulation is made by estrogens, part of which are derived from testosterone. The increase in serum estradiol concentration occurring in both sexes enhances pulsatile GH

secretion and circulating IGF-I, supporting the pubertal growth spurt, which will contribute up to 20% of the final adult height. The high pubertal levels of GH and IGF-I are maintained during several years of rapid growth. There is debate about the respective role of estrogens and androgens during pubertal growth. It is accepted that estradiol secreted by the ovary in girls or derived by aromatization from testosterone in boys is the main factor; the secretion of estradiol is responsible for the growth spurt and its continuous increase is the cause of bone maturation until epiphyseal closure and cessation of growth in both sexes. Numerous genetic diseases, which involve sex steroid secretion or action, provide pathological models of pubertal growth, supporting the view of the prominent role of estrogens.

In addition, androgens may play a role as shown by the growth stimulation obtained with some synthetic androgens that cannot be converted to estrogens. Animal studies have shown that androgens do have direct effects independent of those of estrogens. They promote chondrocyte maturation and bone growth. How these local actions, which are probably similar for estrogens, are integrated with the action of the GH–IGF-I system remains unclear.

In addition to their effect on growth, sex steroids are responsible for the pubertal increase in mineral

bone mass (essentially calcium content) up to the adult level, which is attained in the early twenties. Estrogens contribute substantially, probably up to 25%, to volumetric bone mineral density. However, the 25% greater bone mass in postpubertal boys than in girls is likely due mainly to the pubertal testosterone secretion. Therefore, it is not surprising that the absence or severe delay in puberty in both sexes will lead to persisting growth but insufficient bone mineralization. The opposite occurs if puberty is precocious.

### Cortisol and Glucocorticoids

Cortisol produced by the adrenal gland has little influence on growth under physiological conditions and cortisol deficiency does not impair growth. However, this is finely tuned as demonstrated by the following clinical conditions: (1) when a cortisol-deficient patient requires replacement therapy, adjustment of the dosage will be monitored partly for its effects on growth, with reduced growth velocity induced by a minor overdose, and (2) growth retardation may be the sole symptom, or the first symptom, of chronic hypersecretion of cortisol, as seen in children with Cushing's syndrome. Corticoids have a direct effect on the growth plate and may reduce GH secretion. In addition, they induce a state of IGF-I resistance.

### Vitamin D

Vitamin D is a hormone that is ingested in the diet (as vitamin D) or produced by the skin (as vitamin D<sub>3</sub>). It is transformed into its active biological form by the liver and the kidney. It is a critical regulator of calcium absorption and metabolism and hence of bone

mineralization at all ages. In infants and young children, vitamin D deficiency can impair growth by disrupting chondrocyte maturation and inhibiting mineralization, causing widening of the epiphyseal growth plate with specific changes in bone formation, as can be seen on X-rays of the distal part of long bones. These changes are known as vitamin D-deficiency rickets. This condition may be part of a more severe and complex nutritional disorder, which contributes to the growth impairment in these young children.

### See Also the Following Articles

Body Proportions • Gigantism: Excess of Growth Hormone • Growth Hormone (GH) • Growth, Normal Patterns and Constitutional Delay • Insulin-like Growth Factors • Postnatal Non-Endocrine Overgrowth • Skeletal Development During Childhood and Adolescence

### Further Reading

- Bonjour, J.-P., and Rizzoli, R. (2001). Bone acquisition in adolescence. *In* "Osteoporosis" (R. Marcus, D. Feldman, and J. Kelsey, eds.), pp. 621–638. Academic Press, London.
- Reginato, A. M., Wang, W. F., and Olsen, B. R. (2001). Developmental biology of bone. *In* "Osteoporosis" (R. Marcus, D. Feldman, and J. Kelsey, eds.), pp. 189–212. Academic Press, London.
- Reiter, E. O., and Rosenfeld, R. G. (1998). Normal and aberrant growth. *In* "Williams Textbook of Endocrinology" (J. D. Wilson, D. W. Foster, H. M. Kronenberg, and P. R. Larsen, eds.), pp. 1427–1507. W. B. Saunders, Philadelphia, PA.
- Tanner, J. M. (1994). Auxology. *In* "The Diagnosis and Treatment of Endocrine Disorders" (M. S. Kappy, R. M. Blizzard, and C. J. Migeon, eds.), pp. 137–192. Charles Thomas, Springfield, IL.
- Thorner, M. O., Lee Vance, M., Laws, E. R. Jr., Horwath, E., and Kovacs, K. (1998). The anterior pituitary. *In* "Williams Textbook of Endocrinology" (J. D. Wilson, D. W. Foster, H. M. Kronenberg, and P. R. Larsen, eds.), pp. 249–340. W. B. Saunders, Philadelphia, PA.

changes in extracellular potassium concentration can be life-threatening. Therefore, plasma potassium concentration must be maintained close to its setpoint. Cells act as an important storage pool for potassium, alternatively taking up excess potassium or releasing potassium, and operate to maintain its plasma concentration within the narrow limits allowed for physiologic function.

## EXTRARENAL REGULATION OF POTASSIUM BALANCE

Potassium can be actively transported into and can diffuse out of all cells. Extracellular potassium concentration is therefore regulated by cell membranes that act as effector organs by buffering rapid changes in plasma potassium concentration, such those occurring during the absorptive period. This system is highly effective, particularly in protecting against sudden changes in potassium concentration. For instance, after an intravenous bolus injection of KCl that raises the plasma potassium concentration by 1.1 mmol/liter (a 25% increase), the system is able to return to the setpoint within approximately 7 min of the start of the bolus injection.

The maintenance of distribution of potassium across cells is largely dependent on the activity of the  $\text{Na}^+/\text{K}^+$ -ATPase, which expends energy to pump sodium out of the cell and potassium into the cell. Virtually every cell in the body uses this mechanism to maintain a high intracellular potassium concentration and a low extracellular potassium concentration.

There are several regulatory factors that control the distribution of potassium within body fluid compartments (Table I). Among them, hormones play a prominent role. Insulin drives potassium into cells independent of its hypoglycemic effect, possibly increasing  $\text{Na}^+/\text{K}^+$ -ATPase activity. Also,  $\beta$ -agonists, such as epinephrine, promote rapid and profuse potassium movement from the extracellular to the intracellular fluid by cyclic AMP-dependent activation of the  $\text{Na}^+/\text{K}^+$ -ATPase. Both insulin release and epinephrine release increase after meals, thus facilitating the movement of the ingested potassium to the intracellular space. Hyperkalemia stimulates adrenal aldosterone secretion, which not only affects the renal excretion but also favors a shift of potassium to the intracellular space. Apart from hormonal factors, there are other extrarenal mechanisms that control the distribution of potassium within the body compartments. Changes in the acid-base status can lead to changes in plasma potassium concentration. In particular, a

**Table I Determinants of Transcellular Potassium Distribution**

<b>Lower plasma potassium concentration</b>
Hormones
Insulin
$\beta$ -Adrenergic agonists
Aldosterone
Metabolic alkalosis/respiratory alkalosis
<b>Increase plasma potassium concentration</b>
Cell destruction (hemolysis, crushing)
Hormones
$\alpha$ -Adrenergic agonists
Glucagon
Hyperchloremic acidosis
Plasma hyperosmolality
Drugs
$\beta$ 2-Antagonists
Angiotensin-converting enzyme inhibitors
Digitalis
Prostaglandin synthesis inhibitors
Succinylcholine

decrease in blood pH causes hydrogen ions to enter cells. Consequently, potassium exits cells in exchange for hydrogen entering. Alkalemia generally induces the opposite response. An acute rise in plasma osmolality causes an abrupt increase in plasma potassium, most likely due to the rapid exit of water from the cells, accompanied by an increased amount of potassium leaving the cell. Finally, several drugs (Table I) could significantly affect extrarenal potassium homeostasis.

## RENAL REGULATION OF POTASSIUM BALANCE

Long-term potassium homeostasis is regulated mainly by the kidney, which ultimately controls plasma potassium concentration as well as total body potassium. The balance of potassium intake and potassium excretion depends on the rate of renal excretion, which can range from less than 10 to over 400 mmol per day, accounting for 90–95% of potassium excretion in the body; the remainder is eliminated by the gastrointestinal tract and sweat glands.

Potassium is freely filtered across the glomerular membrane. Although there is a complex process of potassium reabsorption and secretion by the nephron from the proximal tubule through the loop of Henle, the bulk of filtered potassium load is reclaimed in the proximal nephron, mostly in the proximal tubule.

Beyond the early distal tubule, the tubule may modulate potassium delivery by either reabsorbing or secreting it. Hence, urinary potassium excretion is ultimately regulated by changes in potassium secretion at the distal tubule and the collecting duct under the influence of a number of modulators (Table II).

It should be considered that under most circumstances the kidney primarily works to excrete (rather than to reabsorb) potassium to maintain potassium balance against variation in potassium intake from dietary or other sources. Under normal conditions, when ample potassium is consumed in the diet, there is an addition of potassium along the distal nephron; in the presence of a very high potassium intake, the distal nephron effects net secretion, whereas it effects net potassium reabsorption only in the presence of potassium restriction. Cells in the connecting and collecting tubules respond to a number of signals by modulating their rate of potassium secretion over a wide range of values. The  $\text{Na}^+/\text{K}^+$ -ATPase located in the basolateral membranes of the tubules provides the driving force for potassium secretion by pumping potassium into the cells. This enzyme system

responds to changes in extracellular potassium concentration, changes in extracellular pH, and changes in mineralocorticoid hormone secretion.

First, intake of potassium is the most important determinant of potassium secretion, by affecting the tubular excretion rate, probably as a function of intracellular potassium concentration. Second, the feedback control of aldosterone secretion by the zona glomerulosa of the adrenal cortex is needed for regulation of potassium excretion. Aldosterone secretion is under the strict control of two stimuli, i.e., the plasma concentration of angiotensin II and the plasma concentration of potassium itself. Aldosterone, which stimulates  $\text{Na}^+/\text{K}^+$ -ATPase, promotes potassium secretion by increasing basolateral potassium entry into the cells and intracellular potassium concentration. This in turn will favor the delivery of potassium into the renal tubule lumen down its concentration gradient, thereby stimulating potassium secretion. Glucocorticoids also affect potassium secretion primarily by increasing tubule fluid flow rate. Acid-base changes acutely affect secretion. Alkalosis induces greater potassium secretion at the distal tubule by moving potassium into the cells and acidosis moves potassium out of cells as hydrogen ions move in. The urine flow rate itself affects potassium secretion; with diuresis, increased amounts of fluid and salts move to the distal tubule, enhancing potassium excretion. Thus, pharmacological agents, particularly diuretics, may alter potassium excretion. For instance, loop diuretics, such as furosemide, in addition to increasing flow rate to the distal nephron, inhibit potassium reabsorption by the thick ascending limb of the loop of Henle. In contrast, the so-called potassium-sparing diuretics amiloride and triamterene block the apical sodium channel in the late distal tubule and collecting duct. The blockade of the apical sodium channel increases sodium excretion without affecting potassium secretion. Finally, aldosterone antagonists (e.g., spironolactone) also increase sodium excretion without an increase in potassium excretion.

**Table II Determinants of Renal Potassium Excretion**

**Lower potassium excretion**

- Decreased intracellular potassium concentration
  - Decreased dietary intake
  - Acidosis
- Unfavorable electrical or chemical profile
  - Reduction of luminal electronegativity
  - Reduction of distal urine flow
  - Reduction of distal sodium delivery or reabsorption
- Kidney function
  - Decreased tubular filtration rate
  - Tubular damage
- Diuretics
  - Inhibitors of the apical sodium channel (amiloride, triamterene)
  - Inhibitors of aldosterone

**Increase potassium excretion**

- Increased intracellular potassium concentration
  - Increased dietary intake
  - Alkalosis
  - Aldosterone
- Favorable electrical or chemical profile
  - Increase in luminal electronegativity
  - Increase in distal urine flow
  - Increase in distal sodium delivery
  - Low luminal chloride concentration
- Loop diuretics (furosemide)

**NUTRITIONAL AND METABOLIC ASPECTS**

Potassium is largely present in fruit, vegetables, legumes, and nuts, in the form of inorganic (bicarbonate and phosphate) or organic (gluconate and citrate) salts. Potassium is also found in milk and dairy products and in some types of meat (see Table III).

The total, nondiscretionary intake of potassium varies among different populations. In the

Table III Potassium Content<sup>a</sup> of Selected Food Items

<b>Legumes</b>	
Beans (dry)	1445 (37.05)
Canned beans	232 (5.94)
Beans (fresh)	650 (16.66)
Lentils (dry)	980 (25.12)
Chick peas (dry)	800 (20.51)
Broad beans (fresh)	210 (5.38)
Peas (fresh)	202 (5.17)
Soybeans (dry)	1740 (44.61)
<b>Vegetables</b>	
Asparagus	239 (6.12)
String beans	280 (7.17)
Mushrooms	470 (12.05)
Peppers	210 (5.38)
Turnip	240 (6.15)
Squash	210 (5.38)
Carrots	220 (5.64)
Fennel	276 (7.07)
Potatoes	570 (14.61)
Cabbage	260 (6.66)
Artichoke	376 (9.64)
Cauliflower	350 (8.97)
Eggplant	184 (4.71)
Tomatoes	297 (7.61)
Spinach	530 (13.58)
Broccoli	340 (8.74)
Chicory	180 (4.61)
Lettuce	240 (6.15)
Green peppers	129 (3.30)
Prickly lettuce	240 (6.15)
<b>Fruit</b>	
Apricots	320 (8.25)
Cherries	229 (5.87)
Peaches	260 (6.66)
Oranges	200 (5.12)
Orange juice	180 (4.61)
Pears	130 (3.33)
Grapes	192 (4.92)
Apples	120 (3.07)
Plums	190 (4.87)
Bananas	350 (8.97)
Grapefruit	230 (5.89)
<b>Milk</b> (skimmed)	150 (3.84)

<sup>a</sup>mg/100g (mmol/100 g).

International Collaborative Survey of Electrolyte Excretion and Blood Pressure, which analyzed 24h sodium and potassium urinary excretion in over 10,000 individuals from 32 countries, the average potassium intake ranged from 25 mmol/day (American

blacks in Goodman, Wisconsin, United States) to 87 mmol/day (Xingu Indians, Brazil). A few isolated populations participating in this worldwide survey were found to consume much greater amounts of potassium and smaller amounts of sodium than people living in industrialized countries, their sodium/potassium ratio being similar to that estimated by anthropologists for our hunter-gatherer ancestors. As discussed above, major sustained increases in plasma potassium concentration are almost never caused by dietary changes, because the homeostatic mechanisms are highly efficient in healthy individuals. In response to a large oral potassium load, the excess potassium is partly stored in the large cellular compartment and partly excreted in the urine in a matter of a few hours. Thus, hyperkalemia may develop only in the presence of severe pathological conditions that impair the activity of these defense mechanisms (e.g., renal failure, adrenal insufficiency). On the other hand, the control of potassium homeostasis is less effective in preventing the hypokalemia caused by a subnormal potassium intake, particularly in the presence of a concomitant high-sodium diet, which induces an obligatory potassium loss. In fact, severe potassium depletion may occur when habitual potassium intake is less than 25 mmol/day. In particular, two conditions may often lead to potassium depletion. One is the "protein-modified fast," which may be adopted by severely obese patients in the early phase of a weight-reducing program; the other is chronic alcoholism, due to the large amount of "empty calories" ingested from alcohol.

For the reasons just discussed, the collection of accurate information about dietary potassium intake is of primary importance. As powerful homeostatic mechanisms act to minimize changes in plasma potassium, the measurement of plasma potassium concentration is of limited value in estimating the adequacy of potassium intake. Several approaches based on dietary assessment methods have been proposed. The personal recall of food consumption is the least reliable method; history records are slightly better, although their accuracy may be hampered by incomplete record-keeping, variability in the potassium content of foods, and errors in estimation of quantity. The most accurate single technique for estimation of dietary intake of nutrients, including potassium, is the collection and analysis of duplicate food samples; unfortunately, this method is expensive and not applicable to population studies. For everyday purposes, the measurement of urinary potassium excretion provides a reasonably good estimate of dietary intake. Its use is based on the assumption that at least 90% of the potassium ingested is eliminated through

the renal route. Well-recognized limitations of this method include the relatively high day-to-day intra-individual variability in potassium intake (and thus potassium excretion) and also the concomitant variability in sodium intake, which in turn influences potassium excretion. To overcome this problem, a certain number of 24 h urine collections are requested if a reasonably good estimate of an individual's habitual potassium intake is to be obtained. Despite these limitations, the measurement of urinary excretion is as yet the simplest and least expensive way to assess dietary sodium and potassium intakes in population studies as well as in clinical studies.

### See Also the Following Articles

Ion Channels • Peptide Hormones, Intracellular Transport

### Further Reading

- DeFronzo, R. A. (1987). Regulation of extrarenal potassium homeostasis by insulin and catecholamines. *In* "Potassium Transport: Physiology and Pathophysiology. Current Topics in Membranes and Transport" (G. Giebisch, ed.), Vol. 28, pp. 299–329. Academic Press, San Diego, CA.
- Giebisch, G., Malnic, G., and Berliner, R. W. (1986). Renal transport and control of potassium excretion. *In* "The Kidney" (B. M. Brenner and F. C. Rector, Jr., eds.), 3rd ed., pp. 177–205. Raven Press, New York.
- Seldin, D. W. and Giebisch, G. (eds.). (1989). "The Regulation of Potassium Balance." Raven Press, New York.
- Siani, A., and Strazzullo, P. (1996). Why and how to increase dietary potassium intake. *Nutr. Metab. Cardiovasc. Dis.* **6**, 245–254.
- Thier, S. O. (1986). Potassium physiology. *Am. J. Med.* **80**, 3–7.
- Young, D. B. (1990). Potassium homeostasis and blood pressure and sodium-volume regulation. *In* "Hypertension: Pathophysiology, Diagnosis, and Management" (J. H. Laragh and B. M. Brenner, eds.), pp. 1131–1150. Raven Press, New York.

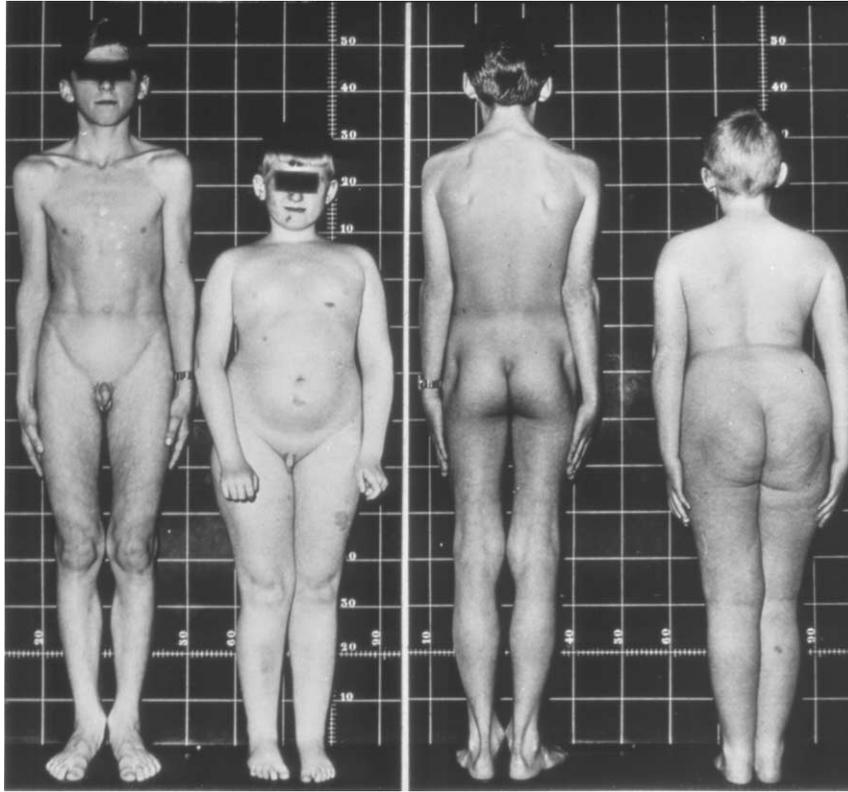


Figure 1 Original photo of twins by Prader: in each photograph, the boy with PWS is on the right and the healthy boy is on the left.

average of 145 to 150 cm. In some individuals, scoliosis, as well as osteopenia and osteoporosis, may be present.

### Hypogonadism

Genital hypoplasia (particularly in boys, who present with a small penis), cryptorchidism, and a bifid or



Figure 2 Severe muscle hypotonia in an infant with PWS.

hypoplastic scrotum are common features. In most cases, puberty remains incomplete, more so in boys than in girls, as a consequence of hypothalamic dysfunction. Although precocious development of pubic and axillary hair is a frequent finding, it is the consequence of premature secretion of adrenal androgens, but the pubertal development of the gonadal axis is delayed, insufficient, or absent. Single cases with complete precocious puberty are reported in boys with adult testicular volumes and in girls with menarche. Two women with PWS have become pregnant. In one case, a woman with maternal disomy gave birth to a healthy girl, whereas the other woman, who had a deletion of 15q11–q13, gave birth to a child with Angelman's syndrome.

### Developmental Delay and Mental Retardation

Developmental delay is a major concern in most cases. In particular, speech and motor development are retarded. Speech and language difficulties become apparent from an early age on. Children with PWS

usually are able to sit up at 12 months, walk between the ages of 28 and 32 months, and talk in short sentences as late as at 42 months. Most individuals with PWS present mild to moderate mental retardation and IQ testing indicates an average IQ of approximately 60 to 70.

### Behavioral Problems

During early childhood, a characteristic behavioral profile emerges. Even at this early stage, young children typically stick to an activity with more persistence than other children and have difficulty with changes in routine. Nevertheless, younger children are happy, affectionate, and cooperative. After the age of 2 years, children with PWS may become obsessed with food and later develop all kinds of food-seeking strategies and atypical behavior, such as gorging on available food, breaking into locked food storage areas, and getting up at night to forage for food. By the time they reach school age, the compulsiveness and obsession of their behavior become more evident. Typical temper tantrums are often observed. Stubbornness and intolerance of frustration relates in the beginning primarily to withholding of food, but may occur also in other situations at a later stage. Adolescents are almost always described as extraordinarily stubborn, clever, manipulative, moody, and prone to temper outbursts. The incidence of psychological disorders is increased in PWS and genuine psychosis occurs in approximately 5 to 10% of patients.

### Respiratory Abnormalities

Respiratory abnormalities in PWS are well known. An increased incidence of sleep-related breathing disorders has been reported in obese adults with PWS and a primary disturbance of central respiratory control has been demonstrated in young, not yet obese children with PWS. The pathogenesis of respiratory problems seems to be multifactorial in origin, including peripheral and central mechanisms, such as muscular hypotonia and facial dysmorphism as well as hypothalamic and chemoreceptor dysfunction.

### Miscellaneous Characteristics

Individuals with PWS have a characteristic face with a narrow bifrontal diameter, almond-shaped eyes, strabismus, and a triangular mouth. Many individuals are hypopigmented with fair hair and blue eyes. Oral

characteristics include thick saliva, hypoplastic enamel, and caries. Pain sensitivity is reduced and a tendency to self-injury is observed, especially skin-picking on arms, hands, and feet.

### Life Expectancy

Life expectancy has been prolonged well into adulthood. Whereas complications of morbid obesity, such as type 2 diabetes and cardiac or respiratory deficiencies have previously doomed affected individuals to an early death, increased awareness of the syndrome and more attention to its management have significantly reduced their likelihood.

### DIAGNOSTIC CRITERIA

Diagnostic criteria for PWS were first proposed by Holm in 1981 and were further developed through a consensus process in 1993. At that time, sophisticated genetic analyses were not yet widely available. Because diagnosis of PWS can be confirmed by genetic testing, clinical diagnostic criteria should be used more often to raise diagnostic suspicion and prompt testing. Therefore, the purpose of clinical diagnostic criteria has shifted from assisting in confirming the diagnosis to raising diagnostic suspicion, thereby prompting testing. Accordingly, revised clinical criteria to help identify appropriate patients for DNA testing for PWS have been suggested by Gunay-Algun *et al.*

### METABOLISM IN PWS

#### Carbohydrate Metabolism

Diabetes mellitus is a common complication, occurring in 7 to 9% of adolescents and 17 to 40% of adults with PWS as a consequence of severe obesity. The variance in the reported prevalence depends mainly on differences in age and the degree of obesity, but also on the ethnicity of the various study groups.

The specific etiology of the type 2-like diabetes in PWS remains unknown. Children and adolescents generally present with decreased fasting insulin levels and a normal insulin sensitivity, but a reduced and delayed insulin response of beta cells in oral glucose tolerance tests. Normal insulin sensitivity seems to be related to the relatively low degree of visceral fat accumulation. Later, the manifestation of a type 2-like diabetes in PWS is assumed to be precipitated by the addition of excessive obesity to impaired insulin

secretion. In fact, also in PWS, insulin sensitivity can be impaired above a critical increase in fat mass.

### Energy Balance

The enormous fat accumulation in PWS is caused by an imbalance of energy intake and energy expenditure, which results in an increase in energy storage. Basal metabolic rate—largely identical with the resting energy expenditure—was found to be decreased by 20 to 50% in PWS, when related to weight or height, reflecting the decrease in lean mass in this syndrome. Activity-related energy expenditure, assessed by deuterium dilution, is also decreased in PWS. The reason that PWS children and adolescents engage less in physical activity has been ascribed to hypothalamic dysfunction.

### Lipid Metabolism

Certain aspects of lipid metabolism in PWS differ from nonsyndromal obesity. In PWS, triglyceride levels are normal, though still correlated with abdominal obesity. Low-density lipoprotein cholesterol levels, however, are elevated and high-density lipoprotein cholesterol levels are decreased, as is found in GH deficiency.

## HYPOTHALAMIC DYSFUNCTION

Despite in-depth knowledge of the genetic condition in PWS, the final link between the chromosomal disorder and the clinical symptoms remains unclear. Hypothalamic dysfunction, as already originally presumed by Prader *et al.*, appears to underlie many of the features of PWS, including hypogonadism, disturbed energy balance, high pain threshold, and sleep disorders, but no overt structural abnormalities of the hypothalamus have been found yet. It has been shown that growth hormone deficiency due to hypothalamic dysregulation contributes not only to the abnormal growth pattern and osteopenia, but also to the excess of body fat and to the deficit of lean body mass, with reduced energy expenditure as a consequence. The decreased growth hormone (GH) secretion in PWS differs from that seen in simple obesity, where GH secretion is not disturbed, but rather is down-regulated and fully reversible by weight loss.

In view of the intriguing similarity of PWS to leptin deficiency, as shown by the ob/ob mouse with decreased satiation, hypoactivity, and obesity as well as short stature with decreased GH secretion and

infertility, some studies have investigated leptin levels in individuals with PWS, but found normal plasma leptin levels for body fat. Therefore, it is tempting to speculate that a disorder further down the cascade of the weight-regulating system may be responsible for the dysfunctional hormonal regulation in the hypothalamus as well as for decreased satiation and hypoactivity.

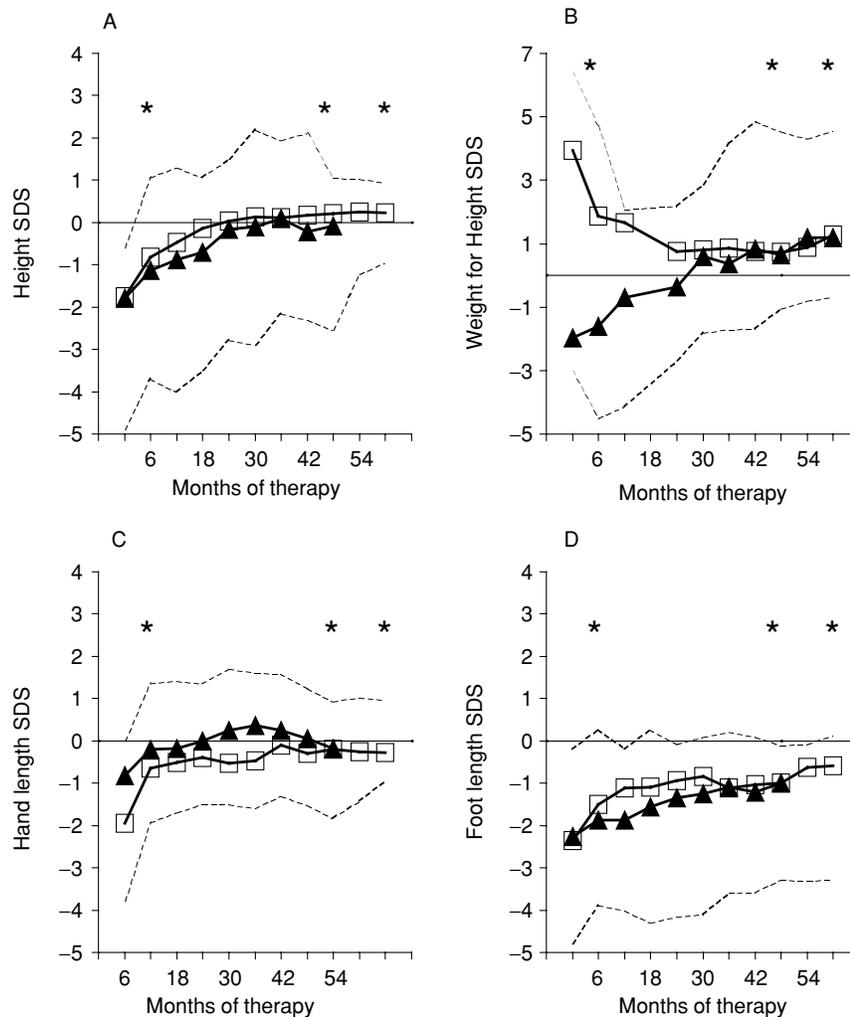
## THERAPY

### Comprehensive Team Approach

Individuals with PWS need a variety of interventions to optimize their growth and development. These physical and occupational therapies include dietary management, GH and sex steroid substitution, speech, language and learning disability services, behavior and family interaction management, support, and care. Successful patient management therefore requires a multidisciplinary team, consisting of a PWS specialist, an endocrinologist, a geneticist, a psychologist, a dietician, a physiotherapist, and a coordinating nurse.

### Growth Hormone Treatment

In prepubertal obese children with PWS, administration of GH has a remarkable impact on growth and, in combination with restriction of food intake, also on body composition, resulting in a dramatic change in the phenotype. GH treatment continuously increases height velocity and normalizes growth and body proportions (Fig. 3). If treatment is instituted early enough, final height prediction will reach the parental target height range after 3 years and short stature as well as small hands and feet will no longer be present. GH has a sustained impact on the net loss of body fat, which nevertheless remains elevated and improves the pattern of serum lipids. In addition to the medical benefits, the disappearance of the obese phenotype of prepubertally GH-treated PWS children relieves the patients and their families of stigmatization. A further benefit of GH treatment in PWS is the increase in lean body mass and, subsequently, resting energy expenditure. This leads to a reduction of energy stores, mainly of body fat, if energy intake is not increased. However, the initial deficit in lean body mass is counteracted by GH only during the first year of therapy. In the long term, GH therapy does not further compensate for this deficit. Further favorable effects of GH treatment in PWS were reported on respiratory muscle function, physical capacity,



**Figure 3** Height standard deviation score (SDS) (A), weight for height SDS (B), hand length SDS (C), and foot length SDS (D) of children with PWS, referring to normative data of the Zurich Longitudinal Study, before and during up to 5.5 years of GH therapy. Values are shown for young, initially underweight children (black triangles,  $n = 10$ ) and overweight prepubertal children (open squares,  $n = 8$ ). Medians are shown as solid lines and the minimum and maximum of the combined groups are shown as broken lines. Significant differences, tested at 6, 48, and 60 months by Wilcoxon test, in each group versus the value before therapy are indicated by asterisks ( $*P < 0.05$ ). Reprinted with permission from Eiholzer, U., *et al.* (2000). *Horm. Res.* 53, 44–53. Karger, Basel.

strength, agility, and activity. However, the benefits of GH treatment also need to be weighed against the risks related to diagnostic procedures or treatment; during narcosis and sedation, the specific risk due to central hypoventilation must be considered. Hypothalamic hypothyroidism may be uncovered during GH therapy and needs adequate substitution. Scoliosis or kyphosis can be aggravated by GH-induced catch-up growth and requires careful orthopedic management. Furthermore, another study has warned that GH treatment may accelerate the manifestation of diabetes in PWS, particularly in children who become extremely obese during therapy.

Regardless of GH treatment, carbohydrate metabolism in PWS needs to be closely and permanently monitored. GH treatment in PWS is a rather new therapeutic option and therefore only a little experience has been gained in controlled studies thus far. Many questions concerning appropriate age at initiation of treatment, optimum dosage, long-term safety, and long-term effects remain unanswered.

### Sex Hormone Substitution

Although hypogonadism is clearly documented in PWS, the substitution of sex hormones in puberty

remains controversial. Most authors recommend androgen substitution in males because of its beneficial effects, such as complete virilization, change in voice, prevention of osteoporosis, and increasing muscle mass as well as activity level, but some authors believe that it may lead to a more aggressive behavior and aggravate temper tantrums. However, there are no studies examining behavior during sex steroid substitution. In young women with PWS, estrogen substitution may prevent osteoporosis and although an increase in obesity is feared, it has not yet been scientifically assessed in these patients.

### Physical Therapy and Exercise

Physical therapy should aim to prevent scoliosis and to improve muscle strength and physical activity. Increased physical activity may lead to increased energy expenditure; it promotes a negative energy balance, raises the postexercise metabolic rate, builds muscle mass, and enhances the overall sense of well-being. Therefore, the crucial importance of regular physical activity should be clearly communicated to patients, parents, and caregivers.

### Dietary Restriction and Psychological Counseling

Even though height and weight are normalized during GH treatment, children with PWS must stick to a strict diet with a reduced energy intake of 75% of that for healthy children to stabilize the weight balance. A food intake restriction of this extent is possible only with close and strict supervision by parents and caregivers. Consequently, there is a lifelong need for environmental modifications to restrain food intake in PWS. This enormous task not only challenges

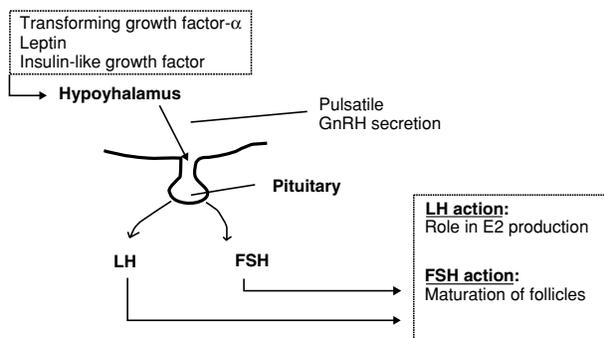
all caregivers, but may also exacerbate behavioral problems in patients. Therefore, there is no doubt about the need for psychological counseling and a comprehensive care team approach.

### See Also the Following Articles

Carbohydrate Metabolism: Diabetes Mellitus, Genomic Aberrations • Congenital Lipoid Adrenal Hyperplasia • Growth Hormone (GH) • Hypothalamic Hypogonadism • Obesity, Childhood and Adolescence • Obesity, Treatment of • Short Stature and Chromosomal Abnormalities • Undescended Testes

### Further Reading

- Burman, P., Ritzen, E. M., and Lindgren, A. C. (2001). Endocrine dysfunction in Prader-Willi syndrome: A review with special reference to GH. *Endocr. Rev.* **22**, 787–799.
- Cassidy, S. B. (1992). “Prader-Willi Syndrome and Other 15q Deletion Disorders.” Springer, Berlin, Germany.
- Cassidy, S. B. (1997). Prader-Willi syndrome. *J. Med. Genet.* **34**, 917–923.
- Eiholzer, U. (2001). “Prader-Willi Syndrome: Effects of Human Growth Hormone Treatment.” Karger, Basel, Switzerland.
- Eiholzer, U., l’Allemand, D., and Zipf, W. B. (2003). “Prader-Willi Syndrome as a Model for obesity.” Karger, Basel, Switzerland.
- Greenswag, L. R., and Alexander, R. C. (1995). “Management of Prader-Willi Syndrome.” Springer, New York.
- Gunay-Aygun, M., Schwartz, S., Heeger, S., O’Riordan, M. A., and Cassidy, S. B. (2001). The changing purpose of Prader-Willi syndrome clinical diagnostic criteria and proposed revised criteria. *Pediatrics* **108**, E92.
- Holm, V. A. (1981). The Diagnosis of Prader Willi syndrome. In “Prader Willi Syndrome” (V. A. Holm and S. J. P. L. Sulzbacher, eds.), pp. 27–44. University Park Press, Baltimore, MD.
- Holm, V. A., Cassidy, S. B., Butler, M. G., Hanchett, J. M., Greenswag, L. R., and Whitmann, B. Y. G. F. (1993). Prader-Willi syndrome: Consensus diagnostic criteria. *Pediatrics* **91**, 398–402.
- Prader, A., Labhart, A., and Willi, H. (1956). Ein Syndrom von Adipositas, Kleinwuchs, Kryptorchismus und Oligophrenie nach myotonieartigem Zustand im Neugeborenenalter. *Schweiz. Med. Wochenschr.* **86**, 1260–1261.



**Figure 1** Schematic representation of gonadotrophin release and action. Examples of possible mediators of gonadotropin-releasing hormone (GnRH) secretion are shown. LH, luteinizing hormone; FSH, follicle-stimulating hormone; E2, estradiol.

idiopathic or organic. In general, the majority of girls with CPP have the idiopathic form. The incidence of CPP in boys is much lower than in girls, and more frequently the cause is organic.

It has been shown that 70–90% of girls with CPP are idiopathic. The number of organic causes will likely increase with the refinement of central nervous system imaging by magnetic resonance imaging (MRI), revealing, for example, hypothalamic hamartoma in children formerly diagnosed with idiopathic CPP. In young girls, CPP must be differentiated from premature thelarche without central activation.

In idiopathic CPP, no organic cause for the precocious onset of puberty is found. In general, the younger the girl at the onset of CPP, the higher the chance of organic pathology.

In organic CPP, a cerebral organic lesion causes the premature onset of puberty. Local pressure on GnRH neurons or disruption of inhibitory fibers may cause GnRH release. [Table I](#) lists the causes of organic CPP. Treatment of CPP in children with organic lesions is primarily aimed at the underlying pathology. However, the progression of puberty is not always halted by this treatment.

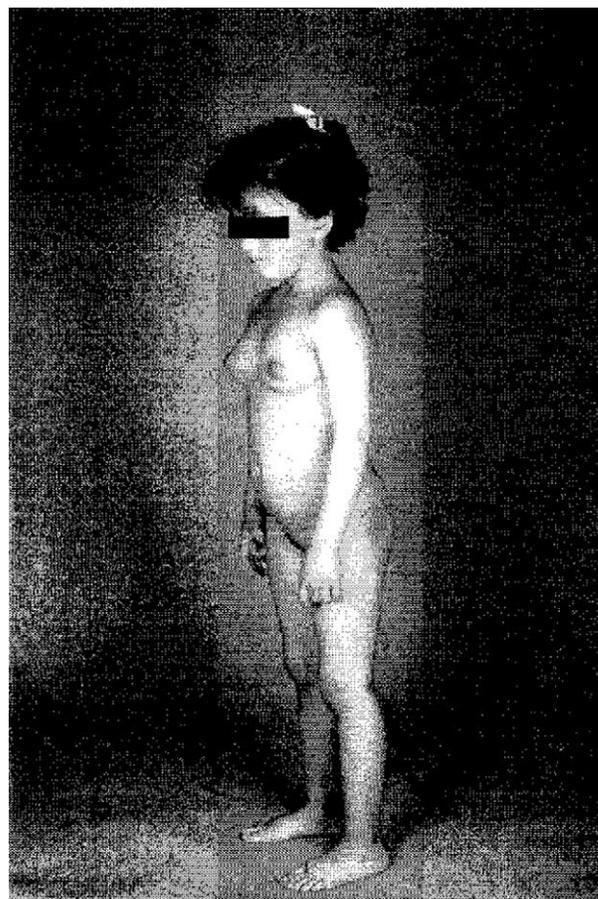
**Table I** Causes of Organic CPP

Hypothalamic hamartoma
Neurofibromatosis I
Brain tumors (e.g., craniopharyngeoma and astrocytoma)
Hydrocephalus and myelomeningocele
Cerebral trauma
Arachnoid cyst

## DIAGNOSIS AND CLINICAL PRESENTATION

Girls with CPP usually present with breast development as the initial sign of puberty ([Fig. 2](#)). Frequently, this is accompanied by acceleration of growth, occurrence of pubic hair, and, sometimes, behavioral changes. In some girls, menarche occurs soon after the onset of breast development due to rapid progression of puberty.

Less visible changes occur in bones and internal genitalia. The progression of bone age accelerates as a result of estrogenic stimulation and may cause premature closure of the growth plate, resulting in compromised final height. The method of bone age assessment in CPP is not uniform. In most studies, the Greulich–Pyle method is used. It has been shown that final height predictions for CPP using the tables of Bayley and Pinneau for accelerated bone age significantly overestimate the achieved final height, and it is more appropriate to use the “average” tables for prediction.



**Figure 2** A 7-year-old girl with clinical presentation of CPP.

Pelvic ultrasonography reveals the changes in aspect and volume of the ovaries and uterus. The ovaries increase in volume, and the number of follicles as well as the diameter of the ovaries increase. Uterine findings include changes in shape and the presence of endometrium.

In young girls presenting with CPP, MRI of the brain (especially the pituitary region) is mandatory to exclude central nervous system (CNS) abnormalities. The role of MRI in girls with signs of puberty at age 6 or 7 is under discussion. Some studies indicate that in these cases the search for pathology should not be performed. On the other hand, MRI in girls with CPP at age 6 or 7 has shown CNS abnormalities. A decision tree is being developed to rationalize the decision process.

An essential part of the diagnosis of CPP is the biochemical confirmation of the central (hypothalamopituitary) stimulation of the gonads by gonadotropins. In premature thelarche, for example, no central activation is present. The demonstration of central activation is a prerequisite for treatment with GnRH agonists. Basal values of LH or follicle-stimulation hormone (FSH) have limited value. Therefore, stimulation tests with native GnRH or a short-acting GnRH agonist (GnRHa) should be used in the diagnostic workup.

The increasing sensitivity of the pituitary for GnRH stimulation in normal puberty is the basis of the test. Depending on the assay, the peak value of serum LH should exceed a particular threshold. Early studies used thresholds  $>10$  IU/liter using radioimmunoassay methods, whereas the use of the more sensitive immunofluorometric assay allows cutoff values of approximately 6 IU/liter. A prepubertal response in a GnRH stimulation test does not completely rule out the presence of CPP since it has been demonstrated that the more potent GnRH agonists, when used in a stimulation test, are able to produce pubertal LH peaks in children with prepubertal LH peaks in the standard GnRH stimulation test. Mainly children with recent onset of puberty demonstrate these varying results in the different stimulation tests.

Some authors have emphasized that a predominant response of LH over FSH should be used in the diagnosis of puberty. However, there is no absolute value for the LH:FSH ratio.

The use of sex steroid serum levels in the diagnosis of CPP is limited as well because of the circadian variation of these levels. However, they may be used in the evaluation of pubertal suppression during treatment. Table II provides a summary of the diagnostics in CPP.

**Table II** Important Elements in the Diagnosis of Central Precocious Puberty in Girls

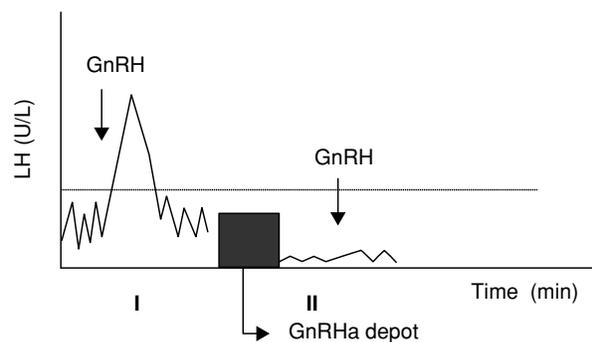
History	Onset of breast development, menarche, behavioral changes, pubic hair development, concomitant disease or complaints, family history
Physical examination	Pubertal staging according to Tanner, height and weight, signs of primary disease
X-ray	Bone age and height prediction
Ultrasonography	Size of ovaries and uterus
Magnetic resonance imaging	Hypothalamic region, pituitary, optic nerves
Laboratory	GnRH stimulation test: Luteinizing hormone and Follicle-Stimulating hormone response to GnRH (agonist) administration

## TREATMENT

### Gonadotropin-Releasing Hormone Agonists

Synthetic agonists of GnRH in depot form are the treatment of choice for CPP. Basically, the native GnRH molecule is modified at least at the glycine 6 position, where it is substituted by another amino acid resulting in a superagonistic effect. Prolonged exposure of the pituitary to a GnRHa paradoxically results in inhibition of gonadotropin secretion (Fig. 3).

The mechanism of this inhibitory effect is not completely clear. For example, it seems that the frequency



**Figure 3** Schematic representation of the inhibiting effect of gonadotropin-releasing hormone agonist (GnRHa) exposure. I, GnRH stimulation before GnRHa treatment. Luteinizing hormone (LH) increases considerably in reaction to GnRH stimulus. II, GnRH stimulation during GnRHa treatment. There is no increase in LH after GnRH stimulus due to the inhibitory effect of prolonged GnRHa exposure.

of the GnRH pulses is not altered after GnRHa exposure and that down-regulation of GnRH receptors occurs to only a limited extent. The depot formulations of GnRHa have been shown to be able to effectively suppress the endogenous pubertal activity in CPP. The clinical representation of this inhibition is arrest of pubertal development, sometimes even regression of physical signs of puberty, decreasing volume of ovaries, and deceleration of bone age maturation.

In Europe, triptorelin and leuprolide acetate are most frequently used as GnRHa depot preparations. The recommended monthly dose is 3.75 mg intramuscularly or subcutaneously. In general, in the United States 7.5 mg is the recommended dose. The optimal dose should be the lowest with which optimal suppression can be achieved. Only a few side effects have been described, mainly sterile abscesses at the injection sites.

### Indications

In the literature, no general view has been presented as to when to start treatment in CPP. Indications for treatment are either auxological or psychological. The aim of treatment concerning growth is to attain a final height in accordance with parental height and thus to prevent height loss due to the early onset of puberty. Psychological issues include reduction of psychological stress resulting from development of puberty inappropriately early for age. Sexual abuse and early pregnancy appear to be rare. However, for patients with mental retardation, early sexual activity and early menarche require special concern of families and peer groups (school). In these patients, GnRHa treatment may be indicated. Before initiation of treatment, central activation of the pituitary has to be demonstrated.

### Monitoring

Several parameters should be used to monitor the suppressive effect of treatment. As mentioned previously, clinical progression of puberty must stop or even regress, and menstrual bleeding should not occur. However, withdrawal bleeding soon after the initiation of treatment is frequently observed. Furthermore, the pubertal growth velocity should decrease to normal prepubertal values. Bone age progression will slow, especially after 6 months of treatment. On ultrasound, ovarian volume decreases to prepubertal values. Biochemically, baseline sex steroid levels will become prepubertal in effectively suppressed patients. Suppression of hypothalamic-pituitary-gonadal axis activity is determined during treatment by intravenously GnRH testing. Alternative methods include

taking a blood sample after subcutaneous GnRHa or after injection of the GnRHa depot. Some children show behavioral changes after the injection interval, which may indicate ineffective suppression during the final days of the injection interval.

### Results of Treatment

Numerous studies have addressed the effect of GnRHa treatment for CPP, and final height data have become available. However, none of these studies used randomized controlled study designs; thus, all need critical review. Most studies have used the difference between height prediction at the start of treatment and attained final height as the most important outcome parameter. It is well-known that there is a wide variation in bone age assessment between observers; thus, height predictions may lack reliability. Study designs should include an assessment of bone age by only one observer. Furthermore, as noted previously, height predictions vary depending on the method and the prediction tables used. Another way to evaluate the effect of treatment is to compare attained final height with midparental height or genetic target height.

Table III presents some of the major studies on final height after GnRHa treatment in girls with CPP. It can be concluded from the table that final height in girls with CPP after GnRHa treatment results in final height gain of approximately 6 or 7 cm. When compared to parental height, final height is 0.5–1.0 SD below target height or midparental height. Many factors influence the results of treatment, but no clear conclusion can be drawn from the different studies about the relative importance of

Table III Effect of GnRHa Treatment on Girls with CPP<sup>a</sup>

Author	Year	No.	FH – PAH/ start (cm) <sup>b</sup>	FH – TH (SD) <sup>c</sup>
Mul	2000	87	7.4	–0.88
Oerter Klein	2001	80	9.8	–0.59
Arrigo	1999	71	2.9	–0.48
Carel	1999	58	4.8	+0.15
Heger	1999	50	5.9	–0.6
Oostdijk	1996	31	3.4	–1.02

<sup>a</sup>Abbreviations used: FH, final height; PAH, predicted adult height; TH, target height; SD, standard deviation.

<sup>b</sup>The difference between initial prediction at the start of treatment and the attained final height.

<sup>c</sup>FH – TH compares the attained final height with the genetic growth potential expressed as target height.

**Table IV Factors That Influence Final Height after GnRHa Treatment**

---

At start of treatment:
Bone age at start of treatment:
higher bone age → better final height.
Chronological age at start of treatment:
lower chronological age → better final height.
At discontinuation of treatment:
Bone age at end of treatment: the higher the age, the less positive effect on final height.
Chronological age at end of treatment: the higher the age, the less positive effect on final height.
Bone age advance at discontinuation of treatment:
The more advanced over chronological age, the less positive effect on final height.

---

each, mainly due to the complex interaction of the variables included.

Multivariate analyses of the results of treatment are probably most appropriate for the development of guidelines. The factors shown in [Table IV](#) have been shown to influence the results of treatment (height gain) in multivariate analysis. Several studies show better results in girls with CPP when treatment is initiated before age 6. In contrast, in univariate analysis, chronological age at the start of treatment is negatively correlated with height gain. These contradictory results emphasize the need for proper analysis of the results of treatment.

The optimal timing with respect to chronological age and bone age for discontinuation of treatment cannot be determined from the literature. Possibly the best indication to discontinue treatment is a decrease of height velocity below prepubertal values during GnRHa treatment. Alternatively, one may assess height prognosis every 6 months and discontinue treatment after a decrease in height prognosis.

Concerning other outcome variables after GnRHa treatment in CPP, endocrine investigation after discontinuation of treatment showed complete reversibility of gonadal suppression. Menarche occurred within 1 year after discontinuation of treatment. The scant data on fertility showed no adverse outcome.

Studies on bone mineral density (BMD) during GnRHa treatment in girls showed normal BMD for chronological age and low BMD when corrected for bone age. At final height, BMD appeared to be normal.

In some children, body mass index (BMI) increases during GnRHa treatment. Pretreatment BMI SD is the strongest predictor of increased BMI SD at final height. GnRHa treatment does not lead to obesity.

## GnRHa and Growth Hormone

During treatment with GnRHa, it is frequently observed that height velocity decreases, even below prepubertal levels. The effects of GnRHa treatment on the growth hormone (GH)–IGF axis remain controversial. However, several groups have studied the effect of the addition of GH to GnRHa in children with precocious puberty. Results suggest the importance of maintaining adequate height velocity during treatment to achieve height gain. In short-term studies, the addition of GH resulted in increased predicted adult height compared to treatment with GnRHa alone. Only one study provides data for final height, confirming the results of the short-term studies.

## Other Treatment Options

Progestational agents (e.g., cyproterone acetate or medroxy progesterone acetate) no longer play a role in CPP treatment. Hormonal suppression has been shown to be inadequate, thus compromising final outcome. Daily administered preparations of GnRHa (e.g., nasally) should not be used since they have been shown to be inferior to depot preparations.

Depot preparations that allow children to receive an injection once every 3 months instead of monthly are being studied. The difference in efficacy between intramuscularly and subcutaneously administered depot preparations is not known.

Aromatase inhibitors may become important in the treatment of CPP. However, no data in children are available.

## CONCLUSION

Effective suppression of pituitary gonadal function is achieved with depot GnRHa treatment in girls with CPP. Hormonal suppression is fully reversed after treatment is discontinued. Further studies should be performed to determine the final indications of treatment.

## See Also the Following Articles

Adrenarche, Premature • Delayed Puberty and Hypogonadism, Female • Gonadotropin-Releasing Hormone (GnRH) Actions • Growth Hormone (GH)

## Further Reading

Bourguignon, J.-P., and Plant, T. M., (eds.) (2000). The onset of puberty in perspective. *In* "Proceedings of the 5th International

- Conference on the Control of the Onset of Puberty, Liege 1999." Elsevier, Amsterdam.
- Chalumeau, M., Chemaitilly, W., Trivin, C., Adan, L., Breart, G., and Brauner, R. (2002). Precocious puberty in girls: An evidence-based diagnosis tree to predict central nervous system abnormalities. *Pediatrics* **109**, 61–67.
- Kaplowitz, P. B., and Oberfield, S. E. (1999). Reexamination of the age limit for defining when puberty is precocious in girls in the United States: Implications for evaluation and treatment. Drug and Therapeutics and Executive Committees of the Lawson Wilkins Pediatric Endocrine Society. *Pediatrics* **104**, 936–941.
- Saggese, G., Stanhope, R., De Sanctis, V., and Bertelloni, S. (2000). "Supplement 1 (Vol. 13, 2000) of the Journal of Pediatric Endocrinology & Metabolism on the European Conference on Diagnosis and Treatment of Central Precocious Puberty: State of the Art 1999." Freund, London.
- Terasawa, E., and Fernandez, D. L. (2001). Neurobiological mechanisms of the onset of puberty in primates. *Endocr. Rev.* **22**(1), 111–151.

**Table I Differential Diagnosis of Central Precocious Puberty in Males**

Idiopathic: Sporadic
CNS lesions
Hypothalamic hamartoma
Malignancies: Astrocytoma, ependymoma, glioma
Optic glioma
CNS infections
Head trauma
CNS irradiation
Congenital malformations: Arachnoid/suprasellar cyst, hydrocephalus, septo-optic dysplasia
Secondary central precocious puberty
Congenital adrenal hyperplasia
Hypothyroidism

## HORMONAL STUDIES

Diagnostic tests include assays of serum testosterone (T), LH, and follicle-stimulating hormone (FSH), preferably by the most sensitive methods available. Depending on method sensitivity, levels higher than 10–20 ng/dl of total T are in the pubertal range. Although in the early stages of puberty T and LH levels are highest during sleep, early morning T may be of diagnostic value. Basal LH levels can be misleading because of the episodic nature of LH secretion and the lack of sufficient sensitivity of most commercial assays. Unless third-generation sensitive methods are employed, a random LH level may not be distinguishable from the prepubertal range. Therefore, in order to adequately assess gonadotropin secretion, dynamic studies should be performed using a GnRH stimulation test. The characteristic pubertal pattern is that of a significant and incrementally dominant LH response compared to FSH response. The reverse is true of the prepubescent pattern, typically illustrated by a relatively blunted LH response and a predominant FSH response to stimulation.

## IMAGING STUDIES

### Skeletal Age

To determine the biological age of the child, an X ray of the nondominant (left) hand and wrist is taken. If the bone age (BA) is advanced by more than 2 standard deviations for chronological age, it is unlikely that the child has a normal variant of pubertal development. BA is also important for estimating the child's final adult height; thus, it is of value in determining if and when therapeutic intervention is indicated.

## Magnetic Resonance Imaging

It is imperative to perform magnetic resonance imaging (MRI) of the brain in every child with CPP in order to exclude a CNS lesion as the cause (Table I). Studies have shown that CPP may be the only presenting sign of an intracranial tumor or malformation.

## DIFFERENTIAL DIAGNOSIS

Table I lists the main causes of CPP. Idiopathic CPP accounts for up to 50% of cases of CPP in boys, and it is usually a sporadic occurrence. Because of the relatively high incidence of underlying organic lesions, idiopathic CPP should be considered a diagnosis of exclusion in boys. Among the organic causes, CNS pathology should be a foremost consideration in the differential diagnosis.

### Organic Causes

#### **CNS Tumors**

CNS tumors are the most important lesions and they can be occult, sometimes necessitating serial MRIs for detection. The most common ones are hamartomas of the tuber cinereum, optic gliomas, and astrocytomas. Hypothalamic hamartomas are congenital, nonneoplastic, tumor-like lesions formed by heterotrophic gray matter, neurons, glial cells, and fiber bundles in the hypothalamus. They tend to grow slowly and are manifested by puberty occurring at an early age. Some hypothalamic hamartomas are associated with gelastic seizures, also called laughing seizures, which are often resistant to anticonvulsive treatment. Optic gliomas may be associated with neurofibromatosis. Although CNS neoplasms eventually manifest localizing signs (e.g., visual impairment and diabetes insipidus), precocious sexual development may be the earliest manifestation. Other CNS lesions, including congenital malformations, head trauma, or infection, may be causes of sexual precocity.

#### **CNS Radiation**

The incidence of precocious puberty following radiation to the brain is related to the radiation dosage. A relatively high dose is likely to cause gonadotropin deficiency, whereas low doses induce precocious sexual development. In contrast, chemotherapy alone is not a causative factor in sexual precocity.

#### **Secondary CPP**

Long-term exposure to endogenous sex steroids (e.g., for inadequately treated congenital adrenal hyperplasia) leads to an accelerated linear growth and BA and

early initiation of puberty. An intriguing observation with respect to environment is the high incidence of CPP in children adopted from developing countries. Primary hypothyroidism may also cause CPP, which is reversible after thyroid hormone treatment.

## NATURAL HISTORY

The natural history of children with CPP is varied. Frequently, the progression of pubertal development is more rapid than that of the normal pattern of maturation (especially in the case of brain tumors), with accelerated growth and skeletal age. This continuous exposure to sex steroids eventually causes premature fusion of the epiphyseal growth plates, leading to an overall decreased adult height. However, often in idiopathic CPP, secondary sexual characteristics progress either very slowly or in short bursts followed by quiescent periods during which growth velocity returns to normal and BA acceleration is halted. In such cases, the final height may not be compromised, and therapy is not required.

## MANAGEMENT

### Indications for Treatment

The principal objective is the attainment of a normal adult target height as genetically determined. A rapidly advancing BA will result in tall stature during childhood but eventually premature epiphyseal fusion leading to short adult stature. Treatment is clearly indicated in such cases. Other indications include psychosocial/behavioral aspects of precocious puberty. Surprisingly, however, only a few serious psychological problems and long-term sequelae have been described. There is a consensus that not all patients with CPP need medical intervention, such as in cases in which pubertal development is slow and the potential for attaining a normal adult height is preserved.

### Treatment Options

#### **GnRH Agonists**

The optimal therapy of CPP is the administration of GnRH agonists. These agents are potent derivatives of the natural decapeptide GnRH, and they are structurally similar. They are available in either aqueous forms given by daily injections or as depot (long-acting) formulations. By binding to the pituitary GnRH receptor, the agonist initially causes a stimulatory effect of short duration. It subsequently induces

suppression of gonadotropin secretion through sustained occupancy of GnRH receptors and desensitization to endogenous GnRH. For patient convenience and compliance, the trend has been to use depot preparations. One commonly used agonist is leuprolide acetate given in monthly doses of 7.5–11.25 mg. To ensure adequate LH suppression, a short GnRH stimulation test should be performed after the first 2 months of therapy.

#### **Combined GnRH Agonist and Growth Hormone Therapy**

An exaggerated deceleration of the growth rate can be observed during GnRH agonist therapy caused by the excessive suppression of the growth-promoting effects of sex steroids and, indirectly, growth hormone secretion. Some authors have advocated the combined use of growth hormone and GnRH agonist therapy in these instances. This is also necessary in cases of brain tumors or postirradiation when growth hormone deficiency coexists with CPP.

### Treatment Effects

Treatment with GnRH agonist may lead to a decrease in testicular volume, and penile erections become less frequent. When present, aggressive behavior may decrease in frequency and severity. Linear growth velocity is decreased, as is the rate of bone maturation, resulting in improvement of adult height potential.

### Monitoring of Treatment

Periodic evaluation of the child's growth, hormone levels, including serum T and LH, ensures that adequate hormone suppression is maintained. An annual assessment of BA is generally sufficient. The aim should be to achieve normalization of the growth rate proportionate to the BA advance, such that the predicted adult height improves progressively. Family counseling is important during this generally long-term therapy.

### Post-Therapy Physiologic State

Complete reversibility of the hypothalamic–pituitary–gonadal axis suppression occurs after GnRH agonist therapy is discontinued. Puberty resumes in the majority of cases within 3–6 months, and testicular growth becomes age appropriate. No cases of infertility have been reported. Treatment with GnRH agonists, particularly when administered as depot preparations, generally preserves a satisfactory adult

height potential, more favorably in idiopathic CPP than in cases in which there is an organic cause.

### Complications

Side effects are usually mild. Sterile abscesses at the injection site have been reported due to an idiosyncratic reaction to the injection vehicle.

Peak bone mass is accrued during the pubertal years. Bone mineral density (BMD) is often increased for age at the time of diagnosis of CPP and declines during agonist treatment. Also of concern is whether defective bone mineralization could occur after several years of suppression of gonadal steroids. Most studies have not shown such a negative outcome. Nonetheless, the effects on ultimate BMD deserve further study. Optimizing calcium and vitamin D intake during therapy is recommended.

### FUTURE DIRECTIONS

Since estrogen is the primary stimulus for skeletal maturation and fusion of the epiphyseal growth plates, the use of aromatase inhibitors for the enhancement of growth potential has been reported in preliminary studies. This may offer a promising therapeutic modality for boys with CPP whose height potential is seriously compromised.

### See Also the Following Articles

Constitutional Delay of Growth and Puberty (CDGP) • Delayed Puberty and Hypogonadism, Female • Delayed Puberty and Hypogonadism, Male • Delayed Puberty, Male • Precocious Puberty, Central (Female) • Precocious Puberty, Gonadotropin-Independent • Pseudoprecocious Puberty, Male • Puberty, Male: Mechanisms of Onset and Progression • Puberty: Physical Activity and Growth

### Further Reading

- Chemaitilly, W., Trivin, C., Adan, L., *et al.* (2001). Central precocious puberty: Clinical and laboratory features. *Clin. Endocrinol.* **54**(3), 289–294.
- Grumbach, M. M., and Styne, D. M. (1998). Sexual precocity. In “Williams Textbook of Endocrinology” (J. D. Wilson, ed.), 9th ed., pp. 1573–1584. Saunders, Philadelphia.
- Lazar, L., Pertzalan, A., Weintrob, N., *et al.* (2001). Sexual precocity in boys: Accelerated versus slowly progressive puberty gonadotropin-suppressive therapy and final height. *J. Clin. Endocrinol. Metab.* **86**(9), 4127–4132.
- Lebrethon, M. C., and Bourguignon, J. P. (2000). Management of central isosexual precocity: Diagnosis, treatment, outcome. *Curr. Opin. Pediatr.* **12**(4), 394–399.
- Lee, P. A. (1999). Central precocious puberty. An overview of diagnosis, treatment, and outcome. *Endocrinol. Metab. Clin. North Am.* **28**(4), 901–918.
- Partsch, C. J., Heger, S., and Sippell, W. G. (2002). Management and outcome of central precocious puberty. *Clin. Endocrinol.* **56**(2), 129–148.

**Table 1 Causes of Gonadotropin-Independent Precocious Puberty**

Latrogenic or exogenous sexual precocity
Human chorionic gonadotropin-secreting tumors
Hypothyroidism
Gonadal origin
<i>Females:</i> Hormone-secreting ovarian tumor, McCune–Albright syndrome
<i>Males:</i> Testotoxicosis, Leydig cell adenoma, aromatase excess syndrome
Adrenal origin
Congenital adrenal hyperplasia
Virilizing adrenal neoplasm
Cortisol resistance syndrome

retroperitoneum and certain hypothalamic pineal tumors cause sexual precocity in boys by secreting hCG rather than activating the hypothalamic GnRH pulse generator. Hepatomas and hepatoblastomas can also secrete hCG, leading to sexual precocity in boys.

### Familial Male-Limited Precocious Puberty

Normally, testicular Leydig cells produce testosterone after LH binds to its receptor, with subsequent signaling through G proteins and the adenylyl cyclase pathway of signal transduction. Genetic mutations may prematurely activate this cascade either at the level of the LH receptor (testotoxicosis) or downstream, at the level of the stimulatory G proteins (McCune–Albright syndrome).

Familial male-limited precocious puberty, also known as testotoxicosis, is an autosomal dominant form of gonadotropin-independent precocious puberty caused by constitutively activating mutations of the human LH receptor gene. Affected boys develop rapid virilization, growth acceleration, and skeletal advancement between 2 and 4 years of age, with elevated levels of testosterone despite prepubertal levels of LH. Testicular biopsy specimens show premature Leydig cell maturation.

## FEMALE GONADOTROPIN-INDEPENDENT PRECOCIOUS PUBERTY

### McCune–Albright Syndrome

The McCune–Albright syndrome is a sporadic disease characterized by polyostotic fibrous dysplasia, café

au lait pigmentation of the skin, gonadotropin-independent sexual precocity, and multiple endocrinopathies, including hyperthyroidism, pituitary adenomas secreting growth hormone, and autonomous adrenal hyperplasia.

The sexual precocity in girls is frequently heralded by menstrual bleeding in the first 2 years of life. The ovaries contain multiple follicular cysts and commonly exhibit asymmetrical enlargement. Serum estradiol is elevated; in contrast, the LH response to GnRH is prepubertal.

Somatic activating mutations of the stimulatory G protein are found in mosaic form in the affected organs in this disorder. These mutations occur early in embryogenesis, resulting in a widespread mosaic of normal and mutant-bearing cells. The constellation of abnormalities is dependent on the specific distribution of mutant-bearing cells.

### See Also the Following Articles

Constitutional Delay of Growth and Puberty (CDGP) • Delayed Puberty and Hypogonadism, Female • Delayed Puberty and Hypogonadism, Male • Delayed Puberty, Male • McCune–Albright Syndrome • Precocious Puberty, Central (Female) • Precocious Puberty, Central (Male) • Puberty: Physical Activity and Growth

### Further Reading

- Brito, V. N., Batista, M. C., Borges, M. F., Latronico, A. C., Kohek, M. B., Thirone, A. C., Jorge, B. H., Arnhold, I. J., and Mendonca, B. B. (1999). Diagnostic value of fluorometric assays in the evaluation of precocious puberty. *J. Clin. Endocrinol. Metab.* **84**, 3539–3544.
- Brunner, H. G., and Otten, B. J. (1999). Precocious puberty in boys. *N. Engl. J. Med.* **341**, 1763–1765.
- Grumbach, M. M., and Styne, D. M. (1998). Development and growth, puberty: Ontogeny, neuroendocrinology, physiology and disorders. In “Williams Textbook of Endocrinology” (J. D. Wilson, ed.), 9th ed., pp. 1509–1626. Saunders, Philadelphia.
- Latronico, A. C., Shinozaki, H., Guerra, G., Jr., Pereira, M. A., Lemos Marini, S. H., Baptista, M. T., Arnhold, I. J., Fanelli, F., Mendonca, B. B., and Segaloff, D. L. (2000). Gonadotropin-independent precocious puberty due to luteinizing hormone receptor mutations in Brazilian boys: A novel constitutively activating mutation in the first transmembrane helix. *J. Clin. Endocrinol. Metab.* **85**, 4799–4805.
- Weinstein, L. S., Shenker, A., Gejman, P., Merino, M. J., Friedman, E., and Spiegel, A. M. (1991). Activating mutations of the stimulatory G protein in the McCune–Albright syndrome. *N. Engl. J. Med.* **325**, 1688–1695.

dehydroepiandrosterone sulfate (DHAS) as precursor for placental production of estrone ( $E_1$ ) and estradiol ( $E_2$ ). The placenta, which lacks  $16\alpha$ -hydroxylation ability, also uses  $16\alpha$ -DHAS produced in the fetal liver as precursor for estriol ( $E_3$ ), but 90% of estriol production results from fetal adrenal DHAS production.  $E_3$  is the E produced in greatest quantity during pregnancy;  $E_1$  and  $E_2$  are derived equally from fetal and maternal precursors.

$E_1$  in the maternal plasma rises from 6 to 10 weeks and reaches 2–30 ng/ml at term.  $E_2$  rises from 6 to 8 weeks and reaches 6–40 ng/ml at term. During pregnancy,  $E_1$  and  $E_2$  excretion is increased approximately 100-fold compared to nonpregnant levels, whereas  $E_3$  excretion increases 1000-fold. The maternal level of  $E_2$  is higher than in the fetus, whereas the level of  $E_3$  is lower.

The maternal cardiovascular adaptations to pregnancy are also regulated by E. Blood volume is increased by E stimulation of the maternal and trophoblastic renin–angiotensin systems and uteroplacental blood flow is influenced by the vasodilatory effects of E. Placental aromatization is so efficient that no androgens from the maternal side reach the fetuses, protecting it from masculinization. The Es presented to the maternal bloodstream are rapidly metabolized by the maternal liver prior to excretion into the maternal urine. Only approximately 8–10% of the maternal blood  $E_3$  is unconjugated.

## THE FETAL ADRENAL CORTEX

By gestational weeks 8–9, the fetal adrenal cortex is differentiated into a thick inner fetal zone and a thin definitive outer zone, the forerunner of the adult cortex. By the end of the first trimester, the gland reaches a size equal to or larger than that of the adjacent kidney. After the 20th to the 24th week, the adrenal glands slowly decrease in size until a second growth spurt at approximately 34–35 weeks. After delivery, the fetal zone (approximately 80% of the bulk of the gland) rapidly involutes to be replaced by simultaneous expansion of the adult adrenal cortex.

The expression and the isomerase activity of  $3\beta$ -hydroxysteroid dehydrogenase are low in the fetal adrenal and therefore dehydroepiandrosterone (DHA) and DHAS are its major products. Fetal DHA and DHAS production rises steadily, concomitant with the increase in the size of the fetal zone and increase in adrenal weight. Maternal E levels follow the increased availability of the fetal DHAS as a precursor. Early in pregnancy, the gland grows and functions without

adrenocorticotrophic hormone (ACTH), perhaps in response to hCG. After 15–20 weeks, fetal ACTH is required. Paradoxically, during the last 12–14 weeks of pregnancy when fetal ACTH levels are declining, the adrenal gland quadruples in size. Because prolactin is the only fetal pituitary hormone to increase throughout pregnancy, paralleling changes in the size of the fetal adrenal, it was proposed as the critical tropic substance. Nevertheless, only ACTH exerts adrenal steroidogenesis. It activates adenylate cyclase, increases LDL receptors, and increases the expression of its own receptors. These cause increased uptake of circulating LDL-cholesterol, largely derived from the fetal liver, and *de novo* synthesis of cholesterol in the adrenals to sustain the high rates of DHAS and E formation. The tropic support of the fetal adrenal gland by ACTH from the fetal pituitary is protected by the placental conversion of cortisol to cortisone by  $11\beta$ -hydroxysteroid dehydrogenase. In late gestation, when E levels are high, less cortisol is transferred to the fetus, fetal ACTH secretion increases, the fetal adrenal gland undergoes greater maturation, and fetal cortisol synthesis from endogenous cholesterol increases.

Corticotropin-releasing hormone (CRH) production and the size of the fetal adrenal gland are closely correlated. CRH augments fetal ACTH secretion in a positive feedback mechanism, producing adrenal growth and cortisol and DHAS secretion. CRH may also directly stimulate DHAS production.

Adrenal steroidogenesis is subject to autocrine and paracrine regulation by various growth factors, such as inhibin A, activin A, transforming growth factor- $\alpha$ , basic fibroblast growth factor, and insulin-like growth factor-I (IGF-I) and IGF-II. The exact nature of these effects is beyond the scope of this article.

## POLYPEPTIDE HORMONES

The placental villus exterior has cytotrophoblasts (CT), separate mononuclear cells that are prominent early in pregnancy and sparse late in pregnancy, and syncytiotrophoblasts (ST), cells that form a continuous multinuclear layer on the surface. CT is the basic placental stem cell from which the ST arise by differentiation. ST is the functional cell type of the placenta. Trophoblast differentiation is influenced by hCG and several other growth factors. Since the surface of the ST is in direct contact with the maternal blood, placental proteins are secreted directly and preferentially to the mother. Several hypothalamic-like peptides originate in the CT and influence the ST to secrete

pituitary-like hormones. In addition, locally produced hormones, growth factors, and peptides work together to regulate placental function.

### Hypothalamic-like Releasing Hormones

Placental gonadotropin-releasing hormone (GnRH) regulates placental steroidogenesis and release of prostaglandins (PGs) as well as hCG. The placental receptors for GnRH have lower affinity than that of GnRH receptors in the pituitary, ovary, and testis. GnRH receptors are present in both CT and ST in a pattern that parallels hCG secretion, suggesting that GnRH regulates hCG secretion. GnRH release is increased by E, activin A, insulin, and PGs and inhibited by P, endogenous opiates, inhibin, and follistatin.

CRH is produced in the trophoblast, fetal membranes, and decidua. Its production is decreased by P and increased by glucocorticoids, which are responsible for the rise in ACTH and cortisol during the last weeks of pregnancy and during labor. The progressive increase in maternal CRH levels during pregnancy is due to the secretion of intrauterine CRH into the maternal circulation.

### Human Chorionic Gonadotropin

hCG is a glycoprotein  $\alpha\beta$ -dimer. A high content of sialic acid prolongs its half-life compared to its analogue, luteinizing hormone (LH). The  $\alpha$ -subunit is identical in hCG, LH, follicle-stimulating hormone, and thyroid-stimulating hormone. The biological activity is specific to the  $\beta$ -subunit. hCG production and secretion are stimulated by GnRH, interleukin-1 $\beta$ , activin, and E and inhibited by endorphins, inhibin, P, and follistatin. The only definitely known function for hCG is support of the corpus luteum, taking over for LH on approximately the eighth day after ovulation. P production and survival of the corpus luteum are dependent on hCG, in turn supporting the implanting conceptus until the seventh week.

It is likely that hCG stimulates steroidogenesis in the early fetal testes, so that androgen production will ensue and masculine differentiation can be accomplished. It is also possible that the function of the inner fetal zone of the adrenal cortex depends on hCG in early pregnancy. The  $\beta$ -hCG gene is expressed in fetal kidney and adrenal gland, suggesting that hCG may affect the development and function of these organs.

hCG is synthesized mainly in the ST. A maternal serum level of approximately 100,000 IU/liter is

reached at 8–10 weeks of gestation and the corpus luteum begins to involute at this time. hCG levels decrease to approximately 10,000–20,000 IU/liter by 18–20 weeks and remain at that level until term. Clearance of hCG is carried out mainly by renal metabolism.

### Human Placental Lactogen

Human placental lactogen (hPL) is a single-chain polypeptide of 191 amino acids. hPL is very similar to human growth hormone (GH) but has only 3% of its activity. Its lactogenic contribution in human pregnancy is uncertain. Its half-life is approximately 15 min. The hPL level in the maternal circulation is correlated with fetal and placental weight, steadily increasing until it plateaus in the last month of pregnancy (5–7 mg/ml). There is no circadian variation and only minute amounts enter the fetal circulation.

hPL's metabolic role is to mobilize free fatty acids from lipids. Relative hypoglycemia in fasting pregnant women ("accelerated starvation") is due to the transfer of glucose to the fetus by facilitated diffusion and due to the diabetogenic effect of placental hormones (E, P, and especially hPL) causing peripheral insulin resistance and hyperinsulinism. As the fasting glucose level decreases, hPL levels rise to stimulate lipolysis and increase free fatty acids, which are an alternative fuel for the mother, sparing glucose and amino acids for the fetus. With sustained fasting, maternal serum ketone levels rise. Due to limited transport of free fatty acids across the placenta, fetal tissues then utilize these ketones, which do cross the placenta. hPL also enhances the fetal uptake of ketones and amino acids. Insulin antagonism by hPL is mediated by the increase in free fatty acid levels, which, in turn, directly interfere with insulin-directed entry of glucose into cells. With a sustained state of inadequate glucose intake, maternal ketosis may impair fetal brain development and function.

hPL, despite its lower levels in the fetus, directly affects fetal tissue metabolism, including synergistic actions with insulin, especially actions on glycogen synthesis in the liver. The failure of fetal growth hormone to affect fetal growth suggests that hPL may be the fetal growth hormone.

### Human Chorionic Adrenocorticotrophic Hormone

The rise in maternal free cortisol, cholesterol, and pregnenolone is due to secretion of placental ACTH and CRH, which are not suppressible by glucocorticoids. ACTH production in the ST is stimulated by CRH

from the CT. CRH levels in maternal plasma rise in the second trimester, increasing to peak values at term. Oxytocin is a potent stimulator of CRH and ACTH placental production. A decrease in CRH-binding protein near term further increases the cortisol availability during labor and delivery.

### Growth Hormone, Growth Hormone-Releasing Hormone, and Somatostatin

Growth hormone-releasing hormone (GHRH) and somatostatin are found in the placenta and somatostatin is present in the decidua. Somatostatin decreases with advancing gestation. Placental GHRH and somatostatin do not contribute to maternal circulating levels. During the second half of gestation, placental GH gradually replaces pituitary GH in the maternal circulation. Placental GH is not present in fetal blood. Maternal IGF-I levels increase during pregnancy in parallel with GH. Placental GH is not regulated by placental GHRH, but responds inversely to maternal glucose levels, securing glucose availability for the fetus. It also stimulates gluconeogenesis and lipolysis in the mother.

### $\alpha$ -Fetoprotein

$\alpha$ -Fetoprotein (AFP) is a glycoprotein derived mainly from fetal liver and partially from the yolk sac, before this degenerates at approximately 12 weeks. It is comparable in size to albumin and may serve as a protein carrier of steroid hormones in fetal blood. AFP may also be a modulator of cell proliferation, synergizing with various growth factors.

Peak values of AFP in the fetal blood are reached at the end of the first trimester; thereafter, levels decrease gradually until a rapid decrease begins at 32 weeks. Maternal blood levels are much lower than fetal levels, rising until week 32 and then declining. Elevated maternal levels of AFP are associated with a variety of fetal anomalies, multiple pregnancies, and increased risk of spontaneous miscarriage, stillbirth, preterm birth, preeclampsia, neonatal death, and low birth weight. Low maternal levels of AFP are associated with trisomies 21 and 18.

### Relaxin

Relaxin is a peptide hormone produced by the corpus luteum of pregnancy and by the placenta, decidua, and chorion. It rises during the first trimester and declines in the second. It may have a role in maintaining early

pregnancy, ripening the cervix, inhibiting uterine contractions, and relaxing the pubic symphysis.

### Prolactin

Following ovulation, the endometrium becomes a secretory organ and remains so throughout pregnancy. It secretes renin, relaxin, and prolactin. Prolactin production requires the combined effect of P and E plus the presence of other placental and decidual factors, including relaxin, IGF-I, and specific stimulatory and inhibitory proteins. Prolactin derived from the decidua is the source of that found in the amniotic fluid. The prolactin in the fetal circulation is derived from the fetal pituitary. Beginning at 8 weeks, prolactin levels rise from the normal level of 10–25 ng/ml to a peak of 200–400 ng/ml at term. There is marked maternal variability and a diurnal variation as in the nonpregnant state.

Amniotic fluid concentrations parallel maternal serum levels until week 10, rise markedly until week 20, and then decrease until delivery. Prolactin reduces the amnion permeability from fetus to mother and contributes to the regulation of fetal water and electrolyte balance by acting as an anti-diuretic hormone. The increase in maternal levels represents pituitary secretion in response to E.

### Cytokines and Growth Factors

Local placental cytokine production is believed to be important for embryonic growth and in the maternal immune response essential for the survival of the pregnancy. A system of communication is present between maternal decidual and fetal tissue to provide growth factor (GF) support for the placenta, which would include fetal hematopoiesis, a known response to colony-stimulating factor-1. Decidual interleukin-1 $\beta$  and placental tumor necrosis factor- $\alpha$  synergistically release placental interleukin-6 to secrete hCG.

IGF-I and IGF-II are involved in placental, fetal, and postnatal growth. IGFs do not cross the placenta. The fetus can influence maternal IGF-I levels by placental secretion of hPL. During pregnancy, IGF-binding proteins (IGFBP)-2 and -3 decrease, thereby promoting the bioavailability of IGF-I in maternal tissues and enhancing nutrient transfer to the fetus. Placental IGF-I production further enhances transfer of nutrients across the placenta. On the other hand, IGFBP-1 produced in the decidua rises during pregnancy, interferes with IGF-I action, and inhibits fetal growth. Other GFs, such as epidermal GF,

platelet-derived GF, nerve GF, fibroblast GF, and transforming GFs, are involved in differentiation, proliferation, and growth associated with pregnancy.

### Inhibin, Activin, and Follistatin

Placental inhibin A rises in the maternal circulation, peaking at 8 weeks and at term. Activin A produced by the placenta also increases. Activin stimulates and inhibin inhibits placental production of hCG, GnRH, and steroids. Follistatin is an activin-binding protein expressed in the placenta, membranes, and decidua. It antagonizes the stimulatory effects of activin.

### Endogenous Opiates

Fetal and maternal endogenous opiates originating from the pituitary glands are secreted in parallel with ACTH, in response to CRH, which is, in part, derived from the placenta. Endorphins, enkephalins, and dynorphins are produced also by the ST in response to CRH. They inhibit oxytocin, vasopressin, and gonadotropins and increase prolactin secretion.

### The Renin–Angiotensin System

During early pregnancy, maternal prorenin is increased as a result of ovarian stimulation by hCG. The possible roles of the ovarian prorenin–renin–angiotensin system are stimulation of androgen production for E synthesis, regulation of calcium and PG metabolism, and stimulation of angiogenesis. Renin and angiotensinogen are expressed by the membranes and the placenta. Increased maternal renin activity is the result of an E-induced increase in angiotensinogen and a compensatory response to maintain blood pressure in the presence of vasodilation.

### Atrial Natriuretic Peptide

Atrial natriuretic peptide is produced in the heart atrium and in the placenta. It is a potent natriuretic, diuretic, and smooth muscle-relaxant peptide that contributes to the regulation of volume and electrolyte changes associated with pregnancy and delivery.

## PROSTAGLANDINS IN PREGNANCY

### Thromboxane and Prostacyclin

Thromboxane (TXA<sub>2</sub>) is the most powerful vasoconstrictor known, whereas prostacyclin (PGI<sub>2</sub>) is a

potent vasodilator. Platelets predominantly synthesize TXA<sub>2</sub> whereas PGI<sub>2</sub> is derived mainly from the endothelium. The endothelial production of PGI<sub>2</sub> plays an important role in the marked vasodilation that occurs during pregnancy. The placenta is a major source of TXA<sub>2</sub>.

### Prostaglandin and the Fetal Circulation

The predominant effect of PGs in the fetal and maternal cardiovascular system is to maintain the ductus arteriosus, the renal, mesenteric, uterine, and placental arteries, and probably the cerebral and coronary arteries in a dilated state. Control of ductus arteriosus patency and closure is mediated through PGs. The arterial concentration of oxygen is the key regulator of the ductus caliber. With increasing gestational age, the ductus becomes increasingly responsive to increased oxygen. Prostaglandin E<sub>2</sub> (PGE<sub>2</sub>) is the most important functional PG in the ductus, whereas PGI<sub>2</sub>, the major product in the main pulmonary artery, is the major factor in maintaining vasodilation in the pulmonary bed. With increasing maturation, the lung shifts to TXA<sub>2</sub> formation. With the onset of pulmonary ventilation at birth, leading to vascular changes that deliver blood to the duct directly from the lungs, TXA<sub>2</sub> serves as vasoconstrictor to the ductus.

### Prostaglandin and Fetal Breathing

Prior to parturition, fetal breathing is very shallow. Placental PGE<sub>2</sub> may suppress breathing by acting in the fetal brain. Occlusion of the umbilical cord is rapidly followed by a loss of this PGE<sub>2</sub> influence and the onset of air breathing.

### See Also the Following Articles

Assisted Reproductive Technology (ART) • Corticotropin-Releasing Hormone, Placenta • Eating Disorders and the Reproductive Axis • Endometriosis • Fertilization • Gonadotropin-Induced Ovulation • Hyperandrogenism, Gestational • Implantation • In Vitro Fertilization (IVF) • Infertility, Overview • Ovarian Failure Treatment Strategies: Egg Donation • Superovulation and Intrauterine Insemination

### Further Reading

August, P., Mueller, F. B., Sealey, J. E., and Edersheim, T. G. (1995). Role of renin–angiotensin system in blood pressure regulation in pregnancy. *Lancet* 345, 896–897.

- Braunstein, G. D. (1996). Evidence favoring human chorionic gonadotropin as the physiological 'rescuer' of the corpus luteum during early pregnancy. *Early Pregnancy* **2**, 183–190.
- Gluckman, P. D. (1995). The endocrine regulation of fetal growth in late gestation: The role of insulin-like growth factors. *J. Clin. Endocrinol. Metab.* **80**, 1047–1050.
- Handwerger, S., and Freemark, M. (2000). The roles of placental growth hormone and placental lactogen in the regulation of human fetal growth and development. *J. Pediatr. Endocrinol. Metab.* **13**, 343–356.
- Hansen, W. R., Keelan, J. A., Skinner, S. J., and Mitchell, M. D. (1999). Key enzymes of prostaglandin biosynthesis and metabolism. Coordinate regulation of expression by cytokines in gestational tissues: A review. *Prostaglandins Other Lipid Mediat.* **57**, 243–257.
- Mendelson, C. R. (2000). Role of transcription factors in fetal lung development and surfactant protein gene expression. *Annu. Rev. Physiol.* **62**, 875–915.
- Mesiano, S., and Jaffe, R. B. (1997). Developmental and functional biology of the primate fetal adrenal cortex. *Endocr. Rev.* **18**, 378–403.
- Mesiano, S., Katz, S. L., Lee, J. Y., and Jaffe, R. B. (1997). Insulin-like growth factors augment steroid production and expression of steroidogenic enzymes in human fetal adrenal cortical cells: Implications for adrenal androgen regulation. *J. Clin. Endocrinol. Metab.* **82**, 1390–1396.
- Pepe, G. J., and Albrecht, E. D. (1995). Actions of placental and fetal adrenal steroid hormones in primate pregnancy. *Endocr. Rev.* **16**, 608–648.
- Petraglia, F., Florio, P., Nappi, C., and Genazzan, A. R. (1996). Peptide signaling in human placenta and membranes: Autocrine, paracrine, and endocrine mechanisms. *Endocr. Rev.* **17**, 156–186.
- Qu, J., and Thomas, K. (1995). Inhibin and activin production in human placenta. *Endocr. Rev.* **16**, 485–507.
- Rutanen, E. M. (1993). Cytokines in reproduction. *Ann. Med.* **25**, 343–347.

First, studying these genes helps elucidate the normal aging process, a subject that has fascinated humans in their continuous quest for the fountain of youth. Second, mutations in genes involved in premature aging frequently lead to an early onset of cancer (a disease of cellular aging). As the average life expectancy of the general population in the developed world has increased from 49 years in 1900 to 77 years in 2003, diseases associated with old age, such as cancer, Parkinson's disease, and Alzheimer's disease, have become more common. Approximately 30% of people in the developed world will have cancer at some stage during their lifetime; this likelihood increases exponentially with advancing age. It is therefore no surprise that one of the main phenotypes of premature aging syndromes is the development of a broad range of tumors. There are a number of genetic changes required before tumors can develop. These genetic changes vary in each type of cancer and so give each tumor type a specific "genetic profile" at each stage of its development. However, there are also a number of key common factors among different tumor types and these represent the defining features of cancer. One of the main defining characteristics of cancer cells is the failure to maintain "genomic integrity" (i.e., the ability to faithfully replicate DNA); interestingly, this characteristic may also be used in defining PMAS. Genomic instability may be detected as an abnormal chromosome number and/or structure (such as translocations and breaks). In the case of PMAS, mutations occur in individual genes essential for replication, repair, and transcription, i.e., the processes required to maintain genomic integrity.

Here, the following eight PMAS are discussed: ataxia-telangiectasia (AT), Werner's syndrome (WS), Bloom's syndrome (BS), Rothmund-Thomson syndrome (RTS), xeroderma pigmentosum (XP), Cockayne's syndrome (CS), Hutchinson-Gilford Progeria syndromes, and trichothiodystrophy (TTD).

## CLINICAL FEATURES OF PREMATURE AGING SYNDROMES

Table I summarizes the phenotype of PMAS.

### Ataxia-Telangiectasia

AT patients show progressive cerebellar ataxia, which is a lack of coordination and balance. Ataxia begins to manifest at the end of the first year of life and patients may be wheelchair-bound by their teens. Ocular and skin telangiectasias (dilated blood vessels) become manifest between 2 and 8 years of age. There is also a

250-fold increased risk of lymphomas and a 70-fold increase in T cell leukemias, hypersensitivity to ionizing radiation, and premature aging, including progeric skin changes, progeric hair changes, and diabetes mellitus. AT cells show chromosomal instability; e.g., rearrangement of chromosome 14 is often involved in AT.

The causative gene for AT is *ATM* (ataxia-telangiectasia mutated), which maps to 11q22.3. *ATM* is a member of the phosphatidylinositol kinase family and becomes activated after induction of DNA double-stranded breaks, e.g., exposure to ionizing radiation. *ATM* has a role in the induction of cell cycle checkpoints, which allow cells to decide whether to repair DNA damage or commit to apoptosis. Mutant *ATM* results in defects in G1, intra-S, and G2 checkpoints. The involvement of *ATM* in DNA damage checkpoints is further emphasized by the identification of *ATM* as part of the BRCA1-associated genome surveillance complex (BASC); this links *ATM* with a number of tumor suppressor and DNA damage repair proteins. BRCA1 is the protein defective in some cases of hereditary breast cancer. The BASC complex includes *MSH2*, *MHS6*, *MLH1*, *ATM*, *BLM*, the *RAD50-MRE11-NBS1* complex, and DNA replication factor C. Many components of this complex have roles in the recognition of DNA damage or unusual DNA structures, suggesting that this complex performs some kind of "sensor" role. *ATM*'s ability as a sensor was further demonstrated in a study by Bakkenist and Kastan, who showed that DNA damage activates *ATM* via intermolecular autophosphorylation of Ser-1981 and dissociation of dimers. *ATM* is inactive in normal cells as a dimer (or multimers) and irradiation causes rapid intermolecular autophosphorylation, leading to dimer dissociation, which initiates cellular kinase activity. Bakkenist and Kastan proposed that rapid *ATM* activation could be due to DNA double-stranded breaks (caused by irradiation), resulting in the rapid change in some aspect of chromatin structure and subsequent *ATM* activation.

There is functional interaction between some of the genes of PMAS, e.g., *ATM* and *BLM* (mutated in BS). *BLM* interacts directly with *ATM*, BS cells exhibit radiosensitivity, and mitotic phosphorylation of *BLM* is partially dependent on *ATM*. *ATM* is also able to phosphorylate *BLM* in a  $\gamma$ -irradiation dose-dependent manner.

### Bloom's Syndrome

BS is a very rare autosomal-recessive disorder. The majority of patients that have been identified are of

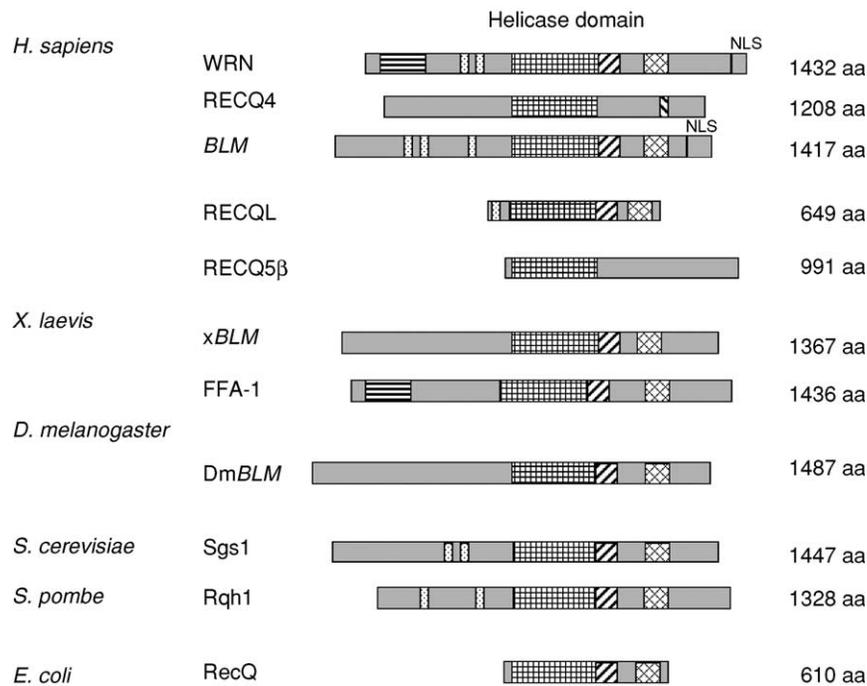
Table I Summary of Premature Aging Syndrome Phenotypes

Syndrome	Phenotype	Mode of inheritance	Frequency	Average age of death
Ataxia-telangiectasia	Progressive cerebellar ataxia; ocular and skin telangiectasia; small stature; hypogonadism; lymphomas; T cell leukemias; reduced immunoglobulin A (IgA), IgE, and IgG levels; increased $\alpha$ -fetoprotein; progeric skin changes; progeric hair changes; diabetes mellitus	Autosomal recessive	1/40,000	~30
Bloom's syndrome	Growth deficiency (proportional dwarfism); sun-induced facial rash; keel-shaped face; areas of abnormal skin pigmentation; infertility in males and subfertility in females; high incidence (12% of cases) of type II diabetes; decreased levels of IgA and/ or IgM; enormous predisposition to cancer	Autosomal recessive	Very rare overall; 2/100,000 in Ashkenazi Jews	~24
Werner's syndrome	Graying and thinning of the hair; bilateral cataract formation; type II diabetes mellitus; hypogonadism; loss of skin elasticity; osteoporosis; atherosclerosis; soft tissue sarcomas; thyroid cancers; osteosarcomas; unusual pattern of loss of subcutaneous fat and skin ulceration	Autosomal recessive	< 1/100,000	~47
Rothmund-Thomson syndrome	Growth deficiency; sunlight sensitivity with accompanying skin atrophy; cataracts; early graying and loss of hair; some increase in cancer incidence (mainly osteogenic sarcomas)	Autosomal recessive	< 1/100,000	Not known
Xeroderma pigmentosum	Highly photosensitive; 1000-fold increase in frequency of cutaneous basal and squamous cell carcinoma and melanomas; progressive neurodegenerative disease	Autosomal recessive	1/250,000	Life expectancy reduced by 30 years
Cockayne's syndrome	Cachectic dwarfism; senile appearance; mental retardation; cataracts; optic atrophy; deafness; skeletal abnormalities; photosensitivity; neurodegeneration	Autosomal recessive	~1/100,000	~20
Trichothiodystrophy	Photosensitive; brittle hair and nails; cataracts; faces have an aged appearance	Autosomal recessive	< 1/100,000	~10
Hutchinson-Gilford progeria syndromes	Much more rapid overall organ aging; midface hypoplasia; premature atherosclerosis; premature coronary artery disease; parental age effect	Undecided	~1/6,000,000	~13.4

Ashkenazi Jewish origin. The gene defective in BS is *BLM*, which maps to chromosome 15q26.1 in humans (Fig. 1). The *BLM* gene encodes a protein of 1417 amino acids that includes 7 conserved amino acid motifs found in many DNA and RNA helicases, including the RecQ family of helicases of which *BLM* is a member. The BS phenotype includes an enormous predisposition to cancer, with a mean age at cancer diagnosis of approximately 24 years. BS patients succumb early in life to the common range of carcinomas associated with the elderly in the general population. Therefore, it may be possible through the study of BS to shed light on at least one crucial aspect of tumorigenesis. BS cells show a 10-fold elevated frequency of homologous recombination events, i.e., an exchange of genetic information between regions of identical or very similar sequence. In BS, this so-called hyper-recombination occurs between both sister chromatids and homologous chromosomes.

The elevated rate of sister chromatid exchange (SCE) is a diagnostic feature of BS. The increased level of interchromosomal recombination may lead to an increase in the rate of somatic loss of heterozygosity, which in turn leads to inactivation of tumor suppressor genes and hence to cancer predisposition in BS.

*BLM* is also part of BASC and works closely with *ATM*. To examine the role of *BLM* within BASC, the subcellular localization of *BLM* and *BRCA1* was analyzed before and after exposure to DNA-damaging agents. In untreated cells, *BLM* and *BRCA1* colocalization was limited to a few bright nuclear foci. However, after treatment with hydroxyurea or ionizing radiation, colocalization was greatly enhanced in those cells that were in mid to late S phase or in G2. This could be indicative of a specific requirement for *BLM/BRCA1* in replication and/or repair of late replicating DNA.



**Figure 1** Schematic representation of the RecQ family of DNA helicases. Members of this family have been identified in bacteria (*Escherichia coli*: RecQ), yeasts (*Saccharomyces cerevisiae*: Sgs1; *Schizosaccharomyces pombe*: Rqh1), fruit fly (*Drosophila melanogaster*: DmBLM), frog (*Xenopus laevis*: the *BLM* orthologue, *xBLM*, and the *WRN* orthologue, FFA-1), and mammals (only information for human is shown) (*Homo sapiens*: *BLM*, *WRN*, RECQ4, RECQL, and RECQ5). The central conserved helicase domain is shown as a grid-filled box. The exonuclease domain of *WRN* and FFA-1 (which is not found in other RecQ helicases) is shown as a box with horizontal lines. The boxes with diagonal lines denote a region of conservation that is C-terminal to the helicase domain. This region is apparently unique to RecQ helicases. The boxes with outlined diamond shapes show the so-called HRDC domain, which is also found in certain enzymes that degrade RNA, and the boxes with dots show highly acidic regions. The gray boxes denote domains with little or no sequence similarity among family members. The nuclear localization signal sequences (NLS) in *BLM* and *WRN* are shown in black. Three variants of the RECQ5 protein may exist, but only the  $\beta$  isoform is indicated. The size of each protein (given as the number of amino acid residues) is shown on the right.

## Werner's Syndrome

WS is caused by mutations in the *WRN* gene, another member of RecQ family of DNA helicases (Fig. 1). WS is a premature aging condition and has a broad phenotype that includes many age-related disorders that develop from puberty onward. Moreover, WS individuals are also cancer-prone, but display a narrower range of tumor types than is seen in BS. Soft tissue sarcomas, thyroid cancers, and osteosarcomas are characteristic of WS. The average age of death for WS patients is 47 years, mainly due to cancer or cardiovascular disease. Some *WRN* features are not seen in normal aging, e.g., an unusual pattern of loss of subcutaneous fat and skin ulceration. WS cells display increased illegitimate recombination and a high frequency of large chromosomal deletions.

## Rothmund-Thomson Syndrome

There has been only limited study of this rare disorder; nevertheless, the gene responsible for RTS has been identified. RTS is caused by mutations in the *RECQ4* gene, another member of the RecQ family of DNA helicases (Fig. 1). In this disorder, affected individuals show growth deficiency, progeroid features, and an increase in cancer incidence. The cancer predisposition in RTS individuals is restricted mainly to osteogenic sarcomas. RTS cells show trisomerization (having three instead of the usual two copies of each chromosome) and an increased frequency of chromosomal aberrations.

## Xeroderma Pigmentosum

XP is an autosomal-recessive disorder, consisting of seven complementation groups (*XPA-G*). The *XPD* and *XPB* proteins are involved in nucleotide excision repair (NER) and transcription initiation by RNA polymerase II. Depending on the type of mutation, different pathways could be impaired, resulting in distinct phenotypes. Mutations destroying *XPD* transcription ability lead to TTD. Some mutations in *XPB* and *XPD* also cause CS and trichothiodystrophy, respectively; hence, xeroderma pigmentosum group D (*XPD*) has been described as one gene, two functions, and three diseases. XP patients are highly photosensitive and have a 1000-fold increase in the frequency of cutaneous basal and squamous cell carcinomas and melanomas. Furthermore, 20% of XP patients (complementation groups A, C, and D) also develop a progressive neurodegenerative disease.

## Cockayne's Syndrome

CS is a rare disorder with two complementation groups; the two causative genes are CSA and CSB. Cells from CS patients are defective in transcription-coupled repair in both the NER and the base excision repair pathways.

## Trichothiodystrophy

There are three complementation groups for this very rare disorder: *XPD*, *XPB*, and TTD-A. *XPD* and *XPB* are helicases of the transcription/repair factor TFIIH (transcription factor IIH). Patients with mutations in *XPD* and *XPB* are defective in NER due to the loss of transcription. *XPD* is a DNA helicase protein involved in repair and transcription and is mutated in some cases of TTD. However, in the TTD-A complementation group, TFIIH is intact and active in transcription and repair, but the levels of TFIIH have decreased. Therefore, TTD can be caused by both a decrease in the level of TFIIH (dose reduction) and a total loss of TFIIH (loss of function).

## Hutchinson-Gilford Progeria Syndrome

This is a rare (approximately 1 in 6 million) developmental disease that affects most of the organs in a manner similar to that which occurs in normal aging, except in these patients, overall organ aging takes place much more rapidly. The mode of inheritance is still unclear, but there are suggestions of both recessive and dominant forms; the causative genes have not been identified.

## MOUSE MODELS OF PMAS

Genetic studies in mice have revealed significant phenotypic differences in PMAS-deficient mutants from those seen in humans. Deletion of *BLM* in the mouse leads to embryonic lethality, deletion of *WRN* does not lead to obvious premature aging, and *ATM* mutants show only modest growth retardation, neuronal damage, and premature aging. One of the major obstacles in creating mouse models of PMAS has been the fact that mice have very long telomeres, a potential reason that many mouse models do not mimic the clinical situation for PMAS. One of the factors that is observed during the aging process in humans is the shortening of telomeres. However, this problem has been circumvented in a number of studies by creating a double knockout (KO) of two PMAS genes. The use

of these double KO mice has yielded mouse models for most human PMAS.

Studies using double-null mice for *ATM* and the telomerase RNA component (*Terc*) have shown that telomere deficiency and *ATM* deficiency compromise organ homeostasis and accelerate aging. These double KO mice show increased telomere erosion and genomic instability, leading to a global proliferation defect seen in all cell types and tissues. The accelerated telomere erosion could form the basis of premature aging seen in these mice. These mice showed increased hair graying, alopecia, and delay in hair regrowth, clot formation, and wound healing.

The *XP* DNA family has some mutations in *XPB* and *XPD* that cause CS and TTD, respectively. Mice with mutations in *XPD*, a DNA helicase protein that is involved in repair and transcription and is mutated in the human disorder TTD, showed an accumulation of DNA damage and symptoms of premature aging. Double mutants of *XPD* and *XPA* (also involved in DNA repair) show an even faster aging process. When accelerated aging in TTD patients occurs, it does so earlier than in WS or BS individuals.

*BLM* mouse models have been generated and the cells from *BLM*<sup>-/-</sup> mice show inherent genomic instability. Although viable animals were recovered in only one case, these mice were cancer-prone. Moreover, embryonic stem cells lacking *BLM* showed an increased frequency of gene targeting, indicative of an increase in homologous recombination (HR). This is consistent with the fact that HR mediates SCEs in vertebrates and that elevated SCEs are seen in BS cells. *WRN*-defective mice have also been generated and crossed with p21 or p53 null mice to generate double mutants. In the first few generations, these

mice did not display premature aging. The p21-*WRN* double mutants also showed no increase in the rate of tumorigenesis. In contrast, the p53-*WRN* double mutants showed an accelerated rate of tumorigenesis associated with the development of a wide variety of tumors not detected in either type of single mutant mice. In a separate study, transgenic mice expressing a mutant *WRN* gene encoding a defective helicase showed none of the premature aging phenotypes characteristic of WS. However, primary cell cultures from these mice did show two of the known characteristics of WS cells: reduced replicative potential and hypersensitivity to 4-nitroquinoline-1-oxide. The expression of the mutant *WRN* also had a dominant-negative effect, resulting in reduced function of the endogenous *WRN* protein.

Studies of *WRN* function in mice and yeast have been very informative regarding the role of telomere maintenance and aging (Table II). The role of Sgs1p in telomeric function (Table II) is apparently conserved in evolution, since mouse *WRN* can partially substitute for yeast *SGS1*. This suggests that in humans *WRN* may play a role similar to that played by Sgs1p in yeast. There are several other lines of evidence suggesting a close link between *BLM* and *WRN* proteins in telomere biology. Both *BLM* and *WRN* interact with telomere repeat-binding protein, TRF2. This *BLM/WRN* interaction with TRF2 occurs via the C-terminal domain in each case and they co-immunoprecipitate and colocalize to alternative lengthening of telomeres-associated PML bodies. This interaction is a functional one, as TRF2 binds and stimulates the catalytic activity of *WRN* and *BLM* helicases.

It is also worth noting that the insulin-like growth factor receptor (IGF-IR) regulates lifespan

Table II Telomere Maintenance and Aging

Yeast	Mammals
<ul style="list-style-type: none"> <li>The <i>WRN</i> homologue in <i>Saccharomyces cerevisiae</i>, <i>SGS1</i>, is also involved in telomere maintenance, at least in strains lacking telomerase.</li> <li>Deletion of <i>EST2</i>, part of the catalytic subunit of telomerase in yeast, leads to premature replicative senescence and a progressive erosion of telomeres.</li> <li>Yeast cells surviving lack of <i>EST2</i> have either amplification of subtelomeric DNA or long and variable telomeric DNA structures reminiscent of human ALT cells.</li> <li>Deletion of <i>SGS1</i> in <i>est2</i> strains leads to accelerated senescence. <i>est2 sgs1</i> strains do not generate "ALT-like" survivors, suggesting that Sgs1p is needed for telomere lengthening in the absence of telomerase.</li> </ul>	<ul style="list-style-type: none"> <li>WS patients' fibroblasts show an accelerated rate of replicative senescence; however, the forced overexpression of telomerase in these cells stops their senescence.</li> <li>The <i>WRN</i> protein colocalizes with telomeric factors in telomerase-independent immortalized (ALT) cell lines.</li> <li>In most tumors, the telomerase gene <i>EST2</i> is reactivated. In ~10% of tumors, instead of telomerase, telomere-stabilizing protein <i>WRN</i> is involved.</li> <li>Sgs1p telomeric function is conserved in evolution, as mouse <i>WRN</i> can partially substitute for yeast <i>SGS1</i>.</li> </ul>

Note. ALT, alternative lengthening of telomeres.

and resistance to oxidative stress in mice, as it does in lower eukaryotes, such as *Caenorhabditis elegans*. IGF-IR KO mice are not viable, but heterozygous mice (*igfr*-/+) are viable and live an average of 26% longer than wild types. Extended longevity is also seen in mice lacking the insulin receptor in adipose tissue, suggesting that a reduction in fat mass without caloric restriction may be associated with increased longevity, possibly via effects on insulin signaling. It has been proposed that the IGF-IR may be a central regulator of mammalian life span. IGF-IR is, in turn, also a regulator of *ATM*, which when mutated leads to AT, one of the main types of PMAS. Studying the nature of the IGF-IR and *ATM* interaction could hold a key to understanding PMAS as well as the process of tumorigenesis.

In summary, model systems are shedding light on the pathogenesis of PMAS, which in turn enhances the understanding of the processes that regulate normal aging and cancer predisposition.

### Acknowledgement

The author thanks Dr. V. M. Macaulay, Dr. S. L. Davies, and Professor I. D. Hickson for helpful comments and Cancer Research UK for financial support.

### See Also the Following Articles

Aging and Longevity of Human Populations • Aging, Animal Models for • Functional Genomics of Aging • Neuroendocrine System and Aging • Oxidative Stress and Aging

### Further Reading

- Bakkenist, C. J., and Kastan, M. B. (2003). DNA damage activates *ATM* through intermolecular autophosphorylation and dimer dissociation. *Nature* **421**, 499–506.
- Bluher, M., Kahn, B. B., and Kahn, C. R. (2003). Extended longevity in mice lacking the insulin receptor in adipose tissue. *Science* **299**, 572–574.
- Bohr, V. A. (2002). Human premature aging syndromes and genomic instability. *Mech. Aging Dev.* **123**, 987–993.
- Brooks, P. J. (2002). DNA repair in neural cells: Basic science and clinical implications. *Mutat. Res.* **509**, 93–108.
- de Boer, J., Andressoo, J. O., de Wit, J., Huijman, J., Beems, R. B., van Steeg, H., Weeda, G., van der Horst, G. T., van Leeuwen, W., Themmen, A. P., Meradji, M., and Hoeijmakers, J. H. (2002). Premature aging in mice deficient in DNA repair and transcription. *Science* **296**, 1276–1279.
- Hickson, I. D. (2003). RecQ helicases: Caretakers of the genome. *Nat. Rev. Cancer* **3**, 169–178.
- Holzenberger, M., Dupont, J., Ducos, B., Leneuve, P., Geloën, A., Even, P. C., Cervera, P., and Le Bouc, Y. (2003). IGF-1 receptor regulates lifespan and resistance to oxidative stress in mice. *Nature* **421**, 182–187.
- Macaulay, V. M., Salisbury, A. J., Bohula, E. A., Playford, M. P., Smorodinsky, N. I., and Shiloh, Y. (2001). Downregulation of the type 1 insulin-like growth factor receptor in mouse melanoma cells is associated with enhanced radiosensitivity and impaired activation of *ATM* kinase. *Oncogene* **20**, 4029–4040.
- Mohaghegh, P., and Hickson, I. D. (2001). DNA helicase deficiencies associated with cancer predisposition and premature aging disorders. *Hum. Mol. Genet.* **10**, 741–746.
- Mohaghegh, P., and Hickson, I. D. (2002). Premature aging in RecQ helicase-deficient human syndromes. *Int. J. Biochem. Cell Biol.* **34**, 1496–1501.
- Uitto, J. (2002). Searching for clues to premature aging. *Trends Endocrinol. Metab.* **13**, 140–141.
- Wong, K. K., Maser, R. S., Bachoo, R. M., Menon, J., Carrasco, D. R., Gu, Y., Alt, F. W., and DePinho, R. A. (2003). Telomere dysfunction and *ATM* deficiency compromises organ homeostasis and accelerates aging. *Nature* **421**, 643–648.

of the actin cytoskeleton are involved in this follicular rescue. Conversely, androgens, tumor necrosis factor  $\alpha$ , IL-6, and Fas ligand all stimulate oocyte apoptosis and follicular atresia. Any alteration in the gene or gene product of these regulators of apoptosis could lead to an inappropriate rate of cell death or cellular dysfunction, premature depletion of oocytes or dysfunctional oocytes, and premature ovarian failure.

## Ovarian Compartments

The ovary can be divided into two functional compartments, the follicle and the stroma. Despite a basement membrane separating the two compartments, the ovary is not an immunologically privileged site, in contrast to the testis. This fact is important to the possible pathogenesis of POF. Resident ovarian lymphocytes and polymorphonuclear granulocytes, macrophages, mast cells, and eosinophils appear to be involved in folliculogenesis and corpus luteum function. Alterations in cellular or humoral immune activity or function have the potential to therefore affect ovarian function. As can be appreciated, the participation of the immune system in ovarian function allows for perturbations that could cause clinical POF.

## Stages of Ovarian Function

Throughout a woman's life cycle, one can discern five stages of ovarian function: (1) ovulatory and fertile, (2) ovulatory and infertile, (3) anovulatory and estrogen producing, (4) anovulatory and androgen producing, and (5) absent hormonal function. These represent a functional continuum, with normal fertility on one end, through seemingly normal function but diminished fertility, through anovulatory states with only estrogen and then only androgen production, to the quiescent, postclimacteric ovary on the other end of the continuum. The stages and transitions may be clinically obvious or they may be subtle. Indeed, clinical reproductive endocrinologists still struggle to differentiate a fertile, ovulatory woman from an infertile, but ovulatory woman. The passage through these stages or levels of function is predictable, but the onset and timing of the transitions are highly variable.

First, as puberty begins, androgen and estrogen are secreted by the ovary as secondary sexual characteristics develop. Soon thereafter, ovulation begins and fertility is established, continuing usually into the fourth decade of life. Metabolic and central nervous

system conditions, such as polycystic ovarian syndrome and hypothalamic amenorrhea, can disrupt this normal state even in the prime fertile years. As women age, the number of oocytes continues to diminish as a result of atresia and ovulation. Due to this ongoing recruitment, selection, and loss of follicles, the quality of the remaining oocytes declines. At this point, the woman is ovulatory, but infertile.

This subfertility generally becomes evident in the population at approximately 35–38 years of age, with 50% of women in this age group being infertile. After 40 years of age, 90% of women are infertile. Despite regular cycles and apparent ovulation, fertility is rare beyond 42 years of age. Further decline in follicle/oocyte number and function results first in oligo-ovulation (with consequent estrogen but not progesterone secretion from the ovary) and then complete anovulation. Once irreversible anovulation is established, it is only a matter of time until estrogen levels decline sufficiently that menopause is noted.

At the final stages of the ovarian life cycle, the ovary continues to produce normal levels of androgens during the first 5 to 7 postmenopausal years of the climacteric. These play a role in maintaining metabolism, libido, and energy, but they may also cause a relative hyperandrogenic effect resulting in hair loss, hirsutism, and metabolic abnormalities such as an atherogenic lipid profile. Thereafter, the androgen levels also decline until the ovary is steroidogenically quiescent.

This is the natural history of the decline in ovarian function and in women with POF it is a process that occurs significantly earlier than the expected age.

## ETIOLOGIES OF PREMATURE OVARIAN FAILURE

Several schemes have been used in the past to classify the etiology of POF. A useful classification advanced by Anasti, Nelson, and Flack recognizes the enhanced understanding of ovarian regulation at the molecular and cellular level, along with the importance of the immune system to normal ovarian function. This classification differentiates those conditions that result in premature or accelerated follicle depletion from those with follicle dysfunction.

## FOLLICLE DEPLETION SYNDROMES

Any inherited or acquired defect that disturbs the process of germ cell migration, oogonial proliferation, or initiation of meiosis to form primordial follicles can

result in an insufficient initial complement of follicles. Similarly, any abnormality in the genes that regulate the rate of early atresia could result in early exhaustion of the oocyte pool.

### Deficient Initial Number

Deficient initial follicle numbers are seen in patients with gonadal dysgenesis and thymic aplasia/hypoplasia. Any mutation of the genes responsible for germ cell migration or mitotic oogonial proliferation may result in an inadequate initial follicle number. Examples of such mutations causing a deficient initial oocyte population of the ovary have been described in mice, such as the *c-Kit* gene at the *W* locus and an insertional transgenic mutation of chromosome 11 A2-3; germ cell deficiency in women with 46,XX gonadal dysgenesis could be the result of autosomal mutations similar to these examples. Thymic hypoplasia or aplasia has also been associated with POF. Both in the thymectomized neonatal mice model and in humans with thymic disorders (thymic aplasia and myasthenia gravis), there are decreased numbers of natural killer and suppressor T cells, leading to cell-mediated autoimmune destruction of the ovarian follicles. In these women, ovarian failure is typically only one feature of their autoimmune polyglandular dysfunction.

### Accelerated Follicle Atresia

Accelerated follicular atresia due to an alteration in the genome is thought to be the basis for one-third to one-half of all POF cases, although only the minority of these cases are identifiable with available molecular techniques. Usually, these alterations involve large chromosomal defects such as 45,X or fragile X (FRAXA) syndromes, although autosomal recessive and autosomal dominant single gene defects such as POF-1, POF-2, and blepharophimosis-associated POF have also been described.

Two competent X chromosomes are required for the proper control and timing of oocyte apoptosis/atresia and follicular depletion. Aberrant control of oocyte atresia may occur when there is an absent X chromosome, a translocation involving the X chromosome, or deletions of portions of the X chromosome. Examples include the ovaries of 45, X stillborn infants in whom germ cells migrate normally to the gonadal ridge with accelerated atresia of the primary oocytes later in fetal life and cohorts of fragile X syndrome patients who have an increased incidence of ovarian dysfunction. Cytogenetic studies of patients with POF that possess

X chromosome deletions have revealed detailed regions of the X chromosome responsible for ovarian function, including the genes POF1 (localized to Xq21.3–q27 or Xq26.1–q27) and POF2 (within Xq13.3–q21.1). Clinically, patients with the POF2 gene deletion have younger onset of ovarian dysfunction than those with the POF1 gene deletion. As molecular genetic studies advance, the exact nature of these genes and their proteins functions will be uncovered.

Specific genes appear to also influence ovarian follicle number, atresia, and the timing of ovarian failure, given the large number of POF patients and families with apparently normal karyotypes. An example is galactosemia, an autosomal recessive condition resulting from deficiency of the enzyme galactose-1-phosphate uridylyltransferase. Most patients with galactosemia eventually develop premature ovarian failure. The premature destruction of primordial follicles seen in these patients is thought to be due to the toxic effects of galactose or one of its metabolites or a deficiency of uridine diphosphate galactose. However, the POF seen in galactosemics could also be the result of a deficient initial follicle number, follicle dysfunction, or abnormal gonadotropins.

Iatrogenic follicle depletion can occur because of ovarian surgery, radiation therapy, or chemotherapy. In a patient over the age of 40, a radiation dose of only 600 cGy results in loss of ovarian function, although younger women have been reported to have a return of ovarian function after higher levels of exposure; a pregnancy has been reported in a 20-year-old woman who had previously received 3000 cGy to the pelvis. Surgical transposition of the ovaries out of the irradiated field has been somewhat successful in preserving fertility, although follicular suppression with gonadotropin-releasing hormone (GnRH) agonists has not.

Chemotherapeutics are designed to destroy rapidly dividing cells, which explains the sensitivity of proliferating granulosa and theca cells of active follicles to these agents. Alkylating agents are non-cell-cycle-specific drugs that result in cross-linking of DNA strands even in nondividing cells and as a result they have been observed to be the most toxic to oocytes. The experience thus far in women treated with chemotherapy has shown that younger women are more likely to regain ovarian function. However, the doses and classes of agents used are also important factors. Based on the observation that prepubertal girls treated with alkylating agents were relatively resistant to oocyte destruction, some work has been performed using GnRH agonists and oral contraceptive pills in an attempt to mimic this prepubertal state

and provide ovarian protection. Data by Blumenfeld demonstrating a protective effect by GnRH agonist in human patients undergoing chemotherapy for lymphoma or autoimmune diseases are encouraging.

Environmental toxicants that affect gonadal function have also been identified. Smoking is a known ovarian toxicant that has been associated with accelerated ovarian atresia. It is well documented that smokers have lower estradiol levels and begin menopause on average 2 years earlier than nonsmokers. In women who are otherwise predisposed to earlier cessation of ovarian function, smoking could accelerate the process such that clinical POF is diagnosed.

Autoimmune disorders have been strongly associated with POF, with up to 60% of POF patients having an associated autoimmune disorder. The most common autoimmune conditions associated with POF are listed in Table I. In each of these conditions, the pathophysiology of POF may involve both premature follicular depletion associated with the inflammatory process and follicular dysfunction; hence, their classification here is arbitrary.

In 119 karyotypically normal POF patients prospectively screened for various autoimmune conditions such as thyroid disease, adrenal dysfunction, hyperparathyroidism, diabetes mellitus, and pernicious anemia, hypothyroidism was the most common comorbid diagnosis with POF (found in 27% of screened POF patients), with 2.5% of patients having diabetes mellitus and 2.5% having Addison's disease. Three percent of POF patients are diagnosed with autoimmune polyglandular failure syndrome types

**Table I Autoimmune Diseases Commonly Associated with Premature Ovarian Failure**

---

Hypothyroidism
Type I diabetes mellitus
Hyperparathyroidism
Addison's disease
Myasthenia gravis
Thymic hypoplasia/aplasia
Crohn's disease
Vitiligo
Pernicious anemia
Systemic lupus erythematosus
Rheumatoid arthritis
Type I autoimmune polyglandular failure syndrome (hypoparathyroidism, adrenal failure, and chronic mucocutaneous candidiasis)
Type II autoimmune polyglandular failure syndrome (adrenal failure and hypothyroidism)

---

I and II. In fact, over 60% of type I patients will have either primary amenorrhea or POF, although the association of POF with type II is more variable. It has also been suggested that bacterial, protozoal, or viral infections may induce a cross-reacting anti-ovarian autoimmune response that could result in POF. For example, oophoritis has been reported in 3–7% of women infected with the paramyxovirus causing mumps. POF has also been noted to occur after varicella, shigellosis, and malaria infection. It also has been associated with cytomegalovirus infections in immunocompromised patients with human immunodeficiency virus and lymphoma. However, the true incidence of ovarian failure related to systemic infections and the exact pathophysiology of POF remain unknown. The infectious relationship with POF discussed here is distinct from that associated with tubo-ovarian abscess formation that actually destroys sufficient ovarian tissue to cause failure of ovarian function.

Lymphocytic oophoritis has been described in some cases of POF. Hoek reviewed 215 cases of POF with ovarian biopsies. Eleven percent of these patients had histologic evidence of oophoritis. Of the patients evaluated for anti-steroid cell antibodies, all of those with oophoritis were positive for these antibodies, suggesting that the inflammatory process was directed, at least in part, at the steroidogenic apparatus of the follicle. Lymphocytic oophoritis has been described clinically as POF associated with tender, palpably enlarged ovaries. Histologically, there is a neutrophilic infiltrate concentrated around developing follicles, with notable sparing of primordial follicles. It has been proposed that the local production of lymphokines and cytokines results in follicular dysfunction, impairing steroid production.

## FOLLICULAR DYSFUNCTION SYNDROMES

Antibodies interfering with normal FSH–receptor interactions have been found in some women with POF, analogous to the blocking antibodies described in myasthenia gravis, primary hypothyroidism, and insulin-resistant diabetes mellitus. For example, an IgG antibody blocking FSH binding to its receptor has been demonstrated in patients with myasthenia gravis who have POF. It is conceivable that the FSH and luteinizing hormone (LH) receptors may act as inciting antigens leading to a localized autoimmune response, since they are unique to the ovary. It has

also been postulated that antibodies against the gonadotropins themselves may be involved in the pathogenesis of POF.

The zona pellucida (ZP) is also unique to the ovary and antibodies against it have been proposed as a mechanism for infertility in women with premature ovarian dysfunction, although not POF. These antibodies are thought to interfere with sperm binding. In several animal models, inoculation with ZP has led to follicular dysfunction and eventual depletion.

Enzyme deficiencies that result in insufficient production of androgens and specifically estrogens can result in follicular dysfunction and clinical POF. Cholesterol desmolase deficiency results in insufficient androgens, estrogens, and corticosteroids and hence is usually incompatible with life, but a partial enzymatic defect may result in ovarian follicle dysfunction secondary to the lack of steroid production in the follicle required for proper oocyte development. Similarly, deficiencies of 17- $\alpha$ -hydroxylase or 17,20-desmolase can result in inadequate estrogen production and follicular dysfunction. Steroidogenic acute regulatory protein is a mitochondrial phosphoprotein that catalyzes the rate-limiting step in steroid synthesis, i.e., translocation of cholesterol from the outer to the inner mitochondrial membrane. The absence or dysfunction of the steroidogenic regulatory protein results in a rate-limiting defect in adrenal and gonadal steroid synthesis. Women with this condition, congenital lipoid adrenal hyperplasia, develop ovarian failure as a result of accumulation of cholesterol in the parenchyma of the ovary.

Although they do not result in elevated gonadotropins, defects in gonadotropin production that can result in secondary amenorrhea and POF have been described. An example is a case of X-linked adrenal hypoplasia in which the DAX-1 mutation results in defective pituitary gonadotropin production.

Any abnormality in gonadotropin structure, gonadotropin receptor, or second-messenger proteins can result in follicular dysfunction, although acquired isolated FSH deficiency causing POF is rare. Examples include FSH receptor abnormalities that have been described in a Finnish cohort with POF; these POF women have follicles present on ultrasound and a normal complement of primordial follicles on ovarian biopsy. Abnormal LH receptors have also been found in POF kindreds. Finally, a patient presenting with secondary amenorrhea and hypergonadotropic hypoenestrogenism was discovered to have a defective cyclic AMP second-messenger system with POF and concurrent pseudo-hypoparathyroidism.

## DIAGNOSIS OF PREMATURE OVARIAN FAILURE

Clinical features of premature ovarian failure, though variable from patient to patient, are usually fairly obvious and hence establishing the diagnosis is not usually difficult. However, because of the impact of this diagnosis on the prognosis for future fertility, general health, and the possibility of coexisting conditions, the diagnosis should always include confirmatory gonadotropin testing. In a woman less than 40 years of age, the findings of amenorrhea for greater than 4 months with significantly elevated gonadotropins ( $> 40$  mIU/ml) on two occasions more than 1 month apart allows diagnosis of POF.

### History

POF is not associated with a characteristic preceding menstrual history. The development of amenorrhea may be acute or it may be insidious. A failure to resume menses after a pregnancy or discontinuing oral contraceptives may be the presenting history. The majority of patients with POF develop secondary amenorrhea after having years of established regular menses and normal fertility. Rarely, secondary amenorrhea may follow menarche without ever establishing regular menses. Most cases of POF are sporadic, but 4% of POF patients do have a family history of the condition. Ten to 15% of POF patients had primary amenorrhea and delayed menarche and, if associated with pubertal delay, 40% of these will have chromosomal abnormalities. Prodromal premature ovarian failure should also be considered in any young woman with menstrual irregularities and elevated gonadotropins, but not yet 4 months of amenorrhea. In the case of prodromal premature ovarian failure, polymenorrhea or oligomenorrhea may occur and can be associated with hot flushing and vaginal dryness. These are the most important patients to diagnose, because a delay in diagnosis may result in irretrievable loss of fertility. A woman with such a diagnosis should be counseled that delaying childbearing, if genetic children are desired, may not be feasible. Young women with amenorrhea, oligomenorrhea, or polymenorrhea who are considering delaying childbearing should have an FSH level measured for diagnostic and prognostic purposes.

Depending on the level of estrogen production, associated symptoms may or may not include vasomotor instability with hot flashes or flushing, vaginal

dryness, genitourinary atrophic symptoms, skin thinning, mild alopecia or hirsutism, insomnia, and emotional or behavioral changes including irritability, anxiety, depression, etc. If the patient has been anovulatory for a time, dysfunctional uterine bleeding may be present.

Patients with suspected premature ovarian failure should be questioned regarding previous ovarian surgery, chemotherapy, or radiation therapy, or a personal or family history of autoimmune disorders such as hypothyroidism, Addison's disease, diabetes mellitus, Graves' disease, vitiligo, systemic lupus erythematosus, rheumatoid arthritis, or inflammatory bowel disease. Sensorineural deafness in association with POF is Perrault's syndrome, a rare familial autosomal disorder

### Physical Examination

A complete physical examination with a specific search for findings linked with the above-noted conditions should be performed. Evidence of genital tract atrophy should be documented. Bimanual examination may contribute to the diagnosis of oophoritis or an enzyme defect if the ovaries are enlarged. A search for physical stigmata of Turner's syndrome should also be noted.

### Laboratory Studies

A karyotype is not useful in establishing the diagnosis of POF. However, because neither parity nor age eliminates the possibility of a chromosomal abnormality, karyotypic analysis should be completed in all

patients diagnosed with POF. Because of the well-documented association between POF and other autoimmune processes, patients should also be screened for other endocrinopathies. Suggestions for the evaluation of patients diagnosed with POF are indicated in [Table II](#).

Although up to 67% of patients with POF have been reported to have anti-ovarian antibodies (depending on the assay system used), they can also be demonstrated in approximately 30% of normal ovulatory women. Specific steroid cell antibodies have been identified in several disease processes including Addison's disease patients who then go on to develop POF within 15 years. However, the incidence remains very low, with only approximately 3% of POF patients having demonstrable steroid cell antibodies. For both anti-ovarian antibodies and anti-steroid cell antibodies, there is no proven therapy that restores ovarian function if they are discovered. As a result, an assay for ovarian or steroid cell antibodies is not thought to be helpful in the evaluation of POF patients.

## THERAPEUTIC CONSIDERATIONS

### Health Maintenance

As described above, the significant association of premature ovarian failure with autoimmune disorders argues for the routine screening of these conditions. After initial screening, periodic reassessment should be considered, especially for Addison's disease and bone demineralization.

Hormone replacement should be strongly considered once the diagnosis of POF is made, since

**Table II Tests Useful in the Evaluation of a Patient Diagnosed with Premature Ovarian Failure**

---

Peripheral blood karyotype
Free thyroxine and thyroid-stimulating hormone (TSH), anti-thyroid antibodies if thyroid abnormalities are discovered
Fasting glucose and electrolytes, including calcium and phosphorous (to evaluate the parathyroids)
Corticotropin stimulation test to exclude adrenal insufficiency [administer 250 µg synthetic corticotropin (Cosyntropin) im or iv, check initial plasma cortisol levels and check levels 30 and 60 min after the administration of the corticotropin; peak cortisol ≥ 20 µg/dl indicates a normal adrenal reserve]
CBC and erythrocyte sedimentation rate
Urinalysis
Antinuclear antibody titer, rheumatoid factor, quantitative serum IgA (IgA deficiency is the most common immunodeficiency associated with autoimmune disorders)
Magnetic resonance imaging of the sella turcica [used to exclude a mass lesion when a rare gonadotropin-secreting pituitary adenoma is considered; a thyrotropin (TRH) stimulation test can be used to confirm this tumor, since such tumors are known to disproportionately secrete free α- and β-LH subunits in response to an infused 500 µg bolus of TRH]
Pelvic ultrasound
Bone densitometry

---

these women are at higher risk for osteopenia, osteoporosis, and accelerated atherogenesis. Cyclic or continuous estrogen and progestogen are useful to relieve symptoms of hypoestrogenism and to maintain bone density. Additional benefits of estrogen therapy include potential cardioprotective effects, delayed onset of Alzheimer's disease, decreased incidence of colorectal cancers, and improved vulvovaginal and urethral tissue structure and function. However, some of these possible benefits of estrogen replacement therapy have been brought into question in postmenopausal women by the recent Women's Health Initiative (WHI) study findings. The implications of these findings for women with premature ovarian failure are unclear; an NIH trial to answer these questions is underway. Calcium supplementation and weight-bearing exercise are also important for skeletal health.

Importantly, psychological support should be provided to patients afflicted with this difficult diagnosis. Frequently, these women feel that fertility potential is taken away from them before they have even had the option to consider their reproductive desires. This can be devastating and threatening to their self-esteem. Thorough education regarding their general health and fertility options is important. A premature ovarian failure support group exists at [www.POFsupport.org](http://www.POFsupport.org). In certain circumstances, individual counseling may also be appropriate.

### Family Building

The natural history of premature ovarian failure includes spontaneous pregnancies in 10% of women after diagnosis. Hormone replacement therapy does not provide birth control and those not desiring pregnancy should be placed on oral contraceptives. Episodic ovarian function has been documented in up to 50% of women with the diagnosis of POF, as demonstrated by intermittently elevated estradiol levels. Approximately 20% of these patients will be found to have elevated progesterones, indicating sporadic ovulation.

Patients with established POF do not respond to ovulation induction using either clomiphene citrate or exogenous urinary or recombinant gonadotropins. Sporadic case reports of successful induction of ovulation with pregnancy are most likely the result of fortuitous spontaneous reversal of POF at the time ovulation induction agents are employed. In fact, some believe that exogenous gonadotropin

administration is harmful, since this might increase the antigen load and stimulate anti-ovarian antibody production.

Ovulation induction after endogenous gonadotropin or ovarian suppression using gonadotropin-releasing hormone agonists, oral contraceptives, danazol, or continuous progestins has been attempted with very limited success. The strategy behind this suppression is to induce a rebound endogenous gonadotropin secretion and/or enhanced ovarian responsiveness to gonadotropins, after the period of suppression. However, despite anecdotal successes with this approach no prospective randomized trials have shown this treatment to be beneficial. There are anecdotal reports of pregnancies occurring after immunosuppressive therapy in patients with autoimmune POF and a trial of every-other-day corticosteroids for autoimmune lymphocytic oophoritis is under way at the National Institutes of Health Clinical Center. However, long-term steroid use has significant risks associated with it, such as osteoporosis, which is already a problem in these patients. Plasmapheresis, thymectomy, and intravenous immune globulin have also been successful anecdotally, although the use of these immunotherapies cannot be recommended since there are no randomized, controlled trials proving their efficacy.

Contemporary assisted reproductive technologies allow new options for these patients. Provided that the woman is otherwise in good health, oocyte donation with *in vitro* fertilization using her partner's sperm is an option in most programs at least to 45–50 years of age; implantation rates using young oocyte donors are very favorable and pregnancy outcome in these women who have donated oocyte-derived embryos is excellent. Other possible strategies that may develop in the future include harvesting of immature oocytes from POF patients with their *in vitro* maturation and co-culture prior to *in vitro* fertilization and prospective identification of women at risk for developing POF with oocyte harvest and cryopreservation for subsequent use before the onset of clinical ovarian failure.

Expectant management with the small possibility of resolution or even temporary remittance may result in pregnancy for some of these women. After the initial diagnosis and education process, these patients should be encouraged to take some time to allow for spontaneous remission or opt for adoption or oocyte donation if they have already suffered from infertility for a number of years or if family building is a pressing issue for them.

## CONCLUSION

As mentioned above, prospective cryopreservation of primordial follicles may become a more practical option in the future for women at risk of developing POF. This availability stresses the importance of early diagnosis and specific testing of women with POF and their younger female relatives. With increasing capabilities in molecular medicine and diagnosis, it may soon become possible to diagnose specific genetic anomalies that will result in follicular dysfunction or premature follicular depletion in the future. With this same advancing technology, genetic therapies to correct the dysfunction or prevent the follicular loss can be envisioned. Improved measures of protecting primordial follicles during cytotoxic therapies and the ability to preserve primordial follicles *in vitro* and mature them successfully are also therapeutic strategies under intense investigation. Specific therapy for autoimmune POF—which could include highly specific blocking antibodies that would potentially prevent the follicular dysfunction and loss and avoid the adverse systemic effects associated with corticosteroids and other immunomodulatory therapies—will be possible in the future.

## See Also the Following Articles

Assisted Reproductive Technology (ART) • Autoimmune Polyglandular Syndrome • Fertilization • In Vitro Fertilization (IVF) • Infertility, Overview • Ovarian Androgen-Producing Tumors • Ovarian Failure Treatment Strategies: Egg Donation • Ovarian-Follicular Apparatus • Resistant

Ovary Syndrome • Superovulation and Intrauterine Insemination

## Further Reading

- Adashi, E. Y., and Hennebold, J. D. (1999). Mechanisms of disease: Single-gene mutations resulting in reproductive dysfunction in women. *N. Engl. J. Med.* **340**, 709–718.
- Adashi, E. Y., Rock, J. A., and Rosenwaks, Z. (1995). “Reproductive Endocrinology, Surgery, and Technology.” Lippincott-Raven, Philadelphia, PA.
- Anasti, J. N. (1998). Premature ovarian failure: An update *Fertil. Steril.* **70**, 1–15.
- Blumenfeld, Z. (2001) Ovarian rescue/protection from chemotherapeutic agents *J. Soc. Gynecol. Invest.* **8**.
- Davis, C. J., and Conway, G. S. (2000). Female sex preponderance for idiopathic familial premature ovarian failure suggests an X chromosome defect: Opinion. *Hum. Reprod.* **15**, 2418–2422.
- Dunbar, B. S., and Skinner, S. M. (2001). The ovary as an immune target *J. Soc. Gynecol. Invest.* **8**.
- Kasteren, Y. M., and Schoemaker, J. (1999). Premature ovarian failure: A systematic review on therapeutic interventions to restore ovarian function and achieve pregnancy. *Hum. Reprod. Update* **5**.
- Nilsson, E., and Skinner, M. K. (2001). Cellular interactions that control primordial follicle development and folliculogenesis. *J. Soc. Gynecol. Invest.* **8**.
- Parrott, E. C., and Santoro, N. (2001). The ovary: Genesis, function, and failure. *J. Soc. Gynecol. Invest.* **8**.
- Premature Ovarian Failure Support Group Web site. (2003). Available at: [www.POFsupport.org](http://www.POFsupport.org).
- Santoro, N. (2001). Research on the mechanisms of premature ovarian failure. *J. Soc. Gynecol. Invest.* **8**.
- Tung, K. S. K., and Bagavant, H. (2001). Autoimmune ovarian disease: Mechanism of induction and prevention. *J. Soc. Gynecol. Invest.* **8**.
- Zinn, A. R. (2001). The X chromosome and the ovary. *J. Soc. Gynecol. Invest.* **8**.

menstrual cycle draws near but remit at or soon after menstruation. There are 7–10 days free of symptoms.

Although this history is characteristic of PMS, it is not sufficient for its diagnosis. Because some women may have selective recall of their symptoms, prospective charting of the symptoms is necessary to establish the relationship of these symptoms to the onset of the menstrual period and to establish a diagnosis. The diagnosis of PMDD also requires prospective daily self-ratings of symptoms for at least two cycles. Though prospective charting is ideal for establishing the diagnosis of PMS/PMDD, some women presenting for evaluation and treatment will be unwilling or unable to prospectively record their daily symptoms and it may be necessary to initiate therapy without adequate prospective charting. In an attempt to establish more objective criteria for the diagnosis of PMDD, some investigators and clinicians have administered psychometric tests to women during both the follicular and the luteal phases of the menstrual cycle. Though the results of psychometric testing in women with PMDD may be normal during the follicular phase, they may be abnormal during the luteal phase. However, the use of psychometric tests is not necessary to make the diagnosis of PMDD. Finally, the fact that the symptoms never last for more than 2 weeks distinguishes PMS and PMDD from other mood or physical disorders.

It is important during the course of the evaluation to rule out other conditions such as anxiety disorders, depressive disorders, anemia, hypothyroidism, chronic fatigue syndrome, chronic fatigue syndrome, central nervous system tumors, or autoimmune disorders. It is also important during the course of evaluation of the patient with PMS/PMDD to evaluate the impact of the symptoms on her self-esteem as well as her relationships with family, friends, and co-workers. Finally, it is also helpful to discuss stresses in the lives of these women for they may exacerbate the symptoms of PMS/PMDD. Medical, physiological, and social stressors that are common in the lives of women with PMS/PMDD include emotional or physical abuse, financial distress, alcohol or substance abuse, or stress from school or work.

## TREATMENT

### Overview

Because no single treatment is universally effective for PMS/PMDD, many have doubted the existence of this disorder and criticized attempts to help women with their premenstrual symptoms. However, in the past

decade there have been over 100 prospective randomized clinical trials of various therapies, which have led to an evidence-based approach to the treatment of PMS/PMDD.

### Initial Nonspecific Therapy

The initial treatment of PMS/PMDD can begin during the phase of evaluation using nonspecific therapies. These nonspecific therapies include nonpharmacologic approaches such as patient education, validation of symptoms through daily prospective charting, relaxation training, and self-help measures such as exercise and nutrition. Because social support is often essential for relief of symptoms, it may be helpful to the women with PMS/PMDD if the clinician involves family members or friends in the process of evaluation so that they may support her and help validate her symptoms. Although data supporting the effectiveness of these nonpharmacologic therapies are limited, these therapies are often used because they lack significant side effects, are very practical, and may benefit the woman's physical and emotional health.

### First-Line Pharmacologic Therapies

Once the diagnosis of premenstrual syndrome has been made, more specific therapy can be initiated for those women who desire pharmacologic intervention. Because of their demonstrated effectiveness in prospective randomized clinical trials, selective serotonin reuptake inhibitors (SSRIs) or agents that inhibit the reuptake of both serotonin and norepinephrine are commonly used. The four commonly prescribed SSRIs (fluoxetine, paroxetine, sertraline, and citalopram) are often effective in their usual starting dose when used daily throughout the menstrual cycle. However, sertraline has been demonstrated to be effective when administered only in the luteal phase (last 14 days) of each menstrual cycle. In contrast to the gradual and delayed response of individuals with chronic depressive disorders to SSRIs, women with premenstrual syndrome may note improvement of their symptoms within days of starting their medication. In addition to the SSRIs, first-line therapies may include venlafaxine or calcium carbonate. Several prospective randomized studies have shown calcium carbonate to be effective in reducing premenstrual syndrome such as depression, water retention, food cravings, fatigue, abdominal bloating, headache, breast fullness, and pain. Finally, mefenamic acid and naproxen sodium have also been shown to be effective

therapies. When administered in the luteal phase of the menstrual cycle, these agents may reduce physical as well as emotional symptoms.

### Second-Line Therapies

When these first-line therapies are either ineffective or unacceptable to the patient, there are other therapies that may be helpful even though scientific support for their use is not as strong as for the first-line therapies. Second-line, nonpharmacologic therapies include biofeedback, bright light therapy, cognitive behavioral therapy, and massage therapy. Second-line pharmacologic agents include the antidepressants fluvoxamine maleate, clomipramine, nefazodone, and phenylpiperazine and the anti-anxiety medication alprazolam. These agents have not been studied to the same extent as the first-line SSRIs, but there are some data to support their use. In addition, the side effects of these medications are infrequent and usually mild. Although obliteration of the menstrual cycle and ovulation by danazol or gonadotropin-releasing hormone (GnRH) agonists has been shown to be effective, their side effects limit their widespread use. Danazol is often associated with fluid retention, acne, facial hair growth, and muscle cramps. GnRH agonists are associated with menopausal symptoms including hot flashes, night sweats, and insomnia.

This wide array of potentially effective therapies offers new hope to women with PMS or PMDD. However, the therapy of PMS/PMDD should be more than just pharmacologic and should include

self-help measures so that the patient becomes an active partner in her therapy. In addition, attempts should be made to recognize and help the patient deal with stresses in her life and to help her establish an effective social support system. The importance of validation, social support, and the elimination of stress cannot be emphasized enough. It is hoped that further research will uncover new and even more effective and safe therapies for premenstrual syndrome.

### See Also the Following Article

Menstrual Cycle: An Integrative View

### Further Reading

- American College of Obstetricians and Gynecologists. (2000). "Premenstrual Syndrome," ACOG Practice Bulletin No. 15. American College of Obstetricians and Gynecologists, Washington, DC
- Dimmock, D. N., Wyatt, K. M., Jones, P. N., *et al.* (2000). Efficacy of selective serotonin-reuptake inhibitors in premenstrual syndrome: A systematic review. *Lancet* **356**, 1131–1136.
- Robinson, R. L., and Swindle, R. W. (2002). Premenstrual symptom severity: Impact on social functioning and treatment-seeking behaviors. *J. Women Health Gender Based Med.* **9**, 757–768.
- Thys-Jacobs, S., Starkey, P., Bernstein, D., *et al.* (1998). Calcium carbonate and the premenstrual syndrome: Effects on premenstrual and menstrual symptoms. Premenstrual Syndrome Study Group. *Am. J. Obstet. Gynecol.* **179**, 444–452.
- Yonkers, K. A. (1997). The association between premenstrual dysphoric disorder and anxiety disorders. *J. Clin. Psychiatry* **58**, 19–25.

stimulation of aldosterone production throughout several days of synthetic adrenocorticotrophic hormone (ACTH) administration (tetracosactrin 1 mg every 12 h by intramuscular injection). These features permitted the clear distinction of this subtype from the non-glucocorticoid-remediable forms of PAL even when they occurred in a familial setting. The familial occurrence of APA and other non-glucocorticoid-remediable forms of PAL was first described by the Greenslopes Hospital Hypertension Unit (GHHU) in Brisbane, Australia in 1991, and was termed familial hyperaldosteronism type II (FH-II) in order to differentiate this new familial form of PAL from the glucocorticoid-remediable variety [familial hyperaldosteronism type I (FH-I)]. Diagnosis of FH-I was simplified even further following the elucidation of the “hybrid” 11 $\beta$ -hydroxylase/aldosterone synthase gene mutation that causes this condition and the development of sensitive and specific genetic tests that followed.

### Seeking the Optimal Approach for the Differentiation of APA from BAH

The differentiation of APA from BAH has proven far less straightforward than the differentiation of FH-I from other subtypes of PAL and has been the subject of intense investigation. Early attempts at imaging APAs involved relatively insensitive and invasive procedures such as adrenal venography, which carried with it a significant risk of contrast extravasation and adrenal hemorrhage. In general, patients with APA have demonstrated somewhat more severe biochemical abnormalities than those with BAH and some investigators have attempted to develop multiparameter biochemical models, incorporating such variables as plasma sodium, potassium, total carbon dioxide, aldosterone, and renin, in order to distinguish the two subtypes. However, the degree of overlap was great and an assessment of the severity of PAL proved to be of little value in determining the subtype in the individual patient.

It was a finding of considerable significance, therefore, that plasma aldosterone levels in patients with APA failed to rise in response to a change in posture from recumbency to an upright position, or to infusions of angiotensin II (Ang II), as was also the case in patients with FH-I, whereas in patients with BAH, responsiveness of plasma aldosterone to these maneuvers was retained, and perhaps even enhanced. In the 1980s, patients with APA or FH-I were further found to differ from those with BAH by demonstrating elevated levels of hybrid steroids, 18-hydroxy-cortisol

and 18-oxo-cortisol. In 1987, however, the GHHU reported unilaterality of aldosterone production on adrenal venous sampling in patients in whom aldosterone was responsive to both posture and Ang II infusion and whose hypertension was subsequently cured by unilateral adrenalectomy. Furthermore, urinary levels of 18-hydroxy-cortisol and 18-oxo-cortisol were usually normal. These patients thus masqueraded biochemically as BAH, yet, like those with classic Ang II-unresponsive APA, could be cured of hypertension following removal of their tumors by unilateral adrenalectomy. The discovery of this new subtype of Ang II-responsive APA meant that APA could not be excluded on the basis of retained aldosterone responsiveness to posture or Ang II infusion or normal hybrid steroid levels.

Newer methods of adrenal imaging, such as scintigraphic scanning following injection of radiolabeled methylcholesterol, computerized tomography (CT) scanning, and magnetic resonance imaging, though relatively noninvasive, have frequently been found to be associated with false-negative studies owing to the small size of many APAs. Furthermore, apparently nonfunctioning adrenal tumors (“incidentalomas”) can masquerade as APA and thereby produce false-positive results on imaging studies.

The procedure that has stood the test of time in terms of reliability in differentiating unilateral, surgically correctable forms of PAL (including APA) from bilateral forms (BAH) is adrenal venous sampling (AVS), in which right and left adrenal venous aldosterone/cortisol ratios are compared to those in simultaneously collected peripheral venous blood. Since imaging studies lack reliability, the argument that AVS should be performed in every case of PAL following exclusion of FH-I by genetic testing, in order to differentiate APA from BAH and lateralize APAs preoperatively, has gathered increasing support.

### Prevalence—The PAL Renaissance

For most of its history, PAL was regarded as being a rare condition, accounting for 1% of hypertension at best. Hypokalemia was thought to be present in almost all cases, so that the search for PAL was considered to be unnecessary in normokalemic individuals.

In the 1990s, and into the 2000s, evidence gathered by the GHHU and subsequently by an increasing number of other investigators suggested that PAL is much more common than previously thought, accounting for at least 5–10% of hypertensives, and that most patients are normokalemic. These findings arose from the availability of improved methods of

screening [and in particular the plasma aldosterone/plasma renin activity (PRA) ratio (ARR)] and the application of the ARR to a widened population of hypertensives to include normokalemic (and not just hypokalemic) patients and those with less resistant forms of hypertension. These measures led to a 10-fold increase in the detection of PAL (including a 4-fold increase in the identification of patients with APA), providing the opportunity for specific and potentially curative treatment to a great many more hypertensive patients than was previously possible.

## PATHOGENESIS

In PAL, excessive production of the salt-retaining hormone aldosterone results in excessive sodium reabsorption via amiloride-sensitive epithelial sodium channels within the distal nephron (leading to hypertension) and continues in the face of renin–Ang II suppression. All grades of hypertension have been observed, although malignant hypertension has only rarely been reported. Family screening, by genetic testing in families with FH-I and ARR testing in those with FH-II, has permitted the identification of normotensive individuals with PAL, suggesting that this condition may evolve through a normotensive “phase,” the duration of which may depend on the presence and efficiency of counter-regulatory (including genetic) mechanisms protecting those individuals against the development of hypertension. Urinary loss of potassium and hydrogen ions, exchanged for sodium at the distal nephron, may eventually result in hypokalemia and metabolic alkalosis if severe and prolonged enough. At the GHHU and Princess Alexandra Hospital Hypertension Unit (PAHHU), where PAL is sought by measuring the ARR in all referred hypertensives, hypokalemia is present in less than one-quarter of patients detected and in less than one-half of those with APA. If hypokalemia does develop, it may be associated with nocturia, polyuria, muscle weakness, cramps, paresthesias, and/or palpitations.

Morbidity in PAL primarily results from hypertension. However, experimental and clinical evidence suggests that aldosterone excess can bring about cardiovascular injury (including inflammation, fibrosis, and remodeling) independently of blood pressure elevation. These effects appear to be preventable by the administration of mineralocorticoid receptor antagonists. Doses of aldosterone used in experimental studies have been very large and the clinical relevance of the resulting cardiovascular changes remains uncertain. Nevertheless, several groups have demonstrated

abnormalities in cardiac morphology or function in patients with PAL that appear to be out of proportion to the elevation in blood pressure and which improved following specific treatment of PAL.

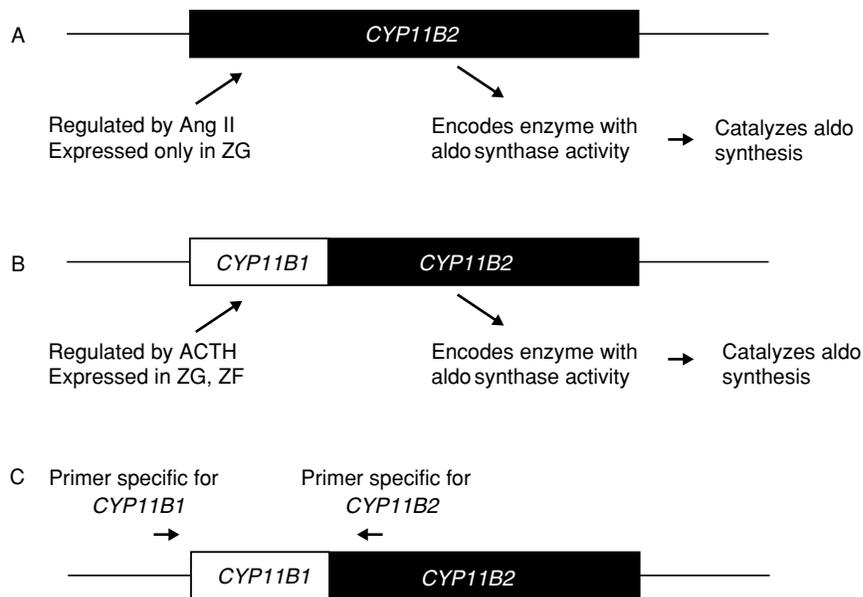
With the exception of the glucocorticoid-remediable variety (FH-I), the etiology of PAL remains unknown. The “hybrid gene” mutation that causes FH-I is composed of 5′ (including regulatory and some coding) sequences derived from the 11 $\beta$ -hydroxylase gene (*CYP11B1*) and 3′ (including all remaining coding) sequences derived from the aldosterone synthase gene (*CYP11B2*) (see Fig. 1). This gene has sufficient *CYP11B2* coding sequences to encode an enzyme that efficiently synthesizes aldosterone from its precursors, deoxycortisone and 18-hydroxycorticosterone. Unlike *CYP11B2*, however, which is predominantly regulated by Ang II, the hybrid gene is regulated by ACTH by virtue of its *CYP11B1* regulatory sequences. Aldosterone production in FH-I is therefore regulated by ACTH rather than by Ang II, which renders it glucocorticoid-suppressible. Unlike *CYP11B2*, the hybrid gene is expressed not only in the zona glomerulosa, but also in the zona fasciculata. Here cortisol is available as a substrate for its 18-hydroxylase and 18-oxidase gene product activities, which probably explains the excessive production of 18-hydroxy-cortisol and 18-oxo-cortisol (“hybrid steroids”) in FH-I.

Other forms of PAL may also have a genetic basis and the familial occurrence of non-glucocorticoid-remediable PAL (FH-II, which includes APAs and BAH) is in keeping with this. Within the GHHU and PAHHU, new families with FH-II have been detected five times more frequently than those with FH-I and affected patients have demonstrated clinical, biochemical, and adrenal morphological features that are indistinguishable from those with apparently sporadic PAL. Mutations responsible for FH-II may therefore be more common, at least collectively, than the hybrid gene and may be operative in patients with apparently sporadic PAL. The genetic bases of FH-II remain uncertain, but in two affected families thus far studied, linkage studies have implicated a locus in chromosome 7p22.

## SUBTYPES

### Pathological

Pathological expression of PAL is highly variable and different histological subtypes frequently coexist in the one patient. The most common is adrenal cortical hyperplasia (which may be diffuse or nodular), followed by adenoma, and adrenocortical carcinoma is



**Figure 1** Regulation and outcome of expression of (A) the wild-type aldosterone synthase gene (*CYP11B2*) and (B) the “hybrid” 11 $\beta$ -hydroxylase (*CYP11B1*)/*CYP11B2* gene. (C) The strategy used for amplification of the hybrid gene by long-polymerase chain reaction. *CYP11B2* sequences are represented as black boxes and *CYP11B1* sequences as open boxes. Ang II, angiotensin II; ACTH, adrenocorticotropic hormone; ZG, zona glomerulosa; ZF, zona fasciculata. See text for detailed description.

fortunately very rare. Cells contained in APAs may resemble those of the zona fasciculata (ZF), zona glomerulosa (ZG), or zona reticularis or may have cytological features of both ZF and ZG (“hybrid cells”). The predominant cell type may help to determine the biochemical behavior of the tumors, with the presence of at least 80% non-ZF (ZG, reticularis or hybrid)-type cells seeming to confer Ang II-responsiveness, whereas unresponsiveness has usually been observed in APAs in which 50–100% of the cells were ZF-like.

### Functional

From a management point of view, it is more useful to subclassify PAL according to functional (rather than morphological) characteristics and in particular whether the PAL is:

1. glucocorticoid-remediable (FH-I, which makes up less than 1% of patients with PAL), in which case treatment with small doses of glucocorticoids would be expected to result in excellent hypertension control;
2. non-glucocorticoid-remediable and lateralizes to one adrenal (consistent with APA, which constitutes approximately one-third of patients with PAL; or with aldosterone-producing carcinoma, which is rare), in which case unilateral adrenalectomy would be expected to bring about cure or marked improvement in hypertension; or

3. non-glucocorticoid-remediable and bilateral (consistent with BAH, which constitutes approximately two-thirds of patients); it usually is managed medically with aldosterone antagonist medications.

## DIAGNOSIS

### Screening for PAL

Because hypokalemia is present in only a minority of patients, measurement of plasma potassium lacks sensitivity for detecting PAL. On the other hand, the occurrence of hypokalemia in a hypertensive patient, in the absence of an alternative explanation (for example, treatment with diuretics or gastrointestinal loss of potassium accompanying vomiting or diarrhea), is highly suggestive of PAL.

Plasma aldosterone levels frequently lie within the wide normal range in patients with PAL, including those with APA, and a normal level therefore does not exclude the diagnosis. Although the demonstration of suppressed PRA levels is highly sensitive for PAL, it lacks specificity, occurring in several other inherited and acquired salt-loaded hypertensive states (including high dietary salt intake, Liddle’s syndrome, inherited deficiency of 11 $\beta$ -hydroxysteroid dehydrogenase type 2 or inhibition of this enzyme by licorice abuse or carbenoxolone treatment, deoxycortisone-secreting tumors, hypertensive forms of congenital

adrenal hyperplasia, and inherited activating mutations of the mineralocorticoid receptor).

Measurement of the ARR has become widely regarded as the most reliable available means of screening for PAL. Many factors, however, including the presence of hypokalemia, medications, dietary salt intake, posture, and time of day, can affect the ARR level and reduce its reliability. Before measuring the ARR, therefore, it is important to: (1) correct hypokalemia (which can lower plasma aldosterone and cause false-negative results); (2) where possible, replace potentially interfering medications (such as beta-blockers,  $\alpha$ -methyl dopa, clonidine, and nonsteroidal anti-inflammatory drugs, which can cause false-positive results; and diuretics, dihydropyridine calcium antagonists, Ang II-converting enzyme inhibitors, and Ang II receptor blockers, which can cause false-negative results) for several weeks (at least 4 weeks for diuretics and at least 2 weeks for the rest) with others that have minimal effects on the ARR (such as verapamil slow-release, hydralazine, or alpha-blockers), or at least be prepared to take the effects of interfering medications into account, and (3) encourage a liberal dietary salt intake (rather than a salt-restricted diet, which can cause false-negative results).

Sensitivity of the ARR is optimized by collecting blood at midmorning (at 9:00 to 10:00 A.M.) rather than later in the day and from seated (rather than recumbent) patients who have been upright (sitting, standing, or walking) for at least 2 h.

Under the above conditions, a ratio of greater than 30 (plasma aldosterone in nanograms per 100 ml; PRA in nanograms per milliliter per hour) is considered suggestive of PAL.

### Confirmation of PAL

As the ARR is a screening test only, further testing is necessary to definitively confirm or exclude the diagnosis of PAL. The most reliable of these, used by the GHHU and PAHHU, is the fludrocortisone suppression test, in which failure of upright (collected at 10:00 A.M.) plasma aldosterone to suppress to less than 6 ng/100 ml at the conclusion of 4 days of fludrocortisone (0.1 mg every 6 h) and slow-release sodium chloride (Slow Na<sup>+</sup>, 1800 mg thrice daily with meals) administration is considered diagnostic. To avoid the potentially confounding effects of abnormal potassium levels on plasma aldosterone, plasma potassium is measured frequently (three to four times daily) over the 4-day test and slow-release potassium chloride (Slow K<sup>+</sup>) is given every 6 h in doses sufficient to maintain levels within the normal range.

An alternative approach, used at the Mayo Clinic, involves the measurement of 24 h urinary aldosterone levels during the third day of oral salt loading (sufficient to achieve a urine sodium excretion of > 200 mmol/day, with enough potassium supplementation to maintain normokalemia). Failure to suppress to 12  $\mu$ g/day or less is regarded as diagnostic.

The saline infusion test involves measurement of plasma aldosterone at the conclusion of an intravenous infusion of 0.9% saline (usually 2 liters over 4 h). Cut-off values for the diagnosis of PAL have varied from > 5 to > 10 ng/100 ml. Although this approach is more convenient than fludrocortisone suppression testing, as it requires only a brief visit, it was found by the GHHU to lack sensitivity for the detection of PAL (including APA).

### Determining the Subtype

#### **Confirmation or Exclusion of FH-I**

The glucocorticoid-remediable form of PAL can be rapidly confirmed or excluded by genetic testing of peripheral blood DNA for the presence of the hybrid gene either by Southern blot or by a faster, polymerase chain reaction-based method. This has rendered as obsolete other, biochemical methods of identifying FH-I (in particular, dexamethasone suppression testing), being far more convenient yet highly reliable.

#### **Differentiating Lateralizing Forms of PAL from Bilateral Forms of PAL**

The great majority of patients with PAL test negative for the hybrid gene. In these patients, AVS is the most reliable means of distinguishing lateralizing forms of PAL, including APA and aldosterone-producing carcinoma, from BAH. This procedure is highly specialized and requires an experienced, highly skilled, and dedicated radiologist to achieve optimal rates of successful cannulation (at GHHU, > 90% for the right adrenal vein and > 95% for the left). Several samples are collected from each side during the procedure to maximize the chances of success, which can be gauged by comparing adrenal and peripheral venous cortisol levels, with gradients of at least 3 indicating an "adequate" sample. AVS should be performed in the morning hours following overnight recumbency in order to avoid the confounding effects of changes in posture on aldosterone levels and to take advantage of the effect of high early morning endogenous ACTH levels on aldosterone production. Provided that adrenal venography is avoided, the risk of adrenal hemorrhage associated with AVS is low.

Calculation of the aldosterone/cortisol ratio for each adrenal and peripheral venous sample corrects for differences in “dilution” of adrenal with nonadrenal venous blood. If the average ratio on one side is significantly (usually two times or more) higher than the simultaneous peripheral venous ratio, and the ratio on the other side is no higher than the peripheral venous ratio, the study is considered to show lateralization, indicating that unilateral adrenalectomy should cure or improve the hypertension.

Fine-cut adrenal CT fails to detect at least one-half of APAs (which average only 1 cm in diameter) and may be misleading by demonstrating nonfunctioning nodules. The use of magnetic resonance imaging does not overcome these limitations. Adrenal CT is nevertheless useful for detecting larger lesions (> 2.5 cm), which may warrant removal based on their malignant potential. Adrenal selenocholesterol scanning has been reported to be useful by some centers, but not by most others (including the GHHU), which have found it to lack sensitivity and specificity for APA.

The demonstration of aldosterone responsiveness (defined as a rise of at least 50% over basal levels) during 2 h of upright posture following overnight recumbency or during a 1 h infusion of Ang II (2 ng/kg/min) was once thought to be specific for BAH, but was later recognized to lack discriminatory value in differentiating BAH from APA. At least one-half of APAs removed at the GHHU and PAHHU have been of the Ang II-responsive variety and at least one-fifth with BAH have demonstrated a lack of responsiveness. Hence, regardless of the presence or absence of Ang II-responsiveness, AVS is required to definitively differentiate APA from BAH and to lateralize APAs preoperatively. Hybrid steroid levels, though elevated in FH-I and Ang II-unresponsive APA, again provide only limited value in differentiating subtypes, since levels are usually normal in both BAH and Ang II-responsive APA.

## TREATMENT

### Patients with FH-I

In FH-I, hypertension is readily controlled by giving glucocorticoids in low doses (e.g., 0.125–0.5 mg of dexamethasone per day). Complete suppression of ACTH-regulated aldosterone production is rarely required and raises the risks of Cushingoid side effects. Alternative approaches include the use of spironolactone and amiloride. Amiloride may be a preferred option for affected children, since it avoids potential problems of growth retardation associated with the

use of glucocorticoids and androgen blockade due to spironolactone. Family screening by genetic testing should be undertaken to identify affected relatives.

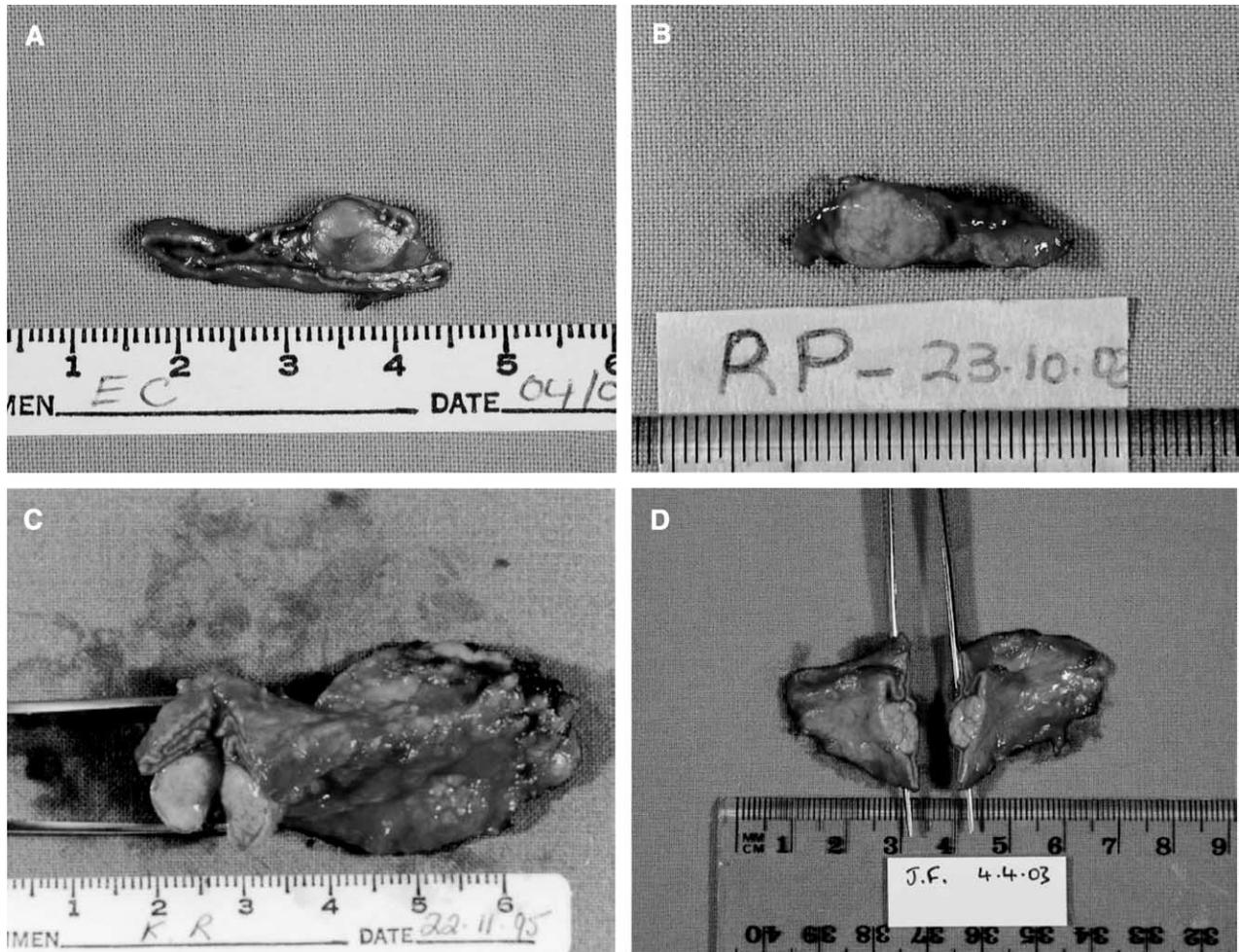
### Patients who Lateralize on AVS

In patients who lateralize on AVS, unilateral adrenalectomy (performed laparoscopically, enabling a faster recovery than the open approach) results in cure of hypertension in 50–60% and significant improvement in the remainder and correction of hypokalemia in virtually all who were hypokalemic preoperatively (see Fig. 2). Operated patients usually report a marked improvement in quality of life. Sometimes no further antihypertensives are required after surgery, but usually more gradual withdrawal of antihypertensive medications is possible over the ensuing 3 to 12 months.

Surgical treatment may be inappropriate in a small number of patients who lateralize, including those in whom surgery is considered unsafe because of coexisting medical conditions or those who would prefer not to undergo surgery for a variety of reasons. These patients can be treated medically with aldosterone antagonists using principles similar to those described below for the medical treatment of patients with nonlateralizing PAL.

### Patients with Bilateral, Non-Glucocorticoid-Remediable Adrenal Aldosterone Overproduction

For patients demonstrating bilateral aldosterone production on AVS, treatment with aldosterone antagonists usually brings about significant and often marked improvement in hypertension control. The results, however, are generally not as gratifying as those in patients with FH-I treated with dexamethasone or in lateralizing patients who undergo unilateral adrenalectomy. Only small doses of spironolactone (12.5mg daily–25mg twice daily) or amiloride (2.5mg daily–7.5mg twice daily) are usually required, provided weeks or months are allowed for the drug to demonstrate its full effect. Even at these low doses, side effects (gynecomastia, menstrual irregularities, and reduced libido) occur in approximately 10% of patients taking spironolactone. A more selective aldosterone antagonist (eplerenone) with less affinity for sex steroid receptors appears to be relatively free of these side effects and has been shown to be effective as an antihypertensive agent in “essential hypertensives.” However, clinical experience with eplerenone in PAL remains limited.



**Figure 2** Aldosterone-producing adenomas (APAs) removed laparoscopically from four patients in whom preoperative adrenal venous sampling lateralized aldosterone production to the side of the subsequently removed adrenal and in whom surgery resulted in cure of hypertension. Preoperative patient characteristics are as follows: (A) male, 57 years, hypokalemic, posture-unresponsive, left APA (0.8 cm) not visible on CT; (B) male, 48 years, hypokalemic, posture-responsive, right APA (0.8 cm) not visible on CT; (C) female, 42 years, normokalemic, posture-unresponsive, left APA (0.9 cm) detected by CT; (D) female, 25 years, normokalemic, posture-responsive, right APA (0.9 cm) not visible on CT.

Aldosterone antagonists must be used with caution in patients with impaired renal function because of their increased tendency to develop hyperkalemia. Concurrent administration of a potassium-wasting diuretic in low doses may help to avoid hyperkalemia, but potassium and creatinine levels still require close monitoring.

### See Also the Following Articles

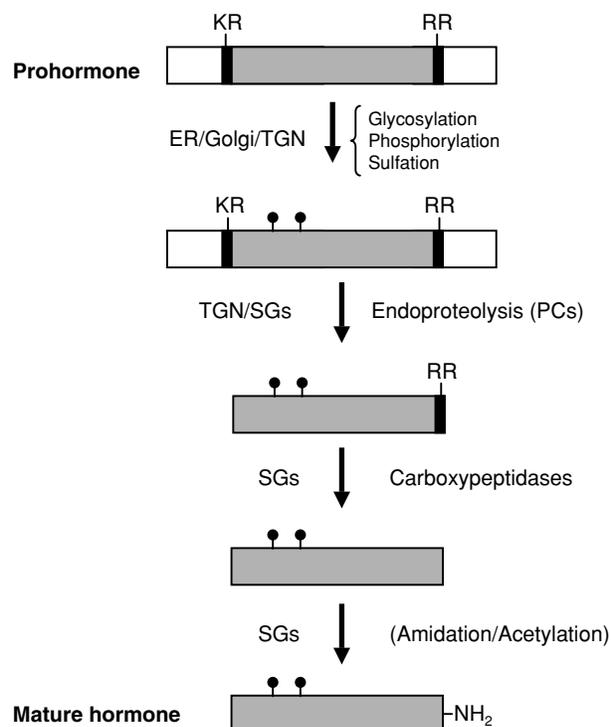
Aldosterone in Congestive Heart Failure • Aldosterone Receptors • Hypertension, Endocrine • Hyporeninemic Hypoaldosteronism • Mineralocorticoids and Mineralo-

corticoid Excess Syndromes • Tissue Renin-Angiotensin-Aldosterone System

### Further Reading

- Conn, J. W. (1968). The evolution of primary aldosteronism: 1954–1967. *Harvey Lectures* 62, 257–291.
- Gordon, R. D., Stowasser, M., Klemm, S. A., and Tunny, T. J. (1994). Primary aldosteronism and other forms of mineralocorticoid hypertension. In "Textbook of Hypertension" (J. D. Swales, ed.), pp. 865–892. Blackwell Scientific, London.
- Hiramatsu, K., Yamada, T., Yukimura, Y., Komiya, I., Ichikawa, K., Ishihara, M., Nagata, H., and Izumiya, T. (1981). A screening test to identify aldosterone-producing adenoma by measuring plasma renin activity. *Arch. Intern. Med.* 141, 1589–1593.

- Lifton, R. P., Dluhy, R. G., Powers, M., Rich, G. M., Cook, S., Ulick, S., and Lalouel, J. M. (1992). A chimaeric  $11\beta$ -hydroxylase/aldosterone synthase gene causes glucocorticoid-remediable aldosteronism and human hypertension. *Nature* **355**, 262–265.
- Stowasser, M., and Gordon, R. D. (2001). Familial hyperaldosteronism. *J. Steroid Biochem. Mol. Biol.* **78**, 215–229.
- Stowasser, M., Gordon, R. D., Rutherford, J. C., Nikwan, N. Z., Daunt, N., and Slater, G. J. (2001). Diagnosis and management of primary aldosteronism. *J. Renin-Angiotensin-Aldosterone System* **2**, 156–169.
- Sutherland, D. J. A., Ruse, J. L., and Laidlaw, J. C. (1966). Hypertension, increased aldosterone secretion and low plasma renin activity relieved by dexamethasone. *Can. Med. Assoc. J.* **95**, 1109–1119.
- Young, W. F., Jr. (1997). Primary aldosteronism: Update on diagnosis and treatment. *The Endocrinologist* **7**, 213–221.



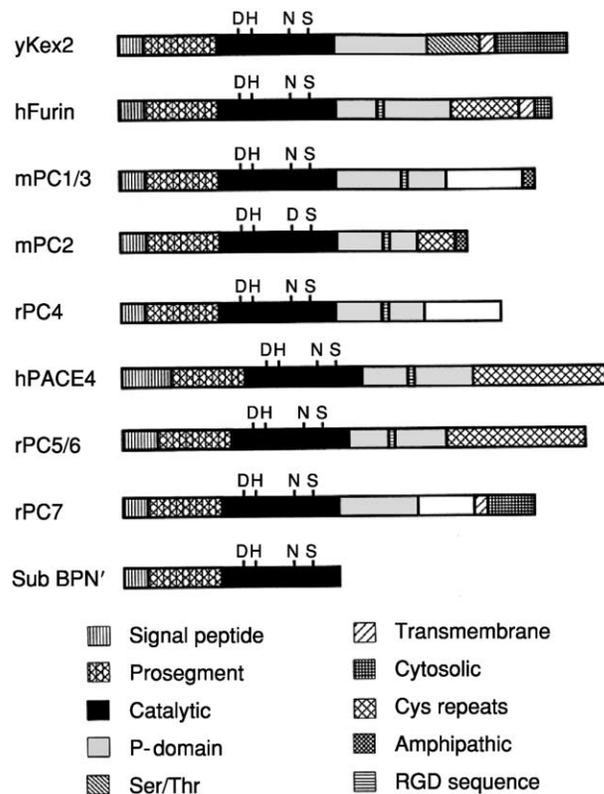
**Figure 1** Prohormone processing by prohormone convertases. The mature peptide hormone sequence (gray) is flanked by pairs of basic amino acids (black) within the prohormone. Prohormones may be glycosylated (black lollipop) in the Golgi en route to secretory granules. The PCs cleave on the carboxyl end of the dibasic cleavage site, and carboxypeptidases remove the C-terminally extended basic amino acids. Finally, peptide hormones may be amidated or acetylated to attain biological activity. K, lysine; R, arginine; ER, endoplasmic reticulum; TGN, *trans*-Golgi network; SGs, secretory granules.

## ENDOPROTEASES IN PROHORMONE PROCESSING

### PC1 and PC2: Biological Activity, Structure, and Cellular Trafficking

#### Biological Activity

Both PC1 and PC2 have been localized within the secretory granules of the anterior and neurointermediate lobes of the pituitary, pancreatic islets, endocrine cells of the heart, and neuropeptide-rich regions of the intestine and brain. Within these tissues, PC1 and PC2 exhibit distinct distributions, and this in turn is responsible for the tissue-specific processing observed for several prohormones. For example, proopiomelanocortin (POMC) is expressed in both the anterior and neurointermediate lobes (NILs) of the pituitary but is processed to a different complement of



**Figure 2** Schematic representations of the structures of the known endoproteolytic prohormone convertases. These enzymes were identified based on sequence homology with the catalytic domain of bacterial subtilisin (Sub BPN'). All enzymes contain the catalytic triad D (aspartic acid), H (histidine), and S (serine) characteristic of all serine proteases. The asparagine (N) residue in the catalytic domain is also conserved throughout the family except in PC2, where it is replaced by an aspartate (D). All prohormone convertases are also glycosylated at multiple sites (not shown). y, yeast; h, human; m, mouse; r, rat.

peptides in each lobe due to the differential distribution of PC1 and PC2. In the anterior lobe, PC1 is highly expressed and cleaves POMC to adrenocorticotrophic hormone (ACTH) and  $\beta$ -lipotropic hormone ( $\beta$ -LPH). In the NILs, both PC1 and PC2 are expressed and cleave POMC in a sequential fashion to yield smaller peptides, melanocyte-stimulating hormone ( $\alpha$ -MSH) and  $\beta$ -endorphin, which are unique to the NILs. Similarly, the tissue-specific processing of proglucagon in the alpha cells of the pancreatic islets and the L cells of the distal small intestine is the result of differential expression of PC1 and PC2. Proglucagon is processed to glucagon largely through the actions of PC2 in the pancreatic islets, whereas PC1 cleaves proglucagon in the intestine to yield the gastrointestinal hormones glucagon-like peptide-1 (GLP-1) and GLP-2. Finally, both PC1 and PC2

cleave proinsulin in a concerted manner to yield insulin and C peptide in the pancreatic beta cell. The B-chain/C-peptide junction is cleaved by PC1, followed by cleavage at the C-peptide/A-chain junction by PC2.

Both PC1 and PC2 require high intracellular calcium concentrations (10–20 mM) and an acidic pH (5.0–5.5) for optimal activity. These conditions exist within the secretory granules, the subcellular compartment to which the PCs are transported and in which prohormone processing takes place.

### Structure

The structure of PC1 and PC2 can be organized into discrete domains (Fig. 2). Each enzyme contains an N-terminal signal peptide, which is removed on translocation of the nascent peptide into the endoplasmic reticulum (ER). Both enzymes contain an N-terminal prodomain, which inhibits the catalytic activity of the PCs. Sequence alignment of PC1 and PC2 shows a 55% amino acid identity within the catalytic domains, which contain the highly conserved active site Asp, His, and Ser. Other sites of sequence identity include Asn-282, which is replaced by Asp-285 in PC2 and is associated with development of the oxyanion hole in bacterial subtilisins. The catalytic domain is followed by the P domain, which contains the conserved tripeptide sequence Arg–Gly–Asp, a recognition sequence by cell surface antigens implicated in cell–cell adhesion. The C-terminal end of the P domain plays a role in regulating the stability, calcium dependence, and pH dependence of the convertases. The C-terminal sequences of PC1 and PC2 contain an amphipathic  $\alpha$ -helical structure that functions in membrane anchoring.

### Subcellular Trafficking and Activation

#### PC1

The subcellular trafficking of PC1 and PC2 has been thoroughly investigated and follows slightly different routes for processing and activation. Both are initially synthesized as proenzymes in the ER. Within this compartment, the route for processing and activation of PC1 diverges from that of PC2.

The 86- to 89-kDa pro-PC1 is quickly converted to an 83- to 86-kDa PC1 by removal of its prodomain at an RSKR cleavage site in an autocatalytic intramolecular manner in the ER. Full-length PC1 is transported through the Golgi in association with a molecular chaperone, pro-SAAS, which inhibits the catalytic activity of PC1. The PC1 inhibitory sequence of pro-SAAS is comprised of a 10- to 12-amino

acid segment near its C terminus and contains a critical Lys–Arg-244 dibasic motif. Targeting of PC1 to the regulated secretory pathway is dependent on sorting information contained within its C-terminal domain, which anchors it to lipid raft microdomains within the membranes of the *trans*-Golgi network (TGN). In the immature granule, PC1 is processed at an Arg–Arg-618 site at its C terminal end in an autocatalytic manner to yield a soluble 64- to 66-kDa form. The catalytic activity of 64- to 66-kDa PC1 is optimal between pH values 5.0 and 5.5, the pH values of the mature and immature secretory granules, respectively. Although PC1 activity has been measured as soluble activity, a significant proportion of PC1 remains membrane associated within the immature and mature secretory granule.

#### PC2

The mechanism of activation of pro-PC2 is more complex than that of PC1. PC2 is unique among the PC family in that it requires an acidic, post-Golgi compartment for processing and activation. Initially synthesized as a 75-kDa proenzyme, pro-PC2 is folded and subsequently associates with 7B2, a 27-kDa neurosecretory protein of the granin family. The 21-kDa N-terminal portion of 7B2 binds to the catalytic domain of pro-PC2 via a polyproline helical structure. The 6-kDa C-terminal peptide of 7B2 inhibits PC2 activity. The pro-PC2–7B2 complex then exits the ER. On reaching the TGN, 7B2 is cleaved by furin to liberate the N-terminal portion of 7B2, allowing for the removal of the prodomain of PC2 in an autocatalytic intramolecular manner. The C-terminal peptide remains associated with pro-PC2; once the prodomain is removed, the inhibitory C-terminal peptide can access the active site of PC2 previously occupied by the propeptide. The C-terminal peptide of 7B2 contains a Lys–Lys pair that is a site of slow cleavage by PC2. Final inactivation of the C-terminal peptide occurs in the secretory granules through the removal of the Lys–Lys pair by CPE.

Both the C-terminal amphipathic  $\alpha$ -helical region and the N-terminal propeptide may serve as sorting signals to direct pro-PC2 to the regulated secretory pathway. Association of the propeptide with lipid rafts in the TGN also directs pro-PC2 into secretory granules. The catalytic activity of active PC2 is optimal between pH values 5.0 and 5.5, the pH values of the mature and immature secretory granules, respectively. Like PC1, a significant amount of PC2 remains bound to membranes of secretory granules.

## Yapsins

In yeast *Saccharomyces cerevisiae* mutants, that lack the subtilisin-like serine protease kexin, a number of aspartyl proteases were able to compensate for its activity, demonstrating that these new enzymes could also be classified as prohormone convertases. These aspartyl proteases are known as yapsins. There are now five members in the yapsin family: three from *S. cerevisiae* (yapsins 1, 2, and 3), one from *Candida albicans* (SAP9), and one from bovine pituitary intermediate and neural lobe secretory granules previously characterized as pro-opiomelanocortin-converting enzyme (PCE, now named yapsin A). Their unique specificity for the basic residue cleavage sites of prohormones such as POMC, proinsulin, cholecystokinin (CCK), and prosomatostatin added a new understanding to the field of aspartyl proteases because it was well known that all previously characterized enzymes of the aspartyl protease family preferred substrates with hydrophobic residues.

## Prohormone Thiol Protease

Another enzyme that has the properties consistent with being a prohormone convertase is the 33-kDa prohormone thiol protease purified and characterized from the chromaffin granules of the adrenal medulla. Within the chromaffin granule, this enzyme is capable of cleaving proenkephalin at either paired or monobasic cleavage sites. It has also been shown to process recombinant proneuropeptide Y but cleaves protachykinin and POMC poorly. It has an optimal pH activity at pH 5.5 and is potently inhibited by thiol-specific protease inhibitors. It has also been found in some lung cancer cells, where it can process parathyroid hormone-related protein (PTHrP) at basic residue cleavage sites to generate bioactive peptides. This enzyme has been identified as cathepsin L.

## Dynorphin-Converting Enzyme

Distinct from the prohormone thiol protease, two other enzymes have been purified from the neurointermediate lobe of the pituitary and from brain and are also thiol sensitive. These enzymes are known as dynorphin-converting enzyme (DCE) and dynorphin A-17-processing enzyme. These enzymes are metalloproteases with molecular masses of 54 and 65 kDa and exhibit isoelectric points of 5.1 and 5.8, respectively. They both have a neutral pH optimum and appear to cleave on the N-terminal side of single basic residue cleavage sites.

## CARBOXYPEPTIDASE AND AMINOPEPTIDASE PROTEASES IN PROHORMONE PROCESSING

Following endoproteolytic cleavage of prohormones by the PCs, exopeptidases such as carboxypeptidases and aminopeptidases remove the basic residue extensions from peptide hormones to yield the active hormone. CPE, found exclusively in endocrine and neuroendocrine tissues, is the primary enzyme that cleaves basic residues from the C-terminally extended peptide hormones. A second carboxypeptidase, carboxypeptidase D (CPD), has the same specificity as CPE and may also play a role, but it is present in endocrine tissues in much lower amounts than CPE. Aminopeptidase activities that specifically remove basic residues from the N terminus of peptides have been found in pituitary and chromaffin granules.

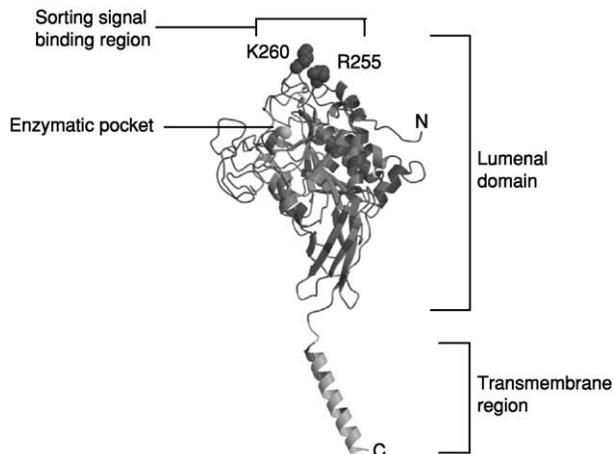
## CPE

### Activity

CPE has been purified and cloned from bovine, human, rat, and mouse endocrine tissue. It is also present in brain. CPE is primarily localized to the secretory granules of neuroendocrine cells where prohormone processing occurs. CPE specifically cleaves C-terminally extended Lys and Arg residues from peptide hormones. It functions optimally between pH values 5.5 and 6.0, the intragranular pH. Its activity is stimulated by  $\text{Co}^{2+}$  and inhibited by metal ion chelators (EDTA, 1,10-phenanthroline) and the active site inhibitor guanidine-ethylmercaptosuccinic acid (GEMSA), which distinguishes it from carboxypeptidase B-like activities in lysosomes.

### Structure

There are two forms of CPE: a 50-kDa soluble form and a 55-kDa membrane-bound form. Both forms are derived from the same mRNA encoding prepro-CPE. Prepro-CPE contains a signal peptide and a prodomain at the N terminus that is separated from the mature enzyme by an RRRRR (where R = Arg) sequence. Mature soluble CPE contains a substrate-binding site, a  $\text{Zn}^{2+}$  binding site, and a catalytic pocket (Fig. 3). In addition, the membrane form contains a 22-amino acid C-terminal domain that forms an amphipathic  $\alpha$ -helix. This domain is transmembrane, with the last 4 to 6 residues forming the cytoplasmic tail (Fig. 3), and functions to anchor CPE to lipid rafts at the TGN. This anchoring is essential for targeting CPE to the regulated secretory pathway.



**Figure 3** Molecular model of carboxypeptidase E (CPE). The three-dimensional structure of CPE shows the orientation of the luminal domain and transmembrane region. The zinc atom is drawn as a sphere to mark the enzymatic pocket. The POMC sorting signal-binding region, R255-K260, is exposed on the surface of the molecule and is spatially distinct from the enzymatic pocket. The transmembrane region forms a partial amphipathic  $\alpha$ -helix at acidic pH that is stabilized through interaction with the lipid bilayer. The C-terminal 4 to 6 amino acids are exposed on the cytoplasmic side of the secretory granule and so comprise a cytoplasmic “tail.”

When anchored to secretory granule membranes, CPE has less enzymatic activity relative to the soluble form that is five to six times more active. Soluble CPE is derived from the membrane form by removal of the C-terminal transmembrane domain. The amount of soluble CPE in the secretory granule ranges from 10 to 70%, depending on the tissue.

#### **Membrane CPE: A Prohormone Sorting Receptor**

In endocrine tissue, the amount of CPE present is in excess of the other prohormone convertases, a significant amount of which is the membrane form. Membrane CPE functions as a sorting receptor for POMC by facilitating the entry of POMC into the regulated secretory pathway. By antisense-mediated depletion of CPE in neuroendocrine cells, Loh was able to show that the sorting of POMC to the regulated secretory pathway was impaired, resulting in constitutive secretion of the prohormone. In addition, her experiments using pituitaries from mutant CPE<sup>fat/fat</sup> mice depleted of CPE also showed constitutive rather than regulated secretion of POMC.

#### **CPD**

CPD has a substrate specificity and pH optimum (5.5–6.5) that is consistent with a role in peptide hormone processing. However, it is more widespread in distribution and has a subcellular localization in the TGN, but not in secretory granules where much of the processing occurs, especially the later steps requiring an exopeptidase action.

#### **Aminopeptidases**

An aminopeptidase activity that cleaves N-terminal basic residues from peptide hormones has been isolated from bovine pituitary secretory granules. It has been characterized as a thiol-metallo-aminopeptidase with a pH optimum of 6.0, is stimulated by  $\text{Co}^{2+}$  and  $\text{Zn}^{2+}$ , and is inhibited by EDTA. Two similar thiol-metallo-aminopeptidase activities have also been found in chromaffin granules. One, termed the Arg-MCA activity, is stimulated by  $\text{Co}^{2+}$ , but the other, termed the Lys-MCA activity, is not. However, it is unclear whether the two activities are derived from one enzyme. Unlike aminopeptidase B-like activities from other tissues, these thiol-metallo-aminopeptidase activities are not inhibited by  $\text{Cl}^-$ .

### **THE PROHORMONE CONVERTASES AND DISEASE**

Mutations in the genes encoding specific prohormones that lead to a defect in, reduction in, or elimination of biologically active peptide hormones may cause certain types of endocrine disorders. For example, some mutations in the POMC gene result in obesity and red hair pigmentation, whereas some mutations in the proinsulin gene result in familial hyperproinsulinemia. In addition, mutations in the genes encoding the prohormone convertases PC1, PC2, and CPE may also result in endocrine disorders. The conversion of proinsulin to insulin and of POMC to ACTH,  $\alpha$ -MSH, and  $\beta$ -endorphin are two obvious endocrine systems that would be greatly disturbed by defects in the PCs. To address the *in vivo* physiological importance of convertases in general, a number of “knockout mice,” in which the gene of a particular PC has been deleted, have been engineered. This approach has been used to study the function of a number of the PCs, including furin, PC4, PC1, and PC2. For CPE, an alternative mouse model is discussed in what follows.

Genetic deletion of furin, a ubiquitously expressed convertase active in the constitutive secretory pathway

of every cell, resulted in lethality and the mice died *in utero*. For PC4, a convertase expressed exclusively in germ cells, the knockout mice survived. The male mice were infertile and the female mice exhibited decreased folliculogenesis in the ovaries, demonstrating a specific function of PC4 in reproduction.

## PC1

Like the furin knockout mouse, the PC1 knockout mouse did not survive, demonstrating an essential role for this neuroendocrine-specific enzyme in embryonic development. However, there are reports of PC1 deficiencies in some human patients. One such patient was characterized as a compound heterozygote, where each allele of the PC1 gene contains a distinct mutation, one that causes expression of a mutant protein that is misfolded and not secreted and one that causes a frame-shift mutation, resulting in a premature stop codon that produces a truncated form of the enzyme. In this individual, PC1 activity was absent and the patient was extremely obese and hypoglycemic. The patient also exhibited endocrine imbalances such as hyperproinsulinemia with very low levels of mature insulin, hypogonadotropic hypogonadism, and increased plasma levels of both POMC and cortisol. Another case report described a patient with isolated congenital ACTH deficiency due to impaired processing of POMC. However, the gene for POMC was normal, and the activities of PC1 and PC2 also appeared to be normal, suggesting the possibility that a defect in another POMC processing enzyme may have been responsible for the observed lack of ACTH. The identity of this enzyme was unknown, but a good candidate is the previously described PCE or yapsin A.

## PC2 and 7B2

The PC2 knockout mice are normal in their appearance besides having a slight decrease in growth rate. However, the mice are chronically hypoglycemic, and at the endocrine level there is impaired processing of proglucagon, proinsulin, and prosomatostatin in pancreatic islets. Histologically, the islets display marked alpha cell hyperplasia and hypertrophy, with a corresponding reduction in the beta cell population. These mice secrete large amounts of unprocessed proglucagon, and the resulting lack of active circulating glucagon explains the observed hypoglycemia. The absence of PC2 also results in the elevation of the des-31,32-proinsulin intermediate, reflecting the preference of PC2 for this substrate.

Endocrine disruptions were also observed in the 7B2 knockout mice. The absence of 7B2, a chaperone necessary for normal PC2 activation, obliterated PC2 enzymatic activity. As a result, the mice developed hypoglycemia, hyperproinsulinemia, and hypoglycagonemia. Interestingly, 7B2 knockout mice also had symptoms of Cushing's disease due to very high circulating levels of ACTH. The overproduction of ACTH originated from the intermediate lobe of the pituitary, in which POMC is normally processed to  $\alpha$ -MSH by PC2. In the absence of active PC2, POMC was instead processed to ACTH, which acted on the adrenal cortex and caused hypercortisolemia. Because the intermediate lobe is under tonic inhibitory control by dopamine, production and secretion of the ACTH are not controlled by the normal inhibitory feedback mechanism of corticosterone. Therefore, production of both ACTH and corticosterone is unregulated, resulting in severe Cushing's disease, and the mice died at 9 weeks of age.

## CPE

In the absence of an engineered CPE knockout mouse, much has been learned from an obese mouse model known as the CPE<sup>fat/fat</sup> mouse. This mouse carries an autosomal recessive mutation that arose spontaneously in the HRS/J inbred strain of mice that maps to the Cpe locus. The CPE gene in these mice showed a T-to-C point mutation within the coding region that resulted in a conversion of Ser-202 to Pro. This mutant CPE lacks enzymatic activity, and the protein is totally degraded in the pituitary and partially degraded in the pancreatic islets and brain. Deficiency in CPE in these mice leads to a phenotype of infertility and obesity. A number of endocrine disruptions are evident, including diabetes due to hyperproinsulinemia, a lack of stimulated secretion of CCK in response to refeeding after fasting, and a severe reduction in intestinal GLP-1 production. These defects are the result of the incomplete processing, or sorting, of gonadotropin-releasing hormone, proinsulin, pro-CCK, and proglucagon.

Several mutations in the human CPE gene have been described in Japanese and Caucasian patients with type 2 diabetes. However, there appears to be no causative relationship between mutations in the CPE gene and type 2 diabetes.

The studies described in this article demonstrate the utility of knockout mice in determining the functions of the PCs in the whole organism. Importantly, these studies highlight the requirement of proper

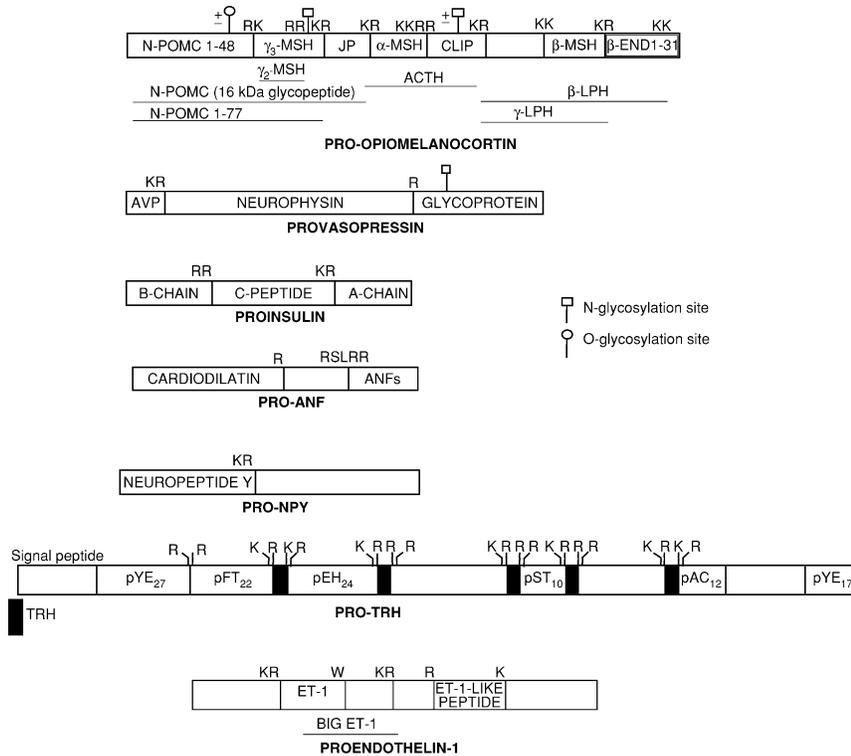
prohormone processing in the maintenance of endocrine function and shed light on the role of PC defects in endocrine disorders. Further studies of other peptide hormones and neuropeptides affected by processing enzyme deficiencies should clarify the extent of the involvement of various endoproteases in prohormone processing in the nervous and endocrine systems.

### See Also the Following Articles

Peptide Hormones, Biological Diversification by Posttranslational Processing • Peptide Hormones, Biosynthesis and Posttranslational Processing of • Peptide Hormones, Intracellular Transport • Peptide Hormones, Regulation and Gene Expression • Prohormones

### Further Reading

- Chrétien, M., Mbikay, M., Gaspar, L., and Seidah, N. G. (1995). Proprotein convertases and the pathophysiology of human diseases: Prospective considerations. *Proc. Assoc. Am. Physicians* **107**, 47–66.
- Cool, D. R., Normant, E., Shen, F. S., Chen, H. C., Pannell, L., Zhang, Y., and Loh, Y. P. (1997). Carboxypeptidase E is a regulated secretory pathway sorting receptor: Genetic obliteration leads to endocrine disorders in *Cpe<sup>fat</sup>* mice. *Cell* **88**, 73–83.
- Hook, V. Y. H. (ed.) (1998). "Proteolytic and Cellular Mechanisms in Prohormone and Proprotein Processing." R. G. Landes, Austin, TX.
- Jackson, R. S., Creemers, J. W., Ohagi, S., Raffin-Sanson, M. L., Sanders, L., Montague, C. T., Hutton, J. C., and O'Rahilly, S. (1997). Obesity and impaired prohormone processing associated with mutations in the human prohormone convertase 1 gene. *Nat. Genet.* **16**, 218–220.
- Steiner, D. F. (1998). The proprotein convertases. *Curr. Opin. Chem. Biol.* **2**, 31–39.
- Westphal, C. H., Muller, L., Zhou, A., Zhu, X., Bonner-Weir, S., Schambelan, M., Steiner, D. F., Lindberg, I., and Leder, P. (1999). The neuroendocrine protein 7B2 is required for peptide hormone processing *in vivo* and provides a novel mechanism for pituitary Cushing's disease. *Cell* **96**, 689–700.



**Figure 1** Examples of prohormones with paired basic residues and glycosylation sites marked. ANF, atrial natriuretic factor.

## PROCESSING OF PROHORMONES

### Prohormone Convertase Discovery

With the discovery of proinsulin and other prohormone precursors and the realization that cleavage occurred at paired basic residues to generate the biologically active peptide, the search for the enzyme capable of making the physiological cleavage began. Kemmler and colleagues proposed that a trypsin-like enzyme would be responsible for cleavage at paired basic residues followed by a carboxypeptidase B-like enzyme that would remove the C-terminal basic amino acids. It has become apparent that trypsin-like limited proteolysis (as opposed to complete protein digestion) of a wide variety of precursors at paired or occasional single basic residues occurs in the *trans*-Golgi network (TGN) and/or secretory granules. The search for these physiological processing enzymes was an arduous task that took more than 20 years and was fraught with false positives due to contaminating enzymes from organelles other than the TGN and secretory granules. The major advance was the finding that the yeast gene kexin (Kex2) encoded an endopeptidase that cleaved pro- $\alpha$ -factor and pro-killer toxin of *Saccharomyces cerevisiae* at paired basic residues similar

to trypsin. Yeast Kex2 was found to cleave mammalian prohormones. Most serendipitously, computer alignment revealed consider homology in the catalytic region of Kex2 with the protein product, called furin, from the gene (*fur*) that was immediately upstream of the tyrosine kinase *fps/fes* oncogene. Shortly thereafter, the proprotein-processing activity of furin at paired basic amino acids was demonstrated. However, the ubiquitous distribution of furin (highest levels in kidney and brain) led investigators to postulate that other prohormone processing enzymes with catalytic sites homologous to Kex2 and furin would have a distribution limited to neuroendocrine cells.

Using polymerase chain reaction (PCR) techniques, two groups (Steiner and colleagues in Chicago and Seidah and colleagues in Montreal) simultaneously isolated two mammalian prohormone processing enzymes with a tissue distribution limited to endocrine and neuroendocrine cells. Steiner and colleagues used the consensus sites from Kex2 for PCR amplification from human insulinoma RNA to yield a probe that was used to screen a human insulinoma library and isolate a full-length cDNA coding for the novel PC, called PC2. Independently, Seidah and colleagues, using primers based on the furin sequence and cDNA from

mouse pituitary RNA, generated a PCR product that was used as a probe to screen mouse pituitary and insulinoma libraries to isolate full-length cDNA clones coding for mouse PC2 and another for a PC named PC1. Both PC1 and PC2 have an endocrine and neuroendocrine distribution and cleave prohormones on the C-terminal side of paired basic residues. It is noteworthy that the same two laboratories that identified prohormones (i.e., those of Steiner and Chretien) identified the PCs more than 20 years later.

### Prohormone Convertase Family Members

Seven members of the PC family have been cloned: furin, PC1 (also referred to as PC3), PC2, PACE4, PC4, PC5/6, and PC7/PC8. Furin, PACE4, PC6B (an isoform of PC5/6), and PC7 are localized in the distal Golgi/TGN, are ubiquitously distributed among cell types, and act predominantly on precursors of growth factors, growth factor receptors, and viral or other secreted glycoproteins. PC1, PC2, PC4, and PC5/6A are localized in the dense core vesicles of the regulated secretory pathway and process neuropeptide and hormonal precursors. Both PC1 and PC2 are widely distributed in the brain. PC1 appears to be present in both the anterior and intermediate lobes of rodent pituitary, whereas PC2 is found predominantly in the intermediate lobe. PC4 is found in the testis, whereas PC5/6A is expressed in the gastrointestinal tract and the brain. Both PC1 and PC2 process a variety of prohormones, including POMC, prosomatostatin, provasopressin, proneurotensin, pro-TRH, proenkephalin, and pro-corticotropin-releasing hormone (pro-CRH). The general consensus is that PC1 is involved in the earlier steps in prohormone processing where a more neutral pH exists, consistent with an earlier activation in the secretory pathway, whereas PC2 is involved in later cleavage steps and requires a more acidic pH, consistent with activation in the maturing secretory granules.

### Prohormone Convertase Knockout Mice

Insights into the role of the PCs in hormonal biosynthesis have been elucidated from studies of mice lacking PC2, PC4, and (more recently) PC1 and in a patient with defective PC1 (discussed later). Mice with a knockout of PC2 developed normally and were fertile but had absent proglucagon processing and impaired proinsulin processing and were hypoglycemic. In the brain, there was impaired biosynthesis of several

neuropeptide precursors, including enkephalin, dynorphin,  $\beta$ -endorphin, melanin-concentrating hormone, neuropeptide E, and neurotensin. Male mice with a knockout of PC4 had impaired fertility, possibly due to altered biosynthesis of PACAP. Interestingly, mice lacking PC1 appear normal at birth but, starting at 3 days, have decreased growth and are about 60% of normal size at 10 weeks. They have multiple defects in processing many hormone precursors, including hypothalamic growth hormone-releasing hormone (GHRH), POMC to ACTH, proinsulin to insulin, and intestinal proglucagon to glucagon-like peptide-1 (GLP-1) and GLP-2. The impaired GHRH processing is thought to contribute to their dwarfism. Most interestingly PC1 heterozygotes are mildly obese and appear to be insulin resistant.

Other enzymes, such as the prohormone thiol protease (PTP), the N-arginine dibasic convertase (NRD convertase), the dynorphin-processing endoprotease and pro-opiomelanocortin-converting enzyme (PCE), have been proposed to be involved in the prohormone process. However, the lack of unexpected cleavage of prohormones so far characterized in PC2 and PC1 knockout mice suggests that enzymes other than PC1 and PC2 do not have an important physiological role in hormonal biosynthesis.

### Prohormone Processing

Thus, hormonal biosynthesis can be viewed as a multistep process (Fig. 2, which uses provasopressin as an example of posttranslational modifications). The initial step in the secretory cells is removal of the hydrophobic N-terminal signal peptide by signal peptidases. The prohormone may then undergo N-glycosylation, protein folding, and disulfide formation in the endoplasmic reticulum as well as phosphorylation, O-glycosylation, and sulfation in the *cis*-Golgi. The initial PC1 or PC2 endopeptidase cleavage at the carboxy side of paired basic residues in the prohormone can occur as early as the TGN. However, for most prohormones, the initial cleavage is thought to take place in the immature secretory granule. Processing can also occur at single basic residues or between or on the N-terminal side of the paired basic residues. The resulting C-terminal basic residue(s) is removed by a member of the metalloprotease family of carboxypeptidases, of which carboxypeptidase E (H) is mostly likely the relevant physiological enzyme. In many but not all hormones, the C-terminal amino acid is  $\alpha$ -amidated by peptidyl  $\alpha$ -amidating mono-oxygenase (PAM), an enzyme that

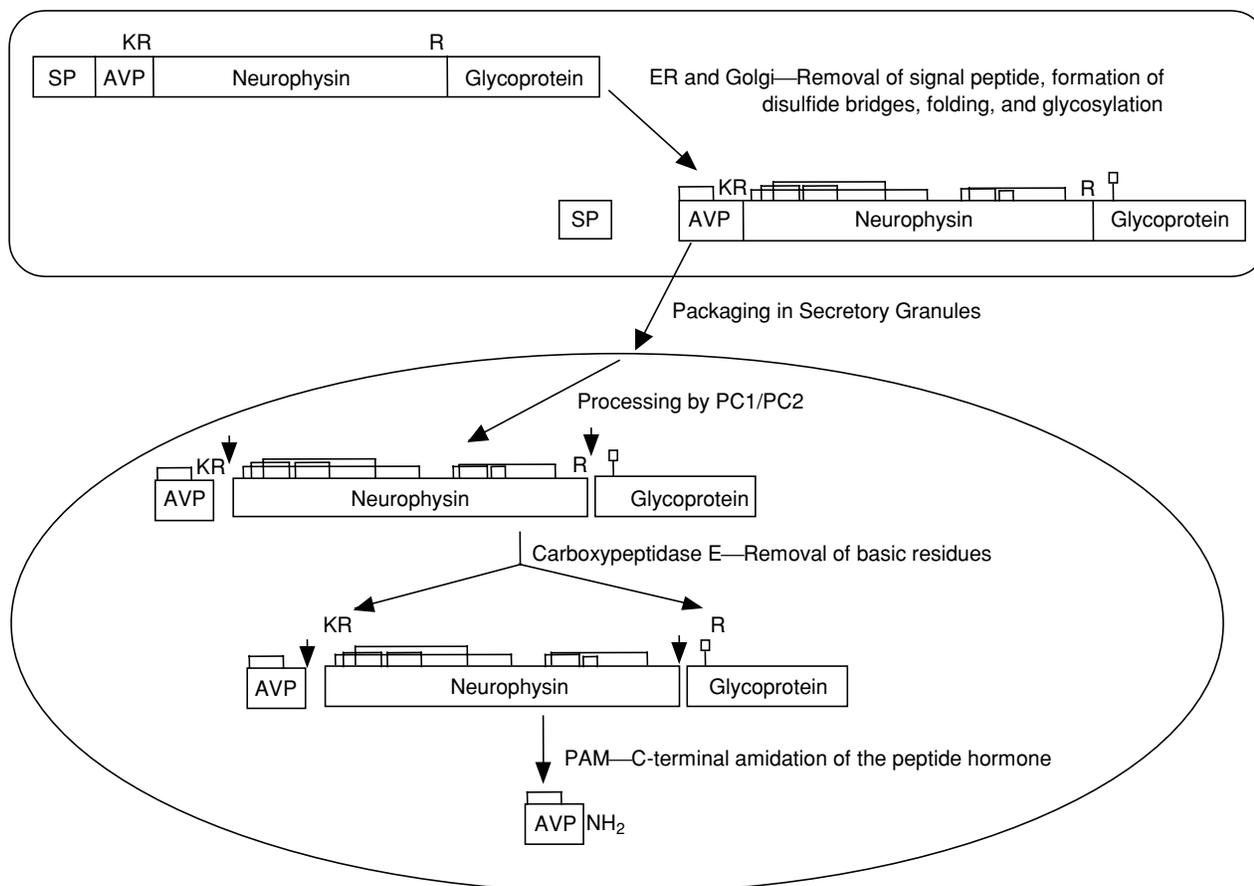


Figure 2 Posttranslational modification scheme of provasopressin. SP, signal peptide.

uses an exposed carboxy-terminal glycine amino acid as the amide donor. Additional posttranslational modifications include removal of free amino acids by aminopeptidases, acetylation of N-terminal amino acids, and conversion of N-terminal glutamate residues to pyroglutamate residues prior to secretion. This pathway is highly regulated to enable the precise amount of active hormone to be synthesized and released.

## PROHORMONE SORTING

Prohormones and peptide hormones follow a unique pathway in their movement through endocrine and neuroendocrine cells. For prohormones to be processed and secreted correctly, they must first be packaged in the same compartment or secretory granule with the processing enzymes. Two popular models have been proposed to explain the sorting process. In the first, the sorting by entry model, prohormones

contain a tertiary structure or “sorting signal” that is responsible for directing the prohormone and some enzymes to the budding secretory granules. The sorting signal has negatively charged amino acids at its core that bind to positively charged regions of membrane-bound proteins or receptors in the TGN and are incorporated into granules budding from the TGN. One such receptor has been identified as the regulated secretory granule processing enzyme, carboxypeptidase E. In contrast, a second model exists in which the prohormones are sorted by retention; that is, aggregates of the prohormone and other proteins are packaged together into immature secretory granules. Once packaged together, nonregulated secretory proteins are removed from the granules by a housekeeping, constitutive-like, secretory pathway, leaving only the prohormones, processing enzymes, and other regulated secretory granule proteins. Both models have gained support and are probably not mutually exclusive. That is, the prohormones first must be sorted and enter the immature secretory granules

and then must be retained in the same granules to be processed. Recent investigations have focused on identifying membrane components to which the sorting signal binds such as cholesterol, a major membrane component of the TGN and the immature secretory granules. Depleting the cells of cholesterol causes mis-sorting of prohormones such as POMC and insulin. Thus, the correct packaging of prohormones in the maturing granules of the regulated secretory pathway is a highly regulated process and a key step in the formation of biologically active peptide hormones from prohormones.

### EXAMPLE OF PROHORMONE

In this section, proglucagon is described as an example of a prohormone and its tissue-specific processing.

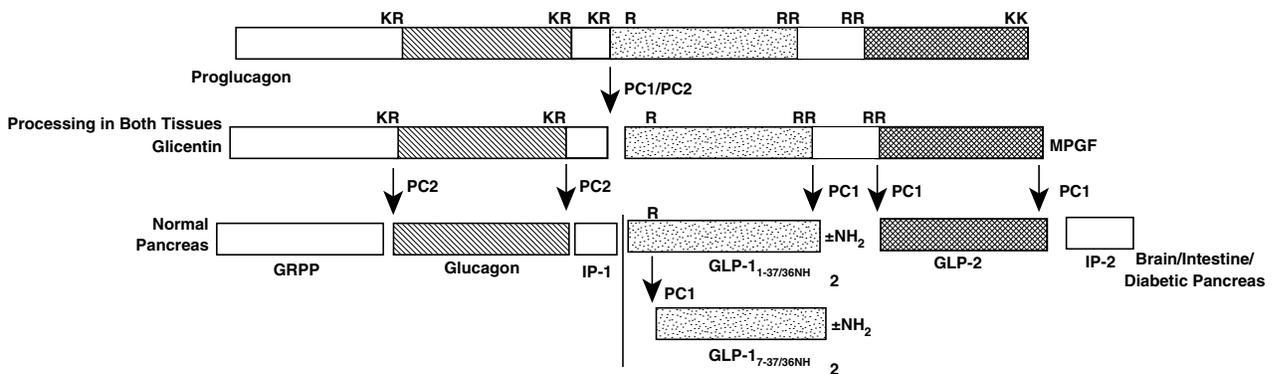
Proglucagon is synthesized in the alpha cells of the pancreas and is processed to bioactive glucagon, glycyntin-related polypeptide (GRPP), intervening peptide-1 (IP-1), and the major proglucagon fragment (MPGF), with the formation of only minimal amounts of GLP-1 (Fig. 3). Glucagon acts to raise plasma glucose levels by stimulating hepatic glycogenesis and gluconeogenesis, whereas the bioactive form of GLP-1, GLP-1<sub>7-36</sub> amide, lowers plasma glucose levels by stimulating insulin release. The processing of proglucagon in the alpha cell of the pancreas differs from that in the L cell of the intestine due to different levels of PC1 and PC2. In the alpha cell of the pancreas, the major hormonal product is glucagons, with only trace amounts of GLP-1 formed, due to the exclusive presence of PC2 in that cell. In contrast, higher levels of PC1 (but not PC2) in the L cells of the intestine lead to GLP-1 biosynthesis. The role of PC2 in generating glucagon in the alpha cell in physiological conditions

and the role of PC1 in generating GLP-1 and GLP-2 in the L cells of the intestine were confirmed by the absence of glucagon in PC2 knockout mice and the absence of intestinal GLP-1 and GLP-2 in PC1 knockout mice, as discussed previously.

In contrast to the normal condition, rats treated with the beta cell-damaging agent streptozotocin (STZ) had a profound increase in both PC1 and PC2 expression in the alpha cell of the pancreas. PC1 and PC2 expression was unchanged in beta cells. Thus, PC1, which is only minimally expressed in normal alpha cells, is induced in the diabetic state (Fig. 3). This was accompanied by an increase in bioactive GLP-1<sub>7-36</sub> amide in the pancreata and serum of the STZ-treated rat, and the ratio of pancreatic amidated GLP-1 immunoreactivity to total glucagon immunoreactivity was also increased. This is the first example of how regulation of processing enzymes can alter the ratio of bioactive hormones. The insulin staining in the remaining beta cells appeared to be increased after STZ treatment. We conclude that in the presence of hyperglycemia, the rat regulates its prohormone-converting enzymes in islet alpha cells, leading to an increase in amidated GLP-1, which can then exert an insulinotropic effect on remaining beta cells.

### HUMAN DISEASE DUE TO PROHORMONE OR PROHORMONE PROCESSING DEFECTS

Patients having frank defects in genes for prohormones or prohormone processing enzymes are probably rare. Recently, a patient with early-onset obesity and postprandial hypoglycemia with elevated glucose,



**Figure 3** Diagrammatic representation of proglucagon and proglucagon-related peptides: general and tissue specific (pancreatic alpha cells vs small intestine L cells). Arrows indicate the sites of cleavage. GRPP, glycyntin-related polypeptide; MPGF, major proglucagon fragment; IP, intervening peptide.

proinsulin, and des-64,65-proinsulin levels, but no detectable insulin, following an oral glucose tolerance test was described. This patient had low levels of ACTH and cortisol, with elevated levels of POMC. The patient was found to have mutations in both alleles of PC1, resulting in complete absence of active enzyme. A second patient with a defect in PC1 and severe diarrhea and obesity has also been reported.

Familial hyperproinsulinemia is a genetic syndrome characterized by markedly increased levels of proinsulin or its conversion intermediates. It is a rare disorder inherited in an autosomal dominant fashion. Earlier studies suggested a defect in the conversion of proinsulin to insulin in these patients. More recent studies using molecular biology techniques have identified six different point mutations in the insulin gene; however, only three of these mutations are associated with hyperproinsulinemia. These defects either lead to proinsulin not being correctly targeted to the regulated secretory pathway or lead to proinsulin not being recognized by the PCs.

Recently, three different mutations in the POMC gene were identified in two unrelated children that led to early-onset extreme obesity, adrenal insufficiency, and red hair pigmentation. The obese phenotype is likely due to the deficiency of melanocortins (most likely  $\alpha$ -melanocyte-stimulating hormone [ $\alpha$ -MSH]), which act in the brain to reduce food intake and are potential mediators of leptin action. Because  $\alpha$ -MSH also influences hair pigmentation, patients deficient in POMC have red hair.

Alterations in the intracellular processing and blood levels of arginine vasopressin (AVP) have been associated with diseases such as familial neurohypophyseal diabetes insipidus (FNDI) and Prader–Willi syndrome. In the autosomal dominant disease FNDI, mutations occur in two general areas of the provasopressin gene. In the first case, mutations are found in genes coding for the signal peptide region that is responsible for entry of the protein into the endoplasmic reticulum (ER), the signal peptide is not cleaved from provasopressin, and the protein is retained in the ER. In the latter case, more than 20 mutations have been identified in the neurophysin region that is responsible for transporting the biologically active peptide hormone, AVP, into the regulated secretory granules. Single amino acid mutations in the neurophysin region cause mis-folding of the prohormone and either mis-sorting of the prohormone to the constitutive secretory pathway or retention of the prohormone in the Golgi. Initially, this causes a lack of biologically active AVP in the blood, resulting in diabetes insipidus. Over a longer time, the magnocellular

neurons responsible for vasopressin synthesis are found to undergo neurodegeneration, most probably due to accumulation of mis-folded provasopressin.

Prader–Willi syndrome is a congenital disease, with more than 70% of the cases associated with a defect in chromosome 15q11-q13. This region is close to the region coding for the neuroendocrine molecular chaperone protein, 7B2, that binds to and prevents premature activation of pro-PC2 in the ER and Golgi. In a subset of these patients, levels of 7B2 are attenuated, resulting in a lack of active PC2 in the regulated secretory granules. Analysis of patients with this defect shows a lack of processing of provasopressin to AVP. How this defect relates to the obese phenotype of Prader–Willi patients is unclear.

## REGULATION OF PROHORMONE PROCESSING: IMPLICATIONS FOR DISEASES

It is unlikely that patients with common multigenetic diseases, such as diabetes and obesity, will have a defect in prohormones or processing enzymes. A more likely scenario is that levels of PCs are altered in various diseases, leading to abnormal levels of at least one hormone. A likely example is type 2 diabetes, in which there is an increased proportion of proinsulin and proinsulin conversion intermediates in the circulation. The elevated ratio of proinsulin to insulin is correlated with the degree of hyperglycemia and also occurs in patients with impaired glucose tolerance, indicating that the defect in conversion of proinsulin to insulin may predate the development of diabetes. In analogy to other endocrine systems, one would expect increased—not decreased—conversion of proinsulin to insulin in the face of high glucose to maintain glucose homeostasis. The fact that there is impaired insulin biosynthesis in the setting of hyperglycemia suggests that this may be a likely underlying defect in some patients with type 2 diabetes. As discussed previously, one of the most intriguing findings from the PC1 knockout mouse is that the heterozygote mouse (decreased but not absent expression of PC1) is mildly obese and has high levels of proinsulin and insulin (indicating insulin resistance), with an increased proinsulin to insulin ratio—all similar to the patient with type 2 diabetes. Proving that a component of human diabetes is dependent on impaired levels of PCs remains difficult and will probably be elucidated by examining proinsulin processing and levels of PCs in the pancreas of deceased diabetic organ donors.

In animal studies, the PCs are regulated by agents that alter cyclic AMP (cAMP) levels. Both the PC1 and PC2 promoters contain consensus cAMP response elements (CREs). Binding to the dopamine D2 receptor decreases intracellular cAMP. Thus, treatment with the dopamine antagonist haloperidol, which would be expected to raise intracellular cAMP levels, increased rat pituitary message levels for PC1, PC2, and POMC but not for furin, whereas the dopamine agonist bromocriptine decreased rat pituitary PC1, PC2, and POMC mRNA levels. Similarly, CRH, which stimulates intracellular cAMP levels, increased PC1 mRNA and promoter activity. Glucose, another agent that stimulates intracellular cAMP levels in islet cells, stimulated PC1, PC2, and proinsulin mRNA levels in insulin-producing  $\beta$ TC3 cells.

Agents that do not alter intracellular cAMP levels have also been found to regulate PC1 or PC2 levels. For example, dexamethasone treatment of AtT-20 cells decreased both PC1 and PC2 mRNA levels, suggesting that excess glucocorticoids can suppress the levels of PCs in corticotroph cells. More recently, Jansen and colleagues demonstrated that the transcription factor, early growth response-1 (EGR-1), stimulates PC2 promoter activity.

Most interestingly, thyroid hormone status regulates PC1 and PC2 levels in rodents, offering an explanation for the protean effects of thyroid hormone. In the pituitary, thyroid hormone directly or indirectly regulates gene expression in all cell types. We demonstrated that 6-*n*-propyl-2-thiouracil (PTU)-induced hypothyroidism stimulated, whereas triiodothyronine ( $T_3$ )-induced hyperthyroidism suppressed, PC1 and PC2 expression in the rat anterior pituitary. Both the PC1 and PC2 promoters were found to have multiple negative thyroid hormone response elements. Many hormones are increased in the anterior pituitary of hypothyroid rodents. The simultaneous alterations of both PC1 and PC2 by thyroid hormone (as compared with the alterations in either PC1 or PC2 alone, in the patient deficient in PC1, and in the mouse with a knockout of PC1 or PC2) may mediate profound alterations in levels of many hormones.

## CONCLUSION

Processing of prohormones by proteolytic enzymes is an important aspect of hormonal biosynthesis. The control of posttranslation processing represents an additional mechanism (in addition to transcriptional and translational regulation) for the organism to alter levels of bioactive hormones. Although patients with

a genetic defect in either their prohormones or their prohormone processing enzymes are likely to be rare, we postulate that many common polygenic diseases, such as diabetes and obesity, may be due in part to impaired prohormone processing. The fact that two closely related processing enzymes, PC1 and PC2, likely process most prohormones leads to the concept that many hormone levels will change in parallel with exposure of the cell to agents that regulate activity of the PCs. For example, an agent that increases intracellular cAMP response element-binding protein (CREB) levels will increase PC1 and PC2 activity (there are CREs on the promoters of both PC1 and PC2). This will lead to increased prohormone processing in that cell and could result in concurrently altered levels of many hormones. We call this the "milieu hypothesis."

## Acknowledgments

T.C.F. is supported by a National Institutes of Health grant (DA-14659) and a Center of Clinical Research Excellence grant (U54 RR14616-01) to Charles R. Drew University of Medicine and Sciences. D.R.C. is supported by grants from the National Institutes of Health (DK58111), Department of Defense (DAMD17-00-C-0020), and National Alliance for Autism Research.

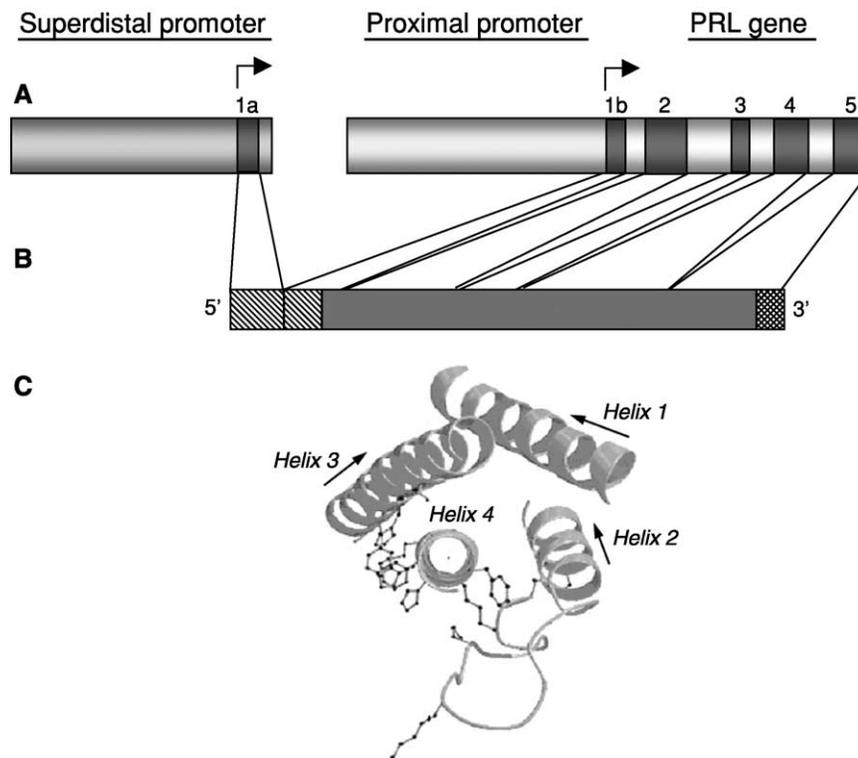
## See Also the Following Articles

ACTH,  $\alpha$ -MSH, and POMC, Evolution of • Peptide Hormones, Biological Diversification by Posttranslational Processing • Peptide Hormones, Biosynthesis and Posttranslational Processing of • Peptide Hormones, Intracellular Transport • Peptide Hormones, Regulation and Gene Expression • Prohormone Convertases

## Further Reading

- Chretien, M., and Li, C. H. (1967). Isolation, purification, and characterization of gamma-lipotropic hormone from sheep pituitary glands. *Can. J. Biochem.* **45**, 1163–1174.
- Cool, D. R., Normant, E., Shen, F., Chen, H. C., Pannell, L., Zhang, Y., and Loh, Y. P. (1997). Carboxypeptidase E is a regulated secretory pathway sorting receptor: Genetic obliteration leads to endocrine disorders in Cpe<sup>fat</sup> mice. *Cell* **88**, 73–83.
- Eipper, B. A., and Mains, R. E. (1980). Structure and biosynthesis of pro-adrenocorticotropin/endorphin and related peptides. *Endocrine Rev.* **1**, 1–27.
- Furuta, M., Yano, H., Zhou, A., Rouille, Y., Holst, J. J., Carroll, R., Ravazzola, M., Orci, L., Furuta, H., and Steiner, D. F. (1997). Defective prohormone processing and altered pancreatic islet morphology in mice lacking active SPC2. *Proc. Natl. Acad. Sci. USA* **94**, 6646–6651.
- Jansen, E., Ayoubi, T. A., Meulemans, S. M., and Van de Ven, W. J. (1995). Neuroendocrine-specific expression of the human prohormone convertase 1 gene: Hormonal regulation of

- transcription through distinct cAMP response elements. *J. Biol. Chem.* **270**, 15391–15397.
- Mains, R. E., Dickerson, I. M., May, V., Stoffers, D. A., Perkins, S. N., Ouafik, L. H., Husten, E. J., and Eipper, B. A. (1990). Cellular and molecular aspects of peptide hormone biosynthesis. *Front Neuroendocrinol.* **11**, 52–89.
- Nie, Y., Nakashima, M., Brubaker, P. L., Li, Q-L., Perfetti, R., Jansen, E., Zambre, Y., Pipeleers, D., and Friedman, T. C. (2000). Regulation of pancreatic PC1 and PC2 associated with increased GLP-1 in diabetic rats. *J. Clin. Invest.* **105**, 955–965.
- O’Rahilly, S., Gray, H., Humphreys, P. J., Krook, A., Polonsky, K. S., White, A., Gibson, S., Taylor, K., and Carr, C. (1995). Brief report: Impaired processing of prohormones associated with abnormalities of glucose homeostasis and adrenal function. *N. Engl. J. Med.* **333**, 1386–1390.
- Seidah, N. G., Gaspar, L., Mion, P., Marcinkiewicz, M., Mbikay, M., and Chretien, M. (1990). cDNA sequence of two distinct pituitary proteins homologous to Kex2 and furin gene products: Tissue-specific mRNAs encoding candidates for pro-hormone processing proteinases. *DNA Cell. Biol.* **9**, 415–424.
- Smeekens, S. P., and Steiner, D. F. (1990). Identification of a human insulinoma cDNA encoding a novel mammalian protein structurally related to the yeast dibasic processing protease Kex2. *J. Biol. Chem.* **265**, 2997–3000.
- Steiner, D. F., and Oyer, P. (1967). Biosynthesis of insulin and probably precursor of insulin by human islet cell adenoma. *Proc. Natl. Acad. Sci. USA* **57**, 473–480.
- Zhu, X., Zhou, A., Dey, A., Norrbom, C., Carroll, R., Zhang, C., Laurent, V., Lindberg, I., Ugleholdt, R., Holst, J. J., and Steiner, D. F. (2002). Disruption of PC1/3 expression in mice causes dwarfism and multiple neuroendocrine peptide processing defects. *Proc. Natl. Acad. Sci. USA* **99**, 10293–10298.



**Figure 1** Diagram of the human PRL gene (A), mRNA transcript (B), and mature PRL (C). The left and right arrows designate extrapituitary and pituitary transcription start sites, respectively. The numbers designate the exons. Note that extrapituitary PRL is longer in the 5' untranslated region.

than the pituitary counterpart in the 5' untranslated region. Both pituitary and extrapituitary sites produce an identical PRL protein.

The PRL protein is made of a single chain of 199 residues that is stabilized by three intramolecular disulfide bridges. A model of PRL predicts its arrangement in four antiparallel  $\alpha$ -helices organized in an “up–up–down–down” fashion (Fig. 1), a motif that is shared with some hematopoietic factors and interleukins. Posttranslational modifications generate PRL variants that differ in size or functional groups and account for the functional diversity of PRL. Larger forms result from dimerization or aggregation, whereas smaller forms are produced by proteolysis. Some variants retain PRL-like activities, whereas others have unique properties or no known functions. A 22-kDa form is important in female reproduction, whereas a 16-kDa product possesses antiangiogenic activity. Glycosylated PRL is detected in serum, amniotic fluid, and milk at variable ratios, but it is unknown whether it possesses unique physiological functions. PRL can also be phosphorylated on

serine and/or threonine residues, resulting in charge variability and altered function.

## REGULATION

PRL is the primary secretory product of the pituitary lactotroph. The lactotrophs comprise 20 to 30% of total anterior pituitary cells and show significant heterogeneity in morphology, basal hormone release, and responsiveness to secretagogues. Lactotrophs retain proliferative capacity during adulthood, and their numbers increase during pregnancy and lactation. Their proliferative potential accounts for a higher incidence of lactotroph tumors (prolactinomas) than of other types of pituitary tumors. Pituitary PRL is subjected to multiple regulators that are classified into four broad categories: endocrine, paracrine, juxtacrine, and autocrine. Endocrine agents originate from the hypothalamus and gonads, and they reach the lactotrophs via the blood. Paracrine factors reach the lactotrophs by diffusion from neighboring

pituitary cells. Juxtacrine interactions emanate from the extracellular matrix of adjacent cells. Autocrine agents are synthesized by the lactotrophs themselves. Consequently, the overall secretory activity of the lactotrophs reflects a balance between local and distant releasing and inhibiting factors.

Pituitary PRL gene expression is affected by multiple hormones, neurotransmitters, and growth factors. Compounds that bind to G protein-linked receptors (e.g., thyrotropin-releasing hormone [TRH], vasoactive intestinal peptide [VIP], dopamine) activate protein kinase A, protein kinase C, and/or calcium/calmodulin-dependent pathways. These are mediated via a variety of transcription factors that bind to consensus sequences within the PRL promoter. Estrogens, on the other hand, diffuse into the nucleus, where they bind to their receptors. The activated receptors act as transcription factors by binding to an estrogen response element (ERE) located within the distal region of the proximal PRL promoter. Among growth factors, insulin and epidermal growth factor (EGF) stimulate, whereas transforming growth factor- $\beta$  (TGF- $\beta$ ) suppresses, the PRL gene. Many of the growth factors bind to transmembrane receptors with intrinsic tyrosine kinase activity and exert pleiotropic actions such as stimulation of PRL gene transcription, increases in hormone storage, and alterations of lactotroph morphology.

The pituitary lactotroph is unique in its capacity for high constitutive PRL secretion, but its activity is tonically suppressed by hypothalamic dopamine. Dopamine binds to a type II dopamine receptor and exerts multiple actions on the lactotrophs, including lowering of intracellular calcium and cyclic AMP (cAMP) levels, inhibition of PRL gene expression and release, and suppression of cell proliferation. PRL itself, acting via a short-loop negative feedback mechanism, is the primary regulator of the dopaminergic system. Although a singular releasing factor for PRL has not been identified, several neuropeptides, including TRH and VIP, are capable of acute stimulation of PRL release under some conditions.

## RECEPTORS AND SIGNAL TRANSDUCTION

PRL exerts its actions by binding to high-affinity plasma membrane receptors. The gene encoding the human PRL receptor is located on chromosome 5 and contains at least 10 exons encompassing more than 100 kb. The PRL receptor belongs to the

cytokine/GH/PRL receptor superfamily, characterized by a single-pass transmembrane stretch that divides the receptor into an extracellular ligand-binding domain and an intracellular domain. The extracellular domain contains two disulfide bonds and a WS motif (Trp-Ser-x-Trp-Ser) that is required for receptor folding and may participate in the formation of a ligand-binding pocket. The cytoplasmic domain has a proline-rich motif ("box 1") near the plasma membrane that couples to intracellular signaling molecules. Humans express primarily one "long" form of the receptor, whereas rats have three isoforms. However, several receptor isoforms with deleted or truncated sequences have been detected recently in normal and malignant human tissues, although their function remains unknown. The PRL receptor is expressed by most tissues, with the highest expression in the liver and the mammary gland. PRL receptor expression is altered in response to changes in circulating PRL and steroid hormones.

Binding of PRL to its receptor induces sequential receptor dimerization. Two binding sites on the PRL molecule, site 1 (made of helices 1 and 4) and site 2 (made of helices 1 and 3), are required for induction of receptor homodimerization and formation of an active complex. The receptor is devoid of intrinsic tyrosine kinase activity and uses the JAK2-STAT pathway as its main signaling cascade. JAK2 (Janus kinase 2) is rapidly phosphorylated on PRL binding and induces the phosphorylation of the receptor itself, other associated kinases, and STAT (signal transducers and activators of transcription) proteins. Of the seven known STAT proteins, STAT1, -3, and -5 are activated by PRL, with STAT5a and -5b being especially important for mammary gland development and function. The activated STAT proteins dimerize, translocate to the nucleus, and bind to specific sequences on target genes. The ras/raf/MAP kinase cascade and fyn, a member of the Src kinase family that phosphorylates phosphatidylinositol 3-kinase (PI3K), are also activated by PRL in a cell-specific manner. All of the preceding pathways have been characterized under acute conditions, whereas the mechanism mediating long-term effects of PRL are not well understood. In addition, human GH and PL also bind to the PRL receptor and mimic some of the actions of PRL.

## PROLACTIN AS A CYTOKINE

Although PRL was initially considered an exclusive pituitary hormone, many nonpituitary tissues were

**Table 1** Extrapituitary Tissues and Cell Types That Synthesize PRL and Fluid Compartments That Contain Detectable Amounts of PRL in Humans

Tissue	Cell type	Fluid compartment
Brain	Neurons	Cerebrospinal fluid
Thymus	T lymphocyte	
Lacrimal glands	Epithelium	Tears
Lymph nodes	Lymphocytes	
Sweat glands	Epithelium	Sweat
Breast	Epithelium, adipose	Milk
Spleen	B lymphocytes	
Skin	Fibroblasts	
Uterus	Myocytes, stroma	Amniotic fluid
Prostate	Epithelium	

later found to contain immunoreactive PRL. The widespread distribution of tissues capable of PRL synthesis as well as those that contain PRL is illustrated in [Table 1](#). The most established extrapituitary sites that produce PRL in humans are the decidua, immune system, brain, and myometrium, with emerging evidence for PRL synthesis by the skin and exocrine glands, including mammary, sweat, and lacrimal. PRL is produced by a wide variety of cells of different embryonic origin, morphology, and physiological functions. Some (e.g., lymphocytes, epithelia) are less differentiated and have a high proliferative capacity, whereas others (e.g., neurons) are postmitotic and terminally differentiated. Immunoreactive PRL is also present in tissues that do not produce PRL but are capable of concentrating PRL from the blood.

Another remarkable feature of PRL is its presence in many body fluid compartments. Although all hormones are present in serum and most are excreted into urine, PRL is also found in cerebrospinal fluid (CSF) and the amniotic fluid and is also secreted into milk, tears, and follicular fluid. Whereas PRL in the amniotic fluid originates from a local source (decidua), PRL in milk and in CSF is derived from both locally produced and circulating PRL. Significant cellular resources must be spent in transporting PRL into these compartments, yet little is known about the functions subserved by PRL in these sites. In contrast to the pituitary, little is known about the regulation of PRL production/release in extrapituitary sites except that they do not respond to dopamine, neuropeptides, or estrogens but are regulated by local autocrine/paracrine factors. A superdistal promoter upstream of exon 1a regulates PRL gene expression in extrapituitary sites and is silenced in the pituitary gland ([Fig. 1](#)).

## BIOLOGICAL FUNCTIONS

PRL is one of the most versatile hormones, fulfilling more than 100 functions and thereby surpassing the number of known actions of all other pituitary hormones combined. PRL functions are associated with reproduction, growth/development, osmoregulation, metabolism, immune regulation, and brain function/behavior. The effects of PRL on reproduction involve multiple systems and tissues that differ in their importance in a species-specific manner. The principal target for PRL is the mammary gland, where it promotes growth and differentiation of the lobuloalveolar structures and is essential for the initiation and maintenance of lactation and for the production of milk proteins. However, in some species, continuous lactogenesis is supported by GH rather than by PRL. In rodents, PRL has both luteotropic and luteolytic actions on the ovary and supports progesterone production. Therefore, in these species, PRL plays a major role in modulating the reproductive cycle and is crucial for pregnancy and lactation. PRL also has well-established mitogenic, secretory, and morphogenic effects on the prostate.

PRL plays a major role in regulating water and electrolyte balance in fish and amphibians, with lesser osmoregulatory actions in birds and mammals. Whereas the control of development and body growth is normally ascribed to GH, there are some functional overlaps between GH and PRL, especially in lower vertebrates. PRL stimulates proliferation of the pancreatic islets and increases insulin secretion primarily during pregnancy. In the immune system, PRL induces proliferation and differentiation of functional activity of various lymphoid cells, but transgenic animals lacking PRL or its receptor have little if any immune disturbances. Within the brain, PRL affects the production and release of several hypothalamic releasing/inhibiting hormones and has significant effects on maternal behavior.

## PHYSIOLOGY AND PATHOPHYSIOLOGY IN HUMANS

Consistent with PRL's function as an adaptive hormone rather than an indispensable one, the profile of PRL release varies under many physiological conditions and is dissimilar among species. There are no sex differences in serum PRL concentrations in children, but PRL levels are twice as high in women than in men after puberty. Adult women do not exhibit marked changes in the PRL secretory profile throughout the menstrual cycle. This is in contrast to rodents, where a clear preovulatory surge of PRL is evident. Stress

conditions, including anesthesia, surgery, electric shock, exercise, and insulin-induced hypoglycemia, also stimulate PRL release in both men and women. PRL release progressively rises throughout gestation, and its elevated levels are essential for the preparation of the breast for the initiation of lactation, which is prevented during pregnancy by progesterone. Pregnancy is also associated with lactotroph hypertrophy and hyperplasia, accounting for enlargement of the pituitary. During pregnancy, large amounts of PRL are independently produced by the decidua, and the amniotic fluid contains large concentrations of PRL whose function remains unknown. Suckling is the most potent physiological stimulus for PRL release. The magnitude of the suckling-induced PRL rise is robust during early lactation but wanes thereafter. Tactile stimuli of the breast can increase serum PRL in nonlactating women but not in men.

Under normal conditions, PRL has permissive effects on human reproduction. However, PRL overproduction (hyperprolactinemia) is a pituitary-related disorder that interferes with reproductive processes. Hyperprolactinemia results from several causes, including PRL-secreting tumors and therapy with certain dopaminergic-blocking drugs. Common manifestations of hyperprolactinemia include amenorrhea (lack of menstrual cycles) and galactorrhea (inappropriate milk production) in women and hypogonadism, decreased sex drive, and impotence in men. The inhibitory effects of excess PRL occur at both central (hypothalamic-pituitary) and peripheral (gonadal) sites. Estrogens are responsible for a higher incidence of prolactinomas in women but have little acute effects on PRL release in humans, in contrast to their effects in rats. Oral contraceptives do not increase serum PRL levels appreciably and do not contribute to prolactinoma initiation or progression. Lactational amenorrhea, used by some postpartum women as a method of contraception, appears to be linked to the frequency and duration of suckling episodes rather

than to a persistent hyperprolactinemia. Recent evidence indicates that PRL may play a role as a mitogen/antiapoptotic factor in breast and prostatic cancer.

### See Also the Following Articles

Acromegaly, Diagnosis of • Pituitary Lactotroph Disorders • Prolactin, Evolution of • Prolactinoma, Clinical Manifestations • Prolactinoma, Diagnosis • Prolactinoma, Pathogenesis • Prolactinoma, Therapy

### Further Reading

- Asa, S. L., and Ezzat, S. (1998). The cytogenesis and pathogenesis of pituitary adenomas. *Endocrine Rev.* **19**, 798–827.
- Ben-Jonathan, N., and Hnasko, R. M. (2001). Dopamine as a prolactin inhibitor. *Endocrine Rev.* **22**, 724–763.
- Ben-Jonathan, N., Mershon, J. L., Allen, D. L., and Steinmetz, R. W. (1996). Extrapituitary prolactin: Distribution, regulation, functions, and clinical aspects. *Endocrine Rev.* **17**, 639–669.
- Blackwell, R. E. (1992). Hyperprolactinemia: Evaluation and management. *Reprod. Endocrinol.* **21**, 105–124.
- Bole-Feysot, C., Goffin, V., Edery, M., Binart, N., and Kelly, P. A. (1998). Prolactin (PRL) and its receptor: Actions, signal transduction pathways, and phenotypes observed in PRL receptor knockout mice. *Endocrine Rev.* **19**, 225–268.
- Dutt, A., Kaplitt, M. G., Kow, L. M., and Pfaff, D. W. (1994). Prolactin, central nervous system, and behavior: A critical review. *Neuroendocrinology* **59**, 413–419.
- Freeman, M. E., Kanyicska, B., Lerant, A., and Nagy, G. (2000). Prolactin: Structure, function, and regulation of secretion. *Physiol. Rev.* **80**, 1523–1631.
- Gourdji, D., and Laverriere, J. N. (1994). The rat prolactin gene: A target for tissue-specific and hormone-dependent transcription factors. *Mol. Cell. Endocrinol.* **100**, 133–142.
- Horseman, N. D. (2001). "Prolactin." Kluwer, Boston.
- Hu, Z.-Z., Zhuang, L., and Dufau, M. L. (1998). Prolactin receptor gene diversity: Structure and regulation. *Trends Endocrinol. Metab.* **9**, 94–102.
- Sinha, Y. N. (1995). Structural variants of prolactin: Occurrence and physiological significance. *Endocrine Rev.* **16**, 354–369.
- Vonderhaar, B. K. (1999). Prolactin involvement in breast cancer. *Endocrinol. Related Cancer* **6**, 389–404.

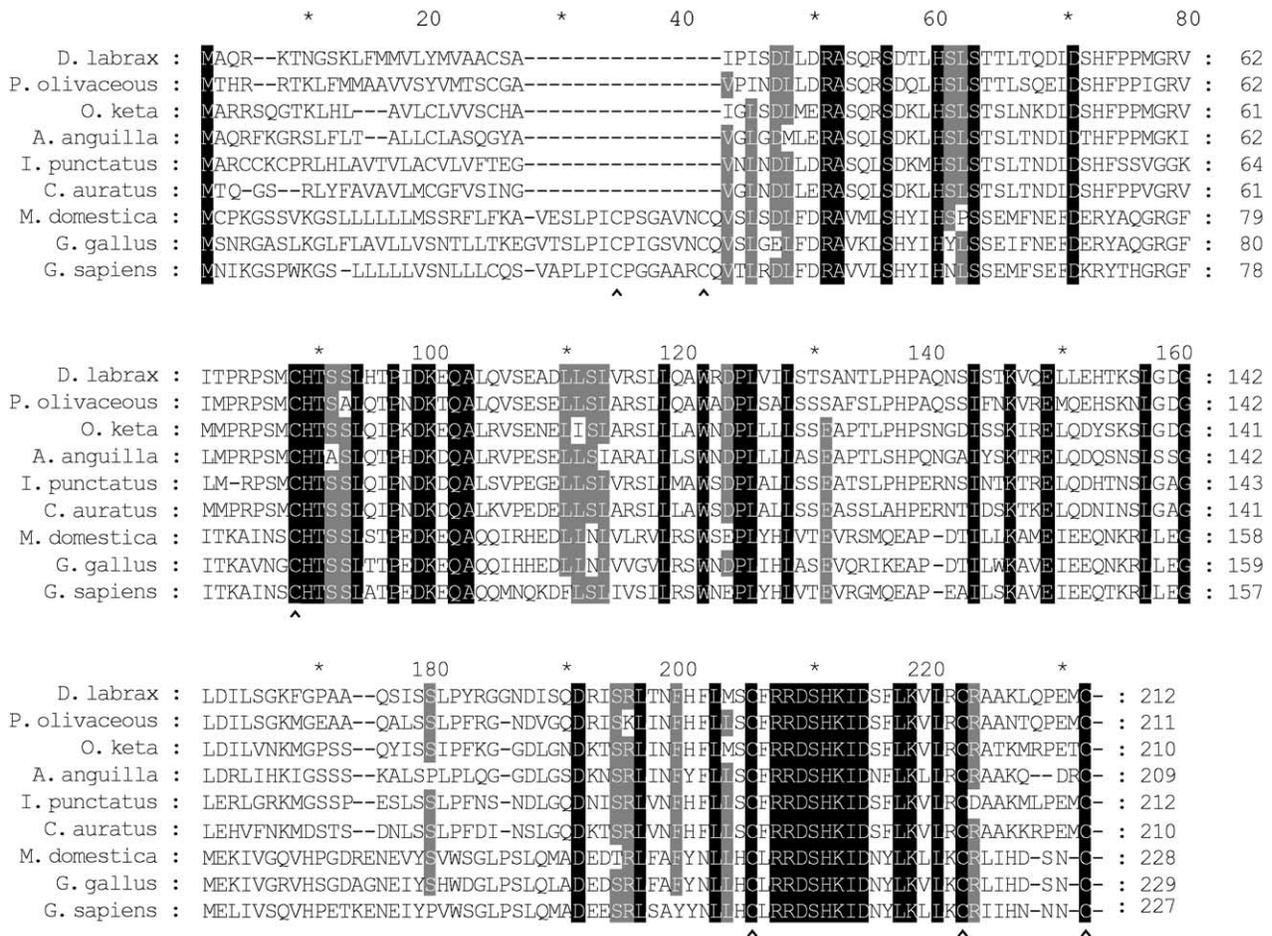
teleost PRL are in equivalent positions (amino acid 46, 160, 177, and 187) to those in the mature tetrapod protein. The difference in structure, caused by the absence of one disulfide bridge, has been proposed to be one of the factors limiting biological activities of PRL in fish compared with those in mammals.

There is a variable degree of sequence homology between PRL molecules in various vertebrate species that tends to reflect their phylogenetic relationship. For example, sequence similarity between fish and higher vertebrate PRL is low at about 25 to 30%, compared with 60 to 95% sequence similarity shared between PRLs of mammals. Comparison of the primary sequence of PRL from representative species of various vertebrates reveals that conserved residues are

found along the entire length of the molecule (Fig. 1). In the mammals and birds, the conserved residues appear to cluster into four relatively distinct regions that are separated by regions of lower homology. Inclusion of PRL from fishes in such analysis results in the identification of only two relatively well-conserved segments of nearly identical or chemically similar sequence.

### PROLACTIN GENES

The GH/PRL/PL/SL gene family has arisen as a consequence of a fairly complex evolutionary process. All of these molecules are proposed to have evolved from a common ancestral gene that underwent at least



**Figure 1** Multiple sequence alignment of the amino acid sequence of PRL, including the signal sequence, from a mammal (*Homo sapiens*), a bird (*Gallus gallus*), a marsupial (*Monodelphis domestica*), and six teleost fishes (*Paralichthys olivaceus*, *Dicentrarchus labrax*, *Oncorhynchus keta*, *Anguilla anguilla*, *Ictalurus punctatus*, and *Carassius auratus*). The conserved cysteine residues are indicated by an arrowhead; totally conserved amino acids in all of the sequences are highlighted by a black background; and conservative substitutions are indicated by a gray background. Note that there are two principal regions of sequence conservation between the fish and the tetrapods.

two successive rounds of gene duplication between 500 and 400 million years ago. Primate PLs belong to the GH lineage, whereas rodent and ungulate PLs are members of the PRL lineage. Additional family members have been identified in several mammals and include mouse proliferin, bovine PRL-related cDNA I, and rat PRL-like protein A.

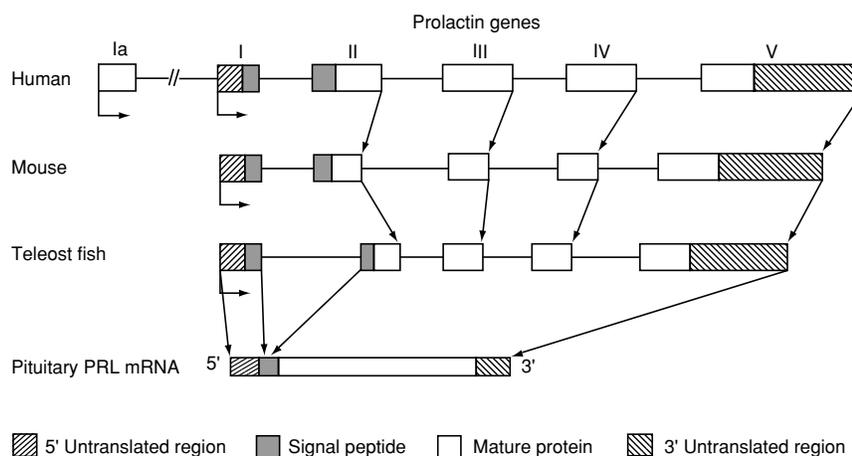
## Gene Organization

A single gene for PRL has been identified in most mammals, although the bovine genome has been shown to contain several genes that encode PRL-like proteins. In addition, in teleosts such as the salmonids and cyprinids, as a consequence of their tetraploid/polyloid nature, at least two PRL genes are known to exist. The PRL gene is present as a single copy on chromosome 6 in humans, chromosome 17 in rats, and chromosome 13 in mice. The chromosome localization of PRL has not been reported for any other species. The PRL gene has been characterized in several mammals and teleosts and is composed of 5 exons and 4 introns, with the human gene being the exception by having an additional exon at the 5' end of the gene that is designated exon Ia. PRL gene organization in the vertebrates has been well conserved, as have the position and type of each exon/intron splice site. In all vertebrate genes characterized, exon I encodes the 5' untranslated region and part of the signal sequence, exon II encodes the remainder of the signal peptide and the N terminus of the mature protein, exons III and IV encode the mature peptide, and

exon V encodes the C-terminal region of the hormone and the 3' untranslated region. In the human gene, the additional noncoding exon, Ia, contains an alternative transcriptional start site whose use is placenta specific and so likely to be a more recent evolutionary development.

The molecular mechanisms underlying the pituitary-specific transcription of the PRL gene in tetrapods has been studied extensively. Two distinct, highly conserved regions are involved in this process: (1) an activating region immediately upstream of the TATAA box and (2) an enhancer region between  $-1.8$  and  $-1.5$  kb distal to the transcription start site. The transcription factor PIT1/GHF1 binds at several sites in both of these regions of the PRL gene and regulates its expression, although binding of other factors is proposed to be required for full promoter activity. Far fewer studies exist characterizing transcriptional activation of teleost PRL genes, and although considerable sequence divergence is found in the promoter region of the fish and mammalian gene transcription is regulated by the same transcription factor. In fact, the sequence conservation between the mammalian and fish PIT1/GHF1 transcription factor is far higher than that found for PRL, and it is also highly conserved both structurally and functionally.

Figure 2 shows a schematic representation (not to scale) of the genomic organization of human, mouse, and teleost fish PRL genes and the respective mRNA transcripts. Exons are represented by numbered boxes, and introns are represented by a line between exons. Arrows between the human, rat, and teleost genes indicate equivalent exons, and the contributions



**Figure 2** Schematic representation (not to scale) of the genomic organization of human, mouse, and teleost fish PRL genes and mRNA transcript. Exons are represented by numbered boxes, and introns are represented by a line between exons. Arrows among the human, rat, and teleost genes indicate equivalent exons, and the contributions of the exons to the mature pituitary PRL mRNA transcript are indicated.

of the exons to the mature pituitary PRL mRNA transcript are indicated.

## EVOLUTION OF PROLACTIN FUNCTIONS

Precursor molecules for GH and PRL probably existed in ancestors of the chordates, and a distinct PRL molecule has not been isolated from agnathans; instead, cells in the proximal pars distalis react with antisera to both fish GH and PRL. The specialization of pituitary cells secreting PRL occurred first in teleost fishes, but in higher vertebrates there is a mosaic of cells producing both GH and PRL as well as each individual hormone.

The functions of PRL fall into the broad categories of hydromineral balance, growth and energy metabolism, development, reproduction, immunomodulation, and behavior, but it is clear that PRL, through its multiple interactions, is an important coordinating factor for seasonal and reproductive cycles as well as for survival.

### Hydromineral Balance

Probably one of the earliest functions of PRL in vertebrates concerned ion and water homeostasis in aquatic species. In freshwater and euryhaline species of fishes, PRL ensures survival in water of low ionic and osmotic content by reducing sodium efflux and water influx. These effects result from impermeabilization and increases in mucus secretion of skin, gill, and buccal epithelia and a reduction in gill  $\text{Na}^+ \text{K}^+$  ATPase activity. In the fish kidney, PRL stimulates  $\text{Na}^+ \text{K}^+$  ATPase activity, urine flow, glomerular size, and sodium absorption. In the gut, water, ion absorption, and  $\text{Na}^+ \text{K}^+$  ATPase activity are reduced by PRL, whereas mucus production is stimulated. In the urinary bladder, water absorption is reduced, sodium absorption and  $\text{Na}^+ \text{K}^+$  ATPase activity are increased, and calcium retention may be promoted.

In the aquatic larval stages of Anuran amphibians and in newts (Urodeles), PRL has functions similar to those in euryhaline fishes in controlling sodium and water fluxes across permeable epithelia such as the skin and bladder. There is a lack of information about a possible role for PRL in hydromineral balance in reptiles, and a postulated effect stimulating salt secretion from nasal salt glands of birds and reptiles has not been substantiated unequivocally.

There is only fragmentary evidence of a role for PRL in hydromineral physiology of mammals in that

it affects movement of sodium, calcium, and chloride ions as well as water across intestinal membranes; however, it may be especially important for uptake of minerals and amino acids across mammary epithelial cells for incorporation into milk. In addition, milk contains PRL that may be important for stimulation of calcium uptake across small intestine epithelia in the suckling young. PRL affects water movement across amniotic membranes, being stimulatory in pig and sheep but inhibitory in humans.

Thus, the physiological actions of PRL that are important for survival in low osmotic aquatic environments have been adapted particularly for use in epithelia concerned with reproduction and nurture of offspring.

### Growth, Development, and Nurture

The similarities between GH and PRL of fishes suggested that PRL is also a growth-promoting hormone, and both Tilapia PRLs (tPRL177 and tPRL188) have shown growth-promoting effects (although they are measured by different criteria). In fish (Tilapia) and mammals (rat), PRL promotes insulin-like growth factor-1 (IGF-1) production by the liver, stimulating sulfate uptake into cartilage. Several epidermal tissues respond to PRL by cell proliferation, including melanocytes in fish and mammals, epidermal tissues in molting reptiles and amphibians, and keratinocytes and hair in mammals.

PRL is regarded as a postembryonic developmental hormone, particularly during metamorphosis of amphibian vertebrates, where it prevents thyroid hormone stimulation of both apoptosis of tail cells and morphogenesis of limbs; thus, the PRL that allows survival of aquatic stages of amphibian development opposes metamorphic progression. In mammalian development, deciduum-derived PRL in the amniotic fluid reduces fluid volume by transfer of water but may also enter the fetal circulation and affect lung and immune system maturation. In the mammalian hypothalamus, PRL influences maturation of the neonatal neuroendocrine system. PRL has been shown to affect development of bone in mice, specifically to inhibit endochondral ossification *in vivo* and to depress alkaline phosphatase activity of osteoblasts *in vitro*.

Secretion of PRL is coordinated with cycles that may be diurnal or seasonal. In fish, these are related to metabolic responses. In general, fish acclimated to long photoperiods respond to PRL injections early during the light period by reducing lipid stores and increasing nonlipid tissue accumulation; however,

similar treatments midway through the light period cause fat deposition. In rats, there is a diurnal cycle of PRL release; however, in mature females, the surge of PRL release during proestrus is linked to the development of the corpus luteum that follows the fall in PRL release.

A role for PRL in fat metabolism has also been established in birds. For example, in penguins, which fast while incubating eggs, circulating PRL levels remain high until fat stores are critically reduced when increased corticosterone production causes a fall in plasma PRL; at this point, penguins return to the sea to replenish the lipid energy store. In mice, leptin from adipose tissue may control release of neural PRL-releasing peptide (PrRP), resulting in reduced food intake and increased body weight; short-term treatments of adipocytes *in vitro* with PRL reduce insulin-stimulated leptin production. In fetal sheep, PRL is required for the up-regulation of the specific uncoupling protein-1 (UCPI) that enables rapid production of heat from brown adipose tissue after birth. So, PRL can affect metabolism of both white and brown adipose tissues, demonstrating conservation of roles in lipid metabolism among fish, birds, and mammals.

PRL effects on epithelia can also be directly important for nutrition of offspring. For example, in cichlid fish, PRL stimulates mucus production by the epidermis and the young fry feed on this; interestingly, the mucus also contains PRL and GH. The crop sac mucosa of pigeons responds to tetrapod PRL by extreme hyperplasia, forming a cheesy mass on which the offspring feed. Regurgitation of food for the offspring is also dependent on PRL. In most mammals, PRL is necessary for mammary gland development, growth, and lactogenesis, although this is not the case in goats and cows.

Thus, in fish, birds, and mammals, PRL is important for providing essential nutrients on which the offspring depend for survival during early development.

## Reproduction

The functions of PRL in controlling and modulating reproduction are numerous but are only clearly defined in mammals. They include growth and differentiation of the mammary gland, luteotropic and luteolytic actions in the ovary, stimulation of gonadotropin receptors in the testis, and stimulation of prostate epithelium activities. From reported studies, it appears that roles for PRL in lower vertebrate reproduction are more closely related to behavioral responses than to

specific interactions with reproductive tissues. From an evolutionary viewpoint, PRL may have been important for reproduction of marine fishes; when the crossopterygian ancestors migrated to fresh water, PRL, already functioning as a reproductive hormone, would have been important for their survival in the low sodium/low osmotic environment.

## Behavior

The effects of PRL on behavior are frequently correlated with successful reproduction; in fish, eggs and fry are protected by fin fanning and PRL stimulates building a nest of foam from secreted mucus by paradise fish. In birds, PRL encourages nesting and nest building, nest attendance, and egg incubation. PRL is needed for the migration of salamanders (red eft) to return to water for breeding and for the movement of sticklebacks to fresh water to breed. Maternal behavior in mammals, including feeding offspring and grooming, may also be dependent on PRL.

## Immunomodulation

PRL has multiple roles in the mammalian immune system through interactions in the spleen and thymus and various differentiated lymphocytes and macrophages. Anemia, leukopenia, and thrombocytopenia caused by hypophysectomy all are reversed by PRL replacement therapy. Macrophages, which are immune cells of ancient lineage, are activated by PRL in both fish and mammals. Augmentation of immune responses by PRL may be due to its antiapoptotic action in lymphocytes. The source of the PRL, acting as a paracrine or autocrine factor, may be leucocytes themselves given that granulocytes contain a high molecular weight PRL of 43 kDa.

## Evolution of Prolactin Function

The brief overview of PRL functions suggests a broad outline of the evolution of the wide range of PRL functions throughout the vertebrates. The initial role, which may well predate the chordates, was probably as a responsive factor to osmotic changes due to variations in environmental ion concentrations of PRL-secreting cells located at interfaces. The specialization of cells of the pituitary to secrete PRL developed during fish evolution as direct contact with the environment decreased, although some fish retain direct environmental contact via the orohypophyseal duct. Nevertheless, fish pituitary PRL cells remain sensitive

to plasma and medium osmotic pressures, resulting in control of hydromineral balance, particularly sodium ion concentrations. This function persists into mammals, where ion uptake into mammary epithelium and extrusion into milk are controlled by PRL. The functions of PRL in the immune system and in bone cells are likely to have evolved from the head kidney cells of teleost fishes; these cells are lymphopoietic and hemopoietic and are considered to be the evolutionary precursors of bone marrow in higher vertebrates.

## CONTROL OF PROLACTIN SECRETION

Because PRL is a very important coordinating factor between an organism and its environment, control of its secretion by the pituitary is multifactorial, including environmental cues such as light, scent, sound, external osmotic pressure, and ion content that interact with factors from the brain (both neurotransmitter and neuroendocrine) and from reproductive organs (especially the ovary).

The most ancient controlling factor for release and synthesis of PRL is probably the external osmotic pressure that would operate in isolated cells of simple metazoa. Against this background of a basic physicochemical control mechanism, numerous external and internal agents have been identified as influencing PRL secretion by pituitary cells. In euryhaline teleosts, low external osmotic pressure, generally due to low sodium ion content, is the most potent stimulating agent for PRL secretion.

However, throughout the vertebrates, pituitary PRL cells are principally under tonic inhibitory control by dopamine from the tuberoinfundibular dopaminergic (TIDA) neurons of the hypothalamus, and when removed from hypothalamic connection, pituitaries continue to secrete PRL—although, somewhat perversely, very low dopamine concentrations (e.g.,  $10^{-12}$  M) are stimulatory. Changes in the pattern of dopamine release from the TIDA neurons in rats correlate with the cyclical peaks and troughs of plasma PRL concentrations. A principal stimulatory factor of both PRL secretion and PRL cell hyperplasia is estrogen, which again affects PRL cells throughout the vertebrates. Other stimulatory agents are mainly hypothalamic hypophysiotropic factors, including thyrotropin-releasing hormone (TRH), vasoactive intestinal polypeptide (VIP), somatostatin, and gonadotropin-releasing hormone (GnRH) in some fish, as well as neurotransmitters, including  $\gamma$ -amino butyric acid (GABA), endorphins, and enkephalins. Two

specific PrRPs, PrRP20 and PrRP 31, have been identified in brains of rats, cattle, and humans, and a homologue of PrRP20 has been identified in carp (Ca-RF) and tilapia (tPrRP), suggesting a conserved function throughout the vertebrates.

Circadian rhythms of PRL secretion in mammals are controlled by neurotransmitter/neuroendocrine inputs from the hypothalamus and correlate with dopamine release, but light patterns control cycles (e.g., in the rat). Extrapituitary PRL secretion appears to be influenced most strongly by steroids such as progesterone (e.g., in the deciduum and estrogens in the brain), but more data are required for comprehensive understanding of these tissue systems. Steroids are ancient “bioregulators” originally found in surrounding water of early life forms, and their interaction with PRL-secreting cells in extant vertebrates was probably established a very long time ago—before the first vertebrates evolved.

## CONCLUSION

PRL genes and proteins have been modified to only a limited degree during evolution from fishes to mammals, but their functions have been adapted for physiological systems as vertebrates progressed from aquatic to terrestrial and aerial forms. However, the earliest precursor genes and peptides for PRL will have to be sought in genomes and tissues of primitive chordates and invertebrates to understand their origins and evolution to the versatile and important endocrine and paracrine factor of vertebrates.

## See Also the Following Articles

ACTH,  $\alpha$ -MSH, and POMC, Evolution of • Cytokines, Evolutionary Aspects and Functions • Insulin and Insulin-like Growth Factors, Evolution of • Natriuretic Peptide System, Evolution of • Prolactin (PRL) • Prolactinoma, Clinical Manifestations • Prolactinoma, Diagnosis • Prolactinoma, Pathogenesis • Prolactinoma, Therapy • Somatostatin, Evolution of • Steroid Receptors, Evolution of

## Further Reading

- Ben-Jonathon, N., Mershon, J. L., Allen, D. L., and Steinmetz, R. W. (1996). Extrapituitary prolactin: Distribution, regulation, functions, and clinical aspects. *Endocrine Rev.* **17**, 639–669.
- Bole-Feysot, C., Goffin, V., Edery, M., Binart, N., and Kelly, P. A. (1998). Prolactin (PRL) and its receptor: Actions, signal transduction pathways, and phenotypes observed in PRL receptor knockout mice. *Endocrine Rev.* **19**, 225–268.
- Freeman, M. E., Kaniyaska, S., Lerant, A., and Nagy, G. (2000). Prolactin: Structure, function, and regulation of secretion. *Physiol. Rev.* **80**, 1523–1631.

Nicoll, C. S., Mayer, G. L., and Russel, S. M. (1986). Structural features of prolactins and growth hormones that can be related to their biological properties. *Endocrine Rev.* 7, 169–203.

Ono, S. (1970). "Evolution by Gene Duplication." Springer-Verlag, Heidelberg, Germany.

Rand-Weaver, M., Kawauchi, H., and Ono, M. (1993). Evolution of the structure of the growth hormone and prolactin family. *In* "The Endocrinology of Growth, Development, and Metabolism in Vertebrates" (M. P. Schreibman, C. G. Scanes, and P. K. T. Pang, eds.), pp. 13–42. Academic Press, New York.

The presence of even minute amounts of milk expressible from one or both breasts justifies the diagnosis of galactorrhea. Its persistence for more than 1 year after normal delivery and cessation of breast-feeding, or its occurrence in the absence of pregnancy, generally is taken as a definition of inappropriate lactation.

The incidence of galactorrhea has been variously reported in normal women as ranging from 1 to 45% of women tested. This variability probably is the result of differences in the techniques used to express milk from the breast and the way in which nonmilky secretions are classified. Lowering the blood prolactin (PRL) level in patients with normal or elevated PRL levels almost always will lead to a marked decrease in, or an abolition of, lactation.

Galactorrhea may be present in approximately 5 to 10% of normally menstruating women, and recent experience has shown that basal PRL levels are normal in more than 90% of these women. When amenorrhea or oligomenorrhea is associated with galactorrhea, approximately 75% of these women will be found to have hyperprolactinemia.

Galactorrhea in men has been reported in 10 to 20% of cases and is virtually pathognomonic of a prolactinoma.

## FEMALE REPRODUCTIVE FUNCTION

Hyperprolactinemia has been found in most studies to suppress luteinizing hormone (LH) pulsatile secretion by decreasing pulse amplitude and frequency. With menopause, hyperprolactinemia can prevent the expected rise in gonadotropins; normalization of PRL levels with bromocriptine results in elevation of gonadotropin levels and hot flashes. The pituitary gonadotroph response to gonadotropin-releasing hormone (GnRH) in hyperprolactinemia has generally been found to be normal, increased, and decreased in humans. Hyperprolactinemia in women has been associated with loss of positive estrogen feedback on gonadotropin secretion. High concentrations of PRL are inhibitory to progesterone and estrogen production by the ovary. PRL can inhibit estrogen formation by antagonizing the stimulatory effects of follicle-stimulating hormone (FSH) on aromatase activity, and direct inhibition of aromatase synthesis has been shown.

Although the amenorrhea caused by hyperprolactinemia usually is secondary, it also can be primary if the disorder begins before the usual age of puberty. In patients with primary amenorrhea due to hyperprolactinemia, estrogen deficiency and failure to develop

normal secondary sexual characteristics may be the presenting problem. Galactorrhea is variable in this setting because the breast might not have been exposed to appropriate priming with estrogen and progesterone. Patients with primary amenorrhea tend to have macroadenomas more commonly than do those with secondary amenorrhea, although the reasons for this are unknown.

A short luteal phase is the first evidence of interference in the normal cycle by hyperprolactinemia. Subsequently, ovulation is inhibited and oligomenorrhea occurs. Finally, amenorrhea may take place. In a summary of 18 series of 1409 women with prolactinomas undergoing transsphenoidal surgery, the frequency of oligomenorrhea was 94.0% and that of galactorrhea was 85.9%.

Infertility may be a presenting symptom of patients with hyperprolactinemia and is invariable when gonadotropin levels are suppressed with anovulation. In three series of women (a combined 367 cases) studied for infertility, one-third were found to have hyperprolactinemia. Most of the women had amenorrhea and galactorrhea as well, but in one series of 113 cases of infertility, 5 of the 22 hyperprolactinemic women had neither amenorrhea nor galactorrhea. That PRL excess may be important in these types of patient is suggested by the finding that treatment of similar patients with bromocriptine restored fertility. In some of these women, transient hyperprolactinemia lasting for 1 to 2 days during the cycle can be documented; this subset usually responds to bromocriptine with increased progesterone during the luteal phase as well as improved fertility.

Reduced libido and orgasmic dysfunction are found in most hyperprolactinemic amenorrheic women when such complaints are specifically elicited. Reduction in PRL levels to normal restores normal libido and sexual function in most of these women.

PRL levels have been found to be elevated in 19 to 50% of women with polycystic ovaries (PCOs). Bromocriptine treatment of hyperprolactinemic patients with PCOs usually results in a reduction of testosterone and LH levels and resumption of ovulatory cycles. Why many patients with PCOs have hyperprolactinemia is not clear. It has been hypothesized that the increased estrogen levels found in PCOs stimulate increased PRL secretion; however, no correlation has been found between estrone levels and PRL levels in these patients. The increased androgen secretion associated with PCOs may cause hirsutism and other excessive androgen action, such as acne and seborrhea, although these features may also be caused by excessive androgen secretion by the adrenal.

## MALE REPRODUCTIVE FUNCTION

In hyperprolactinemic men, there is a decrease in the pulsatile secretion of LH and FSH, and testosterone levels are low or in the lower part of the normal range. The testosterone response to stimulation with human chorionic gonadotropin (hCG) has been reported to be both decreased and normal; in those with decreased responses, there is improvement in the response when PRL levels are lowered with bromocriptine. If there is sufficient normal pituitary tissue, reduction in elevated PRL levels to normal usually results in a return of normal testosterone levels. Although some studies have suggested that drug-induced elevated PRL levels cause a partial block in the enzyme 5 $\alpha$ -reductase, resulting in a decrease in dihydrotestosterone (DHT) levels, this has not been found in studies of men with prolactinomas. Testosterone therapy of hyperprolactinemic men does not always correct the impotence until PRL levels are brought down to normal. Whether this is due to a decrease in DHT levels has not been verified directly.

Chronic hyperprolactinemia in males results in impotence and decreased libido in more than 90% of cases. Other findings of hypogonadism, such as decreased beard growth and decreased strength, are less common.

Sperm counts and motility are decreased with an increase in abnormal forms, and histological studies reveal abnormal seminiferous tubule walls and altered Sertoli cell ultrastructure. The semen analysis does not always return to normal despite a return to normal of testosterone levels with restoration of normoprolactinemia.

A number of surveys have attempted to assess the frequency of hyperprolactinemia among men with complaints of impotence or infertility. Between 2 and 25% of males with impotence have been found to be hyperprolactinemic in various series. However, only 1 to 5% of men with infertility have been found to be hyperprolactinemic.

## EFFECT OF PROLACTIN ON BONES

Hyperprolactinemic women have a decreased bone mineral density, but whether this effect is mediated by estrogen deficiency or is a direct effect of the hyperprolactinemia is controversial. Correction of the hyperprolactinemia results in an increase in bone mass. Studies of hyperprolactinemic women who were not amenorrheic and hypoestrogenic have shown that their bone mineral density is normal, confirming the initial hypothesis that it is the estrogen deficiency that mediates the bone mineral loss. A similar, androgen-dependent loss of bone mineral that is reversible with reversal of the hypoandrogenic state is found in hyperprolactinemic men. However, as determined by recent studies in men with an aromatase gene defect, it is likely that the osteopenia in the hyperprolactinemic, androgen-deficient men is due to reduced estrogen levels as well.

### See Also the Following Articles

Prolactin, Evolution of • Prolactin (PRL) • Prolactinoma, Diagnosis • Prolactinoma, Pathogenesis • Prolactinoma, Therapy

### Further Reading

- Colao, A. M., Loche, S., Cappa, M., *et al.* (1998). Prolactinomas in children and adolescents: Clinical presentation and long-term follow-up. *J. Clin. Endocrinol. Metab.* **83**, 2777–2780.
- Davis, J. R. E., Sheppard, M. C., and Heath, D. A. (1990). Giant invasive prolactinoma: A case report and review of nine further cases. *Q. J. Med.* **74**, 227–238.
- Greenspan, S. L., Oppenheim, D. S., and Klibanski, A. (1989). Importance of gonadal steroids to bone mass in men with hyperprolactinemic hypogonadism. *Ann. Intern. Med.* **110**, 526–531.
- Klibanski, A., Biller, B. M. K., Rosenthal, D. I., Schoenfeld, D. A., and Saxe, V. (1988). Effects of prolactin and estrogen deficiency in amenorrheic bone loss. *J. Clin. Endocrinol. Metab.* **67**, 124–130.
- Ramot, Y., Rapoport, M. J., Hagag, P., and Wysenbeek, A. J. (1996). A study of the clinical differences between women and men with hyperprolactinemia. *Gynecol. Endocrinol.* **10**, 397–400.

**Table 1 Etiologies of Hyperprolactinemia**


---

<b>Pituitary disease</b>
Prolactinomas
Acromegaly
Empty sella syndrome
Lymphocytic hypophysitis
Cushing's disease
<b>Hypothalamic disease</b>
Craniopharyngiomas
Meningiomas
Dysgerminomas
Nonsecreting pituitary adenomas
Other tumors
Sarcoidosis
Langerhans cell histiocytosis
Neuraxis irradiation
Vascular
Pituitary stalk section
<b>Medications</b>
Phenothiazines
Haloperidol
Atypical antipsychotics
Monoamine-oxidase inhibitors
Tricyclic antidepressants
Reserpine
Methyl dopa
Metoclopramide
Amoxepin
Cocaine
Verapamil
Serotonin reuptake inhibitors
Protease inhibitors
<b>Neurogenic</b>
Chest wall lesions
Spinal cord lesions
Breast stimulation
<b>Other</b>
Pregnancy
Hypothyroidism
Chronic renal failure
Cirrhosis
Pseudocyesis
Adrenal insufficiency

---

Tricyclic antidepressants cause modest hyperprolactinemia in approximately 25% of patients. Monoamine oxidase inhibitors may also cause a minimal elevation of PRL levels. These drugs likely facilitate several possible stimulatory pathways, and their effects on dopamine are uncertain. Serotonin reuptake

inhibitors, by increasing synaptic serotonin levels, rarely cause hyperprolactinemia.

Chronic opiate abuse is associated with mild hyperprolactinemia and menstrual dysfunction. Cocaine abuse can also cause chronic mild hyperprolactinemia.

### Antihypertensive Drugs

Alpha-methyl dopa causes moderate hyperprolactinemia by reducing the level of dopamine that reaches the pituitary by inhibiting the enzyme l-aromatic amino acid decarboxylase, which is responsible for converting L-dopa to dopamine, and by acting as a false neurotransmitter to decrease dopamine synthesis. Reserpine, now a little-used antihypertensive drug, causes hyperprolactinemia by interfering with the storage of hypothalamic catecholamines in secretory granules.

Verapamil increases basal PRL secretion acutely and chronically, and patients have been described with galactorrhea associated with sustained hyperprolactinemia. In a survey of patients taking verapamil, PRL levels were found to be elevated in 8.5% of these patients. Verapamil raises PRL by blocking the hypothalamic generation of dopamine by blocking N-type calcium channels. Other classes of calcium channel blockers (e.g., dihydropyridines, benzothiazepines) do not increase PRL levels.

### Protease Inhibitors

Several patients have been reported with galactorrhea and hyperprolactinemia due to protease inhibitors. The frequency of this phenomenon and the mechanism by which this occurs are unknown.

## PATHOLOGICAL CONDITIONS CAUSING HYPERPROLACTINEMIA

### Stress

Physical stress (e.g., physical discomfort, exercise, hypoglycemia) and psychological stress can cause an acute transient rise in PRL levels. Chronic hyperprolactinemia due to prolonged physical or psychological stress has not been reported except for pseudocyesis, in which PRL levels fall with psychotherapy.

### Renal Disease

Hyperprolactinemia occurs in approximately 75% of women and 50% of men with end-stage renal disease. Although metabolic breakdown of PRL is delayed in renal failure, there is also increased production. Therefore, there is disordered regulation of PRL

secretion, although the exact nature of the defect has not been defined. About one-quarter of individuals with renal insufficiency not requiring dialysis (serum creatinine 2–12 ng/ml) have PRL levels in the 25- to 100-ng/ml range. When such patients take a medication known to alter hypothalamic regulation of PRL, such as methyl dopa or metoclopramide, PRL levels may rise to more than 2000 ng/ml. Correction of the renal failure with transplantation causes a return of PRL levels to normal.

### Cirrhosis

Basal PRL levels are increased in patients with alcoholic cirrhosis in frequencies varying from 16 to 100% and in patients with nonalcoholic cirrhosis in frequencies varying from 5 to 13%. In one study, 50% of patients with hepatic encephalopathy were found to be hyperprolactinemic. It is thought that the hyperprolactinemia is due to disordered hypothalamic neurotransmitter regulation, although there are no specifics in this regard.

### Hypothyroidism

Primary hypothyroidism is associated with a modest increase in the level of PRL in 40% of patients, but levels greater than 25 ng/ml are reached in only 10% of patients. The mechanisms involved include increased thyroid-releasing hormone (TRH) production, increased sensitivity of lactotrophs to TRH, and (possibly) increased pituitary vasoactive intestinal peptide (VIP) generation. Because many patients with long-standing hypothyroidism may have evidence of pituitary enlargement on X rays, the finding of hyperprolactinemia, galactorrhea, and/or amenorrhea associated with an enlarged pituitary seen in hypothyroidism may be easily confused with a prolactinoma. Therapy with L-thyroxine will cause the PRL levels to return to normal and can even result in a regression of pituitary size.

### Adrenal Insufficiency

Glucocorticoids have a suppressive effect on PRL gene transcription and PRL release. Rare cases of hyperprolactinemia occurring in patients with adrenal insufficiency in whom the PRL levels returned to normal with glucocorticoid replacement have been reported.

### Neurogenic

Sexual breast stimulation, suckling, and nipple rings cause a reflex release of PRL that is mediated, in part,

by afferent neural pathways going through the spinal cord. Chest wall lesions (e.g., herpes zoster, cancer, burns, surgery) and cervical cord lesions have been reported to result in elevated PRL levels and galactorrhea through stimulation of these afferent neural pathways.

### Ectopic Prolactin Secretion

Ectopic production of PRL is exceedingly rare. Symptomatic hyperprolactinemia due to well-documented PRL production from a renal cell carcinoma, a gonadoblastoma, and ectopic pituitary tissue in ovarian teratomas has been reported. Given the great frequency of prolactinomas, "idiopathic hyperprolactinemia," and other causes of hyperprolactinemia, a search for an ectopic source of PRL secretion is not warranted unless some other tumor shows up coincidentally.

### HYPOTHALAMIC/PITUITARY STALK DISEASE CAUSING HYPERPROLACTINEMIA

Hyperprolactinemia caused by mass and infiltrative lesions of the hypothalamus and the pituitary stalk, such as clinically nonfunctioning adenomas and craniopharyngiomas (Table I), is due to disturbance of the neuroendocrine mechanisms that control PRL secretion. It is generally assumed that this PRL elevation is due to disinhibition of the tonic dopamine inhibitory action at the level of the pituitary lactotrophs. Because other pituitary function may remain normal in many of these patients, it is apparent that there still is significant transmission of hypothalamic-releasing factors to the pituitary in most of these cases despite increased PRL levels. Patients with fairly high levels (e.g., 100–300 ng/ml) associated with hypothalamic/stalk disease likely have continued PRL-releasing activity along with dopamine deficiency, resulting in higher PRL levels.

### IDIOPATHIC HYPERPROLACTINEMIA

When no specific cause is found with the evaluation outlined later, the hyperprolactinemia is of uncertain etiology and has been designated to be idiopathic. It is recognized that, in many such cases, small prolactinomas that are too small to be detected by current radiological techniques may be present. In other cases, the hyperprolactinemia is due to presumed hypothalamic regulatory dysfunction, but no dysfunction specific to idiopathic hyperprolactinemia has

been definitively elucidated. Long-term follow-up has found that PRL levels return to normal in about one-third of such patients, there is a rise in PRL levels to more than 50% over baseline in 10 to 15% of them, and prolactin levels remain stable in the remaining patients. Over a 2- to 6-year follow-up of 199 patients, only 23 developed evidence of microadenomas and none developed macroadenomas.

## PROLACTINOMAS

When there is no obvious cause of the hyperprolactinemia from the routine screening outlined later, a radiological evaluation of the hypothalamic-pituitary area is mandatory to define a possible prolactinoma or other mass lesion. This includes patients with even mild PRL elevations. MRI with gadolinium enhancement generally provides the best anatomic detail. It should be emphasized that it is very important to distinguish between a large nonsecreting tumor causing modest PRL elevations (usually <150 ng/ml) and a PRL-secreting macroadenoma (PRL levels usually >250 ng/ml) because the therapy will be quite different. One potential additional problem in investigating patients with mild hyperprolactinemia is the finding of a false-positive computed tomography (CT) or MRI scan. Because these techniques are now able to pick incidental nonsecreting tumors, cysts, infarcts, and the like, the finding of a “microprolactinoma” on scan in a patient with elevated PRL levels might not always be a true-positive finding, and cases have been reported with the hyperprolactinemia being due to a medication and an apparent adenoma on MRI being a false-positive finding. When there is no obvious cause for the hyperprolactinemia and the MRI or CT scan is normal, the patient is deemed to have idiopathic hyperprolactinemia.

Uncommonly, prolactinomas will secrete growth hormone in addition to PRL. Some of these patients have little in the way of acromegalic features, and their clinical presentation is that of a prolactinoma. Even more rare are tumors secreting PRL and adrenocorticotrophic hormone (ACTH).

## CLINICAL TESTING

PRL is secreted episodically, and some PRL levels during the day may be above the upper limit of the normal established level for a given laboratory. Thus, the finding of minimally elevated levels in blood requires confirmation in several samples. As indicated in the previous section, there are a number of conditions that may cause moderate PRL elevations, although

generally less than 250 ng/ml. A careful history and physical examination, screening blood chemistries, thyroid function tests, and a pregnancy test will exclude virtually all causes except for hypothalamic-pituitary disease.

However, a specific caution is needed when two-site immunoradiometric assays or chemiluminometric assays are used because patients with large prolactinomas with very high PRL levels may appear to have PRL levels that are only moderately elevated or even normal (i.e., on the order of 10–200 µg/L) due to the “hook effect.” This can cause confusion with clinically non-functioning adenomas, which may cause a similar elevation of PRL due to “disinhibition” of PRL secretion, as noted earlier. This confusion can be avoided by always remeasuring the PRL in such patients at 1:100 dilution given that PRL levels in samples with the hook effect will then increase dramatically.

Stimulation and suppression tests have given nonspecific results, and consensus has developed that such tests reveal no more information than just measurement of basal PRL levels and so are worthless.

## See Also the Following Articles

Prolactin, Evolution of • Prolactin (PRL) • Prolactinoma, Clinical Manifestations • Prolactinoma, Pathogenesis • Prolactinoma, Therapy

## Further Reading

- Barkan, A. L., and Chandler, W. F. (1998). Giant pituitary prolactinoma with falsely low serum prolactin: The pitfall of the “high-dose hook effect”—Case report. *Neurosurgery* **42**, 913–915.
- Bevan, J. S., Burke, C. W., Esiri, M. M., and Adams, C. B. T. (1987). Misinterpretation of prolactin levels leading to management errors in patients with sellar enlargement. *Am. J. Med.* **82**, 29–32.
- Hou, S. H., Grossman, S., and Molitch, M. E. (1985). Hyperprolactinemia in patients with renal insufficiency and chronic renal failure requiring hemodialysis or chronic ambulatory peritoneal dialysis. *Am. J. Kidney Dis.* **6**, 245–249.
- Hutchinson, J., Murphy, M., Harries, R., and Skinner, C. J. (2000). Galactorrhoea and hyperprolactinemia associated with protease inhibitors. *Lancet* **356**, 1003–1004.
- Kelley, S. R., Kamal, T. J., and Molitch, M. E. (1996). Mechanism of verapamil calcium channel blockade-induced hyperprolactinemia. *Am. J. Physiol.* **270**, E96–E100.
- Martin, T. L., Kim, M., and Malarkey, W. B. (1985). The natural history of idiopathic hyperprolactinemia. *J. Clin. Endocrinol. Metab.* **60**, 855–858.
- Molitch, M. E. (2001). Disorders of prolactin secretion. *Endocrinol. Metab. Clinics North America* **30**, 585–610.
- Naidich, M. J., and Russell, E. J. (1999). Current approaches to imaging of the sellar region and pituitary. *Endocrinol. Metab. Clinics North America* **28**, 45–79.

receptor-deficient mice develop lactotroph hyperplasia and prolactinomas. During recent years, attention has been focused on the role of various molecular events involved in prolactinoma pathogenesis.

## Genes

Several genes involved in carcinogenesis in other tissues are also overexpressed or underexpressed in prolactinomas. However, the distinct role of the large majority of these genes involved in prolactinoma pathogenesis is unclear. Oncogenes and proto-oncogenes (which have to be converted to active oncogenes to become active) are dominant mutated proteins that can induce cancer. Until now, activating mutations of the *gsp* oncogene have not been detected in prolactinomas, but these mutations seem to be specific for the somatotroph tumor lineage. Activating mutations of the *H-ras* proto-oncogene have been found only in the relatively rare pituitary carcinomas, for example, in only one case of an aggressive prolactinoma.

In contrast to the dominant proto-oncogenes and oncogenes, tumor suppressor genes can initiate tumor cell growth only when both recessive alleles are lost or altered. The most studied and well-known tumor suppressor gene, *p53*, probably plays little or no role in pituitary tumorigenesis given that neither gene mutations nor gene deletions could be detected in pituitary adenomas or carcinomas. Pituitary tumors can also develop as one of the clinical manifestations of the multiple endocrine neoplasia type 1 (MEN1) syndrome. The MEN1 tumor suppressor gene (MIM 131100) is located on chromosome 11q13 and encodes a 610-amino acid protein named MENIN. However, LOH analysis failed to show that inactivation of this gene plays an important role in the pathogenesis of sporadic prolactinomas.

## Growth Factors

Fibroblast growth factor-2 (FGF-2: basic FGF) is expressed within folliculostellate cells in the anterior pituitary and stimulates prolactin secretion by both normal lactotrophs and prolactinomas. FGF-4 is encoded by the heparin-binding secretory transforming gene (*hst*). Like FGF-2, *hst*/FGF-4 stimulates pituitary prolactin secretion, but it also stimulates lactotroph proliferation and development of prolactinomas. Pituitary tumor-derived transforming gene-1 (*pttg*) is another paracrine growth factor gene that is probably involved in prolactinoma pathogenesis. It is highly expressed in pituitary tumors and is a

vertebrate analogue of the yeast securin inhibitors of sister chromatid separation. FGF-2 and estrogens both induce *pttg*, and *pttg* regulates FGF-2 secretion. The *pttg* protein may be involved in pituitary angiogenesis and prolactinoma tumorigenesis. Also, transforming growth factor- $\alpha$  may play a role in the pathogenesis of prolactinomas.

## CONCLUSION

The unraveling of defects and/or processes leading to prolactinoma formation is still in its infancy, and the mechanism of prolactinoma formation grossly remains elusive.

## See Also the Following Articles

Fibroblast Growth Factor (FGF) • Prolactin, Evolution of • Prolactin (PRL) • Prolactinoma, Clinical Manifestations • Prolactinoma, Diagnosis • Prolactinoma, Therapy • Thyrotropin-Releasing Hormone (TRH)

## Further Reading

- Asa, S. L., Kelly, M. A., Grandy, D. K., and Low, M. J. (1999). Pituitary lactotroph adenomas develop after prolonged lactotroph hyperplasia in dopamine D2 receptor-deficient mice. *Endocrinology* **140**, 5348–5355.
- Atkin, S. L., Landolt, A. M., Jeffreys, R. V., Diver, M., Radcliffe, J., and White, M. C. (1993). Basic fibroblastic growth factor stimulates prolactin secretion from human anterior pituitary adenomas without affecting adenoma cell proliferation. *J. Clin. Endocrinol. Metab.* **77**, 831–837.
- Baird, A., Mormede, P., Ying, S. Y., Wehrenberg, W. B., Ueno, N., Ling, N., and Guillemin, R. (1985). A nonmitogenic pituitary function of fibroblast growth factor: Regulation of thyrotropin and prolactin secretion. *Proc. Natl. Acad. Sci. USA* **82**, 5545–5549.
- Bates, A. S., Farrell, W. E., Bicknell, E. J., McNicol, A. M., Talbot, A. J., Broome, J. C., Perrett, C. W., Thakker, R. V., and Clayton, R. N. (1997). Allelic deletion in pituitary adenomas reflects aggressive biological activity and has potential value as a prognostic marker. *J. Clin. Endocrinol. Metab.* **82**, 818–824.
- Beckers, A., Abs, R., Reyniers, E., De Boule, K., Stevensaert, A., Heller, F. R., Kloppel, G., Meurisse, M., and Willems, P. J. (1994). Variable regions of chromosome 11 loss in different pathological tissues of a patient with the multiple endocrine neoplasia type I syndrome. *J. Clin. Endocrinol. Metab.* **79**, 1498–1502.
- Boggild, M. D., Jenkinson, S., Pistorello, M., Boscaro, M., Scanarini, M., McTernan, P., Perrett, C. W., Thakker, R. V., and Clayton, R. N. (1994). Molecular genetic studies of sporadic pituitary tumors. *J. Clin. Endocrinol. Metab.* **78**, 387–392.
- Cai, W. Y., Alexander, J. M., Hedley-Whyte, E. T., Scheithauer, B. W., Jameson, J. L., Zervas, N. T., and Klibanski, A. (1994). Ras mutations in human prolactinomas and pituitary carcinomas. *J. Clin. Endocrinol. Metab.* **78**, 89–93.

- Chandrasekharappa, S. C., Guru, S. C., Manickam, P., Olufemi, S. E., Collins, F. S., Emmert-Buck, M. R., Debelenko, L. V., Zhuang, Z., Lubensky, I. A., Liotta, L. A., Crabtree, J. S., Wang, Y., Roe, B. A., Weisemann, J., Boguski, M. S., Agarwal, S. K., Kester, M. B., Kim, Y. S., Heppner, C., Dong, Q., Spiegel, A. M., Burns, A. L., and Marx, S. J. (1997). Positional cloning of the gene for multiple endocrine neoplasia-type 1. *Science* **276**, 404–407.
- Ezzat, S., Smyth, H. S., Ramyar, L., and Asa, S. L. (1995). Heterogenous *in vivo* and *in vitro* expression of basic fibroblast growth factor by human pituitary adenomas. *J. Clin. Endocrinol. Metab.* **80**, 878–884.
- Ezzat, S., Walpola, I. A., Ramyar, L., Smyth, H. S., and Asa, S. L. (1995). Membrane-anchored expression of transforming growth factor-alpha in human pituitary adenoma cells. *J. Clin. Endocrinol. Metab.* **80**, 534–539.
- Faccenda, E., Melmed, S., Bevan, J. S., and Eidne, K. A. (1996). Structure of the thyrotrophin-releasing hormone receptor in human pituitary adenomas. *Clin. Endocrinol. (Oxf)* **44**, 341–347.
- Friedman, E., Adams, E. F., Hoog, A., Gejman, P. V., Carson, E., Larsson, C., De Marco, L., Werner, S., Fahlbusch, R., and Nordenskjold, M. (1994). Normal structural dopamine type 2 receptor gene in prolactin-secreting and other pituitary tumors. *J. Clin. Endocrinol. Metab.* **78**, 568–574.
- Heaney, A. P., and Melmed, S. (1999). Pituitary tumour transforming gene: A novel factor in pituitary tumour formation. *Clin. Endocrinol. Metab.* **13**, 367–380.
- Herman, V., Drazin, N. Z., Gonsky, R., and Melmed, S. (1993). Molecular screening of pituitary adenomas for gene mutations and rearrangements. *J. Clin. Endocrinol. Metab.* **77**, 50–55.
- Herman, V., Fagin, J., Gonsky, R., Kovacs, K., and Melmed, S. (1990). Clonal origin of pituitary adenomas. *J. Clin. Endocrinol. Metab.* **71**, 1427–1433.
- Karga, H. J., Alexander, J. M., Hedley-Whyte, E. T., Klibanski, A., and Jameson, J. L. (1992). Ras mutations in human pituitary tumors. *J. Clin. Endocrinol. Metab.* **74**, 914–919.
- Kovacs, K., Ryan, N., Horvath, E., Ezrin, C., and Penz, G. (1978). Prolactin cell adenomas of the human pituitary: Morphologic features of prolactin cells in the nontumorous portions of the anterior lobe. *Horm. Metabolic Res.* **10**, 409–412.
- Levy, A., Hall, L., Yeudall, W. A., and Lightman, S. L. (1994). p53 gene mutations in pituitary adenomas: Rare events. *Clin. Endocrinol. (Oxf)* **41**, 809–814.
- Li, Y., Koga, M., Kasayama, S., Matsumoto, K., Arita, N., Hayakawa, T., and Sato, B. (1992). Identification and characterization of high molecular weight forms of basic fibroblast growth factor in human pituitary adenomas. *J. Clin. Endocrinol. Metab.* **75**, 1436–1441.
- Molitch, M. E. (1987). Pathogenesis of pituitary tumors. *Endocrinol. Metab. Clinics North America* **16**, 503–527.
- Molitch, M. E. (1999). Medical treatment of prolactinomas. *Endocrinol. Metab. Clinics North America* **28**, 143–169.
- Pei, L., Melmed, S., Scheithauer, B., Kovacs, K., and Prager, D. (1994). H-ras mutations in human pituitary carcinoma metastases. *J. Clin. Endocrinol. Metab.* **78**, 842–846.
- Poncin, J., Stevenaert, A., and Beckers, A. (1999). Somatic MEN1 gene mutation does not contribute significantly to sporadic pituitary tumorigenesis. *Eur. J. Endocrinol.* **140**, 573–576.
- Prezant, T. R., Levine, J., and Melmed, S. (1998). Molecular characterization of the MEN1 tumor suppressor gene in sporadic pituitary tumors. *J. Clin. Endocrinol. Metab.* **83**, 1388–1391.
- Saeger, W., and Ludecke, D. K. (1983). Pituitary hyperplasia: Definition, light, and electron microscopical structures and significance in surgical specimens. *Virchows Arch. A: Pathol. Anat. Histopathol.* **399**, 277–287.
- Shimon, I., Hinton, D. R., Weiss, M. H., and Melmed, S. (1998). Prolactinomas express human heparin-binding secretory transforming gene (hst) protein product: Marker of tumour invasiveness. *Clin. Endocrinol. (Oxf)* **48**, 23–29.
- Shimon, I., Huttner, A., Said, J., Spirina, O. M., and Melmed, S. (1996). Heparin-binding secretory transforming gene (hst) facilitates rat lactotrope cell tumorigenesis and induces prolactin gene transcription. *J. Clin. Invest.* **97**, 187–195.
- Shimon, I., and Melmed, S. (1997). Genetic basis of endocrine disease: Pituitary tumor pathogenesis. *J. Clin. Endocrinol. Metab.* **82**, 1675–1681.
- Tordjman, K., Stern, N., Ouaknine, G., Yossiphov, Y., Razon, N., Nordenskjold, M., and Friedman, E. (1993). Activating mutations of the Gs-alpha gene in nonfunctioning pituitary tumors. *J. Clin. Endocrinol. Metab.* **77**, 765–769.
- Wenbin, C., Asai, A., Teramoto, A., Sanno, N., and Kirino, T. (1999). Mutations of the MEN1 tumor suppressor gene in sporadic pituitary tumors. *Cancer Lett.* **142**, 43–47.
- Zhang, X., Danila, D. C., Katai, M., Swearingen, B., and Klibanski, A. (1999). Expression of prolactin-releasing peptide and its receptor messenger ribonucleic acid in normal human pituitary and pituitary adenomas. *J. Clin. Endocrinol. Metab.* **84**, 4652–4655.
- Zhang, X., Horwitz, G. A., Heaney, A. P., Nakashima, M., Prezant, T. R., Bronstein, M. D., and Melmed, S. (1999). Pituitary tumor transforming gene (PTTG) expression in pituitary adenomas. *J. Clin. Endocrinol. Metab.* **84**, 761–767.

normal. Some patients have excellent reduction in PRL levels into the normal range but only modest changes in tumor size, whereas others have persistent mild hyperprolactinemia (but >88% suppression from basal values) with nearly complete disappearance of their tumors. A reduction in PRL levels always precedes any detectable change in tumor size, and PRL nonresponders are also tumor size nonresponders. Once maximum size reduction is achieved, the dose of bromocriptine can often be gradually reduced or even tapered off without tumor regrowth.

The most common side effects of bromocriptine are nausea and vomiting; these are usually transient but may recur with each dose increase. Orthostatic hypotension usually is a problem only when initiating therapy and rarely recurs with dose increases. Limiting nausea and vomiting occurs in 3 to 5% of patients, and digital vasospasm, nasal congestion, psychosis, and depression occur rarely. Side effects can be minimized by starting with 1.25 mg daily with a snack at bedtime. The dose is gradually increased to 2.5 mg twice daily with meals over 7 to 10 days, and PRL levels are checked after 1 to 2 months. Intravaginal administration usually abolishes problems of persistent nausea and vomiting. Most patients respond within 1 to 2 months if they are going to respond at all. Between 5 and 10% of patients either do not respond to bromocriptine or have only minimal responses. Doses higher than 7.5 mg daily are usually not necessary except in some patients with very large tumors.

### Pergolide

Pergolide (Permax) has been approved by the U.S. Food and Drug Administration for the treatment of Parkinson's disease. Although such approval for the treatment of hyperprolactinemia is lacking, there is considerable experience with its use in prolactinoma patients. Hyperprolactinemia can be controlled with single daily doses of 50 to 250  $\mu\text{g}$  with tolerance and efficacy comparable to those with bromocriptine. Tumor size change is similar to or slightly better than that seen with bromocriptine. Questions regarding possible cardiac valve abnormalities with the high doses used to treat Parkinson's disease have recently been raised.

### Quinagolide

Quinagolide (CV 205-502) is a nonergot dopamine agonist, with tolerance and efficacy similar to bromocriptine and pergolide in terms of control of hyperprolactinemia and tumor size, and it can also be given once daily. Approximately 50% of patients who

are resistant to bromocriptine respond to quinagolide. Although side effects are similar, some patients appear to tolerate quinagolide better than they do bromocriptine. Quinagolide is not available in the United States.

### Cabergoline

Cabergoline (Dostinex) is different from the other drugs in that it has a very long half-life and can be given orally once or twice weekly. The long duration of action stems from its slow elimination from pituitary tissue, high-affinity binding to pituitary dopamine receptors, and extensive enterohepatic recycling. A number of studies have shown that cabergoline is generally more effective than bromocriptine at lowering PRL levels and reducing tumor size with a substantial reduction in side effects. Rare patients who experience limiting nausea and vomiting can also be treated with intravaginal cabergoline. Comparison of patient groups shows that 96% of those who had never received prior dopamine agonists had tumor shrinkage of more than 50%, whereas similar reductions were seen in 68% of those who had been intolerant of prior bromocriptine, 64% of those who had been resistant to bromocriptine, and 70% of those who had been responsive to bromocriptine.

With any dopamine agonist, in some patients PRL levels decrease but plateau before they reach normal. Often, the dose continues to be increased in such patients in the hope of achieving normal levels; however, once this is recognized, the dose should gradually be decreased until the lowest dose that achieves the maximum effect is reached. Although most patients respond to dopamine agonists with a rapid fall in PRL levels, some experience stepwise reductions with each stepwise increment in dose. As long as there is a continued fall with each increase, the dose can be increased further to quite high amounts, recognizing that much larger doses are used to treat Parkinson's disease without substantial adverse effects.

### Estrogens

In women with idiopathic hyperprolactinemia or microadenomas in whom fertility is not an issue, galactorrhea is not bothersome, and lack of estrogen is the major concern, consideration can be given to using estrogen replacement rather than dopamine agonists. No tumor enlargement has been found with estrogen treatment in such patients. However, individual cases with enlargement of tumors during estrogen therapy have been reported, so that such

patients should be followed carefully with periodic monitoring of PRL levels. Because of concerns regarding tumor size, dopamine agonists are preferred for women with macroadenomas.

## SURGERY

The success rates for transsphenoidal surgery are highly dependent on the experience and skill of the surgeon as well as on the size of the tumor. Approximately 70 to 80% of patients with microadenomas and 25 to 35% of those with macroadenomas can be expected to have PRL levels normalized by 1 to 12 weeks following surgery. Postoperative recurrence of hyperprolactinemia is found in approximately 20% of patients within the first year following surgery. Therefore, the ultimate, long-term, surgical cure rate for microadenomas is approximately 50 to 60% and for macroadenomas is approximately 25%, using a normal PRL level as the criterion. For patients with giant prolactinomas and those with considerable cavernous sinus invasion, the chance of surgical cure is essentially zero.

Complications from transsphenoidal surgery for microadenomas are quite infrequent, with the mortality and morbidity rates being 0.3% to 0.4%. The mortality and morbidity rates for transsphenoidal surgery for all types of secreting and nonsecreting macroadenomas are 0.9% and 6 to 20%, respectively, depending on the experience of the surgeon. Surgery involving craniotomy is much more hazardous.

## RADIOTHERAPY

Because of the excellent therapeutic responses to transsphenoidal surgery and medical therapy, radiotherapy is generally not considered to be a primary mode of treatment for prolactinomas. Normalization of PRL occurs in less than one-third of patients. Gamma knife-focused radiotherapy is used when there is a residual tumor in the cavernous sinus left after surgery is carried on patients who did not respond to medical therapy with somewhat better and faster efficacy compared to conventional radiotherapy. The major adverse effect of all types of radiotherapy is hypopituitarism.

## THERAPY IN THE PREGNANT WOMAN

When a woman harbors a prolactinoma as the cause of the hyperprolactinemia, two major issues arise when

ovulation and fertility are restored: (1) the effects of the dopamine agonist on early fetal development occurring before a pregnancy is diagnosed and (2) the effect of the pregnancy itself on the prolactinoma.

## Effects of Dopamine Agonists on the Developing Fetus

As a general principle, it is advised that fetal exposure to the dopamine agonist be limited to as short a period as possible. Mechanical contraception should be used until the first two to three cycles have occurred, so that an intermenstrual interval can be established and a woman will know when she has missed a menstrual period. Thus, the dopamine agonist can be stopped after being given for only approximately 3 to 4 weeks of the gestation. When used in this fashion in more than 6000 pregnancies, bromocriptine was not found to cause any increase in spontaneous abortions, ectopic pregnancies, trophoblastic disease, multiple pregnancies, or congenital malformations. Experience was limited to just over 100 women with the use of bromocriptine throughout the gestation; however, no abnormalities were noted in the infants except for one with an undescended testicle and one with a talipes deformity. Few data are available on the safety during pregnancy of pergolide or quinagolide. Outcome data available on 265 pregnancies in which cabergoline was administered to facilitate ovulation do not show increased risks of preterm, ectopic, or multiple birth deliveries or malformations. However, these data are relatively sparse compared with the data in more than 6000 pregnancies with bromocriptine, so that bromocriptine is favored when fertility is the major reason for treatment. On the other hand, these data are encouraging, so that mothers may be reassured if they get pregnant while taking cabergoline.

## Effect of Pregnancy on Prolactinoma Size

Estrogens have a marked stimulatory effect on PRL synthesis and secretion, and the hormonal milieu of pregnancy can stimulate lactotroph cell hyperplasia. In women with prolactinomas, this can result in tumor growth. The risk of clinically significant tumor enlargement (e.g., headaches, visual disturbances) is 1.4% for women with microadenomas and 26.2% for those with macroadenomas unless they have had prior surgery or irradiation, which lowers the risk to 3.0%. Reinstitution of bromocriptine is usually successful in reducing symptomatic tumor enlargement rapidly, and no ill effects on the infants have been observed

in these cases. Surgical decompression may be needed (rarely) if bromocriptine is not successful in reducing tumor size.

### Recommendations for Management

For the hyperprolactinemic woman with a microadenoma or macroadenoma that is intrasellar or extends infrasellarly, bromocriptine is preferred as the primary treatment because of its efficacy in restoring ovulation and very low (1.4%) risk of clinically serious tumor enlargement. Such a patient should be followed carefully throughout gestation. PRL levels do not always rise during pregnancy in women with prolactinomas as they do in normal women. However, they may rise substantially without tumor enlargement; conversely, PRL levels might not rise with tumor enlargement. Therefore, periodic checking of PRL levels is of no benefit. Because of the low incidence of tumor enlargement, routine, periodic, visual field testing is not cost-effective. Visual field testing and scanning are performed only in patients who become symptomatic.

In a woman with a larger macroadenoma that may have suprasellar extension, there is a 26% risk of clinically significant tumor enlargement during pregnancy when only bromocriptine is used. There is no clear-cut best therapeutic approach, and this has to be a highly individualized decision that the patient must make after a clear documented discussion of the following therapeutic alternatives. One approach is just to use bromocriptine to allow ovulation, discontinue it when pregnancy is documented, and then observe the patient carefully for evidence of tumor growth. Second, prepregnancy transsphenoidal surgical debulking of the tumor greatly reduces, but does not eliminate, the risk of serious tumor enlargement. After surgical debulking, bromocriptine is still required to restore normal PRL levels and allow ovulation. A third approach, that of giving bromocriptine continuously throughout gestation, has been advocated, but data regarding the effects of continuous bromocriptine therapy on the developing fetus are still quite meager, and such therapy cannot be recommended without reservation.

For pregnant patients with macroadenomas treated with bromocriptine alone or after surgery, careful follow-up with monthly visual field testing is warranted. Repeat scanning is reserved for patients with symptoms of tumor enlargement, evidence of a developing visual field defect, or both. If symptomatic tumor enlargement occurs with any of these approaches, reinstatement of bromocriptine is probably less

harmful to the mother and child than is surgery. However, such medical therapy must be monitored very closely, and transsphenoidal surgery or delivery (if the pregnancy is far enough advanced) should be performed if there is no response to bromocriptine and vision is progressively worsening.

### See Also the Following Articles

Prolactin, Evolution of • Prolactin (PRL) • Prolactinoma, Clinical Manifestations • Prolactinoma, Diagnosis • Prolactinoma, Pathogenesis

### Further Reading

- Bevan, J. S., Webster, J., Burke, C. W., *et al.* (1992). Dopamine agonist and pituitary tumor shrinkage. *Endocrine Rev.* **13**, 220–240.
- Billir, B. M. K., Molitch, M. E., Vance, M. L., *et al.* (1996). Treatment of prolactin-secreting macroadenomas with the once-weekly dopamine agonist cabergoline. *J. Clin. Endocrinol. Metab.* **81**, 2338–2343.
- Colao, A., DiSarno, A., Landi, M. L., *et al.* (2000). Macroprolactinoma shrinkage during cabergoline treatment is greater in naive patients than in patients pretreated with other dopamine agonists: A prospective studying 110 patients. *J. Clin. Endocrinol. Metab.* **85**, 2247–2252.
- Colao, A., and Lombardi, G. (1998). Growth-hormone and prolactin excess. *Lancet* **352**, 1455–1461.
- Corenblum, B., and Donovan, L. (1993). The safety of physiological estrogen plus progestin replacement therapy and with oral contraceptive therapy in women with pathological hyperprolactinemia. *Fertil. Steril.* **59**, 671–673.
- Freda, P. U., Andreadis, C. I., Khandj, A. G., *et al.* (2000). Long-term treatment of prolactin-secreting macroadenomas with pergolide. *J. Clin. Endocrinol. Metab.* **85**, 8–13.
- Landolt, A. M., and Lomax, N. (2000). Gamma knife radiosurgery for prolactinomas. *J. Neurosurg.* **93**(Suppl. 3), 14–18.
- Laws, E. R., Jr., and Thapar, K. (1999). Pituitary surgery. *Endocrinol. Metab. Clinics North America* **28**, 119–131.
- Molitch, M. E. (1999). Management of prolactinomas during pregnancy. *J. Reprod. Med.* **44**(Suppl.), 1121–1126.
- Molitch, M. E. (2002). Prolactinoma. In “The Pituitary” (S. Melmed, ed.), 2nd ed., pp. 455–495. Blackwell, Malden, MA.
- Musolino, N. R. C., and Bronstein, M. D. (2001). Prolactinomas and pregnancy. In “Pituitary Tumors and Pregnancy” (M. D. Bronstein, ed.), pp. 91–108. Kluwer Academic, Norwell, MA.
- Passos, V. Q., Souza, J. J. S., Musolina, N. R. C., *et al.* (2002). Long-term follow-up of prolactinomas: Normoprolactinemia after bromocriptine withdrawal. *J. Clin. Endocrinol. Metab.* **87**, 3578–3582.
- Robert, E., Musatti, L., Piscitelli, G., *et al.* (1996). Pregnancy outcome after treatment with the ergot derivative, cabergoline. *Reprod. Toxicol.* **10**, 333–337.
- Tsang, R. W., Brierley, J. D., Panzarella, T., *et al.* (1996). Role of radiation therapy in clinical hormonally-active pituitary adenomas. *Radiother. Oncol.* **41**, 45–53.
- Verhelst, J., Abs, R., Maiter, D., *et al.* (1999). Cabergoline in the treatment of hyperprolactinemia: A study in 455 patients. *J. Clin. Endocrinol. Metab.* **84**, 2518–2522.

The presence of genetic factors affecting the risk of prostate cancer is also supported by data on the incidence of prostate cancer in different racial groups. The prostate cancer incidence rate is highest for men of African American descent (149/100,000 person-years), with intermediate rates for U.S. Caucasian men (107/100,000) and lower rates for men of Asian descent (39/100,000 for Japanese men and 28/100,000 for Chinese men). For Asian men who migrate to the United States, the rate of prostate cancer increases dramatically, although it remains lower than that observed for U.S. white males. Even when corrected for stage of detection and access to medical care, black men with prostate cancer tend to have lower survival rates than white men with the same disease status.

The increase in prostate cancer incidence for men who migrate from Japan to the United States suggests the presence of environmental factors, especially dietary risk factors, in the development of prostate cancer. However, there is controversy regarding the role of dietary factors in the development of prostate cancer. Several studies have suggested that dietary fat, especially fat from red meat, appears to be an important risk factor for the development of this disease. Other dietary elements may have an acute effect on PSA, a marker for prostate cancer growth. Lycopenes (a substance derived from tomato with antioxidant activity) or soy protein (a phytoestrogen) have apparent anti-tumoral effects in men with prostate cancer. It is not clear whether soy protein acts directly or only through its estrogenic activity against prostate cancer.

Hormonal factors are critical for the development of prostate cancer. Testosterone is necessary for prostate epithelium to grow, and early prostate cancer has been shown to be androgen dependent.

### Prevention of Prostate Cancer Deaths

It has not been definitively established that early detection of prostate cancer will lead to improved survival. However, the clinical approach of the widespread use of PSA blood tests, coupled with regular digital rectal examination (DRE), is an attempt to decrease the potential significant morbidity and mortality of prostate cancer. Prostate cancer, the second leading cause of cancer deaths in men after lung cancer, shortens life expectancy dramatically. In addition, ineffective treatment of this disease and disease progression can cause significant morbidity.

Prostate cancer screening may decrease the risk of subsequent death from this disease. The decreased death rate from prostate cancer in the United States

during the past few years is associated with the introduction of PSA and the widespread application of local treatment. This suggests that prostate cancer surgery coupled with earlier detection using the PSA blood test may have had a positive effect on the rate of prostate cancer death. Labrie *et al.* studied men who were aggressively screened with DRE and yearly serial PSA blood tests compared to men who received routine medical care. They observed a statistically significant decrease in prostate cancer deaths after 7 years for the screened group. Preliminary studies from the Tyrol region of Austria, where PSA screening was introduced in 1987 and has been extensively applied since 1994, also demonstrate a marked decrease in prostate cancer deaths compared to historical control data and prostate cancer death rates in other regions of Austria.

## DIAGNOSIS

### Detection of Prostate Cancer

#### **Suspicion of the Presence of Cancer**

An abnormal DRE of the prostate or an elevated PSA blood test level suggest the presence of prostate cancer. An abnormal exam includes the presence of nodular areas of the prostate or an asymmetric prostate gland. The presence of either of these factors is an indication for prostate biopsy in men who would benefit from the diagnosis of prostate cancer (men with a >10-year life expectancy and the possibility of localized disease or those with symptoms of metastatic disease).

#### **Prostate-Specific Antigen**

PSA is an enzyme normally produced by the glandular tissue of the prostate. It is produced at detectable levels only by prostate tissue. PSA is normally secreted outside of the body in urine or semen. If PSA backs up into the body and is present at elevated levels in the blood, then an abnormal condition is present within the prostate. This abnormality may be caused by trauma to the prostate (such as occurs after biopsy or cystoscopy), infection, benign enlargement, or prostate cancer. Therefore, an elevated PSA level does not diagnose the presence of prostate cancer, and prostate biopsy is required to confirm a clinical suspicion of cancer.

Since PSA levels tend to increase with prostatic enlargement, PSA can be normalized to total prostate volume using the index of PSA (ng/ml)/prostate volume (cc). Prostatic enlargement tends to occur with age; therefore, it is possible to assume increased

prostatic size as men age. This approach allows assessment of age-specific PSA levels. For men younger than 50 years of age, PSA levels should be 0–2.5 ng/ml. For men 50–59 years of age, normal PSA levels are 0–3.5 ng/ml. For men 60–69 years of age, PSA should be 0–4.5 ng/ml. For men 70–79 years old, PSA levels of 0–6.5 ng/ml are considered normal.

The specificity of the PSA blood test has been improved by the use of a derivation of the PSA. PSA may exist free in serum or bound to large proteins such as  $\alpha_1$ -antichymotrypsin inhibitor. The percentage of free PSA is a measurement of the proportion of PSA protein that is free in the bloodstream versus that which is bound to other proteins. Men with low PSA levels (<1.0 ng/ml) rarely have cancer, and the vast majority of men with markedly elevated PSA levels (>30 ng/ml) have cancer. The percentage of free PSA is useful for determining whether biopsies are needed for men with a PSA between 3 and 10 ng/ml. Prostate cancers tend to release PSA into the bloodstream that is bound to proteins, whereas benign prostatic tissue tends to release more free PSA. Therefore, men with a low percentage of free PSA (<25% free) are more likely to have prostate cancer than men with elevated percentage free levels in the presence of an abnormal serum PSA level (between 3 and 10 ng/ml).

### **Prostate Biopsy**

Prostate cancer is detected using biopsies of the prostate. Biopsies should be performed by a transrectal approach using ultrasound guidance. Transrectal ultrasound (TRUS) is used to evaluate prostate size and to ensure that biopsies have an adequate distribution of samples throughout the prostate gland. Typically, at least a sextant pattern of biopsies is performed to effectively detect any tumor in the peripheral zone of the prostate that is at least 1 cm in size. Additional samples of the anterior (transitional zone) are obtained. Although TRUS is used to guide the biopsies and many tumors are hypoechoic, there is no diagnostic appearance of cancers on ultrasound.

### **Latent vs Clinical Prostate Cancer**

More than 90% of prostate cancers are adenocarcinomas. They are present as microscopic lesions in the majority of men older than age 50. These microscopic (latent) tumors cannot generally be detected by standard biopsies and are detectable only at autopsy. Latent tumors progress into clinically detectable prostate cancer in approximately 10% of all men. The mechanisms by which latent prostate cancer progresses to

clinically significant and detectable tumors are poorly understood. All racial groups have similar rates of latent cancer, but the rate of progression to clinically significant cancer and death differs greatly. Therefore, some genetic factors appear to act by accelerating the rate of transformation of latent prostatic tumors into clinically significant tumors.

PSA-based screening for prostate cancer does not significantly increase the frequency of detection of small, presumably latent and clinically insignificant cancers. Therefore, it is inaccurate to state that a cancer found on biopsy is unimportant. Longitudinal series of men from Sweden (where prostate cancer management is usually expectant) who have prostate cancer and who live 10 years or more have shown that the majority who are not provided initial treatment die of the disease.

### **Precursors of Prostate Cancer**

There is no *in situ* carcinoma of the prostate that progresses to an invasive tumor. However, high-grade (grade II or III) prostatic intraepithelial neoplasia (PIN) is associated with invasive carcinomas. For men with high-grade PIN in the absence of an invasive carcinoma on prostate biopsies, a subsequent diagnosis of invasive cancer is made in at least 30–50%. Low-grade PIN (PIN I) is of no clinical significance. It is not associated with subsequent detection or development of prostate cancer.

### **Cancer Location and Volume**

Based on the site of detection of early palpable and nonpalpable tumors, most (85%) prostate cancers develop out of the peripheral part of the prostate, primarily in the posterior part of the gland. At the time of detection, prostate cancers are usually multifocal, with an average of seven sites observed on serial sectioning of the entire gland. Cancers less than 0.2 cc in volume and of low grade (Gleason sum <7) are rarely aggressive and are most commonly found in autopsy series or in prostates removed for benign prostatic hypertrophy (BPH). Cancers less than 4 cc in volume rarely invade the seminal vesicles or metastasize to pelvic lymph nodes.

### **Grade**

Prostate cancers are graded using the Gleason system. This approach evaluates the glandular pattern of the tumor as examined at relatively low-power magnification. Cytologic features of the tumor are not considered in the Gleason grading system. The predominant and secondary patterns of glands within the tumor are identified and each is assigned a grade of 1–5,

with 1 being the least aggressive (most differentiated) and 5 being the most aggressive (least differentiated) pattern. The two grades are added together to provide a Gleason sum score of 2–10. The Gleason score is often the most important prognostic factor of the progression of prostate cancer. For Gleason scores of 2–4, cancers rarely progress early or cause death within 10 years. Localized tumors are most commonly detected with the intermediate grades of 5 or 6. Gleason 7 cancers have a significantly worse prognosis than Gleason 5 or 6 tumors. Most Gleason 8–10 tumors have at least microscopically metastasized prior to detection, although a subset of men with disease are cured with local treatment only. There is good intraobserver concordance within one Gleason sum score; however, low-grade tumors may dedifferentiate over time into more aggressive tumors. The detection of limited amounts of prostate cancer is clinically important because the volume of cancer on prostatic needle biopsy is unrelated to the amount of cancer actually present within the prostate gland.

## Staging

Prostate cancers are staged using the T, N, and M classification system. Localized tumors include clinical T1 or T2 tumors (N0M0), whereas clinical T3 (N0M0) tumors are considered locally advanced, and any nodal or metastatic involvement constitutes systemic prostate cancer. Clinical staging is primarily performed by DRE. The most common stage at which prostate cancers are detected is T1c (a nonpalpable lesion detected on biopsy evaluation of an elevated PSA). Nodal involvement may be detected with computed tomography (CT) of the pelvis, whereas metastases to bone (the next most common site) is primarily performed by radionuclide bone scan. For most men with localized disease (T1c, Gleason sum score 5–6, or PSA <10), there is a <2% risk of finding metastatic disease on bone scan or CT scan.

## TREATMENT

Treatment of prostate cancer is generally limited to those men with a life expectancy >10 years or men with symptomatic metastatic disease. Therefore, it is possible to discuss treatment options for prostate cancer based on the stage of disease diagnosed.

### Localized Disease

The goal of treatment of clinically localized disease is the eradication of all local tumor. Despite apparently

effective treatment, some men with clinically localized disease may have micrometastatic disease that limits local treatment to a less than 100% success rate. Nevertheless, the majority of men with clinically localized disease are potentially curable with effective treatment.

### Watchful Waiting

Many men with prostate cancer have long-term survival without aggressive treatment. For men with Gleason grade 2–4 tumors, the survival rate for more than 5–10 years is nearly identical to that of untreated patients, and aggressive treatment may not be needed. Only if survival >10 years is expected should treatment be considered. Alternatively, for men with Gleason grade 8–10 tumors, the 5-year death rate from prostate cancer is approximately 50%. These observations suggest that watchful waiting may be very appropriate for patients with a life expectancy of <10 years and well-differentiated tumors. For most men with more aggressive tumors or longer life expectancy, effective local treatment is necessary to avoid death from prostate cancer. For younger men with prostate cancer, treatment may be critical to avoid death from this disease.

### Radiation Therapy

Radiation may be delivered to the prostate using external beam therapy, which is commonly applied using a three-dimensional conformal technique, interstitial radiation delivered by seeds placed directly into the prostate gland, or a combination of both treatments. The effectiveness of radiation therapy is limited by the relative insensitivity of prostate cancer to radiation. With increasing radiation doses, there is a lower risk of subsequent positive biopsies. When radiation doses of 60–70 cGy are delivered to the prostate, positive biopsies occur in 30–40% of treated patients. These results are not acceptable for long-term cure, although short-term control of cancer may be provided. Only when doses higher than 81 cGy are applied do positive biopsy rates decline below 10%. This dose of radiation is difficult to apply because the small intestine, bladder, and rectum are all highly sensitive to radiation damage. The observation of locally persistent prostate cancer (on biopsy) after attempted definitive radiation provides a possible explanation for the slightly lower 10-year disease-free rates observed after radiation therapy compared to results obtained with surgical therapy. However, the difference in survival after treatment of clinically localized disease is primarily due to patient selection biases rather than the effectiveness of the treatment. Selection biases in treatment arms can be controlled to some degree by

categorization of disease status based on PSA, clinical stage, and Gleason grade. Unfortunately, disease status has to be categorized to allow comparison of different treatment regimens because randomized trials of treatment modalities for localized prostate cancer have not been performed.

### **Surgical Treatment**

Surgical prostatectomy has been used as a treatment for prostate cancer since 1904. However, successful early results were dependent on the detection of small-volume cancers, which was rare. In addition, a poor understanding of pelvic anatomy resulted in high complication rates (impotence and infection) after surgical prostatectomy. Furthermore, the surgical margins of resection around the prostate are limited by the apposition of the prostate to the bladder and sphincteric muscles as well as the rectum.

Significant advances in the surgical treatment of prostate cancer depended on an improved understanding of the anatomic relationship of the prostate to the nerves that provide erectile function as well as the relationship of the prostate to the sphincteric muscles. These observations were made possible because of the anatomic observation of the relationship of blood vessels to the prostate. With better control of blood loss, identification and preservation of urethral sphincteric muscles and the neurovascular bundles containing the nerves responsible for erectile function became possible. It is now possible to allow preservation of erectile function for 50–80% of men after radical prostatectomy at selected centers, with incontinence rates of 2–10%. The rate of complications is related in large part to the age of the patient.

Despite the potential risks of surgical therapy, prostate cancer appears to be better controlled with surgery than radiation therapy in most series of patients with 10–15 years of follow-up. Thus, radical prostatectomy is the primary choice for treatment of localized prostate cancer for most men who have a life expectancy >10 years and who are candidates for a major surgical operation.

### **Systemic Disease**

Systemic prostate cancer may involve a wide spectrum of disease, from biochemically detected cancer detectable only on PSA blood test with no evidence of lesions on bone scan or CT scan to symptomatic, extensive metastatic lesions easily seen on bone scan with extensive replacement of bone marrow by disease.

The initial treatment of systemic prostate cancer has changed little since the 1940s, when it was demonstrated that hormonal therapy with androgen deprivation can result in dramatic antitumor responses for men with disabling metastatic prostate cancer.

### **Hormonal Therapy**

Androgen withdrawal, by either castration or medical therapy, can result in dramatic responses of PSA levels and clinical disease in the vast majority of men with advanced prostate cancer. Medical therapy usually involves administration of gonadotropin-releasing hormone (GnRH) agonists, which with tonic administration results in ablation of luteinizing hormone (LH) secretion by the pituitary and castrate levels of testosterone production by the testes. GnRH is normally released in a pulsatile fashion from the hypothalamus, in which it directly acts on the pulsatile release of LH (and follicle-stimulating hormone) from the pituitary. Tonic high-level GnRH stimulation of the pituitary (in distinction to pulsatile stimulation) results in a bimodal effect on LH secretion. Tonic pituitary stimulation (as provided by long-acting, potent GnRH agonists) causes an initial surge in LH secretion with increased testosterone, maximally noted approximately 1 week after GnRH agonist treatment. This tonic stimulation results in down-regulation of GnRH receptors on the gonadotrope cells of the pituitary so that the pituitary becomes unresponsive to GnRH and LH secretion stops. The net effect of GnRH agonist treatment is an initial increase in testosterone levels (a flare effect), followed by suppression of LH with a decrease in testosterone to castrate levels within 1 month of continued treatment. No flare effect should be seen with repeated doses of an effective GnRH agonist, usually given as a depot injection or implant.

Obstruction of androgen action by competitive androgen receptor blockers can prevent the action of testosterone and other androgens on prostate cancer cells. Prostate cancer is remarkably androgen dependent and will usually respond to androgen deprivation for a median duration of 18 months in men with extensive prostate cancer. For many men who are treated with androgen deprivation for a prolonged period of time and subsequently have progression of disease, as measured by increased PSA levels, a partial agonist effect of the competitive androgen receptor blocking agents (flutamide and bicalutamide) referred to as antiandrogens may be observed. Cessation of the

antiandrogens will result in a decrease in PSA levels for many patients.

### **Chemotherapy**

Multiple single-agent chemotherapy regimens have been applied in the treatment of prostate cancer with limited success. Objective responses occur in 10–20% of men with extensive prostate cancer treated with single agents. Combinations such as estramustine-based regimens with docetaxol or other microtubule-inhibiting agents may provide objective responses in 40–60% of patients.

### **Immunotherapy**

A series of new agents that are based on immune-based treatments for prostate cancer have been applied on an investigational basis for advanced prostate cancer. Using antibody-based treatments, vaccines, or other immune-stimulating treatments, objective responses in 50% or more of men have been demonstrated. The role of these treatments in the management of men with advanced prostate cancer is yet to be demonstrated.

### **Radiation**

Metastatic lesions typically respond to either spot radiation therapy or systemic radiation directed to bony lesions, such as with intravenous injection of strontium-89. Relief of pain occurs in 60–80% of patients with local bone pain from metastatic prostate cancer. For men with solitary symptomatic deposits, local radiation is most appropriate, whereas patients with multiple sites of symptomatic involvement are best treated with systemic administration of strontium-89.

## **CONCLUSION**

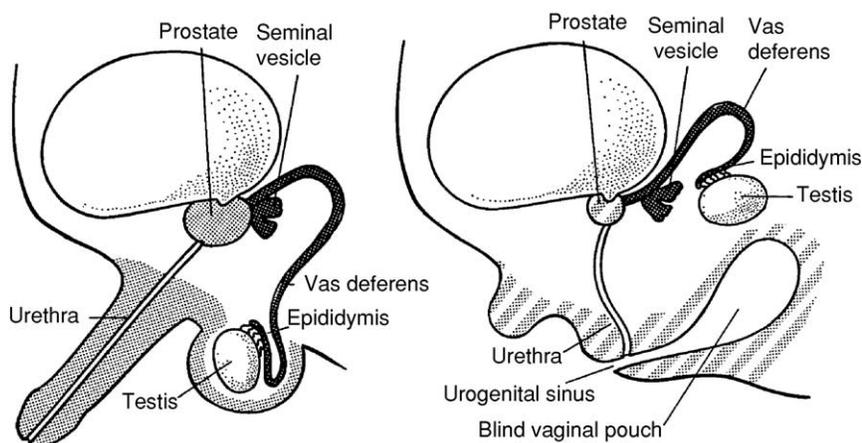
Prostate cancer is a significant health concern for nearly one in nine men in the United States. Although effective local therapies are available, early detection is necessary to allow local treatments such as surgery or radiation. Hormonal therapy is the mainstay of systemic therapy, but it is not a curative therapy since most tumors develop cell lines whose growth is androgen independent. Effective systemic therapies for prostate cancer are still needed. Further understanding of the pathophysiology of prostate cancer is progressing rapidly, in conjunction with improved methods of early detection and improved local treatments.

### **See Also the Following Articles**

Angiogenesis • Benign Prostatic Hyperplasia (BPH) • Estrogen and the Male • GI Hormones in Cancer • Impotence and Aging • Melatonin • Pancreatic Cancer

### **Further Reading**

- Kim, E. D., Grayhack, J. T., Brawley, O. W., *et al.* (2000). Prostate cancer. *In* "Comprehensive Textbook of Genitourinary Oncology" (N. J. Vogelzang, P. T. Scardino, W. U. Shipley, and D. S. Coffey, eds.), 2nd ed. Lippincott Williams & Wilkins, Philadelphia.
- Marks, S. (1999). "Prostate & Cancer." Fisher, Cambridge, MA.
- Pienta, K. J., Epstein, J. I., Brawer, M. K., *et al.* (1998). Prostate cancer. *In* "Campbell's Urology" (P. C. Walsh, E. D. Vaughan, A. B. Retik, and A. J. Wein, eds.), 7th ed. Saunders, Philadelphia.
- Presti, J. C., Jr. (2000). Neoplasms of the prostate gland. *In* "Smith's General Urology" (E. A. Tanagho and J. W. McAninch, eds.), 15th ed. McGraw-Hill, New York.
- Thompson, I., Carroll, P., Coley, C., Sweat, G., McLeod, D., Schellhammer, P., Wasson, J., and Zietman, A. (2000). Prostate-specific antigen (PSA) best practice policy. *Oncology* **14**(2), 267–286.



**Figure 1** A diagram illustrating the roles of testosterone and dihydrotestosterone (DHT) in male sexual differentiation *in utero*. Dark shading indicates regions where male sexual differentiation and development are dependent on testosterone; light shading indicates regions where male sexual differentiation and development are dependent on DHT.

functions as a prehormone, where its conversion to DHT results in differentiation of the external genitalia and prostate. These findings are supported by studies demonstrating that in the human fetus DHT formation occurs in the urogenital sinus, urogenital tubercle, and urogenital swellings at the time of sexual differentiation, but does not occur in the Wolffian anlage until sexual differentiation is completed. Animal studies by using a 5 $\alpha$ -reductase-2 inhibitor provide further evidence to support the differential roles of testosterone and DHT in male sexual differentiation.

With puberty, the affected males have an increase in muscle mass and deepening of the voice. There is growth of the phallus with rugation and hyperpigmentation of the scrotum. Inguinal testes descend into the scrotum at puberty in some patients. Libido is intact and patients are capable of erections. Although patients are generally oligo- or azoospermic, normal sperm concentrations have been reported in some patients who have descended testes. Affected patients have been reported to father children, suggesting that DHT does not play a major role in spermatogenesis and sperm function. These clinical findings suggest that pubertal events, including male sexual function and spermatogenesis, are primarily testosterone mediated, although the role of DHT produced by 5 $\alpha$ -reductase-1 remains to be defined.

The prostate in the affected males is nonpalpable on rectal examination. It is rudimentary on transrectal ultrasound and magnetic resonance imaging visualization and is approximately 1/10 the size of age-matched normal controls. Prostate diseases, such as prostate cancer and benign prostatic hyperplasia, have

not been reported in male pseudohermaphrodites with 5 $\alpha$ -reductase-2 deficiency. The treatment for benign prostate hyperplasia with 5 $\alpha$ -reductase inhibitors evolved in part from the clinical observation that adult male pseudohermaphrodites with 5 $\alpha$ -reductase-2 deficiency have rudimentary prostates.

Affected adult males have less facial and body hair than their nonaffected male relatives; male pattern baldness has never been observed in affected males.

Sebum production is dependent on androgen action. No demonstrable sebum is produced in 46,XY subjects with complete androgen insensitivity due to mutations in the androgen receptor. Although affected males with 5 $\alpha$ -reductase-2 deficiency rarely have acne, they produce normal amounts of sebum, suggesting that sebum production is regulated by the 5 $\alpha$ -reductase-1 isozyme.

## BIOCHEMICAL FEATURES OF 5 $\alpha$ -REDUCTASE-2 DEFICIENCY

The biochemical characteristics of 5 $\alpha$ -reductase-2 deficiency have been well defined over the years. The condition is characterized by the following: (1) normal to elevated levels of plasma testosterone; (2) decreased levels of plasma DHT, with an increased testosterone/DHT ratio at baseline and/or following human chorionic gonadotropin stimulation; (3) a decreased conversion of testosterone to dihydrotestosterone *in vivo*, with conversion ratios of testosterone to DHT of <1%; (4) normal metabolic clearance rates of testosterone and DHT; (5) decreased production of

urinary 5 $\alpha$ -reduced androgen metabolites with increased 5 $\beta$ /5 $\alpha$  urinary metabolite ratios; (6) decreased plasma and urinary 3 $\alpha$ -androstane-2,17-diol glucuronide, a major metabolite of DHT; and (7) a global defect in steroid 5 $\alpha$ -reduction as demonstrated by decreased urinary 5 $\alpha$ -reduced metabolites of C-21 steroids, such as cortisol and corticosterone, in addition to C-19 steroids. Although the defect of 5 $\alpha$ -reduction of steroids is generalized, only the defective reduction of testosterone appears to be of clinical significance.

Increased plasma levels of luteinizing hormone (LH) and an increased LH pulse amplitude with a normal LH frequency have been reported in three subjects. The elevated mean plasma LH level occurs despite normal to elevated mean plasma testosterone, suggesting a role for DHT in the negative feedback control of LH. Plasma follicle-stimulating hormone (FSH) levels may be elevated. Although some of the elevation in FSH is undoubtedly attributable to cryptorchidism and seminiferous tubular damage, a role for DHT in the feedback control of FSH cannot be ruled out.

## BIOCHEMISTRY AND MOLECULAR BIOLOGY OF 5 $\alpha$ -REDUCTASE ISOZYMES

Steroid 5 $\alpha$ -reductase isozymes located in the microsomes of the cell are NADPH-dependent proteins that reduce the double bond at the 4–5 position of a variety of C-19 and C-21 steroids. These isozymes convert testosterone to the more potent androgen, DHT. Testosterone and DHT bind to the intracellular androgen receptor (a member of the nuclear steroid/thyroid hormone receptor superfamily) and interact with a cognate androgen DNA-response element to regulate target gene expression. Although testosterone and DHT interact with the same androgen receptor, they produce distinct biological responses under certain conditions. The molecular mechanism for this is unclear even though DHT has been reported to bind to the androgen receptor more avidly than testosterone and the DHT–receptor complex is more efficiently transformed to the DNA-binding state than is the testosterone–receptor complex.

In the early 1960s, it was theorized that multiple 5 $\alpha$ -reductase isozymes existed. In 1975 and 1976, Moore and Wilson detected different pH optima for 5 $\alpha$ -reductase activity in genital and nongenital skin. In the genital skin, the major enzymatic activity had a narrow, acidic pH optimum of 5.5, which was found to be expressed at a low level in the genital skin of male

pseudohermaphrodites with 5 $\alpha$ -reductase deficiency; another enzymatic activity observed had a neutral to alkaline pH (pH 7–9), was present in both normal genital and nongenital skin, and was shown to be expressed at a normal level in the genital skin of male pseudohermaphrodites with 5 $\alpha$ -reductase deficiency. Kinetic analysis of 5 $\alpha$ -reductase activity in the epithelium and stroma of the prostate also suggested that there were different 5 $\alpha$ -reductase activities.

In the early 1990s, two genes encoding two isozymes were cloned using expression cloning technology: steroid 5 $\alpha$ -reductase type 1 (gene symbol: *SRD5A1*) and steroid 5 $\alpha$ -reductase type 2 (gene symbol: *SRD5A2*). Mutations in the 5 $\alpha$ -reductase-2 gene are responsible for male pseudohermaphroditism due to 5 $\alpha$ -reductase deficiency.

The human 5 $\alpha$ -reductase-2 gene has five exons and four introns, encodes a highly hydrophobic 254-amino-acid protein with a molecular weight of approximately 28.4 kDa, and maps to the short arm of chromosome 2 band 23. The type 2 isozyme has a much higher affinity for testosterone (apparent  $K_m = 4\text{--}50\text{ nM}$ ) than the type 1 isozyme ( $K_m = 1\text{--}5\text{ }\mu\text{M}$ ). However, the apparent  $K_m$  (3–10  $\mu\text{M}$ ) for the NADPH cofactor is similar in both isozymes. The type 2 isozyme is sensitive to finasteride, a 5 $\alpha$ -reductase inhibitor, and has an acidic pH optimum in enzymatic assays. However, it may work at a neutral pH optimum in its native state.

The type 2 isozyme is expressed in the external genital tissues early in gestation. In adulthood, its expression in prostate, genital skin, epididymis, seminal vesicle, and liver is relatively high but is quite low in other tissues. It has been reported that this isozyme may also be expressed in the ovary and hair follicles.

The 5 $\alpha$ -reductase-1 gene is normal in male pseudohermaphrodites with 5 $\alpha$ -reductase deficiency. It also has five exons and four introns and is located on the short arm of chromosome 5 band 15. This isozyme has 259 highly hydrophobic amino acids with a molecular weight of approximately 29.5 kDa. There is approximately 50% homology between human type 1 and type 2 isozymes in amino acid compositions. 5 $\alpha$ -Reductase-1 has a broad alkaline pH optimum, a low substrate affinity, and a low sensitivity to finasteride inhibition. This isozyme is expressed in nongenital skin, liver, and certain brain regions. However, its expression in the prostate, genital skin, epididymis, seminal vesicle, testis, adrenal, and kidney is low. Its expression in the liver and nongenital skin is detected at birth and is present throughout life, whereas its expression in embryonic tissues is low.

The physiological function of 5 $\alpha$ -reductase-1 is still obscure. It may play a significant role in parturition.

### MOLECULAR GENETICS OF 5 $\alpha$ -REDUCTASE-2 DEFICIENCY

More than 33 mutations in the 5 $\alpha$ -reductase-2 gene have been identified, including mutations in the three largest kindreds of male pseudohermaphrodites with 5 $\alpha$ -reductase-2 deficiency in the world—the Dominican, New Guinean, and Turkish kindreds. Subjects from the New Guinean kindred have a large deletion of more than 20 kb in the 5 $\alpha$ -reductase-2 gene. Subjects in the Dominican kindred have a missense mutation in exon 5 of the 5 $\alpha$ -reductase-2 gene, which substitutes thymidine for cytosine and results in a substitution of the nonpolar amino acid tryptophan for the basic, polar amino acid arginine at position 246 of the isozyme. This missense mutation causes a decrease in binding of the cofactor NADPH, an altered pH optimum, and a dramatic loss of enzymatic activity. In the Turkish kindred, a single base deletion in exon 5 of the 5 $\alpha$ -reductase-2 gene has been detected. This single-base deletion (adenine) results in a frameshift at amino acid position 251 and an addition of 23 amino acids at the carboxyl terminus of this 254-amino-acid isozyme. This mutation in the isozyme results in a complete loss of enzymatic activity without an alteration in gene expression.

Mutations in the 5 $\alpha$ -reductase-2 gene are found throughout all five exons of the gene and range from a single point defect to a deletion of the entire gene. These mutations result in various types of enzymatic dysfunction including impaired binding of substrate and cofactor to the isozyme, blocked formation of a functional isozyme (deletions, nonsense mutations, splice-junction alterations), and an unstable isozyme. Although various individual mutations have been characterized, no correlation between the severity of the syndrome and a particular gene defect has been observed.

5 $\alpha$ -Reductase-2 deficiency is an inherited autosomal recessive disease as evidenced by pedigree analysis, biochemical analysis, and molecular genetic analysis. Heterozygotes have a normal male phenotype. It should be noted that approximately 35% of patients with 5 $\alpha$ -reductase-2 deficiency from different families worldwide have been found to be either compound or inferred compound heterozygotes, with mutations in two independent loci resulting in the disease phenotype. This suggests that the carrier frequency of a single mutant allele is higher than previously suggested due to the rarity of the disease phenotype.

### GENDER IDENTITY AND GENDER ROLE IN MALE PSEUDOHERMAPHRODITES WITH 5 $\alpha$ -REDUCTASE-2 DEFICIENCY

Male pseudohermaphrodites with 5 $\alpha$ -reductase-2 deficiency provide a unique genetic model for studying hormonal influence in the development of male gender identity. The psychosexual studies in these subjects demonstrate that, in humans, environmental or sociocultural factors are not solely responsible for the formation of a male gender identity; androgens make a strong and definite contribution. These data support the “hormonal influence theory” in the development of gender identity.

### SUMMARY

Studies of male pseudohermaphrodites with 5 $\alpha$ -reductase-2 deficiency over the past 2–3 decades have provided valuable information about male sexual development and have elucidated the roles of testosterone and DHT in human physiology and pathophysiology. Studies of subjects with this inherited condition have led to the development of specific 5 $\alpha$ -reductase-2 inhibitors for the treatment of benign prostate hyperplasia and male pattern baldness.

### Acknowledgments

This work was supported in part by National Institutes of Health Grant M01-RR-00047 (General Clinical Research Center) and HD-09421-15.

### See Also the Following Articles

Androgen Biosynthesis and Gene Defects • Androgens, Gender and Brain Differentiation • 21-Hydroxylase Deficiency, Classical

### Further Reading

- Andersson, S., Berman, D. M., Jenkins, E. P., and Russell, D. W. (1991). Deletion of steroid 5 $\alpha$ -reductase-2 gene in male pseudohermaphroditism. *Nature* **354**, 159–161.
- Cai, L.-Q., Fratianni, C. M., Gautier, T., and Imperato-McGinley, J. (1994). Dihydrotestosterone regulation of semen in male pseudohermaphrodites with 5 $\alpha$ -reductase-2 deficiency. *J. Clin. Endocrinol. Metab.* **79**, 409–414.
- Imperato-McGinley, J., Guerrero, L., Gautier, T., and Peterson, R. E. (1974). Steroid 5 $\alpha$ -reductase deficiency in man: An inherited form of male pseudohermaphroditism. *Science* **186**, 1213–1216.
- Imperato-McGinley, J., Peterson, R. E., Gautier, T., and Sturla, E. (1979). Androgens and the evolution of male-gender identity

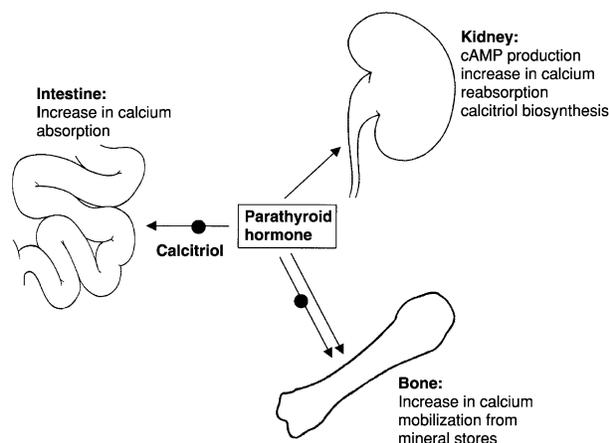
- among male pseudohermaphrodites with 5 $\alpha$ -reductase deficiency. *N. Engl. J. Med.* **300**, 1233–1237.
- Imperato-McGinley, J., Gautier, T., Zirinsky, K., Hom, T., Palomo, O., Stein, E., Vaughan, E. D., Markisz, J., Ramirez, de Arellano, E., and Kazam, E. (1992). Prostate visualization studies in males homozygous and heterozygous for 5 $\alpha$ -reductase deficiency. *J. Clin. Endocrinol. Metab.* **75**, 1022–1026.
- Imperato-McGinley, J., Gautier, T., Cai, L.-Q., Yee, B., Epstein, J., and Pochi, P. (1993). The androgen control of sebum production. studies of subjects with dihydrotestosterone deficiency and complete androgen insensitivity. *J. Clin. Endocrinol. Metab.* **76**, 524–528.
- Imperato-McGinley, J., and Zhu, Y. S. (2002). Gender and behavior in subjects with genetic defects in male sexual differentiation. In “Hormones, Brain, and Behavior” (D. W. Pfaff, *et al.*, eds.), Vol. 5, pp. 303–345. Academic Press, San Diego.
- Russell, D. W., and Wilson, J. D. (1994). Steroid 5 alpha-reductase: Two genes/two enzymes. *Annu. Rev. Biochem.* **63**, 25–61.
- Zhu, Y. S., Katz, M. D., and Imperato-McGinley, J. (1998). Natural potent androgens: Lessons from human genetic models. *Bailliere Clin. Endocrinol. Metab.* **12**, 83–113.



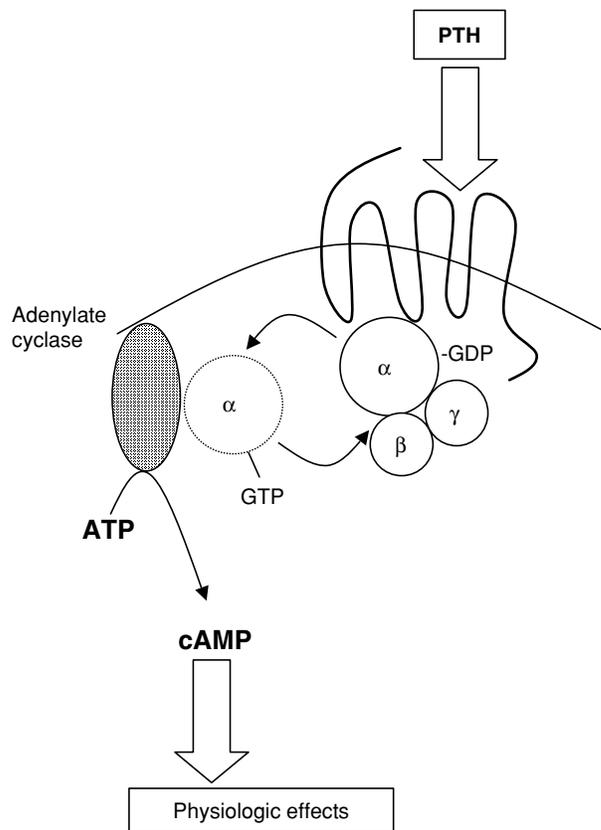
**Figure 1** Brachydactyly in Albright's hereditary osteodystrophy. This female patient with pseudohypoparathyroidism shows the typical foreshortened fourth and fifth metacarpals that result in brachydactyly.

in the renal tubules; it mobilizes calcium from the mineral store in the skeleton; and it increases serum calcium indirectly by promoting the renal conversion of 25-hydroxyvitamin D to calcitriol, which in turn increases intestinal calcium absorption. PTH is also an important regulator of phosphate metabolism, and it acts within the renal tubule to promote phosphate clearance. Calcitriol counterbalances this by promoting phosphate absorption from the gut.

The principal stimulus to PTH secretion is a decline in serum calcium that is detected by calcium-sensing receptors on the parathyroid cell surface. The same calcium-sensing receptor is also present within the renal tubule, where it acts to increase calcium



**Figure 2** Parathyroid hormone action. The receptor for parathyroid hormone can be found in a wide variety of tissues, although the calciotropic actions of PTH are dependent on its actions within the kidney and bone and indirectly via increased serum calcitriol concentrations that promote increased intestinal absorption of calcium.



**Figure 3** The PTH/PTHrP receptor couples to G proteins. PTH signal transduction via its receptor causes  $G_{s\alpha}$  protein to release GDP and bind GTP. Thereafter, GTP-bound  $G_{s\alpha}$  dissociates from its  $\beta\gamma$  subunits and continues to activate adenylate cyclase until hydrolysis of GTP occurs. Inactive  $G_{s\alpha}$  then reassociates with the  $\beta\gamma$  subunits. Not shown is the association of the PTH/PTHrP receptor with other G proteins known to account for some of its actions.

reabsorption when serum calcium decreases. Calcium homeostasis is therefore coordinately regulated by PTH, calcitriol, and calcium to ensure the strict maintenance of serum calcium within a narrow physiologic range.

### PTH Deficiency

A deficiency of PTH as seen in surgical or autoimmune hypoparathyroidism or in patients with congenital absence of parathyroid glands leads to hypocalcemia and hyperphosphatemia. Mild hypocalcemia may be asymptomatic, particularly if it develops slowly. Alternatively, acute development of hypocalcemia may precipitate carpopedal spasm, laryngeal spasm, and/or seizures and may be accompanied by two cardinal physical signs: facial nerve irritability elicited by

tapping over the facial nerve (Chvostek's sign) and carpal spasm that can precipitated by increasing the pressure of a tourniquet to above systolic pressure (Trousseau's sign). Chronic hypocalcemia may paradoxically be accompanied by calcification of soft tissues and/or the basal ganglia and frontal lobes.

## PTH Resistance

Resistance to PTH is defined biochemically as the association of hypocalcemia, hyperphosphatemia, and elevated serum PTH concentrations. Uncommonly, these biochemical features may be due to clinically silent malabsorption. Acquired PTH resistance also occurs in renal disorders (because of calcitriol deficiency, tubular dysfunction, and metabolic acidosis) and may occur in hypomagnesemia. Occasionally, hypocalcemia is seen in severe/catastrophic illnesses such as pancreatitis when the other clinical features overwhelm the biochemical problems. When secondary causes have been excluded, PTH resistance is usually due to congenital "pseudohypoparathyroidism," although some cases lack an identified family history of the condition. An etiological classification of pseudohypoparathyroid states is listed in Table I.

Table I Etiological Classification of Disorders Associated with PTH Resistance

Etiology	Disorder [Mendelian Inheritance in Man (MIM) No.]
Congenital	
Mutation in PTH receptor	Bloomstrand's chondrodysplasia (MIM 215045)
Mutation in <i>GNAS1</i>	
Maternal allele	Pseudohypoparathyroidism type Ia (MIM 103580)
Paternal allele	Pseudopseudohypoparathyroidism (MIM 300800)
Paternal allele	Progressive osseous heteroplasia (MIM 166350)
Methylation defect in <i>GNAS1</i>	
Maternal allele	Pseudohypoparathyroidism type Ib (MIM 603233)
Unknown	Pseudohypoparathyroidism type II
Acquired	
Deficient renal PTH action	Renal failure Magnesium deficiency Malabsorption
Unknown	Pseudohypoparathyroidism type III

## PSEUDOHYPOPARATHYROID STATES

### Concept of Hormone-Resistant States

The major historical advances in understanding pseudohypoparathyroid states are listed in Table II. Albright recognized that the hypoparathyroid state in some patients was not due to PTH deficiency but rather to target organ resistance to PTH. Pseudohypoparathyroidism thus became the first hormone-resistance syndrome recognized in humans. Resistance syndromes have since been described for many hormones. In steroid hormone-resistance syndromes (e.g., androgen-insensitivity syndrome), the causative defect often resides in the hormonal receptor, whereas resistance to peptide hormones (e.g., insulin) is more commonly associated with alterations in postreceptor signaling.

Subsequent efforts to understand PTH action were led by Aurbach, who observed that many of the actions of PTH are mediated by generation of cAMP. In normal subjects and in patients with PTH deficiency, acute bolus infusions of PTH extract or synthetic PTH (1–34) result in 50- to 100-fold increases in urinary cAMP but less dramatic effects on

Table II Historical Milestones in the Study of Pseudohypoparathyroidism

Year	Milestone
1942	Albright <i>et al.</i> describe the first cases of pseudohypoparathyroidism.
1952	Albright recognizes that the skeletal features of pseudohypoparathyroidism can occur without accompanying hypocalcemia and terms this condition pseudopseudohypoparathyroidism.
1957	Rall and Sutherland identify cAMP as a second messenger for several hormonal systems.
1967	Chase and Aurbach show that urinary cAMP increases rapidly in response to purified parathyroid hormone.
1980	Bourne and Aurbach separately show that $G_{\text{sec}}$ activity is reduced by approximately 50% in erythrocyte membranes from patients with pseudohypoparathyroidism type Ia or pseudopseudohypoparathyroidism.
1983	Davis and Hughes propose an imprinting model for inheritance of pseudohypoparathyroidism type Ia and pseudopseudohypoparathyroidism.
1988	The gene encoding $G_{\text{sec}}$ , <i>GNAS1</i> , is cloned.
1989	Patten and colleagues identify <i>GNAS1</i> mutation in a kindred with pseudohypoparathyroidism type Ia.
2000	Liu and colleagues identify methylation defects in the promoter region of <i>GNAS1</i> that are associated with pseudohypoparathyroidism type Ib.

phosphate clearance. In patients with pseudohypoparathyroidism, these responses are blunted or absent. This fundamental test of PTH action in combination with the presence of hypocalcemia and/or Albright's hereditary osteodystrophy became the basis for the phenomenological classification of pseudohypoparathyroid states discussed next.

### Pseudohypoparathyroidism Type Ia

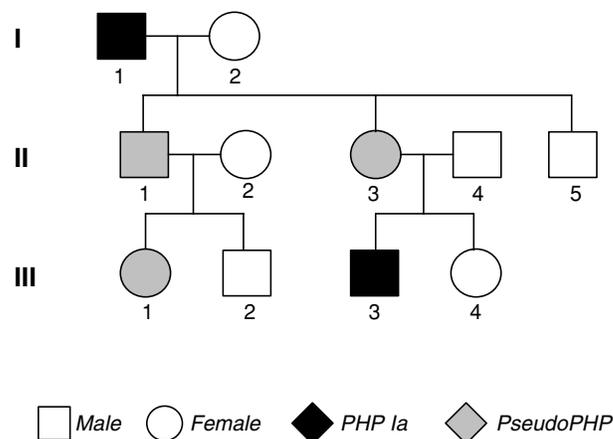
Pseudohypoparathyroidism type Ia [Mendelian Inheritance in Man (MIM) No. 103580] is classically defined as the combination of hypocalcemia and Albright's hereditary osteodystrophy, or "generalized" PTH resistance. This syndrome is known to result from heterozygous mutations in *GNAS1* that cause a reduction in or a malfunction of the  $G_{s\alpha}$  protein. The fact that G proteins are responsible for signaling by a number of other hormones explains the occurrence of hypothyroidism (TSH resistance), hypogonadism (gonadotropin resistance), and other signaling defects in some patients with pseudohypoparathyroidism type Ia. A 50% reduction in  $G_{s\alpha}$  activity may be demonstrated in erythrocyte membranes from these subjects, although this  $G_{s\alpha}$  deficiency appears to have no deleterious effect on red cell function.

Variable serum calcium concentrations in some affected individuals, ranging at times into the normal range, occasionally cause some difficulty in distinguishing this disorder from pseudopseudohypoparathyroidism. However, the response of nephrogenous cAMP to infused PTH is reliably blunted in pseudohypoparathyroidism type Ia, whereas pseudopseudohypoparathyroidism is associated with normal cAMP production.

### Pseudopseudohypoparathyroidism

Pseudopseudohypoparathyroidism (MIM 300800) is defined as the occurrence of Albright's hereditary osteodystrophy without hypocalcemia. The absence of biochemical PTH resistance is consistent with normal renal responsiveness to infused PTH in this condition. Although renal  $G_{s\alpha}$  activity is preserved in pseudopseudohypoparathyroidism, erythrocyte  $G_{s\alpha}$  bioactivity is reduced in a manner comparable to that in pseudohypoparathyroidism type Ia. Moreover,  $G_{s\alpha}$  deficiency in this disorder is associated with similar *GNAS1* mutations that cause pseudohypoparathyroidism type Ia. How does the same *GNAS1* mutation cause pseudohypoparathyroidism type Ia in one patient but pseudopseudohypoparathyroidism in another?

Remarkably, inheritance of  $G_{s\alpha}$  deficiency from an affected mother leads to pseudohypoparathyroidism type Ia, whereas paternal transmission of the defect leads to pseudopseudohypoparathyroidism (Fig. 4). The explanation for this phenomenon was provided by the discovery that some autosomal genes are expressed from only a single allele, determined by its parent-of-origin status, in a process called imprinting. We each inherit two copies of the genes on autosomal chromosomes, one from either parent. For imprinted genes, expression of only the maternal or paternal allele occurs in some tissues, and the other allele is silenced. In the renal tubule, only the maternal *GNAS1* gene is expressed; the paternal gene is imprinted and therefore silenced and not expressed. Thus, in pseudohypoparathyroid syndromes, a paternally inherited mutation is not expressed in the kidney and renal PTH resistance is not present. Conversely, in bone, both paternal and maternal alleles are equally expressed so that an inactivating mutation of *GNAS1* in either allele causes a 50% reduction in  $G_{s\alpha}$  protein. For reasons incompletely understood but probably related to the actions of PTH or PTH-related peptide (PTHrP) during the development of the cartilaginous growth plate, skeletal  $G_{s\alpha}$  deficiency may lead to a shortening of metacarpals and metatarsals as well as other features of Albright's hereditary osteodystrophy listed in Table III. Imprinting of the *GNAS1* locus adequately accounts for the complex inheritance pattern of pseudohypoparathyroidism type Ia and pseudopseudohypoparathyroidism (Fig. 5).



**Figure 4** Complex inheritance of pseudohypoparathyroidism. Hypothetical pedigree illustrating the occurrence of pseudohypoparathyroidism type Ia (PHP Ia) or pseudopseudohypoparathyroidism (pseudoPHP) depending on parent-of-origin inheritance of the *GNAS1* defect. For example, pseudoPHP is inherited from a father affected by either PHP Ia (I.1) or pseudoPHP (II.1), whereas PHP Ia is inherited from a mother with either condition (II.3).

**Table III Features of Albright's Hereditary Osteodystrophy**

Skeletal	
Short metacarpals (usually fourth and/or fifth)	
Short distal phalanx of thumb	
Frontal bossing (hyperostosis frontalis externa)	
Round face	
Short stature (short limb length)	
Advanced bone age	
Heterotopic ossification in skin	
Extraskeletal	
Obesity	
Dental abnormalities (dentin and/or enamel hypoplasia, delayed or absent tooth eruption)	
Developmental delay	

### Pseudohypoparathyroidism Type Ib

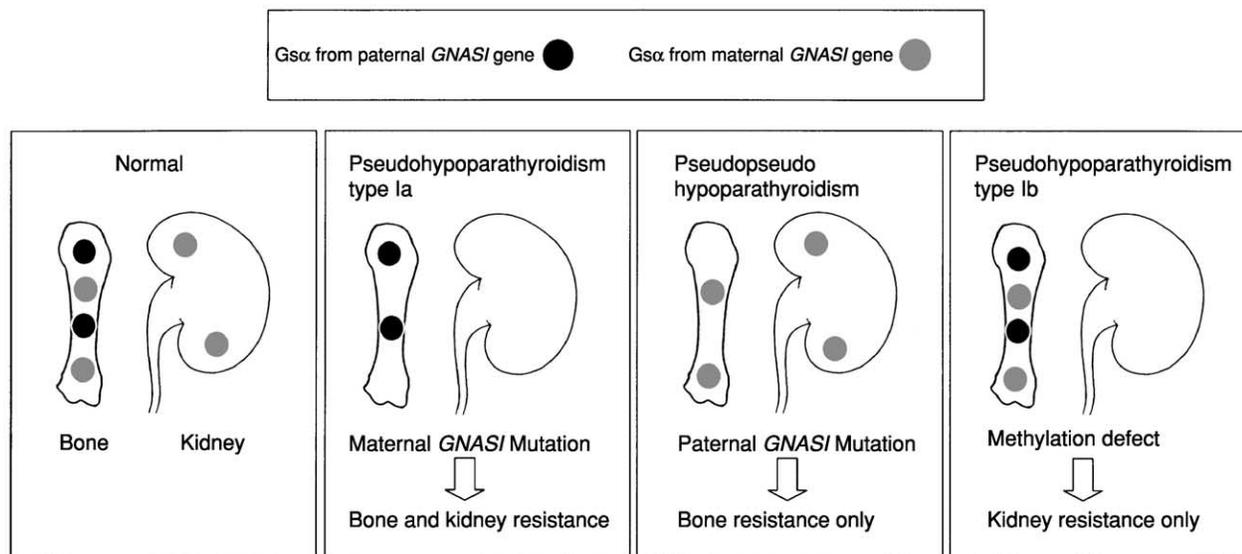
Pseudohypoparathyroidism type Ib (MIM 603233) is defined by selective resistance to PTH in the kidney without the features of Albright's hereditary osteodystrophy. PTH infusion in patients with this syndrome results in deficient stimulation of nephrogenous cAMP or urinary phosphate clearance, whereas  $G_{s\alpha}$  bioactivity in erythrocyte membranes is normal. Furthermore, *GNAS1* gene mutations are not found in pseudohypoparathyroidism type Ib. An early clue to understanding the etiology of this syndrome was the recognition that *GNAS1* is paternally imprinted as described previously, such that only the maternal *GNAS1* allele is

expressed in the kidney. Moreover, when inherited, pseudohypoparathyroidism type Ib is maternally transmitted in a manner similar to that of type Ia. Therefore, this syndrome is thought to result from defective  $G_{s\alpha}$  expression in the renal tubule.

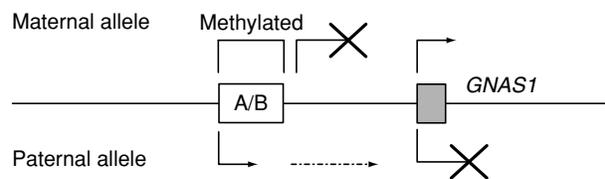
The molecular basis has been clarified with some surprising results. Affected patients show a loss of methylation in an upstream region of the maternal *GNAS1* allele (Fig. 6). In effect, both alleles have the unmethylated pattern usually associated only with the paternal allele. This is perhaps counterintuitive since DNA methylation within gene regulatory elements is typically associated with transcriptional silencing, whereas methylation of the *GNAS1* locus appears to promote renal  $G_{s\alpha}$  expression. The *GNAS1* gene locus is complex and encodes at least three different transcripts from alternative exons upstream of those that otherwise encode  $G_{s\alpha}$ . One possibility is that unmethylated DNA in this region allows expression of alternative mRNA transcripts at the expense of normal  $G_{s\alpha}$  expression in the proximal renal tubule. Attempts to map the causative gene in affected kindreds have identified a candidate locus close to the *GNAS1* gene on chromosome 20. Presumably, this gene, as yet unidentified, regulates the methylation status of *GNAS1* in a tissue-specific manner.

### Pseudohypoparathyroidism Type II

Pseudohypoparathyroidism type II is characterized by hypocalcemia and hyperphosphatemia in the presence



**Figure 5** Tissue-specific expression of  $G_{s\alpha}$  in relation to parent-of-origin and pseudohypoparathyroid phenotype. Only the maternal *GNAS1* allele is expressed in the proximal renal tubules. In bone, both maternal and paternal alleles are expressed.



**Figure 6** Methylation of exons upstream of *GNAS1* determines renal expression of  $G_{s\alpha}$ . Liu *et al.* and Bastepe *et al.* clarified the role of methylation of these alternate exons. In particular, when exon A/B is not methylated (and presumably therefore transcribed), expression of  $G_{s\alpha}$  in the kidney is inhibited. On the other hand, methylation of this exon permits renal  $G_{s\alpha}$  expression. In normal circumstances, only the maternal allele is methylated. In pseudohypoparathyroidism type Ib, neither allele is methylated at this locus (not shown).

of a normal increase in urinary cAMP but blunted phosphaturic response to infused PTH. This disorder has no clear genetic basis, and its cause remains obscure. Defective activation of targets downstream of cAMP production may be responsible, but this has not been established. The features of pseudohypoparathyroidism type II are present in some patients with vitamin D deficiency. Indeed, the ability of PTH to stimulate urinary phosphate clearance is at least partially vitamin D and/or calcium dependent. Vitamin D replacement in some patients with pseudohypoparathyroidism type II has been noted to restore their phosphaturic response to PTH.

### Pseudohypoparathyroidism Type III

The existence of a third subtype of pseudohypoparathyroidism remains controversial. Two independent case reports described hypocalcemia that was postulated to be the result of selective skeletal resistance to PTH. The merit behind this hypothesis is the knowledge that the calcemic actions of PTH are partly due to its actions on the skeleton, and therefore skeletal resistance to PTH might cause hypocalcemia. Nevertheless,  $G_{s\alpha}$  deficiency limited to the skeleton (pseudopseudohypoparathyroidism) is not accompanied by hypocalcemia. Although it is possible that defects in other second messenger pathways may contribute to a separate phenotype, no conclusive evidence for this has been presented.

### Mutations in the PTH Receptor

Although suspected, defects in the PTH receptor have not been implicated in the classical forms of pseudohypoparathyroidism. This receptor also responds to

PTHrP, a hormone that was first identified as the principal cause of humoral hypercalcemia of malignancy but that has important paracrine actions in the control of bone growth. Surprisingly, the first naturally occurring mutations identified in the PTH/PTHrP receptor caused activation of receptor signaling and were associated with a rare skeletal disorder, Jansen's metaphyseal chondrodysplasia (MIM 156400). Subsequently, inactivating mutations in the PTH/PTHrP receptor were identified that were associated with striking effects on skeletal development in Blomstrand chondrodysplasia (MIM 215045), a perinatally lethal condition characterized by advanced bone maturation and chondrocyte differentiation. This lethality probably explains the absence of PTH/PTHrP receptor gene mutations in any form of pseudohypoparathyroidism.

### Progressive Osseous Heteroplasia

Progressive osseous heteroplasia (MIM 166350) is a rare congenital disorder characterized by heterotopic ossification occurring primarily within the dermis, skeletal muscles, and deep connective tissue. Autosomal dominant inheritance has been described in some cases. A connection with pseudohypoparathyroidism was initially suggested by the presence of brachydactyly in some affected individuals and further strengthened by the description of one kindred in which paternal inheritance resulted in progressive osseous heteroplasia, whereas maternal inheritance resulted in Albright's hereditary osteodystrophy. Shore and colleagues identified *GNAS1* mutations in this and other kindreds affected by progressive osseous heteroplasia, and in each case the condition was associated with paternal transmission of the defect. Surprisingly, some of the mutations associated with progressive osseous heteroplasia are the same as those previously identified in patients with either pseudohypoparathyroidism type Ia or pseudopseudohypoparathyroidism. The presence of other genes or epigenetic mechanisms modifying the phenotype caused by *GNAS1* mutations has been invoked to explain this remarkable phenomenon, although the specific details are unclear.

### FUTURE CHALLENGES

Although substantial progress has been made toward determining the pathogenesis of PTH resistance, many uncertainties remain, including the following:

1. The pathophysiology of skeletal defects in Albright's hereditary osteodystrophy: PTH resistance

within bone cells may not necessarily explain the occurrence of skeletal defects in this condition since PTH-stimulated cAMP production from cultured bone cells has been reported to be normal. Perhaps the  $G_{s\alpha}$  deficiency in bone cells interrupts other signaling pathways or variably affects skeletal response to PTH or PTHrP during development.

2. The role of other second messenger systems known to be activated by PTH-receptor signaling in pseudohypoparathyroidism: In particular, it will be interesting to learn whether a specific phenotype is associated with impaired signal transduction via the phospholipase C/inositol triphosphate pathway.

3. The role of other genes modifying the phenotype of pseudohypoparathyroid syndromes, particularly the divergent outcomes of paternal *GNAS1* allele mutation in either pseudopseudohypoparathyroidism or progressive osseous heteroplasia.

4. The mechanism of *GNAS1* genetic imprinting: Further studies of pseudohypoparathyroid type Ib kindreds will likely identify the molecular defect responsible for defective methylation of the *GNAS1* locus. This information will help clarify the mechanism of tissue-specific expression of maternal or paternal *GNAS1* alleles.

Perhaps the most compelling mystery is why imprinting of the  $G_{s\alpha}$  gene occurs at all. It is unclear what teleological advantage could be gained by suppressing the expression of one copy of a gene that is clearly critical for calcium homeostasis. Further study of this syndrome will undoubtedly continue to be rewarding for admirers of the rich tapestry of endocrinology in the future.

## Acknowledgments

We are indebted to Sol Posen for critical review and helpful guidance in the preparation of the manuscript. Moreover, he has shaped our approach to metabolic bone disease both at the bedside and in the laboratory as the founding father of the academic study of calcium metabolism in Australia. He has been a mentor to A.M. for more than 30 years.

## See Also the Following Articles

Hyperparathyroidism, Primary • Hyperphosphatemia • Hypocalcemia, Therapy • Hypoparathyroidism • Parathyroid Cancer • Parathyroid Glands, Pathology • Parathyroid Hormone (PTH)

## Further Reading

- Albright, F., Burnett, C. H., and Smith, P. H. (1942). Pseudohypoparathyroidism: An example of "Seabright-Bantam syndrome." *Endocrinology* **30**, 922-932.
- Albright, F., Forbes, A. P., and Henneman, P. H. (1952). Pseudopseudohypoparathyroidism. *Trans. Assoc. Am. Phys.* **65**, 337-350.
- Bastepe, M., Pincus, J. E., Sugimoto, T., Tojo, K., Kanatani, M., Azuma, Y., Kruse, K., Rosenbloom, A. L., Koshiyama, H., and Juppner, H. (2001). Positional dissociation between the genetic mutation responsible for pseudohypoparathyroidism type Ib and the associated methylation defect at exon A/B: Evidence for a long-range regulatory element within the imprinted *GNAS1* locus. *Hum. Mol. Genet.* **10**, 1231-1241.
- Davis, S. J., and Hughes, H. E. (1993). Imprinting in Albright's hereditary osteodystrophy. *J. Med. Genet.* **30**, 101-103.
- Drezner, M. K., Neelon, F. A., and Lebowitz, H. E. (1973). Pseudohypoparathyroidism type II: A possible defect in the reception of the cyclic AMP signal. *N. Engl. J. Med.* **280**, 1056-1060.
- Farfel, Z., Brickman, A. S., Kaslow, H. R., Brothers, V. M., and Bourne, H. R. (1980). Defect of receptor-cyclase coupling protein in pseudohypoparathyroidism. *N. Engl. J. Med.* **303**, 237-242.
- Jobert, A. S., Zhang, P., Couvineau, A., Bonaventure, J., Roume, J., Le Merrer, M., and Silve, C. (1998). Absence of functional receptors for parathyroid hormone and parathyroid hormone-related peptide in Blomstrand chondrodysplasia. *J. Clin. Invest.* **102**, 34-40.
- Levine, M. A. (2001). Hypoparathyroidism and pseudohypoparathyroidism. In "Endocrinology" (L. J. DeGroot and J. L. Jameson, eds.), 4th ed., pp. 1133-1153. Saunders, Philadelphia.
- Levine, M. A., Schwindinger, W. F., Downs, R. W., Jr., and Moses, A. M. (1994). Pseudohypoparathyroidism: Clinical biochemical and molecular features. In "The Parathyroids" (J. P. Bilezikian, M. A. Levine, and R. M. Marcus, eds.), pp. 781-800. Raven Press, New York.
- Liu, J., Litman, D., Rosenberg, M., Yu, S., Biesecker, L., and Weinstein, L. (2000). A *GNAS1* imprinting defect in pseudohypoparathyroidism type Ib. *J. Clin. Invest.* **106**, 1167-1174.
- Mann, J. B., Alterman, S., and Hills, A. G. (1962). Albright's hereditary osteodystrophy comprising pseudohypoparathyroidism and pseudopseudohypoparathyroidism with a report of two cases representing the complete syndrome occurring in successive generations. *Ann. Internal Med.* **56**, 315-342.
- Patten, J. L., Johns, D. R., Valle, D., Eil, C., Gruppuso, P. A., Steele, G., Smallwood, P. M., and Levine, M. A. (1990). Mutation in the gene encoding the stimulatory G protein of adenylate cyclase in Albright's hereditary osteodystrophy. *N. Engl. J. Med.* **322**, 1412-1419.
- Shore, E. M., Ahn, J., Jan De Beur, S., Li, M., Xu, M., Gardner, R. J. M., Zasloff, M. A., Whyte, M. P., Levine, M. A., and Kaplan, F. S. (2002). Paternally inherited inactivating mutations of the *GNAS1* gene in progressive osseous heteroplasia. *N. Engl. J. Med.* **346**, 99-106.
- Tucci, J. R. (2001). Chronic hypocalcaemia due to selective skeletal resistance to parathyroid hormone. *Clin. Endocrinol.* **55**, 815-818.
- Weinstein, L. S., Yu, S., Warner, D. R., and Liu, J. (2001). Endocrine manifestations of stimulatory G protein  $\alpha$ -subunit mutations and the role of genomic imprinting. *Endocr. Rev.* **22**, 675-705.

elevated in estrogen-secreting tumors. Plasma testosterone levels are high in the case of virilizing tumors.  $17\alpha$ -hydroxyprogesterone, adrenal androgens, and metabolites are increased in virilizing congenital adrenal hyperplasia and adrenal cortex tumors. Some tumors may secrete Müllerian-inhibiting hormone and/or inhibin. Some may be caused by a loss of a tumor suppressor gene, whereas others may result from activating mutations of stimulatory G proteins.

## Etiology

The etiology of pseudoprecocious puberty is presented in [Table I](#). The diagnosis of ovarian tumors is made by radiological imaging of the pelvis. The majority of them are benign (e.g., granulosa cell tumor, teratoma, dermoid cyst) and unilateral. Benign follicular cyst is the most frequent tumor with isosexual pseudoprecocious puberty. Exploratory examination or laparoscopy may be necessary to recognize granulosa cell tumors. Tumors of the granulosa and theca are usually palpable; they secrete large amounts of

estrogens, in particular estradiol, and often progesterone (as in luteomas). The following up of plasma estradiol levels is useful for detecting metastases.

Rarely, ovarian tumors are gonadoblastomas that develop in the fibrous streaks of Turner syndrome or are lipid tumors. Teratomas, teratocarcinomas, and arrhenoblastomas can secrete estrogens, androgens, or both as well as hCG.

Recurrent follicular cysts may be one symptom of McCune–Albright syndrome.

## Management and Therapy

Management of pseudoprecocious puberty is summarized in [Table II](#).

Cancer markers, such as  $\alpha$ -feto-protein and carcinoembryonic antigen, are of great help in the case of malignancy, as they are increased, in particular hCG. Metastases of the lungs can be found in the case of malignancy.

In the case of ovarian or adrenal tumor, management consists of the ablation of the tumor. Ablation of a large cyst is generally indicated. If malignancy is present, radiotherapy and/or chemotherapy should be considered. In the case of late-onset congenital virilizing adrenal hyperplasia, hydrocortisone therapy is indicated.

**Table I Etiology of Pseudoprecocious Puberty**

<b>Isosexual (feminization)</b>
Ovarian tumors
Granulosa cell tumors
Thecal cell tumors
Teratoma
Arrhenoblastoma
Ovarian cysts
Adrenal cortex tumors
Administration of estrogens
hCG-secreting tumors
<b>Heterosexual (virilization)</b>
Congenital virilizing adrenal hyperplasia
Adrenal cortex tumors
Adrenal ectopic tissue in ovary
Ovarian tumors
Dysgerminoma
Yolk sac tumor
Choriocarcinoma
Gonadoblastoma
Teratoma
Arrhenoblastoma
Administration of androgens
Primary cortisol resistance (?)
<b>Partial pubertal precocity</b>
Premature adrenarche
Premature thelarche

## PARTIAL OR INCOMPLETE PRECOCIOUS PUBERTY

### Premature Pubarche or Premature Adrenarche

This condition represents premature and isolated development of pubic and/or axillary hair. This precocious development is due to premature maturation of the androgenic secretions of the adrenal cortex and/or premature changes in the sensitivity to androgens of the target tissue receptors. This condition is observed from 6 years of age. A higher frequency of this condition has been reported in girls born with intrauterine growth retardation. A moderate increase in height and weight is often observed, as is a moderately advanced bone age. The clitoris is usually normal or slightly increased in size. Plasma concentrations of adrenal androgens (e.g., dehydroepiandrosterone [DHEA], dehydroepiandrosterone sulfate [DHEA-S]) are moderately elevated in relation to the age of the girl and usually correspond to the bone age. In some cases, testosterone, androstenedione, and estrone are slightly elevated. Ovarian or adrenal tumors must be

Table II Diagnosis and Management of Pseudoprecocious Puberty

Possible diagnosis	Specific test(s)	E <sub>2</sub> /Adrenal androgens
<b>Pseudoprecocious puberty with feminization</b>		
Ovarian tumor	Imaging	High E <sub>2</sub>
Ovarian cyst	Imaging	High E <sub>2</sub>
Adrenal tumor	Imaging	High E <sub>2</sub>
Exogenous estrogen administration		Low E <sub>2</sub>
Phytoestrogens		Low E <sub>2</sub>
<b>Pseudoprecocious puberty with virilization</b>		
Congenital adrenal hyperplasia	High androgens and 17 $\alpha$ -progesterone	Prepubertal E <sub>2</sub> High androgens
Adrenal cortex tumor	Imaging	Prepubertal E <sub>2</sub> High androgens
Ovarian tumor	Imaging	Elevated E <sub>2</sub> High androgens
Exogenous androgen administration		High androgens Low E <sub>2</sub>

*Note.* E<sub>2</sub>, estradiol; Imaging, detection by ultrasonography, computed tomography, or magnetic resonance imaging. Low levels of pituitary gonadotropins (follicle-stimulating hormone and luteinizing hormone) are observed in all of these conditions.

excluded. A normal plasma level of 17 $\alpha$ -hydroxyprogesterone rules out a late-onset congenital adrenal hyperplasia. Evolution is usually benign, puberty occurs normally, and menarche happens at a normal age. However, in some cases, premature adrenarche leads to the development of polycystic ovary syndrome during adolescence or adulthood. No therapy is needed in the case of premature adrenarche.

### Premature Thelarche

This condition represents a premature isolated development of the breast that is usually observed between 1 and 3 years of age. Signs of estrogenization are always absent; there are no modifications of the areolae, labia minora, or labia majora. The vaginal mucosa has a nonsecreting aspect. There is no sexual hair. Growth is not accelerated, and bone age is not advanced. No sign of estrogenization can be seen on the vaginal smear. Ultrasonographic examination of the uterus and ovaries reveals a prepubertal status. In half of the cases, gynecomastia has been noticed during the neonatal period. Plasma estradiol levels are usually in the prepubertal range. Infrequently, a slightly elevated plasma concentration of estradiol is encountered, suggesting the possibility of some transitional forms of true precocious puberty. In these cases, the follow-up will permit the correct diagnosis. Basal concentrations

of LH and FSH are usually normal, although FSH can be slightly elevated in some cases. An exaggerated response of FSH is observed after GnRH stimulation. LH response remains prepubertal. These abnormalities could explain the premature thelarche. However, an increased sensitivity of the breast tissue to the circulating levels of estradiol at this particular age is not excluded. No therapy should be instituted; the spontaneous evolution is generally excellent, with disappearance of the breasts in most cases. In the vast majority of cases, puberty occurs at the normal age. Follow-up of these cases is necessary because precocious puberty may be observed a few years later in rare cases.

### Premature Menarche

This condition is defined by the occurrence of vaginal bleeding without any sign of pubertal development. In the absence of any pubertal sign, it suggests foreign body, vaginal infection, and/or sexual abuse, but it also suggests possible tumors of the genital tract (Table III). Malodorous vaginal discharge suggests the presence of a foreign body, which is the most frequent cause of vaginal bleeding before puberty. A hymenal opening larger than 5 mm, posterior notches, or lesions of the vulvae is compatible with sexual abuse. In rare cases, the origin of the vaginal bleeding remains unknown.

Table III Causes of Vaginal Bleeding before Puberty

<b>Nonendocrine</b>	
	Vaginitis
	Foreign body
	Vaginal tumor
	Local lesions
	Sexual abuse
<b>Endocrine</b>	
	Estrogen administration
	Ovarian tumor

### See Also the Following Articles

Constitutional Delay of Growth and Puberty (CDGP) • Delayed Puberty and Hypogonadism, Female • Precocious Puberty, Central (Female) • Pseudoprecocious Puberty, Male • Puberty: Physical Activity and Growth • Sexual Maturation, Female

### Further Reading

- Forest, M. G. (1990). Pituitary gonadotrophin and sex steroids secretion during the first two years of life. *In* "Control of the Onset of Puberty" (M. M. Grumbach, P. C. Sizonenko, and M. L. Aubert, eds.), pp. 451–478. Williams & Wilkins, Baltimore, MD.
- Grumbach, M. M., and Styne, D. M. (1998). Puberty: Ontogeny, neuroendocrinology, physiology, and disorders. *In* "Williams Textbook of Endocrinology" (J. D. Wilson, D. W. Foster, H. M. Kronenberg, and P. R. Larsen, eds.), 9th ed., pp. 1509–1626. W. B. Saunders, Philadelphia.
- Rosenfield, R. L. (2002). Puberty in the female and its disorders. *In* "Pediatric Endocrinology" (M. A. Sperling, ed.), pp. 455–518. W. B. Saunders, Philadelphia.
- Sizonenko, P. C. (1993). Precocious puberty. *In* "Pediatric Endocrinology, Physiology, Pathophysiology, and Clinical Aspects" (J. Bertrand, R. Rappaport, and P. C. Sizonenko, eds.), pp. 387–403. Williams & Wilkins, Baltimore, MD.
- Sizonenko, P. C., and Aubert, M. L. (1990). Pituitary gonadotropins, prolactin, and sex steroids: Secretion in prepuberty and puberty. *In* "Control of the Onset of Puberty" (M. M. Grumbach, P. C. Sizonenko, and M. L. Aubert, eds.), pp. 479–499. Williams & Wilkins, Baltimore, MD.

**Table 1** Possible Etiologies of Precocious Pseudopuberty in Boys

Genetic disorders and mutations causing Leydig cell or adrenal hyperplasia
Congenital virilizing adrenal hyperplasia (defects in cortisol biosynthesis)
Constitutively activating LH receptor mutations in Leydig cells
Constitutively activating Gs mutations
Androgen-producing tumors
Leydig cell tumors
Adrenocortical tumors (e.g., adenoma, carcinoma)
Adrenal rest tumors in the testis
Mixed sex cord stromal tumors
Gonadotropin-secreting tumors
Exogenous sex steroids or gonadotropins

choriocarcinoma, chorioepithelioma, dysgerminoma, hepatoblastoma, hepatoma, and teratoma. The secreted gonadotropin is hCG, which acts via the LH receptor and is easy to detect as a tumor marker because healthy boys do not produce hCG. LH and follicle-stimulating hormone (FSH) levels remain low in these boys. Both testes increase in size as in normal puberty. However, if the tumor is testicular, that causes asymmetric growth of the gonad and often irregularity on the surface. Treatment aims at the elimination of the tumor and depends on the location, size, and metastatic status of the tumor.

### Leydig Cell Tumors

Solitary Leydig cell tumors are also rare and account for only 1 to 3% of all testicular tumors. In children, the tumors are most often found between 5 and 10 years of age when the signs of precocious puberty become apparent. Testosterone production is gonadotropin independent; therefore, FSH and LH remain on the prepubertal level. The tumors are usually unilateral, causing asymmetric growth of the testes. Gynecomastia can be found in approximately 10% of these patients because either the tumor itself also produces estrogens or androgens are converted to estrogens by peripheral aromatase activity. Gynecomastia is more common in adults than in children.

Leydig cell tumors are always benign in children, whereas adults may have malignant tumors. The standard treatment is orchidectomy, but if the tumor is encapsulated, surgical enucleation is sufficient.

### Adrenal Rest Tumors

The interstitial compartment of the testis can also harbor adrenal rest tumors that are stimulated by a

chronically elevated adrenocorticotropic hormone (ACTH) level. This can happen in undertreated 21-hydroxylase deficiency or after adrenalectomy (Nelson's syndrome). Adrenal rest tumors closely resemble Leydig cell adenomas, but adrenal rest tumors are bilateral, whereas Leydig cell tumors are usually unilateral. Adequate glucocorticoid substitution causes regression of the adrenal rests in 75% of cases. The tumors are always benign, and glucocorticoid treatment inactivates them even when the tumors are still visible by ultrasound. If the hormonal activity cannot be controlled by glucocorticoids, antiandrogens and aromatase inhibitors can be used to prevent the effects of androgens. Sometimes even testis-sparing surgery can be considered.

### Mixed Sex Cord Stromal Tumors

Mixed sex cord stromal tumors can contain combinations of Leydig, Sertoli, granulosa, and theca cells. These rare tumors can be found at any age, and the most common endocrine manifestation is gynecomastia. These tumors are always benign in children and can be treated by orchidectomy.

## ADRENAL TUMORS

### Virilizing Adrenocortical Tumors

Functioning adrenocortical tumors in children are rare. The peak incidence occurs during the first decade of life, and tumors are found more often in girls than in boys. The neoplasms are usually unilateral. The majority of tumors secrete predominantly androgens, causing virilization and thereby pseudoprecocious puberty. Very rarely, the tumor secretes only glucocorticoids, whereas some tumors secrete both androgens and glucocorticoids, causing both virilization and cushinoid features.

Adrenocortical tumors can be either benign or malignant. Adrenocortical carcinomas in children have a favorable prognosis if they can be removed completely by surgery. However, these tumors are often aggressive, and the outcome is poor when the tumors cannot be eradicated. The adjuvant chemotherapy and radiotherapy have been of little value. Benign tumors are also treated surgically, and their prognosis is good.

Prognosis of the tumor is related to its staging. In stages I and II, the tumor shows neither invasion nor metastasis; therefore, it is easy to excise completely. Size of the tumor distinguishes stages I and II, in which the tumor is less than 5 cm and more than 5 cm, respectively. In stage III disease, there is local invasion without involvement of adjacent organs or

regional lymph nodes. Metastasis or spread to adjacent organs and lymph nodes is a sign of stage IV disease. Children's tumors are usually in stage I or II, thereby explaining the favorable prognosis compared with that of adult patients.

### Diseases Associated with Adrenocortical Tumors

Cancer genetics and studies on molecular mechanisms of tumorigenesis have advanced our knowledge regarding biology of adrenal tumors. [Table II](#) indicates some of the diseases that are associated with adrenocortical tumors. In Li–Fraumeni syndrome, the P53 tumor suppressor gene is mutated, and patients are susceptible to multiple cancers. Deregulation of imprinted genes located at chromosome 11p15.5 causes Beckwith–Wiedemann syndrome, which is characterized by omphalocele, macroglossia, hemihypertrophy, and gigantism. These patients are also susceptible to carcinogenesis in adrenal glands. Patients with Carney complex usually have multiple myxoid skin tumors and may have cardiac myxomas, Sertoli cell tumors, and adrenal tumors. Two genetic loci for Carney complex have been identified: one at chromosome 2p16 and the other at chromosome 17q23–q24. The Carney complex susceptibility gene in chromosome 17 was found to be the type I $\alpha$  regulatory subunit of protein kinase A. In multiple endocrine neoplasia type 1 (MEN1), the molecular defect is in the MEN1 locus in chromosome 11q13, and the syndrome often involves hyperparathyroidism, pituitary tumors, and pancreaticoduodenal tumors. McCune–Albright syndrome, with constitutively activating mutation of the Gs protein, can also manifest in the adrenal gland.

## DIAGNOSIS

### Hormonal and Physical Examination

Diagnostic workup of testicular and adrenal tumors includes analysis of both testicular and adrenal steroids, gonadotropins, and ACTH. Careful measurement

**Table II** Complex Tumor Diseases Associated with Adrenocortical Tumor

Beckwith–Wiedemann syndrome
Carney complex
Li–Fraumeni syndrome
McCune–Albright syndrome
Multiple endocrine neoplasia type 1

of the size of the gonads is important. Increased penile size and virilization in boys who have small testes (longitudinal axis less than 2 cm) indicates an adrenal source of androgens. Symmetric growth of the testes points to gonadotropin stimulation that may be tumor-derived hCG in the case of pseudopuberty or pituitary LH in the case of central precocious puberty. Testicular tumors may cause asymmetric growth of the testes. Testosterone, androstenedione, and dehydroepiandrosterone all are elevated in children with virilizing adrenocortical tumors, whereas the dehydroepiandrosterone sulfate level may overlap with the normal range. 17-hydroxyprogesterone may also be elevated, whereas glucocorticoid levels and estradiol are in the normal range unless the patient has a mixed type of tumor secreting both androgens and cortisol. Testicular Leydig cell tumors may secrete both testosterone and estradiol.

### Imaging Studies

Ultrasonography is a good technique for finding both testicular and adrenal tumors. Visualization of the adrenal glands may be difficult for technical reasons (e.g., obesity, intestinal gas). The use of different positions and oblique scanning planes allows accurate imaging.

Computed tomography (CT) is an excellent method for evaluating adrenal tumors. Thin-section CT scanning finds even the small nodules (3 mm). Homogeneity of contrast enhancement, size of the lesion, and regularity of the borders help in distinguishing between malignant and benign tumors. CT is a sensitive method, but its specificity is limited. Magnetic resonance imaging (MRI) may help in tissue-specific diagnosis of adrenal masses. Chemical-shift MRI has been shown to have good accuracy in distinguishing between benign and malignant adrenal tumors. CT and MRI can be complemented with functional localization techniques such as adrenocortical scintigraphy, which is based on accumulation of radiocholesterol into the adrenal cortex. The radiocholesterol that can be used is  $^{131}\text{I}$ -6- $\beta$ -iodomethyl-19-norcholesterol (NP-59). This technique makes it possible to analyze the functional status of the tumor. There is little use for native X-ray imaging, although malignant tumors in the adrenal gland may sometimes have calcifications.

### See Also the Following Articles

Beckwith–Wiedemann Syndrome (BWS) • Constitutional Delay of Growth and Puberty (CDGP) • Gonadotropin-Secreting Tumors • Gynecomastia • Pseudoprecocious

Puberty, Female • Puberty: Physical Activity and Growth  
• Sexual Maturation, Male

### Further Reading

- Cheville, J. C. (1999). Classification and pathology of testicular germ cell and sex cord-stromal tumors. *Urol. Clinics North America* **26**, 595–609.
- Fragoso, M. C., Latronico, A. C., Carvalho, F. M., Zerbini, M. C., Marcondes, J. A., Araujo, L. M., Lando, V. S., Frazzatto, E. T., Mendonca, B. B., and Villares, S. M. (1998). Activating mutation of the stimulatory G protein (gsp) as a putative cause of ovarian and testicular human stromal Leydig cell tumors. *J. Clin. Endocrinol. Metab.* **83**, 2074–2081.
- Liu, G., Duranteau, L., Carel, J. C., Monroe, J., Doyle, D. A., and Shenker, A. (1999). Leydig-cell tumors caused by an activating mutation of the gene encoding the luteinizing hormone receptor. *N. Engl. J. Med.* **341**, 173–176.
- Mishra, A., Agarwal, G., Misra, A. K., Agarwal, A., and Mishra, S. K. (2001). Functioning adrenal tumours in children and adolescents: An institutional experience. *Austral. N. Zeal. J. Surg.* **71**, 103–107.
- Patil, K. K., Ransley, P. G., McCullagh, M., Malone, M., and Spitz, L. (2002). Functioning adrenocortical neoplasms in children. *BJU Int.* **89**, 562–565.
- Skakkebaek, N. E., Rajpert-De Meyts, E., and Toppari, J. (n.d.). Testicular cancer: Pathogenesis, diagnosis, endocrine aspects. [www.endotext.org](http://www.endotext.org).
- Ulbright, T. M., Amin, M. B., and Young, R. H. (1999). “Atlas of Tumor Pathology: Tumors of the Testis, Adnexa, Spermatic Cord, and Scrotum.” Armed Forces Institute of Pathology, Washington, DC.
- Wajchenberg, B. L., Albergaria Pereira, M. A., Medonca, B. B., Latronico, A. C., Campos Carneiro, P., Alves, V. A., Zerbini, M. C., Liberman, B., Carlos Gomes, G., and Kirschner, M. A. (2000). Adrenocortical carcinoma: Clinical and laboratory observations. *Cancer* **88**, 711–736.



Puberty, Delayed

see Delayed Puberty

which is characterized by an absence of GnRH production, failure to enter puberty, and anosmia.

In the normal male fetus, GnRH, follicle-stimulating hormone (FSH), and LH production begin in the brain by week 10, and the anterior pituitary releases gonadotropins by weeks 11 to 12. During the second trimester of gestation, the GnRH–gonadotrope–testicular axis is very active, leading to enlargement and growth of the already differentiated penis and testes, hence the normal sizes of these structures in the healthy full-term newborn male. Failure of the GnRH–gonadotrope component of the axis during this critical midgestational period results in isolated unambiguous microgenitalia in male infants (i.e., micropenis and hypoplastic testes). The differential diagnoses include absence of the GnRH neuron, birth defects or genetic mutations involving the hypothalamic–pituitary area, and failure of gonadotrope development or function (e.g., Prop-1 mutations). In contrast, primary testicular abnormalities, androgen resistance, or lack of 5 $\alpha$ -reductase will result in ambiguous genitalia or sex reversal in 46,XY infants (i.e., more female-appearing genitalia). Placental human chorionic gonadotropins (hCG) during the first trimester stimulate Leydig cells to produce androgens that require functional androgen receptors and conversion to 5 $\alpha$ -dihydrotestosterone. These mechanisms control the differentiation of the penis. The GnRH neuron returns to a state of quiescence during the third trimester of gestation.

### Postnatal Period

Within a few minutes following delivery, the circulating levels of LH in the human male neonate increase rapidly and are followed by an increase in serum testosterone concentrations during the first 3 to 21 h of postnatal life. FSH levels are slightly elevated in males during the first 3 months of life and then decline. Serum testosterone levels remain high for the first 2 to 4 months of life. By 6 months of postnatal life, the high LH and testosterone levels decline and remain low throughout childhood, indicating that the hypothalamic–pituitary–testicular axis is quiescent during this period because of decreased GnRH production.

The female fetus has lower estrogen concentrations during midgestation, in contrast to the high testosterone levels present in males at the same gestational age. The female fetus has higher pituitary gonadotropin content and higher circulating levels of FSH and LH than do males at a similar age. Castration of male fetal monkeys results in higher serum gonadotropin levels,

similar to those present in the intact fetal female at a similar age. This evidence suggests that gonadal sex steroids do have a negative feedback effect on the active GnRH neuron in fetal life.

Gonadal steroids are not the cause of the inactivity observed in the GnRH neuron during the childhood years. Prepubertal children with gonadal absence or dysgenesis lack gonadal steroids and have low levels of gonadotropins, similar to those present in eugonadal individuals. However, strikingly high levels of LH and FSH are seen in children with primary gonadal failure during the first months of life when the GnRH neuron is in its active phase.

### Childhood

The decreased activity of the GnRH neuron during childhood is a relative phenomenon based on information provided by new highly sensitive assays that have detected low-amplitude pulses of LH and FSH during this period. As the child ages, prior to the appearance of any physical signs of puberty, slightly higher nocturnal gonadotropin pulses have been detected.

### Puberty

The hormonal patterns observed throughout puberty are the result of gradual changes in the development of the GnRH neurosecretory system. Increases in GnRH pulse amplitude initially becomes evident during the interval period between the end of childhood and the onset of puberty. Following the onset of puberty, first there is augmentation of GnRH pulse amplitude, increased FSH and LH pulse amplitude, and enhanced nocturnal gonadotropin release. Soon thereafter, an increase in GnRH pulse frequency develops. Initially, the mean FSH levels increase, and this is followed by an increase in mean LH levels. As puberty progresses, mean basal LH levels increase, the amplitude and frequency of LH pulses increase, and nocturnal LH secretion increases. The nocturnal increase in GnRH release is particularly prominent during midpuberty. Subsequently, GnRH secretion and gonadotropin pulses during daytime hours are similar to the nocturnal pattern.

Attainment of the adult stage of GnRH pulsatile release is a prerequisite for the positive feedback effect of estrogen during puberty. In primates, estrogen stimulates an increase in GnRH release and also directly enhances LH secretion from the gonadotrope.

Gonadarche involves maturation of the testes during puberty, whereas adrenarche refers to maturation of the reticularis zone of the adrenal gland and

production of weaker adrenal androgens such as dehydroepiandrosterone and its sulfated form (DHEA-S). Adrenarche occurs at an earlier age (i.e., beginning at 6–7 years), is not linked to the GnRH–gonadotrope–testicular axis, and is not a prerequisite for pubertal development.

## MECHANISMS CONTROLLING THE ONSET OF PUBERTY: TWO HYPOTHESES

### Gonadostat Hypothesis

The gonadostat hypothesis was initially proposed during the 1930s and was based on the belief that onset of puberty resulted from a differential sensitivity of the pituitary to the feedback effects of ovarian sex steroids. According to this hypothesis, puberty begins when gonadotropin secretion becomes insensitive to the negative feedback effect of gonadal steroids. This hypothesis lacked applicability to primates because it was evident that decreased sensitivity to the negative feedback of estrogens on gonadotropin secretion coincided with the midpubertal stage rather than with the actual onset of puberty. The hypothesis had been based on studies in castrated immature and adult rats; suppression of gonadotropins was achieved with a smaller dose of estrogen in immature castrated rats than was the case in adult castrated rats. The role of the GnRH neuron was not appreciated at that time.

### Central Inhibition of GnRH Release Hypothesis

The central inhibition of GnRH release hypothesis states that childhood is due to central inhibition independent of gonadal sex steroids. Removal of CNS inhibition of the GnRH neuron leads to an increase in pulsatile GnRH that triggers the onset of puberty. The evidence for this hypothesis includes the following. First, gonadectomized monkeys and humans (e.g., gonadal dysgenesis) have low levels of gonadotropins during childhood and very elevated levels that coincide with periods of increased GnRH activity (e.g., during the postnatal and pubertal periods). Second, direct measurement of stalk–median eminence levels of GnRH in rhesus monkeys show low levels of GnRH if gonadectomy is done during the prepubertal stage and high levels if gonadectomy is done during puberty. Third, estrogen injections into gonadectomized prepubertal primates has no effect on GnRH release.

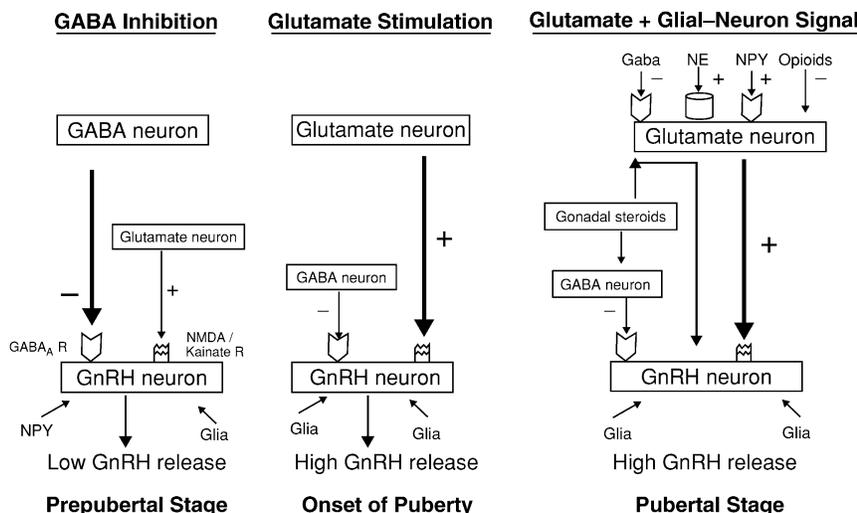
In summary, the onset of puberty is triggered by maturational changes in the hypothalamic mechanisms that control GnRH release independent of gonadal sex steroids. Evidence indicates that the GnRH neurosecretory system is actually more sensitive to ovarian feedback, in contrast to the assumptions proposed in the gonadostat hypothesis. Grumbach and colleagues suggested that gonadal steroids may increase GnRH secretion at the onset of puberty in children, but further investigation is needed to clarify whether subtle species differences exist in humans versus rhesus monkeys.

## NEUROBIOLOGY OF PUBERTY

The GnRH neuron intrinsically is able to secrete GnRH pulses but is incapable of triggering the onset of puberty by itself. GnRH neuronal activity is controlled by inputs from neuronal and astroglial-neuronal networks that are functionally connected. The regulatory processes controlling the input to the GnRH neuron appear to constitute the “central drive” of puberty. Fundamentally, these processes inhibit GnRH production during childhood. Over time, the inhibition is lifted and the dominant input involves stimulatory signals to the GnRH neuron. The central drive is regulated by three major events: two that are transsynaptic and one that involves a glial-to-neuron pathway. The following discussion summarizes these mechanisms.

### Inhibitory Neurotransmitters

Decaval and van der Pol reported that  $\gamma$ -aminobutyric acid (GABA) is the dominant inhibitory neurotransmitter in the hypothalamus that suppresses GnRH release during childhood. Disinhibition involves removal of GABA's negative influence on the GnRH neuron and appears to be a prerequisite for the onset of puberty in primates. GABA is synthesized in GABAergic neurons from glutamate by decarboxylation in the presence of glutamic acid decarboxylase (GAD 65 and GAD 67). It is stored in vesicles and released by exocytosis following depolarization by extracellular  $\text{Ca}^{2+}$ . GAD 65 and GAD 67 are also located in pancreatic beta cells, and elevated GAD 65 antibody titers occur in patients with type 1 diabetes mellitus. GABA synthesis also occurs in perikarya and dendrites but not in glial cells. The latter neurons influence GABA metabolism by providing precursors



**Figure 1** Diagram showing the prepubertal stage, the onset of puberty, and the pubertal stage. During the prepubertal stage, GABAergic mechanisms in primates predominate and the GnRH neurons are inhibited by GABA working directly through the GABA<sub>A</sub> receptors and indirectly by suppressing glutamate neurons. At the onset of puberty, GABA release is inhibited and glutamate neurons are activated, resulting in a dominant glutamatergic stimulation of GnRH neurons. Reduced production of GABA and increased production of glutamate induce the pubertal augmentation in pulsatile GnRH secretion that enhances the pulsatile release of gonadotropins. The ensuing gonadarche results in increased sex steroid production. Glial-to-neuron signaling complements the glutamatergic process. During late puberty, glutamatergic neurons and other signaling pathways (e.g., glial-to-neuron, growth factors, NE, NPY, sex steroids) lead to the attainment of adult GnRH neuronal function.

and degrading overflows of GABA from the synapse. GABA signals through the GABA<sub>A</sub> receptor (Fig. 1).

Other inhibitory neurotransmitters (e.g., opioids, neuropeptide Y [NPY], and melatonin) do not appear to significantly suppress the GnRH neuron during childhood.

### Stimulatory Neurotransmitters

The amino acid glutamate (L-glutamic acid) is the main excitatory neurotransmitter in the hypothalamus, and the GnRH neuron directly receives glutamatergic innervations (glutamatergic system). Enhanced glutamatergic neurotransmission is the main excitatory transsynaptic signal that increases pulsatile GnRH secretion and triggers the onset of puberty. Glutamatergic neurons control GnRH secretion via ionotropic the glutamate receptors NMDA and kainite. The latter is mainly on the GnRH neuron, whereas the former is located on interneurons that are synaptically connected to the GnRH neuron.

In addition, other possible stimulatory neurotransmitters include NPY and norepinephrine (NE), but the evidence is mixed and needs clarification.

### Role of Glial-to-Neuron Signaling and the Influence of Growth Factors

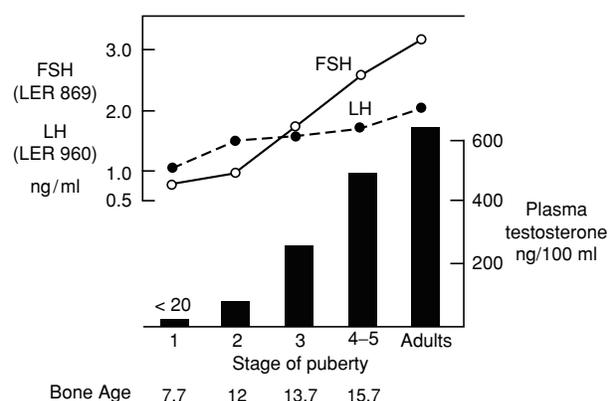
Astroglial cells and the growth factors released by glial cells appear to play an important role in stimulating GnRH release at the onset of puberty. Glia increases the GnRH pulsatility by synthesizing growth factors (transforming growth factors- $\alpha$  and - $\beta$  [TGF- $\alpha$  and TGF- $\beta$ ]), epidermal growth factor (EGF), insulin-like growth factor-1 (IGF-1), basic fibroblast growth factor (BFGF), neuron cell adhesion molecules (NCAM), cytokines (interleukins-1 and -6 [IL-1 and IL-6]), and (NO). Also, astroglia induces plastic rearrangements within the median eminence where there is an abundance of neuroterminals. Lastly, astroglial cells activate specific glial-to-neuron and glial-to-glia signaling pathways that involve a number of the growth factors that increase GnRH release. These signaling pathways are set in motion by ligand-dependent activation of astroglial ErbB tyrosine kinase receptors. Although glutamatergic neurotransmission and astroglial-mediated signaling increase GnRH secretion, the functional relationship of these two regulatory pathways has not been defined.

In conclusion, great progress has been made in our understanding of the complex controls on GnRH release, but the precise mechanisms that initiate the inhibition and stimulation of the GnRH neuron have not been defined.

### PUBERTAL DEVELOPMENT IN BOYS: CLINICAL FEATURES

Tanner and colleagues established standards for the stages of puberty, with Tanner 1 representing the prepubertal phase and Tanner 5 representing adulthood. The relationships between the Tanner stages and serum levels of testosterone, FSH, and LH are illustrated in Fig. 2.

The first physical sign of puberty in boys is growth of the testes, which is defined as a length greater than 2.5 cm or a volume greater than 4 ml. The range of age at onset of puberty is 9 to 14 years (mean age = 11.3 years), and pubertal development is completed in a mean of 3.5 years (range = 2.0 to 4.5 years). Increased testicular volume is brought about mainly by FSH stimulation of the seminiferous tubules that manifest increasing lumen diameter. The interstitial and Leydig cells do not contribute significantly to testicular volume. Prior to onset of puberty, Sertoli cells predominate, whereas later the main cells in the seminiferous tubules are germ cells. African American boys have been reported to begin puberty earlier than white and Hispanic boys. The pace of pubertal progression is variable among healthy boys. Genetics, nutrition,



**Figure 2** Mean circulating levels of FSH, LH, testosterone, and bone age in healthy boys for each stage of puberty. Stage 1 is prepubertal and stage 5 is adult. Baseline levels of FSH and LH increase as puberty progresses, with the FSH level becoming higher than the LH level. However, when GnRH is administered, the peak level of LH is higher than that of FSH, and this finding is evidence of central GnRH control of the gonadotrope.

and freedom from chronic disease are the dominant influences on the age at onset of puberty.

Spermatogenesis appears between 11 and 15 years of age, and sperm is detected in first morning urine at a mean age of 13.3 years. Spermaturia is evidence of spermatogenesis, but the adult stage of sperm maturation is not reached until bone age equals 17 years. Mean age of ejaculation is 13.5 years. The onset of spermatogenesis occurs early in puberty prior to the attainment of peak growth velocity.

Growth of pubic hair begins at a mean age of 11.2 years and reaches adult density at the mean age of 15 years. Axillary hair appears at a mean age of 14 years, although African American boys show this feature at a mean age of 12 years.

Prior to onset of puberty, the growth rates of boys and girls are similar. The growth rate during puberty is greater than that during childhood but is less than the rapid velocities observed during infancy and fetal life. Boys reach peak velocity 2 years later than do girls, and boys are taller at onset of the pubertal growth spurt phase. Peak growth rate occurs at Tanner stage 3 to 4 in boys or at the bone age of 14 years (13.5–14.0). Boys gain approximately 28 cm, and girls gain approximately 25 cm, during puberty. The height difference in adult men and women is due to the taller baseline height, the later age at onset of puberty, and the greater pubertal height gain in boys.

The hormones that control the growth spurt of puberty involve increased production of growth hormone (GH), elevated levels of IGF-1 (GH-IGF-1 axis), sex steroids (mainly estrogens), and nutrition. Estrogens stimulate bone maturation and closure of the epiphyseal growth plate. Bone mineral density increases during puberty if the calcium intake is adequate and if nutrition and pubertal hormones are normal. Peak bone mass is reached at 17 years of age in boys and is reduced in individuals with delayed puberty.

Body composition and energy requirements change greatly during puberty due to the hormonal influences. Men have 1.5 times the lean body mass and 1.5 times the skeletal mass compared with women. The greater strength of men, as compared with women, is due to the increased muscle mass present. Excess weight gain due to overnutrition and lack of exercise are increasing the national prevalence of obesity, type 2 diabetes mellitus, hypertension, and heart disease in young men and women.

Striking changes occur in behavior, mood, and sexuality during puberty. The hormonal changes that are indicative of puberty involve increasing serum levels of testosterone, FSH, and LH. The

GnRH stimulation test is essential for distinguishing gonadotropin-dependent or central puberty from gonadotropin-independent puberty. The pubertal response to GnRH is characterized by a dominance of LH over FSH after administration of GnRH versus the prepubertal pattern that shows a higher FSH peak than LH after GnRH administration. Gonadotropin-independent puberty is characterized by low or negligible rises in FSH and LH. GnRH analogues completely desensitize the gonadotrope and suppress gonadotropin responses because they continuously occupy the GnRH receptor, in contrast to the physiological episodic stimulation of the receptor that occurs in nature.

## CONCLUSIONS

The essential components of puberty in boys involve the following. First, the GnRH neurons migrate from the olfactory placode to the hypothalamus, where they are dispersed even though their activity is functionally synchronized. Second, the GnRH neuron *in vitro* exhibits intrinsic pulsatility of GnRH release. Third, the GnRH neuron is integrated into a functional network that controls its activity throughout life (i.e., the GABAergic, glutamatergic, and glial-to-neuron signaling systems that either suppress or enhance pulsatile GnRH release). Fourth, GnRH and gonadotropin pulsatility are essential for gonadarche. Fifth, the morphologically diffuse yet highly synchronized neuronal network that regulates GnRH release has been termed the GnRH pulse generator. Sixth, the GnRH–gonadotrope–testicular axis is active during midgestation and postnatal life, followed by period of quiescence due to inhibition of the GnRH neuron. Seventh, the onset of puberty is marked by removal of the CNS inhibitory influences on the GnRH neuron and by the dominance of stimulatory signals that have a diurnal pattern during early to mid-puberty, followed by GnRH pulsatility that is similar in the daytime and nighttime periods. Eighth, during puberty, the GnRH pulses develop greater amplitude and frequency.

Ninth, gonadarche is independent of adrenarche and is characterized by increased sex hormone biosynthesis and spermatogenesis.

## See Also the Following Articles

Delayed Puberty and Hypogonadism, Male • Delayed Puberty, Male • FSH (Follicle-Stimulating Hormone) • Gonadotropin-Releasing Hormone (GnRH) Actions • Kallmann's Syndrome and Idiopathic Hypogonadotropic Hypogonadism • LH (Luteinizing Hormone) • Precocious Puberty, Central (Male) • Pseudoprecocious Puberty, Male • Puberty: Physical Activity and Growth

## Further Reading

- Decaval, C., and van der Pol, A. N. (1990). GABA: A dominant neurotransmitter in the hypothalamus. *J. Comp. Neurol.* **302**, 1019–1037.
- Dziedzic, B., Prevot, V., Lomniczi, A., Jung, H., Cornea, A., and Ojeda, S. R. (2003). Neuron-to-glia signalling mediated by excitatory amino acid receptors regulates ErbB receptor function in astroglial cells of the neuroendocrine brain. *J. Neurosci.* **23**, 919–926.
- Grumbach, M. M., and Styne, D. M. (2003). Puberty: Ontogeny, neuro-endocrinology, physiology, and its disorders. In "Williams Textbook of Endocrinology" (P. R. Larson, H. M. Kronenberg, S. Melmed, and K. S. Polonsky, eds.), 10th ed., pp. 1115–1286. W. B. Saunders, Philadelphia.
- Knobil, E. (1990). The GnRH pulse generator. *Am. J. Obstet. Gynecol.* **163**, 1721–1727.
- Krsmanovic, L. Z., Stojilkovic, S. S., and Catt, K. J. (1996). Pulsatile gonadotropin-releasing hormone release and its regulations. *Trends Endocrinol.* **7**(2), 56–59.
- Lee, P. A. (1999). Central precocious puberty. *Endocrinol. Metab. Clinics North America* **28**, 901–918.
- Marshall, W. A., and Tanner, J. M. (1986). Puberty. In "Human Growth" (F. Falkner and J. M. Tanner, eds.), vol. 4, pp. 171–209. Plenum, New York.
- Ojeda, S. R., and Terasawa, E. (2002). Neuroendocrine regulation in puberty. In "Hormones, Brain, and Behavior" (D. Pfaff, A. Arnold, A. Etgen, S. Fahrbach, and R. Rubin eds.), Vol. 4, pp. 589–659. Academic Press, San Diego.
- Styne, D. M. (1991). Puberty and its disorders. *Endocrinol. Metab. Clinics North America* **20**, 43–69.
- Terasawa, E., and Fernandez, D. L. (2001). Neurobiological mechanisms of the onset of puberty in primates. *Endocrine Rev.* **22**, 111–151.

## GROWTH AND MATURATION OF ATHLETES IN WEIGHT-CONTROL SPORTS

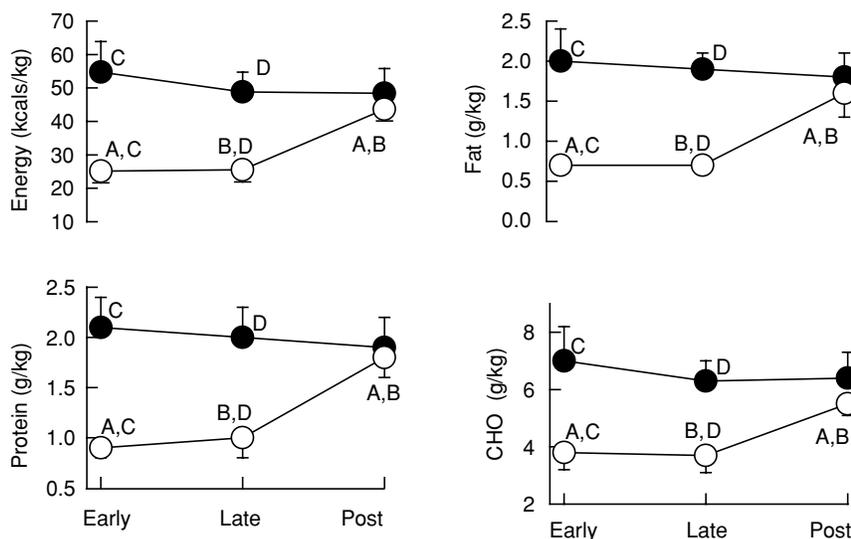
Physical activity is an important component of a healthy lifestyle. It produces many physiologic benefits, including increased strength, bone mineralization, favorable blood lipid concentrations, and improved motor control. It also produces psychological benefits, such as a sense of competence and well-being, increased confidence and self-esteem, and social development. For almost all youth, physical activity and sport training have no effect on growth or maturation. Youth who participate in sports are normal or slightly advanced in their rates of growth and maturation, but these traits are probably self-selected because strength and power advantages due to earlier maturation are major factors that attract children to sports and also influence their success in sports.

Only a small number of athletic youth experience a combination of great energy expenditure and restricted energy intake that slows growth and maturation. The remainder of this article focuses on the growth and maturation of two of these athletic groups, male wrestlers and female gymnasts.

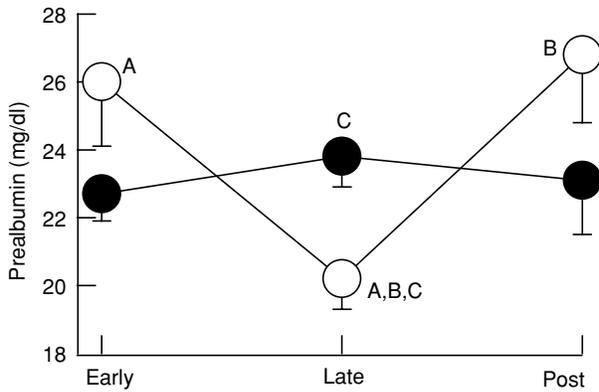
## Wrestling

Wrestlers compete in weight classes and weigh in before each competition to certify that their body weight is within the limits of the weight class. In the hope of gaining a competitive advantage by competing in a lower weight class against a younger and less experienced opponent, some wrestlers reduce their body weight through a combination of exercise and dietary restriction. The typical wrestler practices 2–2.5 h per day, 5 or 6 days per week and, after correction for the basal metabolic rate, expends approximately 800 kcals per practice. During a 3- or 4-month season, wrestlers may have nutrient intakes 50% below recommended amounts (Fig. 1). The amount of nutrient intake is not adequate to meet the combined requirements of resting metabolism, sport training, and growth and results in reduced protein nutritional status (Fig. 2).

This undernutrition produces low testosterone and free testosterone concentrations while maintaining serum luteinizing hormone (LH) concentrations (Fig. 3). In a well-nourished individual, testosterone produces negative feedback on LH secretion. Low testosterone concentrations in wrestlers should result in elevated serum LH concentrations. The lack of



**Figure 1** Relative intake of energy (kcal/kg/day), protein (g/kg/day), fat (g/kg/day), and carbohydrate (CHO; g/kg/day) for wrestler (○;  $n=9$ ; age,  $15.4 \pm 0.3$  years) and control (●;  $n=7$ ; age,  $15.0 \pm 0.4$  years) groups. All subjects were tested in mid-November, 1 or 2 weeks prior to the first wrestling match (Early), 3.5–4 months later (Late) depending on when the wrestlers no longer qualified for tournament competition, and 3.5–4 months after the wrestling season ended (Post). Like letters are significantly different. Data are expressed as mean ( $\pm$  SE). Redrawn from Roemmich, J. N., and Sinning, W. E. (1997). Weight loss and wrestling training: Effects on nutrition, growth, maturation, body composition, and strength. *J. Appl. Physiol.* **82**, 1751–1759.

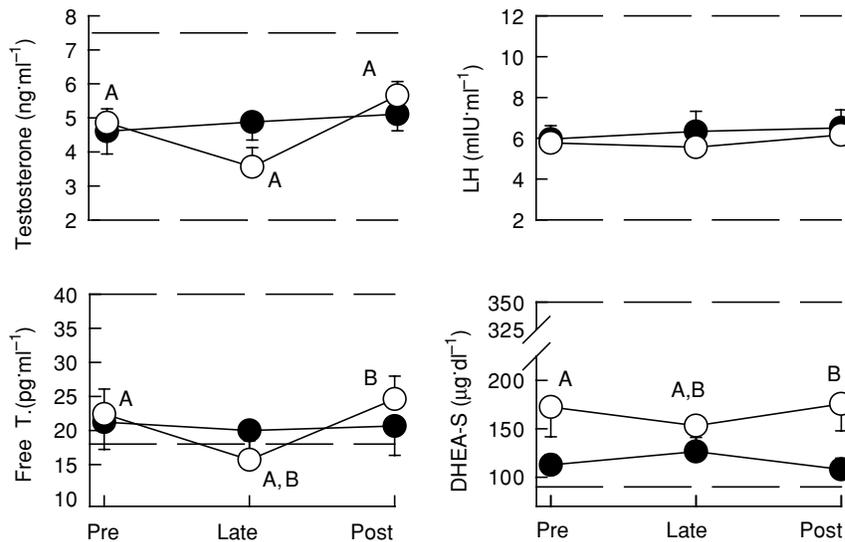


**Figure 2** Serum prealbumin concentration for the wrestler (○;  $n = 9$ ) and control (●;  $n = 7$ ) groups. Like letters are significantly different. Data are expressed as mean ( $\pm$  SE). For definitions, see the legend to Fig. 1. Redrawn from Roemmich, J. N., and Sinning, W. E. (1997). Weight loss and wrestling training: Effects on nutrition, growth, maturation, body composition, and strength. *J. Appl. Physiol.* **82**, 1751–1759.

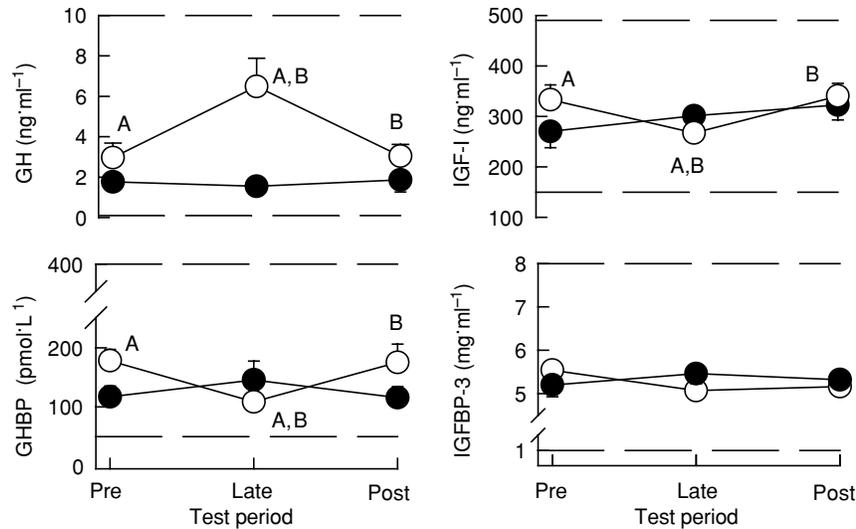
increased LH secretion suggests a central disruption of the HPG axis, most likely a reduction in gonadotropin-releasing hormone secretion from the hypothalamus. GH-IGF-1 axis function is also disrupted (Fig. 4). Partial GH resistance occurs, as indicated by late-season elevations in serum GH concentrations while IGF-1 concentrations are reduced. Decreases

in serum growth hormone-binding protein (GHBP) concentration may also modulate serum GH concentrations. GHBP enhances the growth-promoting effects of GH. When GHBP concentrations are reduced, more GH is released to continue growth at its genetically determined rate. Reductions in GHBP concentrations may also signify a down-regulation of GH receptors since GHBP is an index of GH tissue receptor number. Down-regulation results in GH resistance and reduced growth, so even with increased GH concentrations, growth is slowed to conserve energy for critical body functions.

Although these nutritional and hormonal alterations have the potential to change the rates of growth and maturation of wrestlers during a season, the incremental growth in stature and rate of skeletal maturation of wrestlers are similar to those of controls both during and after the season (Fig. 5). However, growth of fat-free mass and other soft tissues is slowed during the season, followed by catch-up growth in the postseason (Fig. 6). Generally, wrestlers are shorter than average for their age, but this is probably due to self-selection. A cross-sectional study found that a reference group of boys was taller (1.9%) than a large group of wrestlers after the age of 16.4 years, but that the slope values for gain in height did not differ.



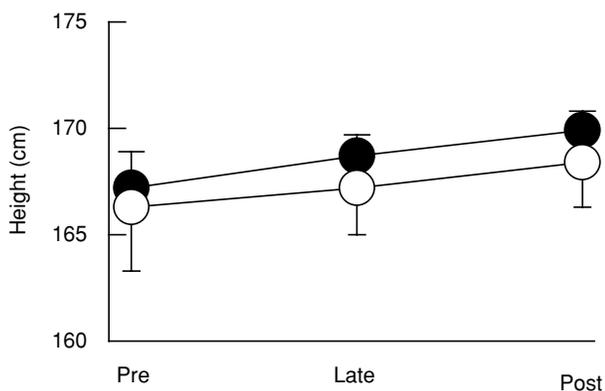
**Figure 3** Mean ( $\pm$  SE) concentrations of testosterone, free testosterone (Free T), luteinizing hormone (LH), and sex hormone-binding globulin (DHEAS) in the wrestler (○;  $n = 9$ ) and control (●;  $n = 7$ ) groups. Like letters are significantly different ( $p \leq 0.05$ ). Serum hormone concentrations are the mean of eight samples drawn every 20 min. Dashed lines indicate the upper and lower normal range for each hormone concentration. For definitions, see the legend to Fig. 1. Redrawn from Roemmich, J. N., and Sinning, W. E. (1997). Weight loss and wrestling training: II. Effects on growth-related hormones. *J. Appl. Physiol.* **82**, 1760–1764.



**Figure 4** Mean ( $\pm$  SE) concentrations of growth hormone (GH), insulin-like growth factor-1 (IGF-1), growth hormone-binding protein (GHBP), and insulin-like growth factor-1-binding protein-3 (IGFBP-3) in the wrestler ( $\circ$ ;  $n = 9$ ) and control ( $\bullet$ ;  $n = 7$ ) groups. Like letters are significantly different ( $p \leq 0.05$ ). Serum hormone concentrations are the mean of eight samples drawn every 20 min. Dashed lines indicate the upper and lower normal range for each hormone concentration. For definitions, see the legend to Fig. 1. Redrawn from Roemmich, J. N., and Sinning, W. E. (1997). Weight loss and wrestling training: II. Effects on growth-related hormones. *J. Appl. Physiol.* **82**, 1760–1764.

## Gymnastics

Gymnasts gain a competitive advantage when they develop muscular strength within a shorter and lighter frame. The performance of female gymnasts often decreases after the onset of puberty due to increases in height and fat mass and the development of secondary sexual characteristics. There has long been concern that gymnasts use sport training and exercise to maintain a lean physique and to delay puberty and growth,

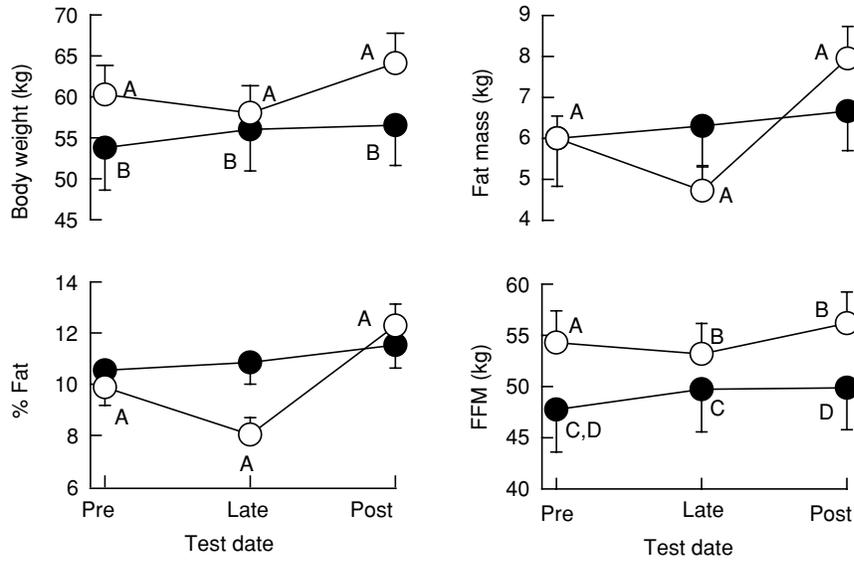


**Figure 5** Height for the wrestler ( $\circ$ ;  $n = 9$ ) and control ( $\bullet$ ;  $n = 7$ ) groups. Like letters are significantly different. Data are expressed as mean ( $\pm$  SE). For definitions, see the legend to Fig. 1. Redrawn from Roemmich, J. N., and Sinning, W. E. (1997). Weight loss and wrestling training: Effects on nutrition, growth, maturation, body composition, and strength. *J. Appl. Physiol.* **82**, 1751–1759.

and a controversial study reported the lack of a normal adolescent growth spurt. Female gymnasts training 18 h/week had a delayed age at menarche, slowed linear growth, and reduced growth potential. Leg length velocity slowed at puberty, with almost no growth after a bone age of 12 years (Fig. 7). Although additional longitudinal studies have not replicated these results, others have observed reduced serum IGF-1 concentrations in gymnasts and relatively short lower limbs of female and male gymnasts. Similar to wrestlers, the growth and maturation of gymnasts may be explained by self-selection because gymnasts tend to be the children of short parents who also had later than average puberty, and shorter limb lengths are present at baseline and do not diminish with training.

## CONCLUSION

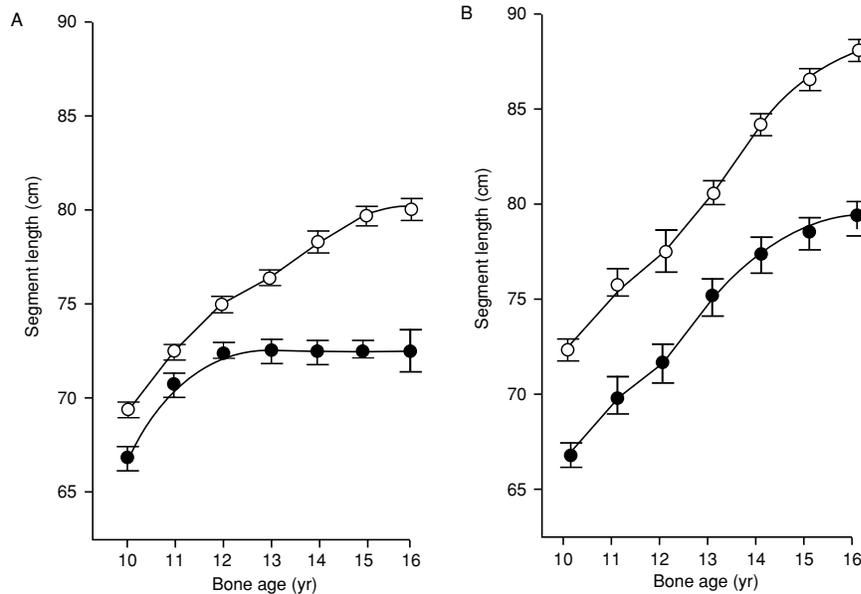
Sport training for most youth does not alter their genetically programmed growth or pubertal development. Normal linear growth encompasses the 3rd to 97th percentile for height and height velocity. When growth continues to track along the same percentile as before initiating training, there is little cause for concern. Individual differences in pubertal timing may cause some athletes to temporarily track along a new percentile. Early maturing athletes will realize an increased height percentile, whereas late-maturing athletes will have a decrease in height percentile.



**Figure 6** Body weight, percentage body fat, fat mass, and fat-free mass (FFM) for the wrestler (○;  $n = 9$ ) and control (●;  $n = 7$ ) groups. Like letters are significantly different. Data are expressed as mean ( $\pm$  SE). For definitions, see the legend to Fig. 1. Redrawn from Roemmich, J. N., and Sinning, W. E. (1997). Weight loss and wrestling training: Effects on nutrition, growth, maturation, body composition, and strength. *J. Appl. Physiol.* **82**, 1751–1759.

Although the average height for participants in some sports may be less than the 50th percentile, this is probably due to self-selection rather than sport training. For example, boys and girls who participated in gymnastics as children and became relatively tall

may become involved in basketball or volleyball during adolescence because greater height can be a disadvantage in gymnastics but beneficial in other sports. Boys and girls with genetically determined delayed puberty and/or short stature may migrate



**Figure 7** Growth in sitting height (○) and leg length (●) from bone age 10 to 16 years. Gymnasts are shown in A and swimmers in B. Note the lack of increase in leg length after bone age 12 years in the gymnasts. Reproduced from Theintz, G. E., Howald, H., Weiss, U., and Sizonenko, P. C. (1993). Evidence for a reduction of growth potential in adolescent female gymnasts. *J. Pediatr.* **122**, 306–313.

toward gymnastics and wrestling because their smaller size is an advantage in these sports.

When wrestlers lose weight to fill empty weight classes or to gain a competitive advantage, or when gymnasts try to delay puberty and growth through dietary restriction, undernutrition can occur. Undernutrition caused by wrestling training and dietary restriction appears to produce a disruption of the pituitary–testicular axis and partial GH resistance. Wrestling training has little effect on linear bone growth or pubertal maturation. The 3- or 4-month period of undernutrition is followed by a long period of adequate nutrition, so the period of undernourishment is not long enough to slow growth or maturation or to reduce the adult height of wrestlers. Female gymnasts who are chronically training and limiting their dietary intake may be at greater risk for limiting their growth. Gymnastics training may slow growth of the lower limbs, but more data are needed to confirm this result, and the effect of sport training and dietary restriction on the neuroendocrine axes of gymnasts has not been adequately studied. The shorter limb lengths of gymnasts may be self-selected because gymnasts have shorter limb lengths when they begin training and the limb length discrepancy does not increase with training.

### See Also the Following Articles

Anorexia Nervosa • Body Composition During Growth • Body Proportions • Constitutional Delay of Growth and Puberty (CDGP) • Delayed Puberty and Hypogonadism, Female • Growth and Chronic Disease • Growth Hormone (GH) • Insulin-like Growth Factors

### Further Reading

- Bass, S., Bradney, M., Pearce, G., Hendrich, E., Inge, K., Stuckey, S., Lo, S. K., and Seeman, E. (2000). Short stature and delayed puberty in gymnasts: Influence of selection bias on leg length and the duration of training on trunk length. *J. Pediatr.* **136**, 149–155.
- Horswill, C. A., Park, S. H., and Roemmich, J. N. (1990). Changes in protein nutritional status of adolescent wrestlers. *Med. Sci. Sports Exercise* **22**, 599–604.
- Nichols, D. L., Sanborn, C. F., Bonnick, S. L., Ben-Ezra, V., Gench, B., and DiMarco, N. M. (1994). The effects of gymnastics training on bone mineral density. *Med. Sci. Sports Exercise* **26**, 1220–1225.
- Roemmich, J. N., and Sinning, W. E. (1996). Sport-seasonal changes in body composition, growth, power and strength of adolescent wrestlers. *Int. J. Sports Med.* **17**, 92–99.
- Roemmich, J. N., and Sinning, W. E. (1997). Weight loss and wrestling training: Effects on nutrition, growth, maturation, body composition, and strength. *J. Appl. Physiol.* **82**, 1751–1759.
- Roemmich, J. N., and Sinning, W. E. (1997). Weight loss and wrestling training: II. Effects on growth-related hormones. *J. Appl. Physiol.* **82**, 1760–1764.
- Rogol, A. D., and Roemmich, J. N. (2000). Puberty as a growth promoted event. In “The Onset of Puberty in Perspective” (J. P. Bourguignon and T. M. Plant, eds.), pp. 97–111. Elsevier, Amsterdam.
- Theintz, G. E., Howald, H., Weiss, U., and Sizonenko, P. C. (1993). Evidence for a reduction of growth potential in adolescent female gymnasts. *J. Pediatr.* **122**, 306–313.
- Veldhuis, J. D., Roemmich, J. N., and Rogol, A. D. (2000). Alterations in growth and body composition during puberty. VII. Gender and maturation-dependent contrasts in the dynamics of growth-hormone (GH) secretion in male and female prepubertal children and late adolescents: Ultrasensitive chemiluminescence-based assay of basal, pulsatile, entropic, and rhythmic modes of GH release. *J. Clin. Endocrinol. Metab.* **85**, 2385–2394.



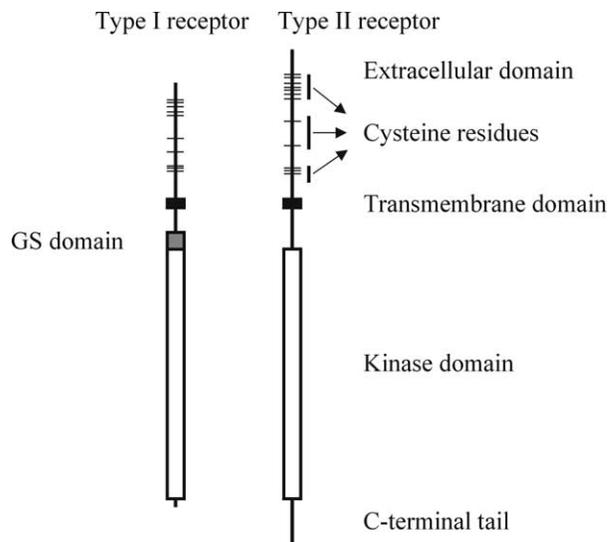
Puberty, Precocious

see Precocious Puberty

## IDENTIFICATION AND STRUCTURE OF RECEPTOR SERINE/THREONINE KINASES

Prior to the cDNA cloning of TGF- $\beta$  family receptors, these receptors had been identified and characterized by cross-linking studies using radiolabeled ligands. By affinity labeling with radiolabeled TGF- $\beta$ , three classes of receptors with distinct sizes were identified on responsive cells: type I (53 kDa), type II (75 kDa), and type III receptor (also termed betaglycan) or endoglin. Whereas type I and type II receptors were both found to be required for signaling, T $\beta$ R-III and endoglin are structurally related proteins with very short intracellular domains lacking signaling motifs that are indirectly involved in signaling; T $\beta$ R-III has been shown to present TGF- $\beta$  to signaling TGF- $\beta$  receptors.

Activins and BMPs were also found to have type I (50–60 kDa) and type II (70–80 kDa) receptors. Expression cloning of receptors for TGF- $\beta$  and activin revealed that TGF- $\beta$  type II and activin type II receptors encode similar transmembrane proteins that contain cytoplasmic domains with predicted serine/threonine kinase activity. Each member of the TGF- $\beta$  superfamily binds to a characteristic combination of type I and type II receptors. Based on sequence similarity, cDNAs for other type I and type II receptors were identified. In mammals, five type II receptors have been identified [TGF- $\beta$  type II receptor (T $\beta$ R-II), activin type II and type IIB receptors (ActR-II and ActR-IIB), BMP type II receptor (BMPR-II), and MIS type II receptor (MISR-II/AMHR-II)] as well as seven type I receptors, also termed activin receptor-like kinases (ALKs) (ALK-1, ALK-2, BMPR-IA/ALK-3, ActR-IB/ALK-4, T $\beta$ R-I/ALK-5, BMPR-IB/ALK-6, and ALK-7). In addition, several splice variants of type I and type II receptors have been reported, but their functions are unknown. Type I and type II serine/threonine kinase receptors are structurally similar; they have small extracellular cysteine-rich domains, single transmembrane domains, and intracellular parts that consist mainly of the serine/threonine kinase domain. The crystal structure of the ActR-II extracellular domain revealed a three-finger fold structure with a scaffold of four disulfide bridges similar to that found in several toxins and one additional disulfide bridge. Type I and type II receptors have similar folds, but sufficient differences exist to suggest that they have different binding modes to the ligand. Interestingly, the crystal structure of T $\beta$ R-I kinase domain indicates that its catalytic center is more reminiscent of tyrosine kinases rather than serine/threonine kinases. Type I

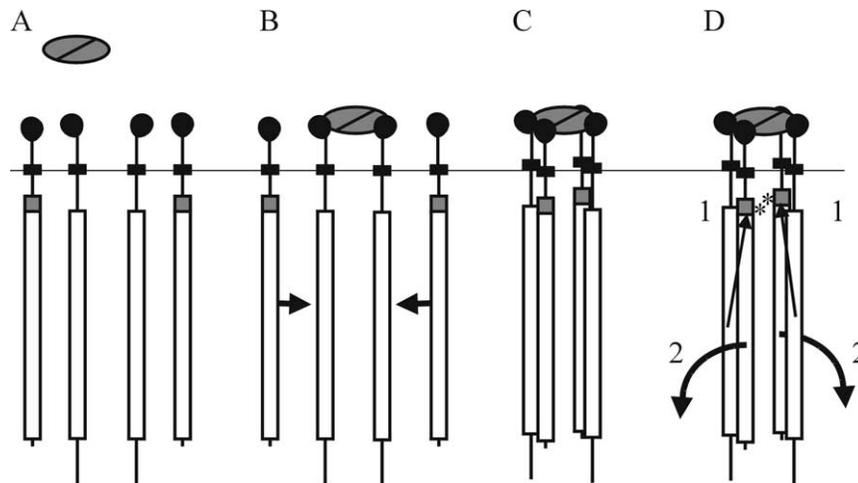


**Figure 1** Structures of TGF- $\beta$  family type I and type II serine/threonine kinase receptors.

receptors, but not type II receptors, have a juxtamembrane region rich in glycine and serine residues (GS domain); this region plays a pivotal role in the activation of type I receptor kinases. Type II receptors, particularly BMPR-II, have longer C-terminal extensions rich in serine residues (Fig. 1), for which the function is unclear.

## MECHANISM OF ACTIVATION OF TGF- $\beta$ SUPERFAMILY RECEPTORS

Genetic and biochemical studies have revealed that both type I and type II receptors are essential for signaling. TGF- $\beta_1$ , TGF- $\beta_3$ , and activins bind to type II receptors and subsequently recruit type I receptors. T $\beta$ R-III is required to enable binding of TGF- $\beta_2$  to T $\beta$ R-II, for which it has only weak affinity. Recent elucidation of the crystal structure of T $\beta$ R-II ectodomain with TGF- $\beta_3$  suggests an assembly mechanism in which the ectodomains of one T $\beta$ R-I and one T $\beta$ R-II bind to adjacent positions on a TGF- $\beta$  monomer and make contact with each other. BMPs have high affinity for type I receptors and low affinity for type II receptors and (like TGF- $\beta_2$ ) display high affinity for preexisting heteromeric type I and type II receptor complexes. The crystal structure of BMP-2 in complex with two BMPR-IA ectodomains revealed that the two receptor ectodomains each contact both BMP-2 monomers on opposite sites of the symmetrical BMP-2 dimer at “wrist” epitopes, without direct contact between the



**Figure 2** Mechanism of TGF- $\beta$  receptor activation. (A, B) TGF- $\beta$  initially binds to T $\beta$ R-II. (C) Subsequently, T $\beta$ R-I is recruited into the complex. (D) Thereafter, T $\beta$ R-I is phosphorylated (asterisks) by the constitutively active T $\beta$ R-II kinase (1). Activated T $\beta$ R-I propagates the signal downstream through phosphorylation of downstream targets (2).

two type I receptor ectodomains. Differences in the mode of ligand binding between type I and type II receptors are, in part, determined by the presence of helix  $\alpha_1$ , unique for type I receptors, which contains a large hydrophobic amino acid residue that sticks out of the helix and that fits into the hydrophobic pocket present in the wrist epitopes of the ligands.

The type II receptor has a constitutively active kinase that, upon ligand-induced heteromeric complex formation between type I and type II receptors, phosphorylates the type I receptor predominantly on serine and threonine residues in its glycine-serine residue-rich (GS) domain. Recent structural studies on the activated T $\beta$ R-I suggest that phosphorylation activates T $\beta$ R-I by converting the GS domain into an efficient substrate recruitment motif and not by increasing the overall kinase activity of the type I receptor. The immunophilin FKBP-12 was identified as a type I receptor-interacting protein that binds to the unphosphorylated GS region and inhibits TGF- $\beta$  signaling by blocking access of phosphorylation sites to the type II receptor kinase. Through this mechanism, FKBP-12 appears to stabilize the inactive conformation of T $\beta$ R-I in the absence of ligand.

The type I receptor acts downstream of the type II receptor and, consistent with this notion, it has been shown to confer signaling specificity within the receptor complex (Fig. 2). The activated type I receptor initiates intracellular signaling by phosphorylating downstream components, including the nuclear effector proteins known as Smads. The L45 loop regions

in the kinase domain of type I receptors were found to be important determinants for signaling specificity.

### ACTIVATION OF DOWNSTREAM EFFECTORS BY TGF- $\beta$ FAMILY RECEPTORS

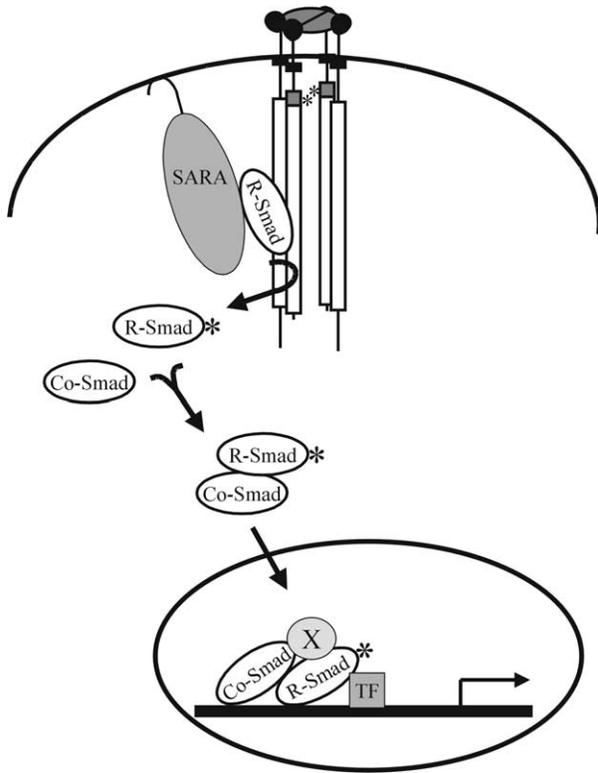
Genetic studies of *Drosophila* and *Caenorhabditis elegans* have led to the identification of a conserved family of proteins, Smads, that play a pivotal role in intracellular signaling downstream of serine/threonine kinase receptors. Smads have two regions of homology at the amino and carboxy terminals—Mad homology domains MH1 and MH2, respectively—that are connected by a proline-rich linker sequence. Based on their functional properties, Smads can be divided into three distinct subclasses:

Receptor-regulated Smads (R-Smads), which transiently interact with activated type I receptors and become phosphorylated on two serine residues in their carboxy-terminal SSXS motif. Whereas Smad2 and Smad3 are phosphorylated by T $\beta$ R-I, ActR-IB, and ALK-7, Smad1, Smad5, and Smad8 are phosphorylated by ALK-1, ALK-2, BMPR-IA/ALK-3, and BMPR-IB/ALK-6.

Common-partner Smads (Co-Smads; e.g., Smad4), which assemble into heteromeric complexes with R-Smads that efficiently accumulate in the nucleus to regulate the transcription of target genes.

Inhibitory Smads (I-Smads; e.g., Smad6 and Smad7), which potentially interfere with signaling by competing with R-Smads for interaction with type I receptors, by competing with Smad4 for binding to R-Smads, or by inducing receptor degradation.

The first step in the TGF- $\beta$ /Smad pathway is the recruitment of R-Smads to the receptor complex. Several proteins with scaffolding, anchoring, and/or chaperone activity have been identified that regulate or facilitate this process. Smad anchor for receptor activation (SARA) presents Smad2 or Smad3 to activated TGF- $\beta$  receptor complexes, whereby SARA interacts directly with the nonactivated R-Smads and the TGF- $\beta$  receptor complexes (Fig. 3). SARA is associated with membranes via its FYVE domain, which interacts with phosphatidylinositol 3-phosphate. The localization of SARA on endosomal membranes



**Figure 3** TGF- $\beta$ /Smad signal transduction. Upon TGF- $\beta$ -mediated heteromeric complex formation of T $\beta$ R-I and T $\beta$ R-II and activation of T $\beta$ R-I by T $\beta$ R-II kinase, the activated T $\beta$ R-I phosphorylates Smad2 and Smad3 (R-Smad) at two C-terminal serine residues. Smad2 and Smad3 are presented to activated receptor complex by SARA. Upon their phosphorylation (asterisks), Smad2 and Smad3 assemble into heteromeric complexes with Smad4 (Co-Smad) that translocate into the nucleus. The nuclear Smad complex, in combination with other transcription factors (TF), coactivators, and corepressors (X), regulates the transcription of target genes.

suggests that TGF- $\beta$  receptors may have to internalize for SARA to present the R-Smads for phosphorylation to the TGF- $\beta$  receptors. Transforming growth factor- $\beta$  receptor-associated protein-1 (TRAP1) binds to inactivated receptor complexes and is thought to act as a Smad4 chaperone. Following receptor activation, TRAP1 leaves the receptor complex, interacts with Smad4, and may facilitate binding of Smad4 to activated R-Smads, after which it dissociates from Smad4.

In the interaction between the activated type I receptor and R-Smad, at least two distinct regions in each protein contribute. First, the L45 loop in the type I receptor interacts with the L3 loop in the MH2 domain of the R-Smad and determines specificity between receptor and R-Smad. Second, the phosphorylated GS domain interacts with a basic surface patch in the L3 loop of the R-Smad, which ensures that only active type I receptors bind R-Smads. Thus, phosphorylation of T $\beta$ R-I introduces a conformational change in the GS region, which switches this domain from a binding site for the inhibitor FKBP-12 into a binding surface for substrate. Phosphorylation of R-Smads on the last two serine residues by the type I receptor kinase triggers their assembly into trimers, mediated via conformational changes in the R-Smad MH2 domains and stabilization of the trimeric complexes through interaction of the phosphoserine binding pockets in the L3 loops with the phosphorylated C-terminal tails of neighboring R-Smad MH2 domains. The activated type I receptors and the phosphorylated C-terminal tails of R-Smads compete for the same L3 loop region of R-Smads, thus providing a mechanism by which activated R-Smads can dissociate from type I receptors. The R-Smad homotrimer is energetically less stable than the R-Smad/Co-Smad heteromeric complex, thus favoring the formation of this heteromeric complex. The stoichiometry between R-Smad and Co-Smad in the heteromeric complex is controversial.

In their basal state, R-Smads and Co-Smads are predominantly localized in the cytoplasm, and ligand stimulation induces their nuclear accumulation. R-Smads and Co-Smads were found to have a nuclear localization sequence (NLS) in their N-terminal regions. Smad4 was also found to have a nuclear export signal (NES) in its linker region. Smad4 is proposed to shuttle constitutively between the nucleus and the cytoplasm. The nuclear accumulation of the heteromeric complex may be stimulated by the shielding of the NES and/or unmasking of the NLS on R- and/or Co-Smad.

Within the nucleus, R- and Co-Smads are transcription factors; transcriptional activation is mediated

via their MH2 domains, which can recruit transcriptional coactivators CBP/p300, whereas their MH1 domains (except for Smad2) can bind to specific DNA sequences. The affinity of Smads for DNA is weak; thus, Smads need to cooperate with other DNA-binding factors to bind efficiently to promoters of target genes (Fig. 3). The first identified Smad transcriptional partner was forkhead activin signal transducer-1 (FAST-1). Smad2 also interacts with the paired-like homeodomain proteins Mixer and Milk. Interestingly, a common Smad interaction motif, containing the PPNK amino acid sequence, was identified in FAST-1 and Mixer/Milk. A large number of transcription factors, of which the activity and expression are also regulated, have been found to interact with Smads to positively or negatively regulate transcriptional responses. This explains, in part, the tissue-specific regulation of target genes by TGF- $\beta$  family members.

#### PHENOTYPES OF TGF- $\beta$ FAMILY RECEPTOR KNOCKOUTS AND TRANSGENIC MICE OVEREXPRESSING DOMINANT NEGATIVE RECEPTORS

As found for mice deficient in TGF- $\beta_1$  that are embryonic lethal, mice deficient in T $\beta$ R-I, T $\beta$ R-II, ALK1, or endoglin died at approximately midgestation due to defects in angiogenesis. TGF- $\beta_1$  knockout mice that survived embryogenesis through maternal supply of TGF- $\beta$  died 3 or 4 weeks after birth from excessive inflammatory response. Studies using transgenic mice with conditional loss of T $\beta$ R-II expression in B cells support a critical role for TGF- $\beta$  in homeostasis of particular B-cell populations as well as in IgA class switching. Transgenic mice with targeted expression of dominant negative (dn) T $\beta$ R-II in T cells showed impaired T-cell homeostasis, manifested by spontaneous differentiation of T cells into effector T cells, and developed autoimmune disease with inflammation of multiple organs. The effect of overexpression of dnT $\beta$ R-II in transgenic mice has also been investigated in several tissue contexts. Expression of dnT $\beta$ R-II in the epidermis induced a hyperplastic and hyperkerotic epidermis and led to a thickened and wrinkled skin. Overexpression of dnT $\beta$ R-II in mammary glands combined with carcinogen treatment resulted in enhanced tumorigenesis in the mammary gland, whereas dominant negative interference of T $\beta$ R-II signaling in the gastrointestinal tract enhanced tumor incidence after challenge with *Helicobacter pylori* or carcinogen. Overexpression of

dnT $\beta$ R-II in osteoblasts resulted in an increase in trabecular bone mass due to reduced bone resorption by osteoclasts. Expression of dnT $\beta$ R-II under control of a metallothioneine promoter revealed a role for TGF- $\beta$  in maintaining epithelial homeostasis and the differentiated phenotype in the exocrine pancreas.

Transgenic mice lacking expression of ActR-II were viable but displayed reduced follicle-stimulating hormone levels and reproductive defects. Mice deficient in ActR-IIB died after birth with severe cardiac defects due to abnormalities in axial patterning and lateral asymmetry. Intercrossing of ActR-II<sup>-/-</sup> and ActR-IIB<sup>-/-</sup> mice caused embryonic lethality as a result of impaired egg cylinder formation, failure to undergo gastrulation, and impaired primitive streak formation, a phenotype that is very similar to that of mice deficient in ActR-IB. Lack of ALK2 expression leads to embryonic lethality soon after onset of gastrulation and is further characterized by defective primitive streak elongation. BMPR-II and BMPR-IA appear to play similar roles during embryonic development since mice lacking expression of either of these receptors are arrested at the egg cylinder stage, do not form mesoderm, and lack development of organized structures. Lack of proper epiblast differentiation might contribute to these developmental defects. In contrast, mice deficient in expression of BMPR-IB are viable but show defects in chondrogenesis and, as a result, in appendicular skeleton formation. Conditional loss of BMPR-IA in limb ectoderm indicates that signaling through this receptor is important for formation of the apical-ectodermal ridge and for dorsal-ventral patterning during limb development. Targeted deletion of BMPR-IA in cardiac myocytes reveals an important role for this receptor in heart morphogenesis and endocardial cushion formation at midgestation of mouse embryonic development.

#### HUMAN ABNORMALITIES CAUSED BY MUTATIONS IN TGF- $\beta$ FAMILY RECEPTORS

Tumor cells have been shown to escape the potent growth inhibitory effects of TGF- $\beta$  by decreasing receptor expression levels or by inactivating mutations in the TGF- $\beta$  receptor genes. T $\beta$ R-II is frequently mutated in an inherited form of colorectal cancer with a microsatellite instability phenotype. In addition, microsatellite instability in the T $\beta$ R-II gene has been reported in atherosclerotic and restenotic vascular cells, providing a mechanistic explanation for the inability to inhibit their cell growth. Loss-of-functional

expression of the T $\beta$ R-II receptor was found to correlate with resistance to TGF- $\beta$ -mediated growth inhibition in chronic lymphocytic leukemia.

Hereditary hemorrhagic telangiectasia, a disease that manifests with vascular dilatation leading to severe hemorrhage, is caused by mutations in endoglin or ALK-1, suggesting that both receptors are involved in the same signaling pathway. Mutations in BMPR-II have been identified in patients with primary pulmonary hypertension, characterized by occlusion of small pulmonary arteries. Mutations in BMPR-IA frequently occur in patients with juvenile polyposis syndrome, indicating a role for BMPR-IA in growth control of colon epithelial cells. The majority of patients with persistent Müllerian duct syndrome, characterized by lack of regression of Müllerian derivatives leading to the development of the female reproductive tract in males, harbor mutations in either AMH or AMHR-II.

## CONCLUSION

Since the identification of serine/threonine kinase receptors in animals in the early 1990s, our knowledge of them has increased dramatically; their ligands have been identified, the molecular mechanisms for activation of these receptors have been revealed, and their intracellular substrates have been identified. TGF- $\beta$  family members regulate the transcriptional program of cells by specific binding and heteromeric complex formation of transmembrane serine/threonine kinase receptors, followed by phosphorylation of intracellular effector proteins called Smads. We have gained our first insights into the three-dimensional structures of the extracellular and kinase domains of TGF- $\beta$  family receptors. However, further studies in which different interacting proteins are cocrystallized with the receptors will provide important information on the structural basis for the activation of serine/threonine kinase receptor complexes and the structural determinants of binding and specificity between ligand–type I/type II receptor, type I–type II receptor, and type I receptor–R-Smad interactions in various TGF- $\beta$  superfamily signaling pathways. This information will perhaps make it possible to rationally design antagonists that block specific responses of TGF- $\beta$  family members.

Perturbed TGF- $\beta$  family signaling has been implicated in or associated with many diseases. In the near future, the translation of our findings of TGF- $\beta$  signaling pathways to clinical applications will be intensely investigated. Animal models with perturbed

TGF- $\beta$  signaling that mimic certain human diseases will increasingly be used to obtain new insights and methods for diagnosis, prognosis, and therapeutic intervention. The generation of floxed TGF- $\beta$  receptors that allow for conditional loss of TGF- $\beta$  receptors in particular tissues and at specified times will be extremely useful in this respect. Moreover, a specific TGF- $\beta$  receptor kinase inhibitor was recently identified. The availability of this inhibitor not only for *in vitro* studies but also for *in vivo* studies in mouse and man and the development of specific inhibitors for type II and other type I receptors are eagerly awaited.

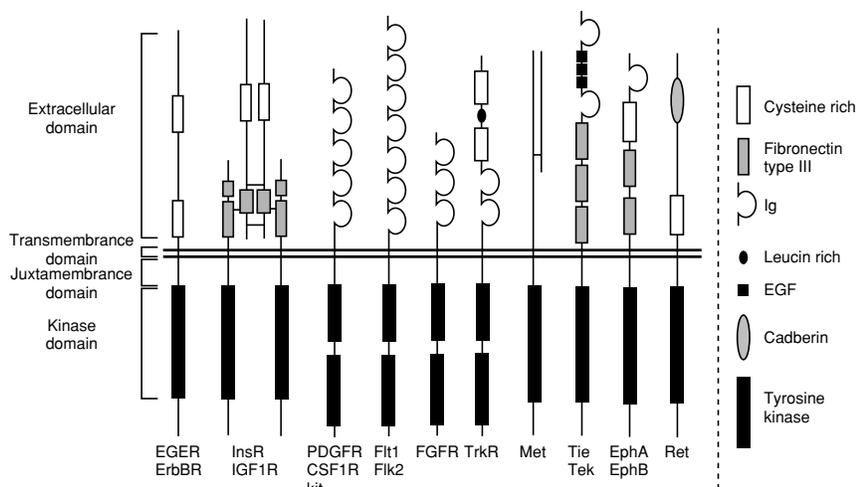
## See Also the Following Articles

Janus Kinases and Cytokine Receptors • Lipid Second Messengers and Receptors • Mitogen-Activated Protein (MAP) Kinases and Receptors • Müllerian Inhibiting Substance: New Insights

## Further Reading

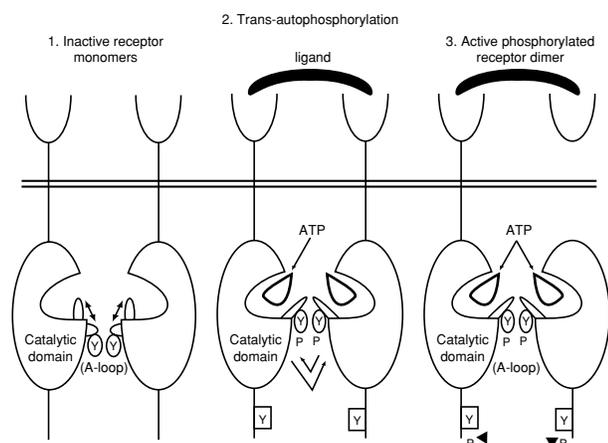
- Callahan, J. F., Burgess, J. L., Fornwald, J. A., Gaster, L. M., Harling, J. D., Harrington, F. P., Heer, J., Kwon, C., Lehr, R., Mathur, A., Olson, B. A., Weinstock, J., and Laping, N. J. (2002). Identification of novel inhibitors of the transforming growth factor- $\beta$ 1 (TGF- $\beta$ 1) type 1 receptor (ALK5). *J. Med. Chem.* **45**, 999–1001.
- Franzén, P., ten Dijke, P., Ichijo, H., Yamashita, H., Schulz, P., Heldin, C.-H., and Miyazono, K. (1993). Cloning of a TGF $\beta$  type I receptor that forms a heteromeric complex with the TGF $\beta$  type II receptor. *Cell* **75**, 681–690.
- Greenwald, J., Fisher, W. H., Vale, W. W., and Choe, S. (1999). Three-finger toxin fold for the extracellular ligand-binding domain of the type II activin receptor kinase. *Nat. Struct. Biol.* **6**, 18–22.
- Hanks, S. K., Quinn, A. M., and Hunter, T. (1988). The protein kinase family: Conserved features and deduced phylogeny of the catalytic domains. *Science* **241**, 42–52.
- Heldin, C.-H., Miyazono, K., and ten Dijke, P. (1997). TGF- $\beta$  signaling from cell membrane to nucleus through Smad proteins. *Nature* **390**, 465–471.
- Huse, M., Muir, T. W., Xu, L., Chen, Y.-G., Kuriyan, J., and Massagué, J. (2001). The TGF $\beta$  receptor activation process: An inhibitor- to substrate-binding switch. *Mol. Cell* **8**, 671–682.
- Kirsch, T., Sebald, W., and Dreyer, M. K. (2000). Crystal structure of the BMP-2–BRIA ectodomain complex. *Nat. Struct. Biol.* **7**, 492–496.
- Lin, H. Y., Wang, X.-F., Ng-Eaton, E., Weinberg, R. A., and Lodish, H. F. (1992). Expression cloning of the TGF- $\beta$  type II receptor, a functional transmembrane serine/threonine kinase. *Cell* **68**, 775–785.
- Massagué, J., Blain, S. W., and Lo, R. S. (2000). TGF $\beta$  signaling in growth control, cancer, and heritable disorders. *Cell* **103**, 295–309.
- Mathews, L. S., and Vale, W. W. (1991). Expression cloning of an activin receptor, a predicted transmembrane serine kinase. *Cell* **65**, 973–982.

- Qin, B. Y., Chacko, B. M., Lam, S. S., de Caestecker, M. P., Correia, J. J., and Lin, K. (2001). Structural basis of Smad1 activation by receptor kinase phosphorylation. *Mol. Cell* **8**, 1303–1312.
- Wrana, J. L., Attisano, L., Wieser, R., Ventura, F., and Massagué, J. (1994). Mechanism of activation of the TGF- $\beta$  receptor. *Nature* **370**, 341–347.
- Wu, J.-W., Hu, M., Chai, J., Seoane, J., Huse, M., Li, C., Rigotti, D. J., Kyin, S., Muir, T. W., Fairman, R., Massagué, J., and Shi, Y. (2001). Crystal structure of a phosphorylated Smad2: Recognition of phosphoserine by the MH2 domain and insights on Smad function in TGF $\beta$  signaling. *Mol. Cell* **8**, 1277–1289.



**Figure 1** Schematic representation of structural components of RTK subfamilies. From Pawson, T. (2002). Regulation and targets of receptor tyrosine kinases. *Eur. J. Cancer* **38**, S3–S10.

kinases. Phosphorylation of tyrosine residues in the activation loop induces vigorous conformational change, leading to making the space in the kinase domain where adenosine triphosphate (ATP) and the peptide

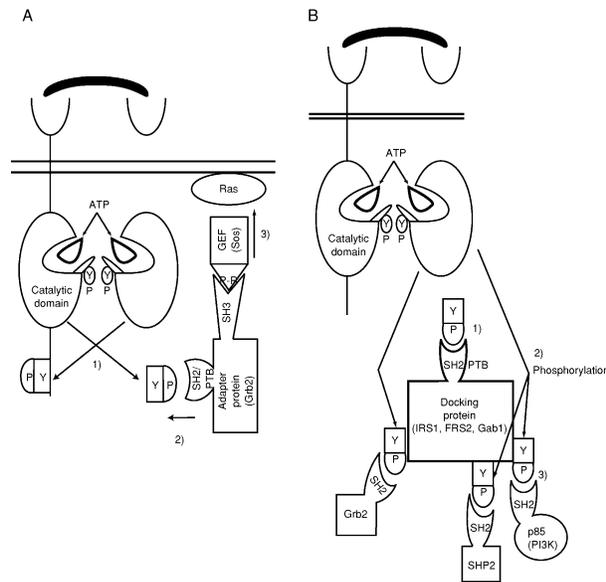


**Figure 2** Model for stimulation of RTK activity by receptor dimerization. Conformation of RTK in the basal state interferes with protein substrate binding and perhaps ATP binding. The A-loop of RTK is relatively mobile. In the absence of ligand, the probability of a trans-autophosphorylation event occurring between randomly colliding receptor is low (step 1). Binding of ligand to the extracellular domain substantially increases local receptor (substrate/enzyme) concentration, providing sufficient opportunity for trans-autophosphorylation of the A-loop to occur (arrow). Autophosphorylation of A-loop tyrosine(s) shifts the A-loop equilibrium toward the active conformation, accommodating substrate binding and facilitating the proper position of residues involved in Mg-ATP binding (step 2). Autophosphorylation occurs on additional tyrosines (arrowhead) that serve as binding sites for downstream signaling proteins (step 3). From Hubbard *et al.*, 2002.

as substrate can enter. This status is the activated condition of the tyrosine kinase.

Following RTK activation, the initial intracellular signaling events occur. Dimerized RTKs phosphorylate the mutual tyrosine residues in the cytoplasmic region of the coupled receptors (Fig. 3A, step 1). In the next step, the phosphorylated tyrosine, with a few amino acids in the activated RTK, is capable of binding to the proteins that are able to bind to phosphorylated tyrosine residue with a specific module. This module is the SH2 (Src homology 2) domain and the PTB (phosphorylated tyrosine binding) domain (Fig. 3A, step 2). In response to extracellular stimuli, RTK transmits the signals to the small G protein Ras, which works as the molecular “on/off switch” for cell proliferation/differentiation signals. Also, Ras is activated by the guanosine diphosphate (GDP)/guanosine triphosphate (GTP) exchange protein Sos. However, RTKs do not bind directly to Sos. The adapter protein Grb2 mediates the interaction between RTK and Sos. Grb2 contains the SH2 domain, which binds to phosphorylated tyrosine of the activated RTK (Fig. 3A, step 3).

The SH2 domain consists of approximately 100 amino acids. The distinct SH2 domain recognizes different amino acid motifs. For example, the SH2 domain of p85 (phosphatidylinositol 3-kinase [PI3K]) binds to the pY-X-X-M motif (where pY = phosphorylated tyrosine and X = arbitrary amino acids). The SH2 domain of Grb2 binds to the pY-X-N-X motif. Also, the SH2 domain of phospholipase C $\gamma$  (PLC $\gamma$ ) binds to the pY-(V/I/L)-X-(V/I/L) motif, and the SH2 domain of SHP2 binds to the pY-(I/VB)-X-(V/I/L)



**Figure 3** Sequence of events following RTK activation. (A) Adapter molecules mediated signaling events. Receptor dimerization leads to autophosphorylation of the noncatalytic region of the cytoplasmic domain, creating docking sites for downstream cytoplasmic targets (step 1). Phosphorylated tyrosine of RTK is recognized by the SH2 or PTB domain of adapter proteins such as Grb2 (step 2). Adapter protein constitutively associates with GDP/GTP exchange proteins such as Sos. Sos, which is recruited to RTK with Grb2, activates Ras (step 3). (B) Docking molecules mediated signaling events. Phosphorylated tyrosine of RTK is recognized by the SH2 domain of docking molecules such as IRS1, FRS2, or Gab1 (step 1). Subsequently, docking molecules are also phosphorylated (step 2) and phosphorylated tyrosine is further recognized by the SH2 domain of various effectors such as PI3K, SHP2, and Grb2 (step 3).

motif (Fig. 4A). Also, the F-L-V-R-E-S sequence in the SH2 domain is required for binding to the phosphorylated tyrosine residue. In particular, the positive charge of the basic amino acid arginin (R) in the F-L-V-R-E-S sequence is critical to electrically bind to the negative charge of phosphorylated tyrosine (Fig. 4B). In addition, other major signaling pathways of activated RTK for cell growth or differentiation are the PI3K–Akt pathways and the PLC $\gamma$ –calcium pathways. PI3K and PLC $\gamma$  also contain the SH2 domain and bind directly to activated RTK, leading to their activation. The PTB domain is the other module to recognize phosphotyrosine. The PTB domain consists of approximately 150 amino acids. The consensus-binding motif of the PTB domain is N-P-X-pY. However, some of the PTB domain binds to N-P-X-Y (nonphosphorylated tyrosine). On the other hand, the group of “docking molecules” works as the signal amplifier. Docking molecules bind

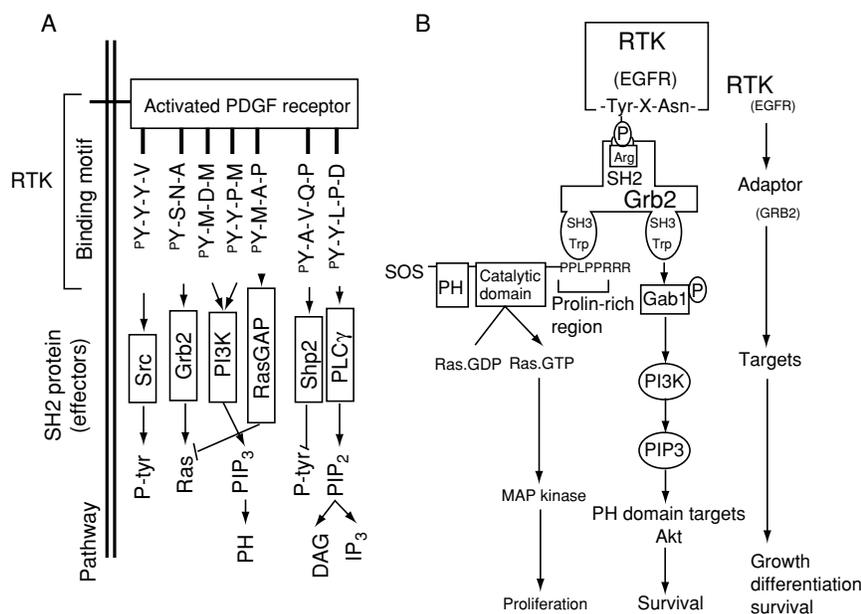
to activated RTK with its SH2 domain, and this is followed by phosphorylation at its multiple tyrosine residues. Subsequently, multiple phosphorylated tyrosine residues of docking molecules recruit multiple effector proteins with their SH2 domain, enabling amplification of the RTK signaling (Fig. 4B).

Furthermore, physical association between Grb2 and Sos is mediated by the interaction between the SH3 (Src homology 3) domain of Grb2 and the prolin-rich motif of Sos (Fig. 3A). The prolin-rich motif is B-X- $\phi$ -P-X- $\phi$ -P or  $\phi$ -P-X- $\phi$ -P-X-B (where B = basic amino acid and  $\phi$  = aliphatic amino acid). An additional important domain for RTK signal transduction is the PH (pleckstrin homology) domain. The PH domain binds to PI. PIs are the products of PI3K such as PI-(3,4,5)-P $_3$  and PI-(3,4)-P $_2$ . Because PIs are produced on the plasma membrane, the function of the PH domain is to recruit proteins to the plasma membrane. Such translocation to the plasma membrane is critical for the activation of many signaling proteins.

## DOWNSTREAM SIGNALING MOLECULES OF RTK

One of the final signaling targets of the RTK pathway is activation of the transcription factors. The main pathway leading to the activation of transcription factor is divided into two parts: one to induce Ras activation and the other to induce activation of the transcription factor. Activation of Ras requires the exchange of Ras-bound GDP to GTP that is catalyzed by the GDP/GTP exchange factor (GEF), Sos. Activated RTK recruits a Grb2–Sos complex. Translocation of Sos at the RTK on the plasma membrane enables interaction with and activates Ras. Meanwhile, many RTKs phosphorylate the docking proteins, which recruit multiple Grb2–Sos complexes. For example, on stimulation with fibroblast growth factor (FGF), the FGF receptor binds to and phosphorylates the docking protein FRS2 (Fig. 3B). Subsequently, phosphorylated FRS2 recruits multiple Grb2–Sos molecular complexes, leading to enhanced Ras activation. Also, FRS2 recruits SHP2 (Src homology 2 containing phosphatase 2). SHP2 also works as adapter molecules for the Grb2–Sos complex, leading to Ras activation. Thus, docking molecules amplify the signals to activate Ras. Activated Ras leads to activation of the serine/threonine kinase cascade following activation of the transcription factor.

The other final target of the RTK signaling pathway is the cytoskeleton, whose regulation changes the



**Figure 4** Illustration of phosphotyrosine sites of RTK being recognized by SH2 domains of different effector proteins. (A) Stimulation of PDGF receptor leads to autophosphorylate multiple tyrosines. A distinct phosphotyrosine motif recruits the SH2 domain of different effectors. (B) Arginin in the F-R-V-R-E-S motif in the SH2 domain plays a critical role in interaction with phosphorylated tyrosine of RTK. Grb2 binds to Sos, leading to Ras-MAP kinase activation and then to cell proliferation. Also, the SH3 domain of Grb2 binds to docking protein Gab1, which binds to PI3K, followed by activation of Akt, eventually leading to cell survival. Thus, activated RTK binds to adapter/docking molecules that associate with and activate multiple effectors. From Pawson, T. (2002). Regulation and targets of receptor tyrosine kinases. *Eur. J. Cancer* 38, S3-S10.

cell shape. Rho family small G protein plays a critical role in regulation of cytoskeleton structure. RTK also activates the GEF for Rho family GTPase. For example, the guanine nucleotide exchange factor (GEF) for Rho family small G protein, ephexin, binds to RTK (EphA). Subsequently, EphA activates Rho, leading to a change of cytoskeleton such as the regression of the growth cone of neurite.

## POSITIVE EFFECTORS IN THE RTK SIGNALING PATHWAY

### Ras

Ras works as the on/off switch in the signaling of molecules leading to cell proliferation. Ras is farnesylated at the C-terminal cysteine residue, which anchors Ras to the cytoplasmic membrane. This anchoring is critical for activation of Ras. Activated Ras recruits the serine/threonine kinase, Raf, to the plasma membrane following autophosphorylation of Raf at serine 388 and its activation. Activated Raf elevates the activity of the MEK-KAPKK-MAPK cascade, leading to activation of the transcription factor Elk1. Other

effectors of Ras are PI3K and RalGDS, which regulate cytoskeleton structure.

### Rho Family Small GTP-Binding Protein

On stimulation with growth factor, the change of the cytoskeleton structure, including plasma membrane ruffling, lamellipodia protrusion, and eventually cell migration, occurs. Rho promotes the formation of F-actin stress fiber and focal contact. Rac promotes lamellipodial formation, and Cdc42 promotes filopodia formation. Stimulation of RTK induces activation of Rho family small G protein, leading to change of the cytoskeleton structure. For example, activated platelet-derived growth factor (PDGF) receptor associates with the molecular complex Ras-GTPase-activating proteins (GAP) with RhoGAP. Also, the GEF for Rac/Rho, Vav, is activated by tyrosine phosphorylation. Furthermore, the RTK Eph receptor is involved in cell migration and formation of neuronal and vascular network. Eph binds to ephexin, the GEF for Rho family GTP-binding protein. Ephexin activates RhoA and, in contrast, inactivates Rac as well as Cdc42. Hence, tyrosine kinases regulate Rac, Rho,

and Cdc42 through their GAP or GEF, leading to cytoskeleton regulation.

The effector of Rho for the change of cytoskeleton structure is the serine threonine kinase ROCK. Another effector of Rho is mDia, which leads to promotion of actin polymerization. The effector of Cdc42 is the neural Wiskott–Aldrich syndrome protein (N-WASP). Also, the effector of Rac is the WASP family verprolin-homologous protein (WAVE). The WASP family protein contains a VCA (verprolin, cofilin, and actin) domain that binds to the actin monomer and actin-related proteins 2 and 3 (Arp2 and Arp3), leading to promotion of F-actin polymerization with the role of Arp2 and Arp3 as the core.

### PI3Kinase

PI3K is the heterodimer of the p85 subunit with the p110 subunit. The p85 subunit contains the SH2 domain, which recognizes the p-Y–M–X–M motif (where p-Y = phosphorylated tyrosine, M = methionine, and X = arbitrary amino acid) of an activated receptor. The interaction of the p85 domain with the phosphorylated receptor activates the kinase activity of the p110 subunit. The p110 subunit phosphorylates the 3-position of PI, PI-(4)-P, and PI-(4,5)-P<sub>2</sub> to produce PI-(3)-P, PI-(3,4)-P, PI-(3,4,5)-P<sub>3</sub>, and PI-(3,4)-P<sub>3</sub> that recruit the PH domain-containing proteins such as serine/threonine kinases PDK and Akt as well as tyrosine kinase Btk. Akt phosphorylates the pro-apoptotic protein Bad, which traps it to 14-3-3 protein on the plasma membrane.

On the contrary, the phosphorylated 3-position of PI-(3,4,5)-P<sub>3</sub> is dephosphorylated by phosphatase and tensin homologue (PTEN). Actually, Akt or PI3K works as an oncogene, whereas PTEN works as a tumor suppressor gene.

### PLC $\gamma$

PLC $\gamma$  also contains the SH2 domain, which makes its translocation to the activated RTK. Then, PLC $\gamma$  is phosphorylated and activated. PLC $\gamma$  hydrolyzes PI, leading to the production of diacylglycerol (DAG) and inositol-(1,4,5)-triphosphates (IP<sub>3</sub>). The former activates protein kinase C, whereas the latter elevates cytoplasmic calcium concentration by its mobilization from the endoplasmic reticulum (ER). Both molecules activate versatile signaling pathways. The vascular endothelial growth factor (VEGF) receptor activates Raf–MEK–MAPK pathways by activation of PLC $\gamma$  without Ras activation.

### STAT

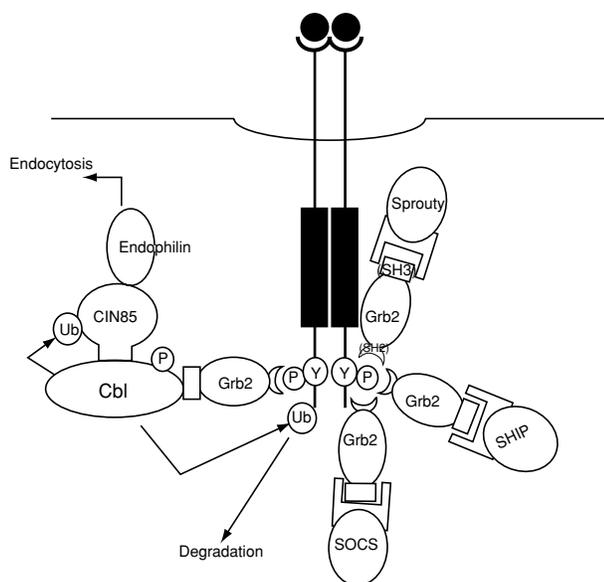
Signal transducers and activators of transcription (STAT) family proteins are activated by the non-receptor-type tyrosine kinase JAK in the cytokine receptor signaling pathway, and subsequently it translocates to the nucleus as transcription factors. Also, STAT activation is required for the epidermal growth factor (EGF) receptor signaling pathway to lead MAP kinase activation and gene expression.

### NEGATIVE EFFECTORS IN THE RTK SIGNALING PATHWAY

Although plasma membrane receptors transmit “positive signaling” involved in cell stimulation and induction of various cellular responses, specific mechanisms have evolved to ensure that appropriate thresholds of signals are achieved and maintained for the right length of time. These mechanisms are referred to as “negative signaling.” In most cases, the same receptor simultaneously induces positive and negative pathways that appear to be functionally connected by numerous feedback mechanisms. A delicate balance between positive and negative signals is critical for normal cell homeostasis, and its deregulation is often implicated in the development of human diseases. Transient inhibition interferes with the strength and duration of the signal in a defined window of time, thereby fine-tuning signal transduction. Several regulatory mechanisms for RTK signaling pathways have been clarified. One example is degradation of the receptor and its ligand by endocytosis followed by degradation in lysosome (Fig. 5). Also, there are negative regulatory molecules specific for each signaling molecule (Fig. 5). Many protein tyrosine phosphatases (PTPs) dephosphorylate and inactivate the protein tyrosine kinases. RasGAP inactivates Ras. MAP kinase phosphatase (MKP) dephosphorylates and inactivates MAP kinase. This section discusses the main negative regulatory molecules.

### Protein Tyrosine Phosphatase

Dephosphorylation of RTK by PTPs is the most elucidated mechanism for inactivation of RTK signaling pathways. Dephosphorylation of activation loop sites in RTK leads to inactivation of the kinase domain, whereas phosphate removal from docking tyrosine blocks activation of specific signaling pathways. The function of PTP is tightly regulated by the change of its localization within the cell. For example,



**Figure 5** Illustration of Grb2 playing role as negative regulator in RTK signaling pathway. The adapter protein Grb2 serves as a regulated scaffold for the binding of other signaling proteins that include both positive (Sos) and negative (Cbl, Sprouty, SHIPs, and SOCS suppressors of cytokine signaling) regulators. In that way, Grb2 can link RTKs with activation of the Ras–MAP kinase pathway via its association with Sos while it promotes receptor ubiquitination and endocytosis by binding to the Cbl–CIN85–endophilin complex. Furthermore, Grb2 recruits Sprouty, SHIPs, and SOCS to further attenuate receptor signaling pathways. Ub, ubiquitin. From Dikic, I., and Giordano, S. (2003). Negative receptor signalling. *Curr. Opin. Cell Biol.* 15, 128–135.

the PTP SHP1 is recruited to the activated RTK Kit (the receptor for the stem cell factor), with interaction of SH2 domains of SHP1 with phosphorylated tyrosine of Kit, and subsequently inactivates it by dephosphorylation. Also, the PTP 1B (PTP1B) colocalizes with the activated PDGF receptor, which is internalized at the ER, followed by its inactivation. Inactivation of RTK reduces its biological function. For example, PTP1B dephosphorylates the insulin receptor, leading to reduced insulin action or so-called insulin resistance. On the other hand, SHP2 associates with the phosphorylated receptor with its SH2 domain and works as an adapter for transmitting the positive signals into the Grb2–Sos–Ras pathway. For example, insulin-induced binding of SH-PTP2 to IRS-1 generates the signal to activate the Ras protein.

### SHIP and PTEN

Both the SHIP (Src homology inositol phosphatase) and PTEN molecules block the PI3K-signaling pathway by degradation of the PI3K products,

PI-(3,4,5)-triphosphate. SHIP dephosphorylates the 5'-position of PI-(3,4,5)-P<sub>3</sub>, leading to the production of PI-(3,4)-P<sub>2</sub>. SHIP is recruited to the phosphorylated immune tyrosine-based inhibitory motif (ITIM) of the immunological receptors with its SH2 domain and inhibits RTK–PI3K signals. PTEN negatively dephosphorylates the 3'-position of PI-(3,4,5)-P<sub>3</sub>, leading to the production of PI-(4,5)-P<sub>2</sub>.

### Cbl

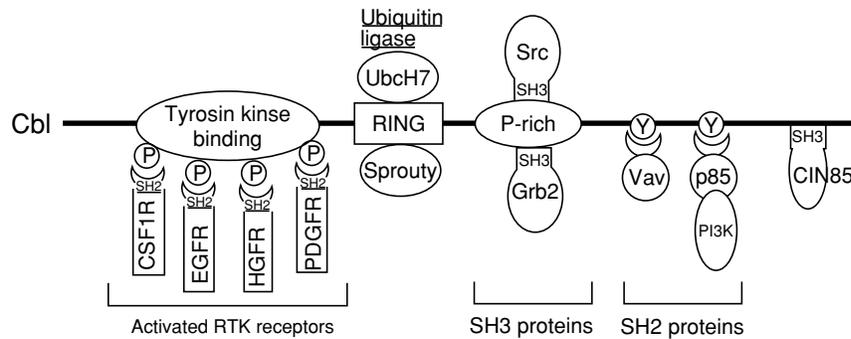
Cbl is the cytoplasmic protein that consists of the SH2 domain and RING finger domain at the N-terminal region as well as the proline-rich motif. Cbl works as negative regulatory molecules for RTK. Cbl binds to activated RTKs through its SH2 domain and subsequently works as an E<sub>3</sub> ubiquitin ligase and interacts with enzymes of the E<sub>2</sub> type such as UbcH7 by the RING finger domain, leading to ubiquitination of tyrosine kinase and eventually its degradation. In addition, Cbl regulates endocytosis of RTK by binding to the CIN85–endophilin complex. Interaction between CIN85 and Cbl is mediated by binding of the SH3 domains of CIN85 to the C-terminal region of Cbl (Fig. 6). The proline-rich motif in CIN85 mediates its constitutive association with the SH3 domain of endophilins, a family of proteins able to regulate negative curvature of the plasma membrane during the early steps of endocytosis. Inhibition of Cbl–CIN85–endophilin interaction was sufficient to block endocytosis of RTK following its degradation (Fig. 5).

### SOCS

Suppressor of cytokine signaling (SOCS) is the inhibitory molecules for the cytokine receptor–JAK–STAT signaling pathway. SOCS1/JAB1 (JAK-binding protein) directly binds to JAK. SOCS2 binds to the cytokine receptor, or SOCS3/CIS3 binds to both the receptor and JAK. Subsequently, SOCS family proteins inhibit the cytokine receptor signaling pathway. Besides cytokine receptor signaling molecules, SOCS2 inhibits the insulin-like growth factor (IGF) receptor signaling pathway.

### Negative Regulation for RAS

Ras is one of the most important molecules for cell proliferation. This is assumed from the evidence that the constitutive active mutant of Ras was detected in 30% of cancer. Therefore, to protect excessive activation of Ras, its activation should be tightly regulated. Although Ras itself has intrinsic GTPase activity that converts GTP-bound Ras (activated form) to



**Figure 6** Cbl, a ubiquitin ligase and a multiadapter protein. Cbl consists of several distinct domains. The amino-terminal portion contains the tyrosine kinase-binding domain that interacts with activated RTKs. Cbl is an E<sub>3</sub> ubiquitin ligase and interacts with enzymes of the E<sub>2</sub> type such as UbcH7. The RING finger domain is responsible for Cbl enzymatic activity. A prolin-rich (P-rich) motif is critical for interaction with SH3-containing proteins such as Grb2 and Src. The Cbl carboxyl terminus contains several tyrosine residues (Y) that, on phosphorylation (P), interact with SH2-domain-containing proteins such as Vav and the regulatory subunit of PI3K (p85). The binding site for the SH3 domains of CIN85 is found in the distal carboxyl terminus of Cbl. From Dikic, I., and Giordano, S. (2003). Negative receptor signalling. *Curr. Opin. Cell Biol.* **15**, 128–135.

GDP-bound Ras (inactivated form), its intrinsic activity is not enough for full inactivation of Ras. Therefore, RasGAP is responsible for full extinction of Ras activity. Meanwhile, the adapter molecule Dok, which binds to RasGAP, also works as a negative inhibitor for the Ras–MAP kinase pathway. In *Drosophila*, RasGAP associates with pY918 of torso RTK and works as a negative effector for signaling molecules of torso. Also, PTP corkscrew (CSW), which associates with pY630, specifically dephosphorylates the negative pY918 torso signaling site, and CSW serves as an adapter protein for DRK (Grb2 homologue) binding, physically linking torso to Ras activation. Thus, opposing actions of CSW and RasGAP modulate the strength of the torso RTK signals. Also, PI3K-dependent membrane recruitment of Dok is essential for its negative effect on mitogen-activated protein. Thus, RasGAP negatively regulates the RTK signaling pathway *in vivo* (Fig. 7).

Rap-1 also works as an inhibitor for Ras. Although Rap-1 is the small GTP-binding protein and has a structure similar to that of Ras, it inhibits activation of Ras. The balance between Rap-1 and Ras is important for cell condition. In the immune-ergic or immune-tolerant condition of the T cell, Rap-1 is dominant. Also, insulin regulates the dynamic balance between Ras and Rap1 signaling by coordinating the assembly states of the Grb2–SOS and CrkII–C3G complexes.

Grap has a structure similar to that of Grb2 and the SH2–SH3–SH2 domain. Grap couples with tyrosine kinases and suppresses the RAS/MAP kinase pathway.

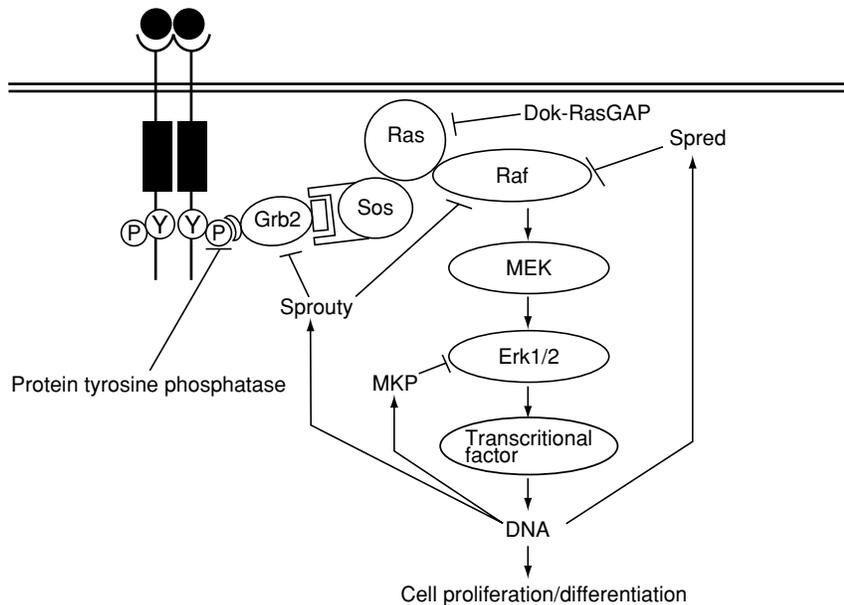
### Negative Regulator for the Ras–MAP Kinase Pathway and Sprouty

Sprouty was recently found to be an inhibitory factor for Raf kinase in the FGF receptor signaling pathway that is involved in formation of bronchial branching in *Drosophila*. Four homologous proteins in the Sprouty family were found. The transcription of Sprouty was induced by MAP kinase on stimulation with EGF as well as FGF. Also, Sprouty inhibited VEGF-induced endothelial cell proliferation and angiogenesis. Although the inhibitory mechanism of Sprouty in the growth factor receptor signaling pathway has not been elucidated in detail, in *Drosophila* it has been shown to inhibit the Grb2 homologue Drk and may also inhibit Raf.

Also, Spred (Sprouty-related EVH1-domain containing) proteins are identified. Spred proteins are able to inhibit Ras–MAP kinase signaling by interfering with Raf kinase activation without reducing Ras activity.

### CONCLUSION

Stimulation of RTK leads to multiple cellular functions that depend on protein–protein, protein–phospholipid, and protein–nucleic acid interactions. Protein interaction domains and motifs provide a basic mechanism to organize signaling pathways and other forms of information of transfer in cells (e.g., cell cycle, protein trafficking, gene expression, DNA repair). The combination use of interaction domains



**Figure 7** Negative regulators for the RTK–Ras–MAP kinase pathway. Signaling molecules of the RTK–Ras–MAP kinase pathway are negatively regulated by several inhibitors. Phosphorylated tyrosine is dephosphorylated by protein tyrosine phosphatase. Activated (GTP-bound) Ras is converted to inactivated (GDP-bound) Ras by RasGAP. Also, RasGAP is recruited to RTK by the adapter molecule Dok. Sprouty and its related molecule, Spred, recently have been shown to possibly inhibit Grb2, Ras, or Raf. MAP kinase is dephosphorylated by MKP. Expression of Sprouty, Spred, or MKP is increased under the MAP kinase pathway.

has allowed the creation of novel signaling pathways and networks. Although much progress has been made in understanding the outline of the RTK signaling pathway, it is still difficult to explain diversity of the cell biological phenomenon. More effort is required to link findings in signal transduction to the cell biological phenomenon.

### See Also the Following Articles

Adenylyl Cyclase • G Protein-Coupled Receptors • Janus Kinases and Cytokine Receptors • Peptide Hormones and Growth Factors: Cellular Signaling Mechanisms • Phosphatidylinositol Turnover and Receptors • Transforming Growth Factor (TGF) Alpha

### Further Reading

- Amano, M., Fukata, Y., and Kaibuchi, K. (2000). Regulation and functions of Rho-associated kinase. *Exp. Cell Res.* **261**, 44–51.
- Arakawa, Y., Bito, H., Furuyashiki, T., Tsuji, T., Takemoto-Kimura, S., Kimura, K., Nozaki, K., Hashimoto, N., and Narumiya, S. (2003). Control of axon elongation via an SDF-1-alpha/Rho/mDia pathway in cultured cerebellar granule neurons. *J. Cell Biol.* **161**, 381–391.
- Bustelo, X. R. (2001). Vav proteins, adaptors, and cell signaling. *Oncogene* **20**, 6372–6381.
- Dikic, I., and Giordano, S. (2003). Negative receptor signalling. *Curr. Opin. Cell Biol.* **15**, 128–135.
- Donovan, S., Shannon, K. M., and Bollag, G. (2002). GTPase activating proteins: Critical regulators of intracellular signaling. *Biochim. Biophys. Acta* **1602**, 23–45.
- Haj, F. G., Verveer, P. J., Squire, A., Neel, B. G., and Bastiaens, P. I. (2002). Imaging sites of receptor dephosphorylation by PTP1B on the surface of the endoplasmic reticulum. *Science* **295**, 1708–1711.
- Horvat, S., and Medrano, J. F. (2001). Lack of *Socs2* expression causes the high-growth phenotype in mice. *Genomics* **72**, 209–212.
- Klaman, L. D., Boss, O., Peroni, O. D., Kim, J. K., Martino, J. L., Zabolotny, J. M., Moghal, N., Lubkin, M., Kim, Y. B., Sharpe, A. H., Stricker-Krongrad, A., Shulman, G. I., Neel, B. G., and Kahn, B. B. (2000). Increased energy expenditure, decreased adiposity, and tissue-specific insulin sensitivity in protein-tyrosine phosphatase 1B-deficient mice. *Mol. Cell Biol.* **20**, 5479–5489.
- Levkowitz, G., Waterman, H., Ettenberg, S. A., Katz, M., Tsygankov, A. Y., Alroy, I., Lavi, S., Iwai, K., Reiss, Y., Ciechanover, A., Lipkowitz, S., and Yarden, Y. (1999). Ubiquitin ligase activity and tyrosine phosphorylation underlie suppression of growth factor signaling by c-Cbl/Sli-1. *Mol. Cell.* **4**, 1029–1040.
- Pawson, T. (2002). Regulation and targets of receptor tyrosine kinases. *Eur. J. Cancer* **38**, S3–S10.
- Plotnikov, A. N., Hubbard, S. R., Schlessinger, J., and Mohammadi, M. (2000). Crystal structures of two FGF–FGFR complexes reveal the determinants of ligand-receptor specificity. *Cell* **101**, 413–424.
- Sasaki, A., Taketomi, T., Kato, R., Saeki, K., Nonami, A., Sasaki, M., Kuriyama, M., Saito, N., Shibuya, M., and Yoshimura, A. (2003). Mammalian Sprouty4 suppresses Ras-independent ERK activation by binding to Raf1. *Nat. Cell Biol.* **5**, 427–432.

- Schlessinger, J. (2002). Ligand-induced, receptor-mediated dimerization and activation of EGF receptor. *Cell* **110**, 669–672.
- Seminario, M. C., and Wange, R. L. (2002). Signaling pathways of D3-phosphoinositide-binding kinases in T cells and their regulation by PTEN. *Semin Immunol.* **14**, 27–36.
- Shamah, S. M., Lin, M. Z., Goldberg, J. L., Estrach, S., Sahin, M., Hu, L., Bazalakova, M., Neve, R. L., Corfas, G., Debant, A., and Greenberg, M. E. (2001). EphA receptors regulate growth cone dynamics through the novel guanine nucleotide exchange factor ephexin. *Cell* **105**, 233–244.
- Soubeyran, P., Kowanetz, K., Szymkiewicz, I., Langdon, W. Y., and Dikic, I. (2002). Cbl–CIN85–endophilin complex mediates ligand-induced downregulation of EGF receptors. *Nature* **416**, 183–187.
- Takenawa, T., and Miki, H. (2001). WASP and WAVE family proteins: Key molecules for rapid rearrangement of cortical actin filaments and cell movement. *J. Cell Sci.* **114**, 1801–1809.
- Takahashi, T., Yamaguchi, S., Chida, K., and Shibuya, M. (2001). A single autophosphorylation site on KDR/Flk-1 is essential for VEGF-A-dependent activation of PLC-gamma and DNA synthesis in vascular endothelial cells. *EMBO J.* **20**, 2768–2778.
- Taylor, M. L., and Metcalfe, D. D. (2000). Kit signal transduction. *Hematol. Oncol. Clinics North America* **14**, 517–535.
- Wakioka, T., Sasaki, A., Kato, R., Shouda, T., Matsumoto, A., Miyoshi, K., Tsuneoka, M., Komiya, S., Baron, R., and Yoshimura, A. (2001). Spred is a Sprouty-related suppressor of Ras signalling. *Nature* **412**, 647–651.
- Zhao, M., Schmitz, A. A., Qin, Y., Di Cristofano, A., Pandolfi, P. P., and Van Aelst, L. (2001). Phosphoinositide 3-kinase-dependent membrane recruitment of p62(dok) is essential for its negative effect on mitogen-activated protein (MAP) kinase activation. *J. Exp. Med.* **194**, 265–274.



## Recklinghausen's Disease

see Neurofibromatosis

during the early 1980s and served thereafter as a paradigmatic signal transduction pathway for “Ca<sup>2+</sup>-mobilizing” receptors. The subsequent years have seen the emergence of many additional phospholipases that were found to be activated in a signal-dependent manner. The phospholipid substrates of SAPLs include not only quantitatively minor membrane constituents, such as PIP<sub>2</sub> and phosphatidic acid (PA), but also major structural building blocks of the bilayer, such as phosphatidylcholine (PC), phosphatidylethanolamine (PE), and sphingomyelin (SM). Thus, PI-PLC turned out to be a model for a large number of SAPLs, including phospholipases of the types A<sub>2</sub>, C, and D, which use various phosphoinositides, PC, PE, PA, and SM as substrates. SAPLs are coordinately activated by given receptor agonists in a defined cell type-specific sequence and play important roles in cell regulation. This article reviews the major classes of SAPLs, emphasizing established facts pertaining to their structure, activity, localization, receptor activation, targets, and functions.

## PHOSPHOINOSITIDE-SPECIFIC PHOSPHOLIPASE C

### Overview

The first and best characterized among the SAPLs, PI-PLC catalyzes the hydrolysis of PIP<sub>2</sub> to form two messengers, inositol-1,4,5-trisphosphate (IP<sub>3</sub>) and 1,2-*sn*-diacylglycerol (DAG), that target Ca<sup>2+</sup> storage organelles and protein kinase C (PKC), respectively. The PI-PLC family is now known to include 11 mammalian isozymes in four classes: PLCβ(1–4), PLCγ(1,2), PLCδ(1–4), and PLCε. PI-PLC isozymes exhibit only subtle differences in their catalytic properties but have distinct tissue-specific expression and mode of activation. The emerging picture is one where enzymes that belong in different PI-PLC classes are regulated by different mechanisms. The mechanisms of activation of two isozymes, PLCβ<sub>1</sub> and PLCγ<sub>1</sub>, are particularly well understood and involve interactions with stimulatory guanine nucleotide-binding (G) proteins and tyrosine kinases, respectively. PLCβ<sub>1</sub> is activated by members of the G<sub>q</sub> class of pertussis toxin-insensitive G proteins. PLCγ<sub>1</sub> is activated by phosphorylation on specific tyrosine residues by receptor tyrosine kinases as well as by the cytoplasmic, receptor-activated Src family tyrosine kinases.

### Structure and Catalytic Properties

All mammalian PI-PLC genes share a catalytic core consisting of two conserved, noncontiguous domains

designated X and Y. In addition, all PI-PLCs have an N-terminal pleckstrin homology (PH) domain and an EF-hand domain as well as a C-terminal C2 domain. PLCδ, considered to be an archetypal PI-PLC isozyme, contains only these five structural elements. PLCβ has an additional C-terminal “tail” domain with a PDZ-binding motif. PLCγ has a large insert between domains X and Y that includes a second split PH domain, two SH2 domains, and an SH3 domain. PLCε has an N-terminal RasGEF domain and two C-terminal RA domains. The function of these domains in regulation of PLC activity of the different isozymes is discussed later. All PI-PLC isozymes hydrolyze PIP<sub>2</sub> exclusively in a catalytic mechanism that requires Ca<sup>2+</sup> ions. The membrane association of PLCδ by a “tether and fix” mechanism may be generalized for the binding of the catalytic core of all PI-PLCs. According to this theory, PLCδ is initially tethered to PIP<sub>2</sub> by the PH domain, and this is facilitated by binding of Ca<sup>2+</sup> to the EF domain. The enzyme is then fixed to the bilayer by formation of a ternary complex among Ca<sup>2+</sup>, phosphatidylserine (PS), and the C2 domain, allowing interaction of the catalytic domains with the PIP<sub>2</sub> substrate. The hydrolysis of PIP<sub>2</sub> then occurs by a sequential reaction mechanism that involves a general acid–base catalysis with formation of a cyclic inositol phosphate intermediate. A Ca<sup>2+</sup> ion within the catalytic pocket plays an important role in catalysis, and the hydrophobic rim of the active site is necessary for penetration of the active site into the bilayer. In PLCβ and PLCγ, the tethering of the enzyme to the membrane involves interaction of the PH domain with phosphatidylinositol 3-P and phosphatidylinositol 3,4,5-trisphosphate (PIP<sub>3</sub>), respectively. This interaction may be part of these isozymes’ activation mechanism.

### Receptor-Mediated Activation

PI-PLC isozymes are normally cytosolic proteins that are translocated to the plasma membrane on receptor activation. PLCβ<sub>1</sub> and PLCβ<sub>3</sub> are relatively widely expressed, whereas the expression of PLCβ<sub>2</sub> and PLCβ<sub>4</sub> is limited to hematopoietic cells and the brain, respectively. All PLCβ forms are stimulated by G protein-coupled receptors (GPCRs) through direct interaction with the subunits of activated G<sub>q</sub> subfamily proteins. Ligand binding to relevant GPCRs results in exchange of guanosine diphosphate (GDP) for guanosine triphosphate (GTP) on the Gα-subunit and subsequently dissociation of the Gβγ dimer. The active, GTP-bound α-subunit interacts with the C-terminal tail region of PLCβ, where it binds the C2 domain and

two clusters of basic residues. The G $\beta\gamma$  dimer is also involved in activation of PLC $\beta$  isozymes, especially of PLC $\beta$ 2, with which it interacts with high affinity via the latter's PH domain. PLC $\gamma$ 1 is expressed ubiquitously, and PLC $\gamma$ 2 is expressed mainly in the hematopoietic system. PLC $\gamma$ 1 is recruited via its N-terminal SH2 domain to specific phosphotyrosine residues on activated receptor tyrosine kinases. PLC $\gamma$ 1 is then phosphorylated on three tyrosine residues. Both association and tyrosine phosphorylation are essential for activation of PLC $\gamma$ 1 by growth factors receptors. Full activation also requires the production of PIP $_3$ , which participates in membrane association of PLC $\gamma$ 1 by interaction with its PH domain and/or the C-SH2 domains. The activation of PLC $\gamma$ 1 and PLC $\gamma$ 2 by immune cell receptors is similar in principle, except that PLC $\gamma$  is recruited to phosphotyrosine residues on adapter/scaffold proteins that are phosphorylated by nonreceptor Src-like kinases. PLC $\delta$ , the smallest and likely the archetypal PI-PLC, is most sensitive to Ca $^{2+}$  ions and is believed to be regulated primarily by changes in intracellular Ca $^{2+}$  concentrations, secondary to activation of other Ca $^{2+}$ -mobilizing pathways. The most recently discovered isozyme, PLC $\epsilon$ , has both Ras activation and Ras effector domains, strongly suggesting that Ras is involved in its regulation, but the details of the bidirectional relationship between these two enzymes are not clear. It should be noted that in addition to PI-PLC-mediated signaling in the plasma membrane, PIP $_2$  hydrolysis may occur intracellularly, for example, in the nucleus and the Golgi apparatus. The mechanisms of PI-PLC activation in endomembrane compartments and its function(s) remain to be fully elucidated.

### Targets of IP $_3$ and DAG

IP $_3$  mobilizes Ca $^{2+}$  ions from intracellular stores in the endoplasmic reticulum (ER) by interacting with a family of IP $_3$  receptors, ion channels that allow efflux of sequestered Ca $^{2+}$  to the cytosol. IP $_3$  may be converted to other inositol phosphates by one or more inositol ring phosphorylation/dephosphorylation reactions, but the specific functions of the resulting inositides are not fully established. DAG is involved in membrane translocation and activation of PKC isozymes. DAG interacts with an N-terminal C1 domain found in conventional (PKC $\alpha$ ,  $\beta$ , and  $\gamma$ ) and novel (PKC $\delta$ ,  $\epsilon$ ,  $\eta$ , and  $\theta$ ) isoforms as well as in protein kinase D. The C1 domain is also found in nonkinase proteins such as RasGRPs, chimaerins, and Munc13 isoforms that act as phorbol ester receptors and potentially also serve as downstream targets of DAG.

## PHOSPHOLIPASE D

### Overview

Phospholipase D (PLD) hydrolyzes phospholipids to PA and a free polar headgroup. PLD activity is rapidly and dramatically activated in response to extracellular stimuli. The signal-dependent activation of PLD was demonstrated in numerous cell types stimulated with various hormones, growth factors, cytokines, neurotransmitters, adhesion molecules, drugs, and physical stimuli. Studies into the mechanisms of receptor-PLD coupling have implicated multiple pathways, including small GTPases, Ser/Thr kinases, and tyrosine kinases. A number of eukaryotic PLD genes have been identified, including numerous plant genes, a yeast PLD gene, and two mammalian genes: PLD1 and PLD2. Significant advances were made in understanding the structure, regulation, and localization of PLD1 and PLD2, but less is known about their isoform-specific functions. Mammalian PLD1 and PLD2 hydrolyze PC preferentially. It is generally assumed that PA is the major messenger molecule generated by PLD. However, the molecular targets of PA *in situ* have not been identified.

### Structure and Catalytic Properties

Most mammalian PLD genes cloned so far belong to an extended gene superfamily, designated the "HKD" superfamily, that includes the eukaryotic PLDs, certain bacterial PLDs, phosphatidyltransferases/phospholipid synthases, endonucleases, and certain viral envelope proteins and their mammalian homologues. All members of this superfamily share a conserved motif (HxKx $_4$ Dx $_6$ GG/S) that confers a similar mechanism of catalysis. Mammalian PLD1 and PLD2 have a catalytic core region with significant internal homology between its N- and C-terminal halves. Noncatalytic N-terminal sequences include a Phox (PX) domain and a PH domain, both implicated in phospholipid binding and protein-protein interactions. An essential C-terminal region is also present.

### Subcellular Localization and Receptor-Mediated Activation

The precise localization of PLD1 and PLD2 is cell type and cell state specific. However, some common themes have emerged. Subcellular fractionations have demonstrated the presence of mammalian PLD1-like activities in multiple cellular membranes, including the nuclear envelope, ER, Golgi apparatus, transport/

secretory vesicles, and plasma membrane. In further studies, PLD1 localization was found predominantly in a perinuclear endomembrane system, including the Golgi apparatus and vesicles identified by organelle-specific markers as comprising the endosomal/lysosomal compartment. PLD1 is transiently translocated to the plasma membrane on stimulation of exocytosis in secretory cells. In contrast to PLD1, in resting cells, PLD2 is localized mainly in the plasma membrane, where it is associated in part with lipid rafts and caveolae. On stimulation, PLD2 is internalized to an endocytic compartment. Additional evidence suggests a role for PLD2 in receptor-mediated endocytosis. Finally, like PLD1, PLD2 also was localized to the Golgi apparatus. Receptor-mediated activation of PLD is complex and still incompletely understood. The activity of PLD1 is strictly receptor dependent and involves its interaction with proteins of the ARF, Rho, and PKC families. GPCRs activate PLD1 via a mechanism that requires PKC $\alpha$  and RhoA as upstream regulators, whereas its activation by immunoglobulin receptors requires Rac1 and ARF6. The activation of the Rho and ARF GTPases depends on stimulation of upstream exchange factors that, in turn, are recruited and stimulated by products of phosphoinositide 3-kinase and PI-PLC. However, much further work is required to fully elucidate these pathways in different cells stimulated by different receptor types. The regulation of PLD2 activity is less well understood and is certainly less signal dependent as compared with PLD1.

### Targets of PA and PA Metabolites

Numerous studies have described *in vitro* effects of PA on various enzyme activities. PA stimulates Ser/Thr kinase activity in cell extracts, although the specific kinases were not yet identified. PKC isoforms that are activated by PA *in vitro* include PKC $\eta$  and PKC $\zeta$ . More recently, PA was reported to interact with Raf1, a MEK kinase in the Erk1/2 cascade. A PA-activated tyrosine kinase activity present in neutrophil cytosol was identified recently as the proto-oncogene c-Fgr. Conversely, a protein tyrosine phosphatase (PTP1C/SHP1) was also shown to bind to, and be activated by, PA. Similarly, both a lipid kinase involved in biosynthesis of PIP<sub>2</sub> (PI 4-phosphate 5-kinase- $\alpha$ ) and a PIP<sub>2</sub>-specific phospholipase C (PLC $\gamma$ ) are regulated by PA *in vitro*. In addition, PA regulates the activity of several GTPases (e.g., ARF, dynamin) and GTPase-activating proteins (GAPs) as well as at least one cyclic AMP (cAMP) phosphodiesterases (PDE4D3 and PDE4A1).

Recently, PA was implicated as an activator of mTOR, a PI kinase-like protein involved in mitogenic signaling and reorganization of the actin cytoskeleton. Although PA-sensitive enzymes are candidate PA targets, direct evidence showing interaction with and regulation by PA *in situ* is still lacking and signaling pathways wherein PA produced by PLD plays an essential role have not been elucidated. Finally, it should be noted that PA may be converted into other signaling molecules (e.g., lyso-PA, DAG) that may act as additional downstream mediators of PLD activation. DAG may activate PKC isozymes and other targets, whereas lyso-PA is a ligand of the Edg GPCR family.

## SPHINGOMYELINASE

### Overview

Certain cytokines, neurotropic factors, apoptosis-inducing stimuli (e.g., cytotoxic drugs, loss of survival factors), and other environmental stresses trigger the cellular accumulation of ceramide. Receptor agonist-dependent production of ceramide occurs predominantly via activation of sphingomyelinase (SMase), a specific phospholipase C-type enzyme that catalyzes SM hydrolysis to ceramide and phosphocholine. However, *de novo* synthesis of ceramide has also been implicated in some cases. Ceramide regulates the action of target protein kinases and phosphatases, leading to cell growth arrest, senescence, differentiation, and apoptosis, depending on cell type and physiological context. Several distinct SMase activities have been identified in mammalian cells. An acid sphingomyelinase (A-SMase) and a Mg<sup>2+</sup>-dependent neutral sphingomyelinase (N-SMase) both are strongly implicated in signal transduction. A-SMases were purified and cloned and are implicated in the Niemann–Pick disease types A and B. Using A-SMase null mice, A-SMase activation was clearly implicated in radiation-induced apoptosis in lung and brain endothelial cells, in Fas-induced apoptosis in hepatocytes, and in apoptosis of mammalian oocytes. Similarly, A-SMase-deficient human lymphoblasts derived from Niemann–Pick patients were protected from radiation-induced apoptosis. An N-SMase activity has also been implicated in apoptosis triggered by death receptors, but the N-SMase gene(s) involved have not been identified conclusively.

### Structure and Catalytic Properties

The A-SMases comprise a group of 70- to 75-kDa proteins that have an N-terminal saposin B domain

and a C-terminal metallo-phosphoesterase domain. The former domain is likely to be involved in substrate-binding/membrane translocation, whereas the latter is the catalytic domain. As expected from its endosomal/lysosomal localization, A-SMase is optimally active at approximately pH 5. A secreted form of A-SMase, derived from the same gene, requires physiological concentrations of  $Zn^{2+}$  for activity. Several putative N-SMase genes were cloned, but it is still uncertain which one represents the signaling, plasma membrane-associated form. Two of these genes, N-SMase1 and N-SMase2, are related to bacterial neutral SMase and contain a catalytic endo/exonuclease/phosphatase domain. However, N-SMase1 is localized to the ER and it exhibits lyso-PAF activity *in vivo*. In contrast, recent work has shown that N-SMase2, which is localized mainly to the Golgi apparatus, exhibits a receptor-stimulated SMase activity *in vivo*.

### Receptor-Mediated Activation

A-SMase is a lysosomal/endosomal enzyme that is translocated to the plasma membrane outer leaflet on activation of the death receptor Fas. Although A-SMase is required for death receptor-induced apoptosis in some systems, it is not known how A-SMase translocation and activation occur except that recruitment of death receptor-associated adapter proteins (e.g., FADD, TRAF) are required. The activation of N-SMase presents a similar enigma. However, the picture that emerges is one where N-SMase activation is downstream to stimulation of cPLA<sub>2</sub> activity, formation of reactive oxygen species, and activation of upstream caspases.

### Targets of Ceramide and Ceramide Metabolites

Bona fide targets of ceramide should bind ceramide specifically and with high affinity *in vitro*, be recruited or activated in response to ceramide generation *in vivo*, and be implicated directly in ceramide signaling pathways. A relatively large number of putative ceramide targets have been suggested, among which are PKC $\zeta$ , cathepsin D, and protein phosphatases PP1 and PP2A. A canonical ceramide-binding domain has not yet been identified, although C1 domain variants have been proposed to act as ceramide-binding modules. However, ceramide may also act as a regulator of membrane dynamics by directly affecting membrane physical properties and thereby modulating intermembrane vesicle trafficking and intramembrane microdomain aggregation.

## PHOSPHOLIPASE A<sub>2</sub>

### Overview

Phospholipase A<sub>2</sub> (PLA<sub>2</sub>) hydrolyzes the *sn*-2 ester bond in PC and PE, releasing a free fatty acid and generating a lyso-phospholipid. The *sn*-2 position is often occupied by arachidonic acid (eicosatetraenoic acid [C<sub>20:4</sub>]), a precursor to a diverse group of important lipid mediators, termed eicosanoids, that include prostaglandins, thromboxanes, leukotrienes, and lipoxins. The release of arachidonic acid by PLA<sub>2</sub> is rate limiting for eicosanoid production and is a major receptor-regulated event catalyzed by a cytosolic, Ca<sup>2+</sup>-dependent enzyme designated cPLA<sub>2</sub> $\alpha$  that, together with cPLA<sub>2</sub> $\beta$  and cPLA<sub>2</sub> $\gamma$ , comprise the group IV PLA<sub>2</sub> gene family. The regulation of cPLA<sub>2</sub> $\beta$  and cPLA<sub>2</sub> $\gamma$  by receptors and their role in arachidonic acid mobilization have not been clarified. Another form, a Ca<sup>2+</sup>-independent PLA<sub>2</sub> (iPLA<sub>2</sub>), appears to be a constitutive housekeeping enzyme. Several secretory PLA<sub>2</sub> (sPLA<sub>2</sub>) forms comprise low-molecular weight proteins (13–15 kDA) that are released from the cells into the extracellular space. Although these enzymes play important roles in digestion and inflammation, they are not involved in receptor-regulated intracellular signal transduction pathways.

### Structure and Catalytic Properties

cPLA<sub>2</sub> $\alpha$  is a 749-amino acid protein that is composed of four distinct domains: an N-terminal C2 domain implicated in membrane binding; a catalytic domain A, containing a lipase consensus sequence; a non-conserved, isoform-unique linker sequence; and a C-terminal catalytic domain B. cPLA<sub>2</sub> $\alpha$  exhibits a preference for phospholipids with a polyunsaturated fatty acid at the *sn*-2 position but shows little selectivity among linolenic acid, arachidonic acid, or eicosapentaenoic acid. However, because of the low abundance of other polyunsaturated fatty acids, arachidonic acid is the major product *in vivo*. Although it was initially designated as a Ca<sup>2+</sup>-dependent PLA<sub>2</sub>, cPLA<sub>2</sub> $\alpha$  does not require Ca<sup>2+</sup> for catalysis. Instead, Ca<sup>2+</sup> is involved in membrane interaction through the C2 domain. The catalytic center of cPLA<sub>2</sub> $\alpha$  includes at least three essential amino acid residues (Arg<sup>220</sup>, Ser<sup>228</sup>, and Asp<sup>549</sup>) identified by homology to other lipases and by site-directed mutagenesis.

### Membrane Translocation and Activation

cPLA<sub>2</sub> $\alpha$  is normally localized in the cytosol. Activation of appropriate cell surface receptors results

in translocation of the enzyme to a nuclear and/or perinuclear compartment, enabling the enzyme to gain access to its membrane substrates. Translocation is dependent on persistent elevation of intracellular  $\text{Ca}^{2+}$  concentrations, and it can be mimicked by a  $\text{Ca}^{2+}$  ionophore. The C2 domain of cPLA<sub>2</sub>α is both necessary and sufficient for mediating its  $\text{Ca}^{2+}$ -dependent binding to membrane phospholipids. However, the selective perinuclear localization of cPLA<sub>2</sub>α indicates that an additional determinant(s) may participate in mediating its translocation. Indeed, the C2 domain is able to interact with vimentin in the presence of  $\text{Ca}^{2+}$ , whereas vimentin expression was shown to modulate arachidonic acid release. In addition, a plasma membrane-resident complex of p11/annexin II may prevent cPLA<sub>2</sub>α from translocating to the plasma membrane on activation, limiting cPLA<sub>2</sub>α translocation to the perinuclear region. Thus, the translocation of cPLA<sub>2</sub>α may readily provide arachidonic acid to the eicosanoid-producing enzymes, most of which are either constitutive or stimulus-dependent residents of the nuclear envelope and/or the ER.

### Receptor-Mediated Activation of cPLA<sub>2</sub>α

The current model for cPLA<sub>2</sub>α activation requires two independent events, namely a  $\text{Ca}^{2+}$ -dependent membrane translocation and phosphorylation by the MAP kinases Erk1 and p38. Persistent elevation of cytosolic free  $\text{Ca}^{2+}$  concentrations resulting from receptor-dependent activation of  $\text{Ca}^{2+}$  channels or mobilization of  $\text{Ca}^{2+}$  from intracellular stores is absolutely required for translocation to perinuclear membranes. The activation of cPLA<sub>2</sub>α also involves its phosphorylation on a critical Ser<sup>505</sup> residue found within a PXSP MAP kinase consensus motif. It is phosphorylated by several MAP kinases, including Erk1 and p38, and this is essential for maximal activation of cPLA<sub>2</sub>α. Phosphorylation of Ser<sup>727</sup> also appears to be required for maximal activation. However, the kinase responsible for Ser<sup>727</sup> phosphorylation has not been identified with certainty. The MAP kinases may be stimulated by  $\text{Ca}^{2+}$ -mobilizing receptors via activation of PKC or activation of Src-like kinases.

### Targets of Free Arachidonic Acid and Eicosanoids

Activation of cPLA<sub>2</sub>α generates free arachidonic acid, which may itself regulate downstream targets or be

further converted to bioactive eicosanoids. Among the possible direct targets of arachidonic acid are PKC isozymes, a N-SMase, PLC-γ, and phospholipase D. However, conversion of arachidonic acid to eicosanoids is probably the major physiological outcome of cPLA<sub>2</sub>α activation, as indicated by compartmentalization of the enzymes that produce prostaglandins, thromboxanes, leukotrienes, and lipoxins in the perinuclear region. Many eicosanoids are released from the cells and then act as extracellular agonists of specific cell surface receptors in an autocrine and/or paracrine mode of action. However, eicosanoids are now also recognized as ligands of nuclear receptors (e.g., peroxisome proliferator-activated receptors [PPAR]-α, -γ, and δ) that act intracellularly to transcriptionally regulate gene expression.

## OTHER PHOSPHOLIPASES

### PC-PLC

Soon after the second messenger function of DAG was established, it became apparent that a second phase of DAG production is generated by hydrolysis of PC. Whereas the source of “early” DAG is PIP<sub>2</sub>, “late” DAG is probably derived by hydrolysis of PC, either directly via PC-phospholipase C or indirectly via PLD and phosphatidic acid phosphohydrolase. Direct PC-phospholipase C-mediated generation of DAG has been suggested to occur in the absence of PI-PLC and PLD activation in a number of cells, especially in response to cytokines and growth factors. A related enzyme may be activated by certain GPCRs in fibroblasts. Unfortunately, the enzyme(s) involved in these responses have not been purified or cloned; consequently, their mechanism(s) of activation by receptors and function(s) are completely obscure.

### Lysophospholipase D

A plasma lysophospholipase (LPLD/autotaxin) catalyzes the hydrolysis of lyso-PC to lyso-PA, an important ligand of the Edg subfamily of GPCRs involved in regulation of cell growth and differentiation. Unlike the other phospholipases discussed herein, LPLD/autotaxin has not been shown to be acutely regulated by receptors, but its expression varies under different physiological and developmental conditions. Future studies may shed more light on its receptor-mediated regulation.

## CONCLUSION

This article has provided an overview of well-established SAPLs. Although much has been learned in the 20-plus years since the discovery of the first lipid-signaling pathway, a complete understanding of the role of phospholipases and their lipid messengers remains an elusive goal. It should be borne in mind that these signaling effector enzymes would always operate within a signaling network that involves a host of other pathways, all engaged in massive parallel processing of the receptor-initiated signal. Complete understanding of such networks, and in particular the role of SAPLs within them, is a future goal that requires much further research to attain. Another important issue is that receptor-regulated phospholipases, unlike other effector enzymes involved in signal transduction, generate messengers that are lipids. Such molecules reside in the lipid bilayer; thus, they may act not only to recruit and modulate activity of downstream target proteins but also to modify the membrane's biophysical properties and thereby membrane dynamics. Finally, unlike most other messengers, lipid messengers are interconvertible; that is, they can be metabolically converted from one biologically active form into another. This would add additional spatial and temporal complexities to any attempt to define, in molecular terms, the logic of phospholipase-dependent signaling pathways.

## See Also the Following Articles

G Proteins and Effectors • Lipid Second Messengers and Receptors • Phosphatidylinositol Turnover and Receptors • Receptor Tyrosine Kinase

## Further Reading

- Capper, E. A., and Marshall, L. A. (2001). Mammalian phospholipases A<sub>2</sub>: Mediators of inflammation, proliferation, and apoptosis. *Prog. Lipid Res.* **40**, 167–197.
- Fukami, K. (2002). Structure, regulation, and function of phospholipase C isozymes. *J. Biochem. (Tokyo)* **131**, 293–299.
- Goni, F. M., and Alonso, A. (2002). Sphingomyelinases: Enzymology and membrane activity. *FEBS Lett.* **531**, 38–46.
- Hannun, Y. A., Luberto, C., and Argraves, K. M. (2000). Enzymes of sphingolipid metabolism: From modular to integrative signaling. *Biochemistry* **40**, 4893–4903.
- Hirabayashi, T., and Shimizu, T. (2000). Localization and regulation of cytosolic phospholipase A<sub>2</sub>. *Biochim. Biophys. Acta* **1488**, 124–138.
- Irvine, R. F. (2003). Nuclear lipid signalling. *Nat. Rev. Mol. Cell Biol.* **4**, 349–360.
- Kazanietz, M. G. (2002). Novel “nonkinase” phorbol ester receptors: The C1 domain connection. *Mol. Pharmacol.* **61**, 759–767.
- Levade, T., Andrieu Abadie, N., Segui, B., Auge, N., Chatelut, M., Jaffrezou, J. P., and Salvayre, R. (1999). Sphingomyelin-degrading pathways in human cells: Role in cell signalling. *Chem. Phys. Lipids* **102**, 167–178.
- Liscovitch, M., Czarny, M., Fiucci, G., and Tang, X. (2000). Phospholipase D: Molecular and cell biology of a novel gene family. *Biochem. J.* **345**, 401–415.
- Mellor, H., and Parker, P. J. (1998). The extended protein kinase C superfamily. *Biochem. J.* **332**, 281–292.
- Rhee, S. G. (2001). Regulation of phosphoinositide-specific phospholipase C. *Annu. Rev. Biochem.* **70**, 281–312.
- Six, D. A., and Dennis, E. A. (2000). The expanding superfamily of phospholipase A<sub>2</sub> enzymes: Classification and characterization. *Biochim. Biophys. Acta* **1488**, 1–19.
- Taylor, C. W., and Laude, A. J. (2002). IP<sub>3</sub> receptors and their regulation by calmodulin and cytosolic Ca<sup>2+</sup>. *Cell Calcium* **32**, 321–334.
- van Blitterswijk, W. J., van der Luit, A. H., Veldman, R. J., Verheij, M., and Borst, J. (2003). Ceramide: Second messenger or modulator of membrane structure and dynamics. *Biochem. J.* **369**, 199–211.

interactions are possible sources of the observed alteration in ligand binding and signaling (e.g., the site of modification may be a significant distance from the ligand binding site). Despite these concerns, results from site-directed mutagenesis identify the amino acids of the receptor that play a role (although not necessarily a direct role) in the functioning of the receptor.

In specific cases, evidence of a direct interaction between ligand and receptor can be obtained by a combination of receptor modification via site-directed mutagenesis and alteration of the ligand. With mutation of either the ligand or the receptor alone, function is lost, but when there is mutation of both, a fully functional receptor results; this is often called rescue of function. Often, this is achieved by switching the amino acids of the ligand and receptor. For example, replacing a negatively charged residue with a positively charged residue of the ligand leads to loss of activity in the natural receptor, but when a positively charged residue is replaced by a negatively charged amino acid the receptor is fully functional. This is a clear indication of a direct interaction between the ligand and the receptor.

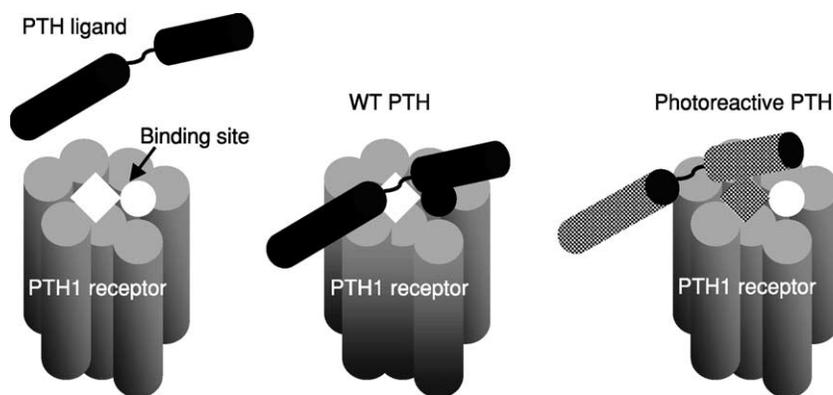
## PHOTOAFFINITY LABELING

An alternative method for determining direct interactions is photoaffinity labeling. Here, the ligand is modified to incorporate a photoreactive amino acid (or moiety). Upon reaction with a light source of the appropriate frequency, this amino acid forms a covalent bond with a residue of the receptor. Upon purification of the ligand–receptor photo adduct (covalent

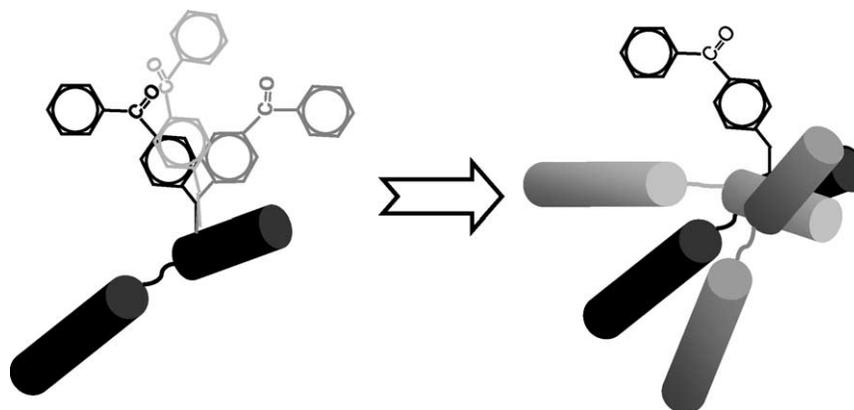
ligand–receptor complex) and enzymatic cleavage to break the receptor into small fragments, the site of the receptor labeled by the ligand can be identified.

This approach provides an exact point of contact between the receptor and its ligand, but caution must be used in the design of the experiment and in the interpretation of the result. The biological properties of the modified ligand should be very similar to those of the wild-type ligand and should always be taken into account in the interpretation of the results. Unfortunately, only a limited number of photo-labile amino acids are available; therefore, they often deviate in physical properties (e.g., hydrophobicity and size) from those of the amino acids that they replace. Electrostatic and hydrophobic contributions and steric complementarity play a major role in defining the receptor–ligand interaction, and alteration in any of these properties may lead to an altered mode of binding. An indication may be the modification of binding affinity or signaling properties. [Figure 1](#) illustrates how a change in bulkiness may affect the specific point of interaction.

Early studies by Rosenblatt on the interaction of PTH with PTH-1 used a photoreactive PTH agonist in which residue 13 (a positively charged lysine) was modified with a photoreactive benzoylphenone moiety. The agonist cross-links specifically to the PTH-1 receptor, but the modification introduced alters both the electrostaticity (positive charge substituted by hydrophobic moiety) and the size (a bulky benzoylphenone added to the lysine side chain) of the targeted site of interaction (position 13). The agonist displays reduced binding affinity to the receptor, and the effect of the chemical modifications must be taken into account in the interpretation of results. [Figure 2](#) illustrates the



**Figure 1** Map of ligand–receptor interactions by photoaffinity labeling. The binding pockets in the PTH1 receptor are denoted by circles [pocket of amino acid in wild-type (WT) PTH] and squares (binding pocket of the photolabile residue). The physical properties of the photolabile residue differ from those of the natural residue of the ligand; therefore, the photolabile residue may not bind to the exact location of the natural amino acid.



**Figure 2** Resolution obtained from photoaffinity-labeling experiments resulting from the inherent flexibility of the ligand containing the photolabile amino acid. Superposition of the possible orientations of the photolabile residue (left) leaves uncertainty in defining the position of the ligand within the receptor (right) as determined by the photoaffinity contact points.

uncertainty at the structural level introduced by the use of an extended side chain at the cross-linking site: The conformational flexibility of the side chain (which links the ligand to the receptor) exemplified on the left results in uncertainty in the position of the ligand (right). The practical consequence is that the characterization of the interaction is an approximation, as could be derived by a picture with a resolution of approximately 10 Å.

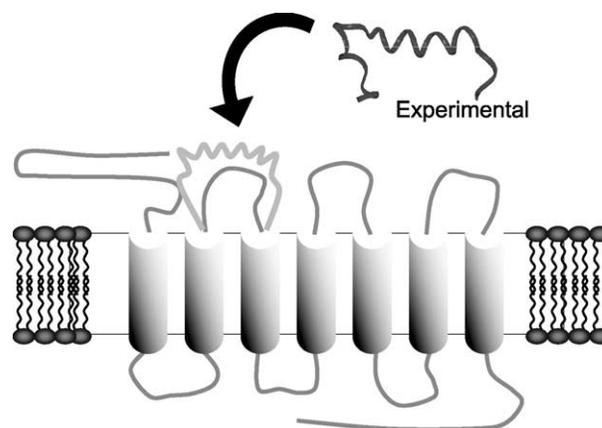
## STRUCTURAL INSIGHT

The results from the site-directed mutagenesis and photoaffinity-labeling studies provide loosely defined distance restraints between the ligand and receptor. As noted previously, care must be used when applying these restraints in the generation of a model of the ligand–receptor complex. Nonetheless, these loose restraints are important guides for the description of ligand–receptor interactions in the context of existing models or experimental structures of the ligand and receptor.

Confidence in the results of combining the ligand and receptor is greatly enhanced by knowledge of the structures of the two players; experimental data defining the conformational preferences of the ligand and the receptor is vital for the development of a meaningful ligand–receptor complex. The more detailed the structural features of the ligand and receptor, the more detailed the resulting model of the ligand–receptor complex.

Unfortunately, in the case of the PTH-1 receptor, no experimental structure of the intact receptor is available. Based on sequence homology (a measure of similarity of the amino acid sequences of two protein receptors), it is assumed that PTH-1 is similar in structure to the transmembrane receptors bacteriorhodopsin and rhodopsin, for which limited

structural data are available. Indeed, the PTH-1 receptor has been shown to contain seven transmembrane helices, an extracellular N terminus, and an intracellular C terminus (Fig. 3). One of the major limitations of the structural features of the bacteriorhodopsin and rhodopsin receptors is that the extracellular region, most important for ligand binding as indicated by both site-directed mutagenesis and photoaffinity labeling, is poorly defined. The temperature factors (measures of the definition of the structure, with the lower factor being the more defined) of the extracellular portion in the X-ray structures of these receptors are very high. Therefore, in the development of the model for the PTH-1



**Figure 3** Incorporation of experimentally determined structural features into the ligand–receptor model. The nuclear magnetic resonance structure of the first extracellular loop of the PTH1 receptor is seamlessly inserted into the model of the seven-transmembrane helices of the receptor. Such structural insight greatly enhances the reliability of the computational models of the ligand–receptor complexes.

receptor, only the topological display and orientation of the seven transmembrane helices could be used. No knowledge of the extracellular domains, the N terminus, and three extracellular loops could be ascertained.

To build a model structure in our laboratory, we used the known characteristics of the receptor (e.g., the presence of seven transmembrane helices), experimental structures derived by NMR studies on receptor fragments, and extensive computer refinement. Figure 3 illustrates how the NMR structure of the first extracellular loop of PTH-1 contributes to the overall model.

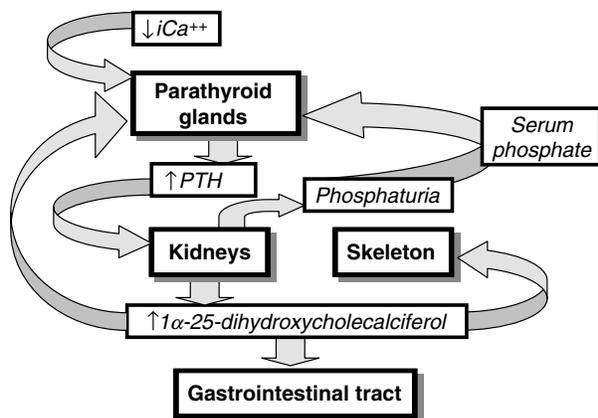
Using a similar approach for the other extracellular domains of the PTH-1 receptor, a high-resolution picture, containing the structural preferences of these domains important for the interaction of PTH, can be developed. The insight obtained from this model should afford the development of novel therapeutic agents for the control of this important receptor.

### See Also the Following Articles

G Protein-Coupled Receptors • Receptor Serine/Threonine Kinases

### Further Reading

- Bisello, A., Adams, A. E., Mierke, D. F., Pellegrini, M., Rosenblatt, M., Suva, L. J., and Chorev, M. (1998). Parathyroid hormone–receptor interactions identified directly by photocross-linking and molecular modeling studies. *J. Biol. Chem.* **273**, 22498–22505.
- Lee, C., Luck, M. D., Juppner, H., Potts, J. T., Jr., Kronenberg, H. M., and Gardella, T. J. (1995). Homolog-scanning mutagenesis of the parathyroid hormone (PTH) receptor reveals PTH-(1–34) binding determinants in the third extracellular loop. *Mol. Endocrinol.* **9**, 1269–1278.
- Mannstadt, M., Luck, M. D., Gardella, T. J., and Juppner, H. (1998). Evidence for a ligand interaction site at the amino-terminus of the parathyroid hormone (PTH)/PTH-related protein receptor from cross-linking and mutational studies. *J. Biol. Chem.* **273**, 16890–16896.
- Piserchio, A., Bisello, A., Rosenblatt, M., Chorev, M., and Mierke, D. F. (2000). Characterization of parathyroid hormone/receptor interactions: Structure of the first extracellular loop. *Biochemistry* **39**, 8153–8160.
- Roelz, C., Pellegrini, M., and Mierke, D. F. (1999). Molecular characterization of the receptor–ligand complex for parathyroid hormone. *Biochemistry* **38**, 6397–6405.
- Zhou, A. T., Bessalle, R., Bisello, A., Nakamoto, C., Rosenblatt, M., Suva, L. J., and Chorev, M. (1997). Direct mapping of an agonist-binding domain within the parathyroid hormone/parathyroid hormone-related protein receptor by photoaffinity crosslinking. *Proc. Natl. Acad. Sci. USA* **94**, 3644–3649.



**Figure 1** Pathogenic factors in the development of secondary hyperparathyroidism in patients with chronic renal disease.

integrin  $\alpha$ -V- $\beta$ -3. Local chemotactic factors, such as osteopontin and bone sialoprotein, may act as the natural ligands for this integrin and thus direct chemotaxis and organize the osteoclast for its resorptive role in bone turnover.

Osteoclastic activity is inhibited by products of bone resorption, namely calcium and transforming growth factor- $\beta$  (TGF- $\beta$ ). This is achieved by separation of osteoclasts from the bone surface. Osteoblastic growth factors, including TGF- $\beta$ , basic and acidic fibroblast growth factor, and bone morphogenic protein, are released during bone resorption as well.

Bone formation is initiated by the release of osteoblastic precursors from the marrow stromal cells. The precursor pool mobilizes to the active resorptive site, forming cuboidal osteoblasts, resulting in matrix synthesis and mineralization. Under direct stimulation by PTH, insulin-like growth factor-1 (IGF-1), along with calcitriol, promotes osteoblast proliferation.

In adults, the remodeling cycle results in a discrepancy between bone formation and bone resorption, with the latter predominating. This difference increases with age, predisposing the elderly to osteopenia.

## PATHOPHYSIOLOGY

### High-Turnover Bone Disease (Secondary Hyperparathyroidism)

Osteitis fibrosa is the most common form of ROD. In this form, both bone resorption and remodeling are increased. However, the remodeled bone is mostly nonlamellar, or non-nutrient supplied, rendering it weak with a predisposition to osteopenia and fractures.

The underlying pathophysiologic mechanism is a combined effect of a reduction in 1-hydroxylation of

25-hydroxyvitamin D<sub>3</sub> in the kidney and hyperphosphatemia due to the presence of renal disease. Diminished circulating levels of calcitriol and its associated hypocalcemia, along with (and perhaps even more important) an elevated serum phosphate concentration, lead to the development of secondary hyperparathyroidism (Fig. 1). This phenomenon is usually seen once the GFR reaches 50–70% of the normal GFR of 100–125 ml/min.

In normal circumstances, calcitriol inhibits PTH synthesis at the pre-pro-PTH mRNA level by binding to calcitriol receptors located in the parathyroid gland. In the presence of renal disease, however, the density of these calcitriol receptors is diminished. Some authors suggest that this leads to a loss of feedback inhibition and a lower threshold for PTH secretion. Furthermore, it has been postulated that the parathyroid gland may lack efficient control of PTH synthesis and secretion in response to the plasma ionized calcium concentration in the presence of uremia. Using *in situ* hybridization and immunohistochemistry techniques, the expression of the calcium-sensing receptor (CaSR) in parathyroid tissue from nonuremic and uremic patients with hyperparathyroidism has been compared. In patients with uremia, the expression of both receptor mRNA and protein is substantially depressed in nodular areas of parathyroid tissue. Furthermore, the presence of hyperphosphatemia stimulates release of PTH by the parathyroid glands in the absence of changes in serum ionized calcium or calcitriol concentrations. The combined effect of hypocalcemia, hyperphosphatemia, and calcitriol deficiency leads to parathyroid glandular hypertrophy and hyperplasia (Fig. 1).

As observed in animal studies, there appears to be end-organ resistance to the effects of PTH in the presence of uremia. It has been suggested that downregulation of PTH receptors plays a potential role in the expression of secondary hyperparathyroidism in uremia. A blunted calcemic response to PTH stimulation in both uremic dogs and rats has been observed. Using *in situ* hybridization techniques, the expression of cellular PTH/PTH-related protein (PTH/PTHrP) receptor (PTH1R) mRNA in human osteoblasts in the presence and absence of renal disease has been examined. Studies show a lower density of PTH1R mRNA in uremic patients when compared to normal bone and high-turnover bone in those without renal disease. Another potential mechanism for skeletal resistance to PTH in uremia is the overproduction of osteoclastogenesis inhibitory factor (OCIF) or osteoprotegerin. OCIF acts as a “decoy” receptor for osteoclast differentiation factor, produced by osteoblasts in

response to PTH stimulation, and inhibits osteoclastic differentiation and maturation. In patients with advanced renal disease, serum OCIF levels increased as renal function decreased. An approximately four-fold elevation in OCIF levels was observed in ESRD patients. Furthermore, OCIF levels have been shown to return to normal as early as 2 weeks following renal transplantation. It is therefore possible that OCIF may act at the skeletal level to produce resistance to the action of PTH.

Additional studies have focused on the potential role of both circulating and locally produced cytokines and growth factors in the development of ROD. TGF- $\beta$ , in particular, is synthesized by both osteoblasts and osteoclasts and stored in bone matrix. Through *in situ* hybridization techniques, TGF- $\beta$  mRNA was primarily localized to osteoblasts and appeared to promote osteoblastic activity and thus bone formation. The degree of TGF- $\beta$  mRNA expression correlated with markers of bone formation, such as bone formation rate (BFR), trabecular apposition rate, and mineralizing surfaces. Furthermore, PTH has been shown to stimulate TGF- $\beta$  expression in cultures of normal human osteoblastic-like cells.

In addition to TGF- $\beta$ , IGF-1 and IGF-2 have been studied as potential mediators of ROD. IGF-1 and IGF-2 play important roles in bone formation by inhibiting the degradation of collagen and promoting osteoblastic recruitment and matrix deposition. It has been suggested that patients with uremia have resistance to the effects of IGF-1, which may be due to a defect at the receptor site. Using *in situ* hybridization techniques, a reduction in IGF-1 mRNA in bone of

subjects with renal disease compared to that in normal and high-turnover bone in the absence of renal disease has been observed.

### Low-Turnover Bone Disease (Adynamic Bone Disease and Osteomalacia)

Many patients with renal disease have normal or near-normal circulating PTH levels, often as a result of the treatment of abnormalities of calcium and phosphate metabolism. As many as 40% of patients with ROD have low-turnover bone disease. In both adynamic bone disease (ABD) and osteomalacia, bone formation and turnover are decreased (Table I). The distinction, however, is in the quality of the bone formed. In osteomalacia, newly formed bone collagen, or osteoid, is largely unmineralized. This defect has historically been attributed to vitamin D deficiency and, recently, to the use of aluminum and other heavy metal compounds in patients with renal disease.

Aluminum affects the skeleton and its mineralization in several ways. Chronic intoxication leads to sustained inhibition of osteoblastic differentiation and osteoclastic activity. Furthermore, *in vitro* studies have shown impaired secretion of PTH from parathyroid cells in the presence of aluminum. The role of calcitriol deficiency in the pathogenesis of osteomalacia is uncertain. Other factors may play a role in osteomalacia associated with uremia because this disease remains problematic despite the dramatic reduction in the use of aluminum-based phosphate binders during the past decade.

Table I Clinical Characteristics of Renal Osteodystrophy<sup>a</sup>

	Ca	Phos	Alk phos	PTH	Histology
<i>Osteitis fibrosa</i>	Variable	↑	↑	↑↑	Peritrabecular and marrow fibrosis ↑ osteoclasts and osteoblasts ↑ BFR
<i>Osteomalacia</i>	Variable	Normal or ↑	Normal or ↑	Slightly ↑	↓ osteoclasts and osteoblasts ↑ osteoid ↑ osteoid seam width ↓ BFR ↑ aluminum
<i>Adynamic bone disease</i>	Normal or ↑	Normal or ↑	Normal or ↑	Less than twice upper limit of normal	↓ osteoclasts and osteoblasts ↓ osteoid seam width ↓ BFR
<i>Mixed disease</i>	Variable	Normal or ↑	Normal or ↑	Normal or ↑	Mixed components of histology detailed above

<sup>a</sup>Abbreviations used: Alk phos, serum alkaline phosphatase; Phos, serum phosphorus; Ca, serum calcium; PTH, serum parathyroid hormone; BFR, bone formation rate.

ABD represents an emerging problem in patients with renal disease. The disease is characterized by a low bone turnover rate with a decrease in both the number and activity of osteoclasts and osteoblasts (Table I). The pathogenesis of ABD is unclear; however, various potential mechanisms have been proposed. Several bone-suppression factors have been implicated, including IL-4, IL-11, nitric oxide, and fragments of PTH-related protein. A deficiency of bone morphogenetic protein-7, a member of the family of TGF- $\beta$  cytokines normally produced by renal tubular cells, may play a role in the development of ABD. Patients who have undergone parathyroidectomy are particularly prone to the development of ABD.

The presence of near-normal or normal levels of circulating PTH is thought to be the predominant predisposing factor for the development of ABD. Use of calcium supplementation, either in ingested form as a phosphate binder or in aqueous form in dialysate solution, when combined with decreased skeletal uptake of calcium increases the propensity to develop hypercalcemia and thus decrease secretion of PTH (Fig. 1). Perhaps even more important, the widespread use of vitamin-D analogues in patients with renal disease leads to a secondary reduction in circulating PTH levels. The combination of hypercalcemia and near-normal or normal levels of circulating PTH is classically associated with the presence of ABD on bone histology.

## HISTOLOGIC CHARACTERISTICS OF ROD

Bone biopsy is the definitive method of establishing the type of osteodystrophy present in a patient with renal disease. Although not universally performed in patients with renal disease, primarily due to patient discomfort and the relatively few pathology centers in the United States that interpret bone biopsies, the histologic characteristics provide valuable insight into the potential pathophysiologic mechanisms involved in the development of bone disease in patients with uremia.

In general, bone histology provides the clinician with important structural determinants of both trabecular bone and marrow. Particular static bone histomorphometric parameters include the number of osteoclasts and osteoblasts, the degree of marrow fibrosis, the amount of unmineralized bone matrix (osteoid), the length and thickness of osteoid seams, the amount of aluminum staining, and the extent of iron stores in the marrow. In addition, dynamic properties such as BFR, estimated by tetracycline

labeling, can be determined and provide valuable insight into the rate of bone turnover.

## High-Turnover Bone Disease

In osteitis fibrosa, significant fibrosis is observed within the peritrabecular areas and marrow, in part due to the secretion of fibrous tissue by fibroblast-like cells. There is an increase in both the number and the size of osteoclasts, with an associated increase in the number of resorptive areas within cancellous bone. In addition, the number of osteoblasts and amount of newly formed osteoid are increased. There is derangement of collagen fibrils, forming a woven straw basket appearance within osteoid seams. As calculated from tetracycline labeling, the BFR is increased. There is only mild staining for aluminum, covering less than 30% of the trabecular surface.

## Low-Turnover Bone Disease

There is an overall reduction in the number of osteoblasts and osteoclasts. The number of bone resorptive sites is decreased. There is either a normal or reduced quantity of unmineralized bone, with relatively few sites of active bone formation. In osteomalacia, defective bone mineralization leads to excessive osteoid formation, with a widened osteoid seam width. In ABD, however, bone mineralization may exceed collagen formation, leading to a normal or narrower osteoid seam width. The characteristic changes seen in secondary hyperparathyroidism, such as peritrabecular and marrow fibrosis, are absent. The rate of bone turnover, as assessed by tetracycline labeling, is dramatically decreased and may approach zero. In aluminum-associated osteomalacia, aluminum staining covers more than 30% of the trabecular surface, with the degree of staining correlating with the degree of histologic changes.

## CLINICAL CHARACTERISTICS OF ROD

Although the definitive diagnosis of ROD is made by bone biopsy, there are several important clinical features that may help the clinician distinguish high-turnover from low-turnover bone disease. In general, the following laboratory markers are useful when assessing bone disease in patients with renal disease: serum alkaline phosphatase, serum calcium, serum phosphate, and serum PTH levels (Table I). It is important to follow such laboratory markers once the reduction in GFR has reached approximately 50% of normal. Furthermore, roentgenographic

features may provide insight into the type of ROD present in a particular patient.

### High-Turnover Bone Disease

The most common symptoms associated with osteitis fibrosa are bone pain and pruritis. Joint discomfort and stiffness, often localized to the lower back, hips, and legs, may be present and can be attributed to arteriolar and periarticular calcification. These symptoms may mimic acute arthritis and must be distinguished from other arthropathies that might be present in patients with chronic renal disease. Symptoms are usually insidious in onset and frequently nonspecific and generalized. The discomfort may be aggravated by either movement or change in posture. On physical examination, bony tenderness, palpable parathyroid glands, and skin excoriations with palpable subcutaneous calcium deposits may be seen. In extreme cases of secondary hyperparathyroidism with an elevated calcium phosphate product (usually more than 65–70), ischemic necrosis of the skin, muscles, and subcutaneous tissues, called calciphylaxis, may ensue. Furthermore, fractures and pseudofractures of the rib cage may lead to pulmonary insufficiency (restrictive lung disease).

In osteitis fibrosa, serum alkaline phosphatase levels are elevated due to the presence of high bone turnover (Table I). Serum calcium levels (either total or ionized) are usually normal or slightly low, attributable to the reduction in 1-hydroxylation of 25-hydroxyvitamin D<sub>3</sub> by the kidneys. Serum phosphate levels are usually elevated, often 6 or 7 mg/dl or higher. Levels of circulating intact PTH (iPTH) are elevated, often >300 pg/ml. Serum PTH levels >1000 pg/ml suggest marked parathyroid glandular hyperplasia (Table I).

Radiologic findings in osteitis fibrosa are highly variable. In mild or early disease, radiologic findings may be absent. However, in advanced disease, subperiosteal bone loss can be seen, which can be best detected on the radial side of the second and third phalanges, the distal ends of the clavicles, and at the sacroiliac joints. Occasionally, blunting of the fingertips is present. In addition, the presence of disorganized high bone turnover may present as osteosclerosis on radiographic examination, producing a “rugger jersey” appearance of the thoracic spine. Bony derangement of the skull may produce a characteristic “salt and pepper” appearance. Calcifications in blood vessels and subcutaneous tissues are occasionally seen. Technetium bone scintigraphy shows increased uptake in various areas of the skeleton.

### Low-Turnover Bone Disease

Adynamic ROD was originally described in the 1970s and 1980s as a consequence of aluminum toxicity related to the use of dialysate water and aluminum-based phosphate binders. Patients presented with bone and muscle discomfort and had significant aluminum staining on bone biopsies. The recognition of aluminum toxicity led to improvements in water purification and the use of calcium-based phosphate binders. As a result, the incidence of aluminum toxicity and its associated bone histologic characteristics of osteomalacia has declined in recent years. However, the development and widespread use of newer vitamin D analogues, with the concurrent use of calcium-based phosphate binders, have led to a greater proportion of renal disease patients with suppressed circulating PTH levels and an increased incidence of ABD. Osteomalacia persists in patients with renal disease, suggesting that factors besides aluminum may play a role in its pathogenesis.

Patients with osteomalacia tend to experience a greater degree of both bone pain and fractures compared to those with osteitis fibrosa. Skeletal deformities, such as scoliosis and kyphosis, are seen and are predominantly localized to the thoracic and lumbar spine. There is also an increased incidence of anemia, which is thought to be due, in part, to an impaired response to recombinant human erythropoietin in patients with renal disease and aluminum toxicity. Neurologic signs, including dyspraxia, asterixis, seizures, and dementia, have been noted in cases of severe toxicity.

In contrast to those with osteomalacia, patients with ABD are less likely to complain of muscular pain and weakness. Joint stiffness is also less common than in patients with osteitis fibrosa.

The risk of vertebral and hip fractures is increased in patients with ABD compared to patients with other types of ROD. In a retrospective review of 1270 ESRD patients observed during a 10-year period, the risk of hip fracture was greatest in those with lower serum PTH levels. Furthermore, patients with PTH values >195 pg/dl did not have an increased risk of hip fracture. In this study, bone biopsies were not systematically performed. In a different study, the effects of circulating iPTH levels on bone mineral density (BMD) were examined in 187 male hemodialysis patients. Patients in the lowest tertile for serum iPTH (mean, 32.9 ± 16.4 pg/ml) had a 2.4-fold increase in vertebral fractures compared to patients in the middle tertile group (mean iPTH, 116.2 ± 40.9 pg/ml). Again, bone histology was not examined.

However, the optimal modality for determining BMD in patients with renal disease is unknown and remains a topic of intense debate.

Serum calcium concentrations in patients with low-turnover bone disease are often normal but may be elevated, primarily due to the sometimes large prescribed amount of daily calcium ingestion in patients with renal disease. Despite the presence of vitamin D deficiency in uremia, passive or diffusional calcium transport across the intestinal membrane exists in relation to the amount of ingested calcium. Coupled with the use of vitamin D analogues in the presence of adynamic bone, this often leads to the development of hypercalcemia (Table I). Serum alkaline phosphatase levels are often normal or slightly elevated. Serum PTH levels are usually <150 pg/ml and may be <100 pg/ml (normal range, 10–65 pg/ml) (Table I). In fact, the combination of a serum PTH level <150 pg/ml and a serum calcium concentration >10 mg/ml has at least an 82% positive predictive value for ABD.

Serum aluminum concentrations poorly reflect the quantity deposited in bone and are mostly a measure of daily ingestion. In a study of 258 peritoneal dialysis and hemodialysis patients, a poor correlation between serum aluminum levels and histologic findings on bone biopsy was observed. Only 50% of patients with serum aluminum levels >40 µg/liter were found to have histologic evidence of low bone turnover associated with at least 25% bone surface aluminum staining. Due to the shortcomings of noninvasive screening for aluminum bone toxicity, other measures such as the deferoxamine infusion test have been used diagnostically. Deferoxamine is a chelating agent that mobilizes and binds aluminum, and in the setting of bone toxicity it will produce dramatic increases in serum aluminum levels after its infusion. However, the sensitivity and specificity of this test have been questioned.

Radiologic features of low-turnover bone disease are less specific than those for osteitis fibrosa. Rib and vertebral fractures are the most common abnormalities. Low skeletal uptake of tracer is observed on nuclear bone scintigraphy.

## TREATMENT STRATEGIES

The primary goal of minimizing the severity of ROD involves the following general approaches: control of serum phosphate, maintenance of normal levels of serum calcium, and the prudent use of vitamin D analogues. Promising therapeutic agents, such as the calcimimetic agents, have been evaluated in patients with renal disease. In advanced cases of secondary

hyperparathyroidism, parathyroidectomy may be indicated.

## Control of Serum Phosphate

The most important intervention in the management of bone disease in uremic patients is the strict adherence to a low-phosphate diet (400–800 mg/day). In uremic rats, dietary phosphate restriction prevented the development of parathyroid hyperplasia, independent of serum ionized calcium or calcitriol concentrations. Avoidance of foods high in phosphorus content, such as dairy products, corn, whole-grain breads and cereals, broccoli, legumes, nuts, beef or chicken liver, and dark colas, is recommended. Consultation with a dietician is often beneficial.

In addition to a low-phosphate diet, use of phosphate binders is often necessary. Phosphate binders decrease intestinal phosphate absorption. Calcium-based binders, such as calcium acetate and calcium carbonate, are ingested with each meal or snack and can be dosed depending on the phosphate intake of a particular meal. The total daily dose of the phosphate binder is highly variable from patient to patient and must be titrated accordingly. It is a reasonable approach to target high-normal levels of serum phosphate while avoiding the development of hypercalcemia. The efficacy and the number of tablets required daily are equivalent for both drugs. Use of aluminum-based phosphate binders should be avoided.

Sevelamer hydrochloride (RenaGel), a relatively new synthetic nonabsorbable phosphate binder that is both calcium and aluminum free, binds phosphate in the intestinal tract and prevents its absorption. Sevelamer has been shown to be as effective as calcium carbonate and calcium acetate in suppressing serum phosphate levels in ESRD patients. Furthermore, the incidence of hypercalcemia, as determined from several clinical trials, is reduced with the use of sevelamer. In addition to its use as a phosphate binder, sevelamer has been shown to decrease total cholesterol and low-density lipoprotein cholesterol and increase high-density lipoprotein cholesterol levels in ESRD patients. Its high cost is a drawback to its widespread use.

## Maintenance of Normal Levels of Serum Calcium

Due to the reduction in serum levels of 1,25-dihydroxyvitamin D<sub>3</sub>, hypocalcemia is often seen in patients with renal disease. In addition to strict control

of serum phosphate levels, it is equally important to maintain normal serum calcium levels in an attempt to minimize oversecretion of PTH and the development of parathyroid hyperplasia. This is achieved in a variety of ways, such as by use of calcium-based phosphate binders, adjustment of the dialysate calcium concentration in hemodialysis patients, and use of vitamin D analogues. If hypocalcemia becomes problematic, it may be necessary to at least temporarily use calcium supplements given in between meals (rather than with meals as a phosphate binder).

### Vitamin D Analogues

The use of vitamin D analogues has been primarily limited to the treatment of osteitis fibrosa in patients with renal disease. The PTH-lowering effect of intravenous administration of calcitriol three times per week in hemodialysis patients was first described in 1989. In the 1990s, second-generation vitamin D analogs [19-*nor*-1,25-(OH)<sub>2</sub>D<sub>2</sub> (paricalcitol, Zemplar) and 1 $\alpha$ -(OH)-vitamin D<sub>2</sub> (doxercalciferol, Hectorol)] became commercially available and were developed in an attempt to decrease intestinal calcium and phosphate absorption. Although these agents have dramatically affected the treatment of secondary hyperparathyroidism in patients with renal disease, their overzealous use in the ensuing years has led to the emergence of ABD.

Given the deficiency of calcitriol in patients with renal disease, the use of calcitriol and other vitamin D analogues is almost universally employed in the treatment of secondary hyperparathyroidism. Vitamin D analogues are often prescribed intravenously during dialysis treatments, improving patient compliance; however, oral forms are commercially available for use in patients with chronic kidney disease.

Due to the deleterious effects of secondary hyperparathyroidism on bone mineral metabolism, treatment with vitamin D analogues is primarily undertaken to lower circulating PTH levels and facilitate calcification of osteoid. The primary objective is suppression of both PTH gene transcription and PTH secretion. However, therapeutic doses of calcitriol and other vitamin D analogues promote intestinal calcium absorption, increasing the propensity to develop hypercalcemia and metastatic calcifications. Furthermore, intestinal phosphate absorption is increased, primarily with the use of calcitriol, leading to the development of hyperphosphatemia. Thus, the therapeutic index for these agents is relatively narrow,

requiring the clinician to closely monitor the levels of serum PTH, calcium, and phosphate.

In general, the use of vitamin D analogues is recommended when the level of circulating PTH is higher than 200–250 pg/ml. In the setting of hypercalcemia or an elevated calcium phosphate product (usually >60–65), the use of these agents is contraindicated. The target level for circulating PTH is unknown; however, most clinicians seek a value of approximately 150–200 pg/ml.

The use of vitamin D analogues has been shown to decrease bone pain while improving both muscle strength and posture in patients with uremia. Furthermore, abnormal bone histology may be reversed with such therapy. A recent observational study of ESRD hemodialysis patients suggests improved survival of these treated with paricalcitol compared to those on calcitriol. To our knowledge, however, the use of vitamin D analogues has not been shown in a randomized controlled trial to improve mortality in patients with renal disease.

### Calcimimetic Agents

The CaSR, located on the parathyroid cell membrane, regulates the release of PTH in response to changes in serum ionized calcium concentration. Calcimimetic agents have similar effects as calcium on parathyroid gland physiology. Currently being investigated as potentially beneficial in patients with renal disease, calcimimetic agents diminish the effect of decreases in serum ionized calcium concentration on the release of PTH by the parathyroid glands. Thus, in patients treated with calcimimetic agents, there is the potential for the development of hypocalcemia, thought to be due to a decrease in PTH secretion and altered calcium flux between bone and the extracellular space. The use of calcimimetic agents has been proposed in the setting of hyperparathyroidism and hypercalcemia, when vitamin D analogues cannot be employed. However, the variable bioavailability of the agents and the propensity of patients to develop hypocalcemia present difficult challenges regarding their widespread use in patients with renal disease. Further studies on the clinical application of calcimimetic agents in patients with renal disease are needed.

### Parathyroidectomy

Parathyroidectomy is usually reserved for patients with severe secondary hyperparathyroidism in whom both dietary and medical treatments fail to improve

either clinical or biochemical parameters. Newer techniques, such as percutaneous ethanol ablation of the parathyroid glands, have been described. Patients in need of surgical treatment of secondary hyperparathyroidism usually have extremely high circulating PTH levels and often complain of intractable pruritis and bone pain. Profound hypocalcemia often develops in the postoperative period and requires close observation.

## CONCLUSION

The presence of renal disease evokes complex alterations in mineral metabolism. Such alterations lead to the development of ROD, increasing morbidity for this patient population. The use of aluminum-based phosphate binders and vitamin D analogues has led to an increased prevalence of osteomalacia and ABD, respectively. Definitive diagnosis of the type of ROD present depends on the evaluation of bone histomorphometric parameters; however, because of the limited use of bone biopsy, the clinician must rely on both clinical and biochemical parameters to make a provisional diagnosis. In general, treatment strategies involve the control of both serum phosphate and calcium, in conjunction with the appropriate use of vitamin D analogues. Parathyroidectomy is reserved for patients with severe secondary hyperparathyroidism for which medical management is ineffective. Promising therapeutic agents, including the calcimimetic agents, are being evaluated in patients with renal disease.

## See Also the Following Articles

Bone Remodeling, Dynamics of • Bone Turnover Markers • Kidney Disease in Diabetes • Kidney Stones • Parathyroid Surgery • Vitamin D • Vitamin D Deficiency, Rickets, and Osteomalacia

## Further Reading

Andress, D. L., Norris, K. C., Coburn, J. W., Slatopolsky, E. A., and Sherrard, D. J. (1989). Intravenous calcitriol in the treatment of refractory osteitis fibrosa of chronic renal failure. *N. Engl. J. Med.* **321**, 274–279.

- Atsumi, K., *et al.* (1999). Risk factors for vertebral fractures in renal osteodystrophy. *Am. J. Kidney Dis.* **33**, 287–293.
- Chertow, G. M., *et al.* (1999). A randomized trial of sevelamer hydrochloride (RenaGel) with and without supplemental calcium: Strategies for the control of hyperphosphatemia and hyperparathyroidism in hemodialysis patients. *Clin. Nephrol.* **51**, 18–26.
- Coco, M., and Rush, H. (2000). Increased incidence of hip fractures in dialysis patients with low serum parathyroid hormone. *Am. J. Kidney Dis.* **36**, 1115–1121.
- Dahl, E., Nordal, K. P., Halse, J., and Flatmark, A. (1990). The early effects of aluminum deposition and dialysis on bone in chronic renal failure: A cross-sectional bone-histomorphometric study. *Nephrol. Dial. Transplant.* **5**, 449–456.
- Felsenfeld, A. J., and Llach, F. (1993). Parathyroid gland function in chronic renal failure. *Kidney Int.* **43**, 771–789.
- Fukagawa, M., Kazama, J., and Shigematsu, T. (2001). Skeletal resistance to PTH as a basic abnormality underlying uremic bone disease. *Am. J. Kidney Dis.* **38**, S152–S155.
- Goodman, W. G. (2001). Recent developments in the management of secondary hyperparathyroidism. *Kidney Int.* **59**, 1187–1201.
- Goodman, W. G., Coburn, J. W., Slatopolsky, E., and Salusky, I. B. (1996). Renal osteodystrophy in adults and children. In “Primer on the Metabolic Bone Diseases and Disorders of Mineral Metabolism.” Lippincott–Raven, Philadelphia.
- Goodman, W. G., *et al.* (2002). The calcimimetic agent AMG 073 lowers plasma parathyroid hormone levels in hemodialysis patients with secondary hyperparathyroidism. *J. Am. Soc. Nephrol.* **13**, 1017–1024.
- Gupta, A., *et al.* (2000). Race is a major determinant of secondary hyperparathyroidism in uremic patients. *J. Am. Soc. Nephrol.* **11**, 330–334.
- Hoyland, J. A., and Picton, M. L. (1999). Cellular mechanisms of renal osteodystrophy. *Kidney Int.* **56**(Suppl. 73), S8–S13.
- Hruska, K. A., and Teitelbaum, S. L. (1995). Renal osteodystrophy. *N. Engl. J. Med.* **333**, 166–175.
- Hutchison, A. J., *et al.* (1993). Correlation of bone histology with parathyroid hormone, vitamin D<sub>3</sub>, and radiology in end-stage renal disease. *Kidney Int.* **44**, 1071–1077.
- Lindberg, J. S., and Moe, S. M. (1999). Osteoporosis in end-stage renal disease. *Sem. Nephrol.* **19**, 115–122.
- Llach, F., *et al.* (1986). The natural course of dialysis osteomalacia. *Kidney Int.* **29**, S74–S79.
- Salusky, I. B., and Goodman, W. G. (2001). Adynamic renal osteodystrophy: Is there a problem? *J. Am. Soc. Nephrol.* **12**, 1978–1985.
- Silver, J., *et al.* (1999). Regulation of the parathyroid hormone gene by vitamin D, calcium and phosphate. *Kidney Int. Suppl.* **73**, S2–S7.
- Stehman-Breen, C. O., *et al.* (1999). Racial differences in bone mineral density and bone loss among end-stage renal disease patients. *Am. J. Kidney Dis.* **33**, 941–946.
- Teng, M., Wolf, M., Lowrie, E., Ofshun, N., Lazarus, J. M., and Thadhani, R. (2003). Survival of patients undergoing hemodialysis with paricalcitol or calcitriol therapy. *N. Engl. J. Med.* **349**, 446–456.

of sodium reabsorption that occurs as a result of both the direct tubular effect of Ang II and the aldosterone secretagogue effect of the peptide, tends to reduce renin secretion back toward normal values. Furthermore, the profound effects of Ang II in inducing small artery hypertrophic remodeling, e.g., by increasing the wall-to-lumen ratio in these vessels, enhance the pressor effect of Ang II. From the clinical standpoint, all of these facts indicate that the chances of detecting an activated renin-angiotensin system and high plasma renin are high when the patient is seen soon after a given renal artery obstructive lesion has become hemodynamically relevant, but tend to wane afterward. This may explain why at the time of diagnosis as many as one-third of patients with renovascular hypertension (RVH) do not have high plasma renin, either in the peripheral blood or in venous blood from the affected side.

## METHODOLOGY

### Patient Selection

Measurement of RVR is an invasive procedure and therefore patients should be selected for this test based on a high degree of suspicion (pretest probability) of RVH, which can be established by using the criteria of Mann and Pickering or recommendations from updated guidelines such as those from the European Society of Hypertension–European Society of Cardiology.

### Preparation of Patients for RVR Measurements

Several conditions can profoundly affect RVR levels (Table I). Therefore, the patients selected to undergo this invasive test should be adequately prepared in order to achieve the best diagnostic accuracy. The patient should be on a normal sodium intake and this should be verified by measuring 24 h urinary sodium excretion on the day of the test. Furthermore, since renin secretion is under the control of the sympathetic nervous system, which follows a circadian pattern of activity, ideally RVR studies should be performed between 8:00 A.M. and 12:00 P.M., after the subject has fasted overnight and has been kept lying quietly in the supine position for at least 1 h before and during the test.

Several drugs (listed in Table I) alter renin secretion and therefore must be withdrawn for at least 2 weeks prior to the RVR testing. If there is concern that the patient may be put at high risk by temporarily

**Table I Preparation of the Patient for RVR Studies**

1. Allow a normal Na<sup>+</sup> intake for at least 7 days (100–200 mmol/day) prior to the study
2. Measure Na<sup>+</sup>/K<sup>+</sup> excretion (24 h urine collection) the day of the RVR study
3. Perform the test between 8 AM and 12 PM
4. Keep the patient supine for at least 1 h prior to and during the test
5. Try to avoid causing any stress to the patient during the test
6. Withdraw for at least 2 weeks prior to the RVR study:
  - Diuretics<sup>a</sup>
  - Beta-blockers
  - Direct vasodilators
  - Angiotensin-converting enzyme inhibitors
  - Ang II type 1 (AT-1) receptor antagonists (sartans)
7. Use a long-acting calcium entry blocker and/or an  $\alpha$ 1-adrenoceptor blocker, if necessary

<sup>a</sup>Due to their long half-life, aldosterone antagonists (spironolactone, kanrenone, and kanrenoate) must be withdrawn for at least 6 weeks before the RVR study.

withdrawing his or her antihypertensive treatment, as is often the case in patients with suspected RVH that can be resistant to therapy, it is wise to allow the use of a long-acting calcium entry blocker, such as verapamil, a long-acting dihydropyridine compound, or an  $\alpha$ 1-adrenoceptor blocker. Under chronic conditions, these agents negligibly affect renin secretion. If necessary, they can be combined together to achieve adequate blood pressure control before and during the test.

### Catheter Placement

It must be pointed out that several factors, which are listed in Table II, can lead to inaccurate findings in RVR studies, thus increasing the number of false-negative and false-positive results. Utmost care should, therefore, be exercised in obtaining blood selectively from the renal vein, a task that can be challenging since radiographic demonstration of venous anatomy is difficult to obtain because blood flows toward the catheter lumen, often leaving uncertainty as to the exact positioning of the catheter tip. Accordingly, it is not uncommon to draw diluted renal blood, e.g., blood not draining solely from the kidney. Additional problems relate to the fact that (1) multiple renal veins are present in a substantial proportion (20–28%) of patients on the right side and in much lower (1–3%) proportions on the left side, thus setting the stage for the collection of less “pure” renal vein blood on the right side than on the left side; (2) the

**Table II Factors Responsible for the Failure of RVR Measurements to Predict the Response to Revascularization Accurately**

*Mix-up of samples*

*Problems related to renal lesions*

Bilateral disease

Branch or segmental disease

*Problems related to renal vein catheterization*

Renal vein not catheterized

Multiple renal veins

Dilution of renal vein blood

Short renal vein on the right

Gonadal vein enters renal vein on the left

*Problems related to renin secretion*

Secretion inactive

Secretion declining

*Nonsimultaneous sampling*

*Assay variability*

right renal vein is shorter than the left renal vein and thus the catheter will more commonly collect an admixture of blood from the vena cava on the right side than on the left side; and (3) plasma drawn from the lower pole vein might not identify an upper pole (segmental) ischemia.

### Nonsimultaneous Sampling

The use of a single catheter for the sampling of both renal veins can run the risk of detecting an artifactual gradient between the two kidneys if an abrupt change in renin secretion occurs between the time of sampling on the first side and the time of sampling on the contralateral side. Therefore, bilateral catheterization with simultaneous sampling seems to be advisable unless the time elapsing between sampling from the two sides can be kept within 5 min. Furthermore, any maneuver that can abruptly increase renin secretion, such as stressful stimuli, should be avoided.

### Assay Variability

There is no question that the measurement of plasma renin, either as a direct immunoradiometric assay of active renin or, even more so, as plasma renin activity, shows some variability (ranging between 8 and 20%), even in the same laboratory and in experienced hands. Obviously, this should be kept in mind when interpreting the results of RVR measurements. However, it has been estimated that an RVR ratio greater than 1.4

**Table III Recommendations for RVR Studies**

- Meticulous labeling of specimen
- Placement of catheter tip selectively (deeply) into the renal vein
- Collection of segmental venous blood, when clinically indicated
- Performance of the test during steady-state renin secretion
- Simultaneous renal vein sampling (use of separate catheters)
- Proper consideration to renin assay variability in interpretation of results
- Frequent interaction between clinician and interventional radiologist

would almost never occur simply on the basis of within-assay variability.

### Summary of Recommendations

In summary, to gain maximal diagnostic information from RVR studies, protocols should take into consideration the principles listed in [Table III](#).

### CRITERIA FOR INTERPRETATION OF RESULTS OF RVR MEASUREMENTS

Four different indexes derived from RVR measurements can be used to evaluate the results of RVR studies, as shown in [Table IV](#).

According to seminal studies from Dr. J. H. Laragh's group, typically, if a patient with unilateral hemodynamically relevant renal artery stenosis is seen during the first stage of the disease, e.g., when renin is high, the renal venous-arterial plasma renin activity (PRA) difference relative to arterial levels from the affected kidney  $[(V_{\text{isch}} - V_{\text{iive}})/V_{\text{iive}}]$  would range between 0.24 and 0.50; furthermore, values  $>0.50$  would indicate reduced renal blood flow. Conversely, values of the renal venous-arterial PRA difference relative to arterial levels from the unaffected kidney  $[(V_{\text{ctl}} - V_{\text{iive}})/V_{\text{iive}}]$  close to 0 would indicate an absence of renin secretion, e.g., contralateral suppression. Importantly, a value of  $[(V_{\text{isch}} - V_{\text{iive}})/V_{\text{iive}}] + [(V_{\text{ctl}} - V_{\text{iive}})/V_{\text{iive}}] < 0.50$  should suggest either incorrect sampling or a segmental disease. Therefore, it should mandate a repeat of the study with segmental sampling.

These important studies have provided a framework for the proper interpretation of RVR studies. However, it should be understood that there are several situations in which this simple scheme cannot be applied, such as in those patients with bilateral hemodynamically relevant renal artery obstruction and

Table IV Indexes Derived from RVR Studies

## Indexes of lateralization of renin secretion

- RVR is the ratio of the PRA<sup>a</sup> level on the affected side to that on the unaffected side, e.g., the PRA value in renal vein blood from the ischemic kidney/PRA value in renal vein blood from the contralateral kidney
- $(V_{isch} - V_{iivc})/V_{iivc}$  is the renal venous-arterial PRA difference relative to arterial<sup>b</sup> levels from the affected kidney

## Indexes of contralateral suppression of renin secretion

- $V_{ctl}/V_{iivc}$  is the ratio of the PRA level on the unaffected (contralateral, ctl) side to the PRA level in the infrarenal inferior vena cava blood, e.g., the PRA value in renal vein blood from the contralateral kidney/PRA value in renal vein blood from the inferior vena cava
- $(V_{ctl} - V_{iivc})/V_{iivc}$  is the simultaneous renal venous-arterial<sup>b</sup> PRA difference relative to arterial levels from the unaffected kidney

<sup>a</sup>Direct measurement of active renin can be used instead of PRA to calculate these indexes.

<sup>b</sup>The values measured in the inferior vena cava blood  $V_{iivc}$  are usually taken as a surrogate measure of the level in arterial blood.

those with segmental renal artery stenoses. Furthermore, as discussed below, the usefulness of the  $(V_{isch} - V_{iivc})/V_{iivc}$  and  $(V_{ctl} - V_{iivc})/V_{iivc}$  indexes did not provide better diagnostic information than the simpler renal vein renin ratio (RVR).

It must also be acknowledged that the usefulness of the indexes derived from the measurement of RVR has been prospectively investigated by up-to-date statistical procedures, such as the receiver operator characteristics (ROC) curve analysis, in large populations of patients only in a study by the authors' group in 2002. Until that study, this lack of information resulted in some confusion regarding the optimal cutoff values. As an example, even the popular cutoff value of 1.5 for the RVR was empirically derived and never tested with the ROC curve analysis. Furthermore, only limited information existed on the usefulness of RVR measurements in patients with total renal artery occlusion and with bilateral renal artery stenoses. Thus, it remained unclear which of these indexes furnishes the best diagnostic accuracy and at which cutoff values the optimal combination of sensitivity and false-positive rate can be achieved.

## CLINICAL RELEVANCE

The measurement of renin in renal vein plasma was introduced in the clinical arena 3 decades ago to demonstrate a unilateral overproduction of renin and thus to establish a pathophysiological link between ischemia-triggered activation of renin synthesis and high BP. The usefulness of RVR for predicting the blood pressure response to angioplasty or stenting has been also confirmed; nonetheless, most investigators would support the concept that lateralization of renin secretion is not a prerequisite for the cure of hypertension on revascularization. There is, indeed, a general consensus among experts that RVH can be

diagnosed only retrospectively, e.g., on demonstration of either normalization of a significant fall of BP values after correction of renal ischemia, due to the lack of accuracy of all tests that have been proposed over the years to predict the outcome of renal revascularization. For RVR, this might be due to the fact that at least one-third of patients with proven RVH do not have evidence of an activated renin-angiotensin system at the time of diagnosis.

It should also be noted that PTRAs and stenting are being used as effective alternatives to surgery for the treatment of RVH, because they incur substantially lower risks and lower costs than surgery and can be performed on the same occasion when diagnostic angiography is carried out. Furthermore, they can be effective and safe even in patients for whom they were formerly deemed to be contraindicated, such as those with atherosclerotic ostial renal artery stenosis. It has also been pointed out that avoidance of RVR measurements would result in a substantial savings of time and money due to cutting of catheters and assay costs. Accordingly, RVR measurements have largely been abandoned.

Yet there are patients, such as those with unilateral small kidney due to total renal artery occlusion and those with renin-producing tumors, in whom this issue is still relevant. In the former patients, nephrectomy remains the best therapeutic option and RVR measurements might be useful to pose the indication to surgery, as discussed below. In the latter patients, a clear-cut gradient of renin secretion may be the only clue to the identification of a small tumor. In truth, these tumors can be quite small and therefore may escape detection with computed tomography, magnetic resonance imaging, and even angiography. Thus, in patients with biochemical evidence of secondary (renin-dependent) aldosteronism and no renal artery stenosis, RVR studies with segmental sampling are mandatory.

## RECOMMENDATIONS FOR USE OF RVR TESTS IN CLINICAL PRACTICE

Several relevant issues concerning RVR were investigated in a prospective study by Rossi and colleagues (2002) in consecutive patients undergoing both digital subtraction angiography and RVR measurements because of a high pretest probability of RVH. The specific aims of the study were to establish the diagnostic accuracy of these indexes in the subsets of RVH patients with and without total renal artery occlusion and with unilateral and bilateral renal artery disease. Based on this study, answers to the following questions can be put forward.

### Which Is the Best Index Derived from RVR Measurements?

The RVRR, e.g., the ratio of PRA in the ischemic side to that in the contralateral side, is the index that provides the best discrimination between patients with and without RVH. The ratio  $(V_{isch} - V_{iivc})/V_{iivc}$ , albeit on average significantly different between patients with and without RVH, shows a greater overlap of values. In contrast, the two indexes of contralateral suppression,  $(V_{ctl} - V_{iivc})/V_{iivc}$  and  $V_{ctl}/A$ , do not provide satisfactory discrimination between patients with and without RVH. When the usefulness of the RVRR index for diagnosing RVH by ROC curve analysis was examined, it was found that the area under the ROC curve is significantly ( $P < 0.0001$ ) different from the identity line, thus confirming that RVRR provides an incremental gain for the diagnosis of RVH.

### Can RVRR Identify RVH Due to Total Renal Artery Occlusion?

Approximately one-fourth of RVH patients who are seen at hypertension clinics can have a totally occluded renal artery. Practically all of them also have markedly elevated PRA levels in the renal vein plasma of the affected side, which translates into a significantly higher ratio of the affected side to the unaffected side (RVRR), compared to both the nonoccluded RVH and the non-RVH patients. Thus, a totally occluded artery is consistently associated with unequivocal evidence of renin lateralization. Accordingly, RVR could be useful not only for distinguishing vascular (occluded) from nonvascular unilateral small kidney, information also attainable by echo-color Doppler assessment of intrarenal artery blood flow, but also for establishing the indication to nephrectomy.

### Can RVRR Identify RVH without Total Renal Artery Occlusion?

The incremental diagnostic value of RVRR for diagnosing RVH was mostly due to the contribution of the minority of RVH patients with total renal artery occlusion. The authors examined this hypothesis and could show a clear-cut shift to the left of the ROC curve, thus unequivocally confirming the usefulness of RVRR for diagnosing RVH due to total renal artery occlusion. The authors also demonstrated a shift to the right, e.g., a decrease in the urea under the curve (AUC), when patients with angiographic evidence of renal artery occlusion were excluded. These findings indicated that RVR identified far more accurately RVH with renal artery occlusion than RVH without renal artery occlusion.

### Can RVRR Identify RVH in Patients with Bilateral Renal Artery Lesions?

The authors investigated this question using the ROC curve analysis in the subsets of patients with unilateral and bilateral renovascular disease. It was found that RVRR, as well as the other indexes, was more accurate in the former patients than in the latter patients, particularly for the identification of RVH caused by total occlusion of the renal artery. Of note, the AUC of the ROC curve for all indexes did not differ significantly from the identity line when patients without total occlusion of the renal artery were excluded. This finding indicates that, in the presence of bilateral lesions, lateralization of renin secretion is detectable by RVR measurements only if there is total renal artery occlusion.

### What Is the Best Cutoff Value of the RVRR?

With the ROC curve analysis, it was also found that the optimal cutoff value, e.g., the value providing the best tradeoff between sensitivity and specificity, for the identification of RVH, was 1.55, e.g., close to the empirically derived value of 1.50. However, even in a selected population with a high prevalence of RVH, this cutoff value corresponded to a low degree of sensitivity (54%) and a relatively high false-positive rate (15%).

In contrast, for identification of RVH due to total renal artery occlusion, the cutoff value of RVRR that warranted the best combination of sensitivity (87%) and false-positive rate (22%) was 1.70, e.g., higher

than that for the identification of RVH. Accordingly, it can be concluded that (1) RVR measurements provide an unequivocal diagnosis in the vast majority of patients with RVH due to total renal artery occlusion and (2) a higher cutoff value is needed to optimize the diagnostic accuracy of RVR measurement-derived indexes in these patients. The fact that, even with a cutoff value that is higher than the commonly used value of 1.5, there is a 22% false-positive rate is certainly not reassuring from the clinical standpoint and mandates some caution, particularly when the RVR is just above this value. This is of interest because the detection of a unilateral small kidney on abdominal ultrasound or renal scintigraphy in hypertensive patients is not uncommon. These patients pose a special challenge from the therapeutic standpoint because, compared to nephrectomy, surgical revascularization of a totally occluded renal artery can provide a better preservation of renal function when undertaken in cases with angiographically demonstrated distal reconstitution of the vessels and evident nephrogram. However, when associated with a unilateral small kidney, total renal artery occlusion is usually best treated by nephrectomy. The latter normalizes blood pressure in the vast majority of such patients, whereas both PTRAs and surgical revascularization are often unsuccessful and lead to a high rate of secondary nephrectomy, because of the coexistence of marked obliterative changes in the intrarenal vascular bed at microangiography. Thus, prior knowledge of the RVR measurements can help to establish the indication for nephrectomy.

## CONCLUSIONS

Evidence indicates that RVR measurements are most useful for identifying patients with a small unilateral kidney due to total renal artery occlusion and those with renin-producing tumors. In contrast, RVR measurements are of limited value for deciding on whether to revascularize a kidney with renal artery stenosis. Since nephrectomy provides the best chances of cure of hypertension in patients with a totally occluded renal artery, RVR measurements are useful for deciding on the indication for nephrectomy. Therefore, they should be performed whenever the clinical picture is indicative of a small unilateral kidney due to total renal artery occlusion and/or of a renin-producing tumor.

## Acknowledgement

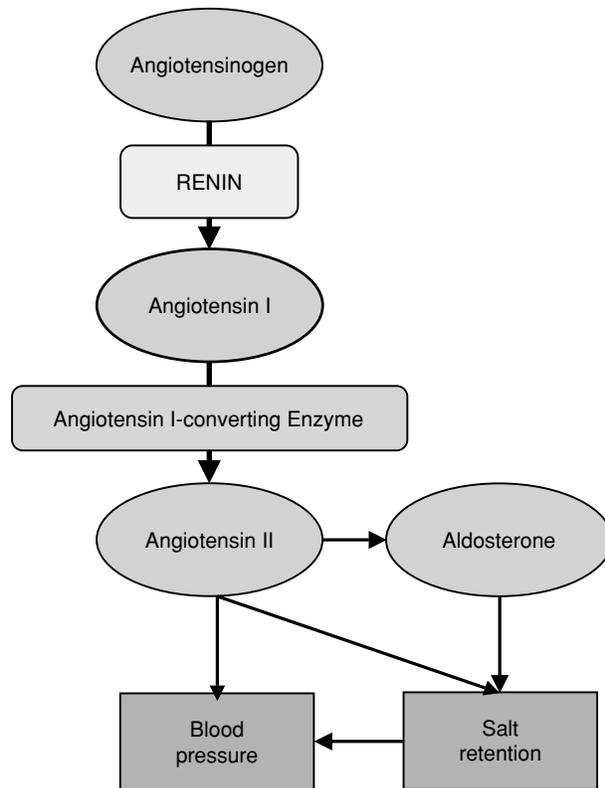
The authors acknowledge the expert secretarial assistance of Mrs. Carla Franceschin.

## See Also the Following Articles

Hyperreninemia • Hypertension, Overview • Hypertension, Renin and • Hyporeninemic Hypoaldosteronism • Renin • Tissue Renin-Angiotensin-Aldosterone System

## Further Reading

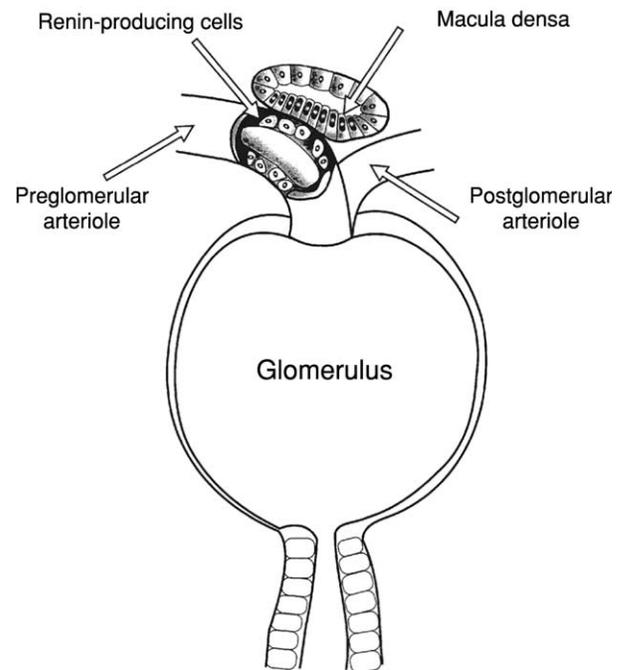
- European Society of Hypertension–European Society of Cardiology Guidelines Committee (2003). 2003 European Society of Hypertension–European Society of Cardiology guidelines for the management of arterial hypertension. *J. Hypertens.* **21**, 1011–1053.
- Harden, P. N., MacLeod, M. J., Rodger, R. S., Baxter, G. M., Connell, J. M., Dominiczak, A. F., Junor, B. J., Briggs, J. D., and Moss, J. G. (1997). Effect of renal-artery stenting on progression of renovascular renal failure. *Lancet* **349**, 1133–1136.
- Luscher, T. F., Greminger, P., Kuhlmann, U., Siegenthaler, W., Largiadere, F., and Vetter, W. (1986). Renal venous renin determinations in renovascular hypertension: Diagnostic and prognostic value in unilateral renal artery stenosis treated by surgery or percutaneous transluminal angioplasty. *Nephron* **44**, 17–24.
- Marks, L. S., and Maxwell, M. H. (1975). Renal vein renin: Value and limitations in the prediction of operative results. *Urol. Clin. N. Am.* **2**, 311–325.
- Pickering, T. G., Sos, T. A., Vaughan, E. D., Jr., and Laragh, J. H. (1986). Differing patterns of renal vein renin secretion in patients with renovascular hypertension, and their role in predicting the response to angioplasty. *Nephron* **44**, 8–11.
- Radermacher, J., Chavan, A., Bleck, J., Vitzthum, A., Stoess, B., Gebel, M. J., Galanski, M., Koch, K. M., and Haller, H. (2001). Use of Doppler ultrasonography to predict the outcome of therapy for renal-artery stenosis. *N. Engl. J. Med.* **344**, 410–417.
- Rossi, G. P., Pavan, E., Chiesura-Corona, M., Bader, M., Paganini, G., Cesari, M., De Toni, R., Feltrin, G. P., Ganten, D., and Pessina, A. C. (1997). Renovascular hypertension with low-to-normal plasma renin: Clinical and angiographic features. *Clin. Sci.* **93**, 435–443.
- Rossi, G. P., Zanin, L., Dessi-Fulgheri, P., Savastano, S., Cavazzana, A., Prayer-Galetti, T., Rappelli, A., and Pessina, A. C. (1993). A renin-secreting tumour with severe hypertension and cardiovascular disease: A diagnostic and therapeutic challenge. *Clin. Exp. Hypertens.* **15**, 325–338.
- Rossi, G. P., Cesari, M., Chiesura-Corona, M., Miotto, D., Semplicini, A., and Pessina, A. C. (2002). Renal vein renin measurements accurately identify renovascular hypertension caused by total occlusion of the renal artery. *J. Hypertens.* **20**, 975–984.
- van Jaarsveld, B. C., Krijnen, P., Pieterman, H., Derckx, F. H., Deinum, J., Postma, C. T., Dees, A., Woittiez, A. J., Bartelink, A. K., Man in 't Veld, A. J., and Schalekamp, M. A. (2000). The effect of balloon angioplasty on hypertension in atherosclerotic renal-artery stenosis. Dutch Renal Artery Stenosis Intervention Cooperative Study Group. *N. Engl. J. Med.* **342**, 1007–1014.
- Vaughan, E. D., Buhler, F. R., Laragh, J. H., Sealey, J. E., Baer, L., and Bard, R. H. (1973). Renovascular hypertension: Renin measurements to indicate hypersecretion and contralateral suppression, estimate renal plasma flow and score for surgical curability. *Am. J. Med.* **55**, 402–408.



**Figure 1** Schematic illustration of the renin–angiotensin–aldosterone system.

the juxtaglomerular portion of the preglomerular arteries. In response to prolonged stress situations, renin expression in vascular smooth muscle cells in renal arterioles can be reactivated.

In human, a single-copy gene on chromosome 1 encodes the enzymatically inactive prorenin. After cleavage of the leader presequence, prorenin is packaged into storage granules, which mature to secretory granules. During vesicle maturation, the enzymatically inactive prorenin is shortened by proteolysis to active renin, which is released from the JGE cells by regulated exocytosis. Some of the prorenin escapes packaging into secretory granules and is constitutively secreted into the circulation. Therefore, plasma contains both prorenin and active renin, with prorenin dominating over active renin in human plasma. Renin is variably glycosylated and has an average molecular weight of approximately 40 kDa. The renin concentration in the plasma can be estimated directly by radioimmunoassay or more indirectly through its enzymatic ability to generate angiotensin I, which in turn can be measured by immunoassay. Plasma renin activity in human is in the range of 1–5 ng angiotensin I per milliliter of plasma.



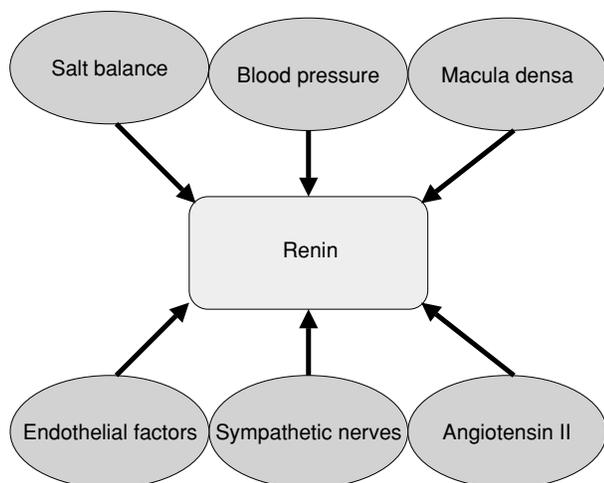
**Figure 2** Localization of renin-producing cells within the kidney, which are the main source of renin.

## PARAMETERS REGULATING RENIN SECRETION AND RENIN SYNTHESIS

Renin secretion and renin synthesis are coordinated, but they are regulated on different time scales. Whereas renin secretion from the kidneys can change within seconds, changes in renin synthesis require several hours. This time gap creates no problem, however, since the kidneys contain enormous stores of renin, sufficient to maintain a normal renin secretion rate over days, even if *de novo* synthesis of renin ceased.

Renin secretion and synthesis are essentially triggered by sympathetic nerve activity with noradrenaline as the main stimulatory transmitter acting via  $\beta_1$  receptors on JGE cells. Renin secretion and synthesis are further controlled by several negative feedback loops (see Fig. 3). A very short feedback loop mediates an inhibitory action of angiotensin II via AT1 receptors. Furthermore, the sodium content in the body, which is determined by the activity of the RAAS, exerts a negative effect.

Finally, the blood pressure, which is influenced by the activity of the RAAS, exerts an inhibitory effect by a direct action on renal preglomerular vessels. Apart from these systemically acting factors, locally generated signals influence the function of JGE cells. Thus,



**Figure 3** Factors controlling renin synthesis and renin secretion in the kidney.

the tubular macula densa cells, which are directly adjacent to the JGE cells, influence renin secretion in an inverse fashion to their salt resorption rate, establishing the so-called macula densa control of renin secretion. Renin secretion and synthesis are also influenced by endothelial factors, such as prostaglandins, nitric oxide, and endothelins. Furthermore, pro-inflammatory cytokines inhibit renin synthesis.

At the cellular level, renin secretion and synthesis are controlled by the cyclic AMP pathway, which acts in a stimulatory manner, and by the calcium pathway, which acts in an inhibitory manner.

### ORGANISMS WITH GENETIC DEFECTS OF COMPONENTS OF THE RAAS

Mice, in contrast to humans, can possess two renin genes (*ren-1* and *ren-2*) and become hypotensive if the *ren-1* gene, corresponding to the human renin gene, is disrupted. Mice lacking functional angiotensinogen as the substrate for renin have low blood pressure and problems in the development of a normal urinary tract system. Mice lacking ACE also display hypotension and the males exhibit a markedly reduced fertility, due to the occurrence of ACE in sperm cells. A novel ACE isoform (ACE-2) is essential for heart function, since mice lacking ACE-2 develop a severe cardiac contractility defect. Mice lacking AT1 receptors have low blood pressure values and show changes in kidney structure. Mice lacking functional AT2 receptors are more sensitive to the hypertensive effect of angiotensin II mediated via AT1 receptors. Mice lacking the

mineralocorticoid receptor mediating the effect of aldosterone have severe problems with sodium balance and blood pressure regulation. These animals die early after birth.

### RENIN FOR PATHOPHYSIOLOGY AND DISEASE

The clinical observation that inhibitors of the RAAS have a positive blood pressure-lowering effect in hypertensive patients suggests that the RAAS is critically involved in the development and maintenance of hypertension. However, renin plasma concentrations are often low during hypertension, suggesting that a hypersecretion of systemic renin is not as often the cause of hypertension. It is possible that local renin-angiotensin systems could be relevant in this context. A clear causal role for renin in pathophysiology is seen in the course of renal hypoperfusion or during ischemic kidney disease, which are associated with hyperreninemia. In addition, perturbations in sodium balance and extracellular volume (sodium accumulation and volume expansion) after prolonged stimulation of sympathetic outflow, such as during chronic heart failure, are causally related to the hypersecretion of renin.

Activation of the local renin-angiotensin system in particular is believed to contribute to changes in blood vessel wall structure, such as media thickening, and is also believed to contribute to the development of cardiac hypertrophy.

### DRUGS INTERFERING WITH THE RAAS

A variety of drugs are able to interfere with the activity of the RAAS. Renin inhibitors have been developed to directly inhibit the enzymatic activity of renin. The frequently used ACE inhibitors block the processing of angiotensin I to angiotensin II, thus lowering the circulating levels of angiotensin II and reducing its biological actions. Similar effects are achieved with the commonly used AT1 receptor blockers. The effects of aldosterone can be inhibited by spironolactone, which is a structural antagonist for aldosterone at the mineralocorticoid receptor.

### See Also the Following Articles

Angiotensin, Evolution of • Atrial Natriuretic Factor and Family of Natriuretic Peptides • Captopril • Hyperreninemia • Hypertension, Renin and • Hyporeninemic

Hypoaldosteronism • Renal Vein Renin • Tissue Renin-Angiotensin-Aldosterone System

### Further Reading

- Bader, M., Peters, J., Baltatu, O., Muller, D. N., Luft, F. C., and Ganten, D. (2001). Tissue renin-angiotensin systems: New insights from experimental animal models in hypertension research. *J. Mol. Med.* **79**, 76–102.
- Fitzsimons, J. T. (1998). Angiotensin, thirst, and sodium appetite. *Physiol. Rev.* **78**, 583–686.
- Kurtz, A., and Wagner, C. (1999). Cellular control of renin secretion. *J. Exp. Biol.* **202**, 219–225.
- Nicholls, M. G., Robertson, J. I., and Inagami, T. (2001). The renin-angiotensin system in the twenty-first century. *Blood Press.* **10**, 327–343.
- Robertson, M. J. (2000). Angiotensin antagonists. *J. Recept. Signal Transduct. Res.* **20**, 211–234.
- Tharaux, P. L., and Coffman, T. M. (2001). Transgenic mice as a tool to study the renin-angiotensin system. *Contrib. Nephrol.* **135**, 72–91.
- Zaman, M. A., Oparil, S., and Calhoun, D. A. (2002). Drugs targeting the renin-angiotensin-aldosterone system. *Nat. Rev. Drug Discov.* **1**, 621–636.

## TH Regulation

Several mechanisms are responsible for modulating the supply of TH to tissues: (1) the hypothalamus–pituitary–thyroid axis controls TH synthesis and release from the thyroid gland via stimulation by TSH; (2) the generation of T3 from T4 at the tissues via the action of deiodinases may modulate tissue responsiveness to TH; and (3) autoregulation of the expression of the TH receptor (TR) can also modulate the effect that TH has on a tissue, e.g., the more receptor, the larger the response, and the less receptor, the smaller the response, for a given amount of hormone. TSH synthesis and secretion are stimulated by TRH, a tripeptide derived from the hypothalamus. TSH originating from the pituitary gland in turn stimulates TH synthesis and secretion. Although T4 is the predominant form of TH released from the gland, and T3 generated at the tissue level is the biologically active hormone, T4 can also bind to the TR, but with 100-fold less affinity. T3, generated from the conversion of T4 in the pituitary and hypothalamus by type II deiodinase, controls the synthesis and secretion of TSH and TRH through a negative feedback loop (Fig. 1).

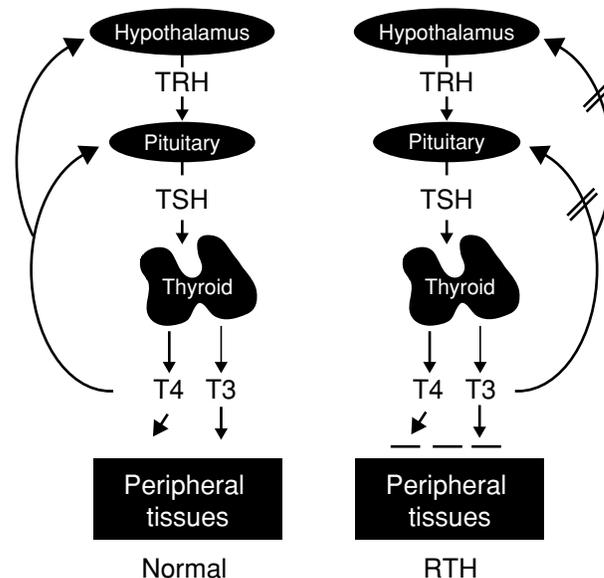
The signs, symptoms, and laboratory abnormalities observed in RTH are due to reduced peripheral tissue and pituitary responsiveness to TH, resulting in impaired negative feedback of T3 on TSH/TRH and impaired peripheral tissue action of T3.

## Mechanisms of TH Action

TH enters the cell by a process of concentration-dependent passive diffusion after dissociation from the serum hormone-binding proteins. Several molecules have been identified as putative TH transmembrane transporters, but none have been shown to be unique for TH. TH action, for the most part, is mediated through specific TRs in the cell nucleus, although pathways of nongenomic action of TH have been described.

TRs are members of the nuclear hormone receptor superfamily together with the retinoic acid receptors (RXRs), 9-*cis*-retinoic acid receptors, vitamin D<sub>3</sub> receptors, and peroxisome proliferator-activated receptors. TRs act as ligand-dependent transcription factors that increase or decrease the expression of target genes depending on whether the gene promoters contain, respectively, positively or negatively regulated TH-response elements.

There are two genes, *TRα* and *TRβ*, located on chromosomes 17 and 3, respectively, that encode TRs.



**Figure 1** Schematic representation of the hypothalamic–pituitary–thyroid axis. Under normal conditions, the hypothalamus secretes TRH, which stimulates the pituitary to produce TSH, which stimulates the thyroid to produce mainly T4. T4 is metabolized in the peripheral tissues to T3 and, in turn, THs feed back in an inhibitory fashion on the hypothalamus and pituitary. Thyroid hormone action on pituitary and peripheral tissues occurs when T3 binds to its receptor; in RTH, the feedback is inhibited due to decreased sensitivity of the pituitary to TH, resulting in elevated TSH with increased T4 secretion. The peripheral tissues also have decreased sensitivity to TH in RTH.

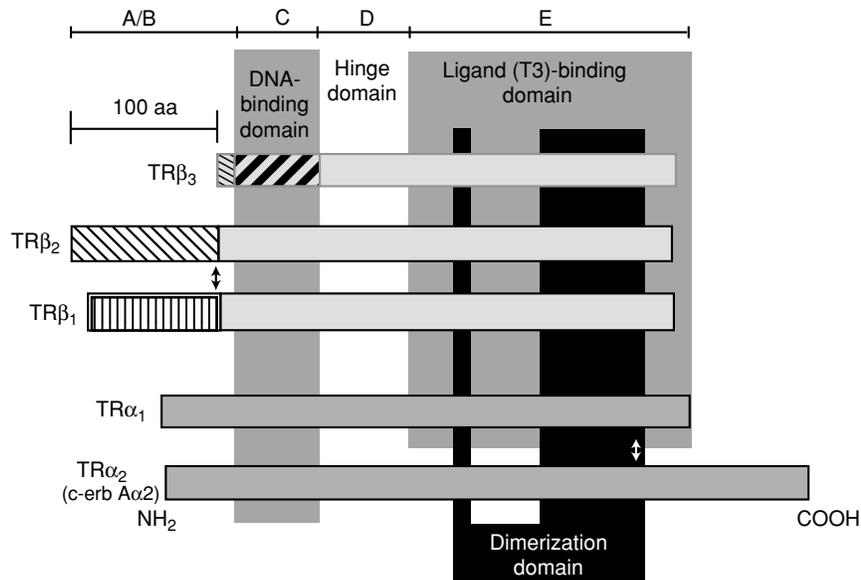
TRs are cellular homologues of the viral oncogene *erb A*. Each gene generates at least two proteins by alternative splicing: TR $\alpha_1$ , TR $\alpha_2$ , TR $\beta_1$ , and TR $\beta_2$ .

The relative expression of the two TR genes and the distribution of their products vary from tissue to tissue and during different stages of development. Furthermore, TH differentially regulates the expression of the two TR genes and their isoforms.

The  $\alpha$  and  $\beta$  TR isoforms have a well-conserved DNA-binding domain (or C domain) separated from the ligand (T3)-binding domain (or E domain) by a hinge domain (D domain). TR $\alpha_2$ , due to a carboxyl-terminal sequence difference, does not bind TH and the molecule does not function as a TH receptor (Fig. 2).

TRs form homodimers (TR/TR) or form heterodimers with RXRs (TR/RXR) and bind to specific DNA sequences termed TH-response elements (TREs), which are located near the transcription start point of TH-regulated genes.

On genes up-regulated by TH, the unliganded TRs suppress basal gene transcription activity by interacting with a class of nuclear proteins known as



**Figure 2** Schematic representation of the principal TR forms and their functional domains. Each of the two TR genes (*TR $\alpha$*  and *TR $\beta$* ) expresses two major protein forms (1 and 2) by alternative splicing (arrows). The shading pattern of the bars corresponds to the degree of amino acid sequence similarity (identical, similar, or completely different). Differences are most marked in the amino termini (NH<sub>2</sub>) of the TRs. The different carboxyl terminus (COOH) of the  $\alpha$ 2 form completely abolishes T3 binding and this product is not considered to be a proper TR. The functional domains are also identified by a letter code and by the striped backgrounds.

corepressors, such as the nuclear receptor corepressor and the silencing mediator for retinoid and TH receptor. Conformational changes in TR, produced by T3 binding, enhance the occupation of TREs by TR/RXR heterodimers, release the corepressor complex, and recruit coactivators such as steroid receptor coactivator 1 (SRC-1), among others, all exhibiting histone acetyl transferase activity. SRC-1 modifies the structure of the chromatin and, by loosening the nucleosome at the site of the promoter, activates gene transcription (Fig. 3). The mechanisms involved in the genes that are down-regulated by TH, such as TSH, are less well understood, but are likely to involve the same classes of molecules on a different TRE that causes corepressors to become “activators” and coactivators to become “repressors.”

## MOLECULAR BASIS OF RTH

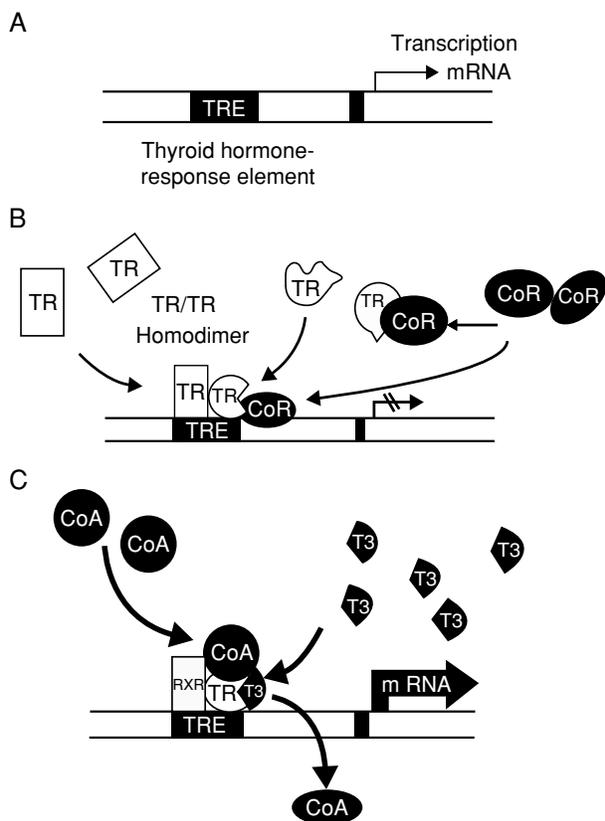
### Location of Mutations Associated with RTH

Eighty to 85% of subjects with RTH have mutations in the *TR $\beta$*  gene, whereas no mutations have been described in the *TR $\alpha$*  gene. One hundred fourteen mutations have been described in 261 different families. The majority of mutations have occurred in three “hot spots” located in the ligand-binding

domain (exons 8–10) of *TR $\beta$*  (Fig. 4). Whereas the majority of these mutations are single nucleotide substitutions, 5 families had nucleotide insertions and 7 families had from one to three nucleotide deletions. Eleven mutations were present in 5 or more families and the R338W mutation was reported to occur independently in 23 different families (Table I). The mutations commonly occur in a guanine immediately following a cytosine in 73% of the cases of frequently appearing mutations compared to only 5% of the other 81 point mutations described in 1 to 4 families each. This finding suggests that the “GC” regions are mutational hot spots in the *TR $\beta$*  gene.

The RTH index family had a homozygous deletion of the *TR $\beta$* <sub>1</sub> allele, representing the only example of recessive inheritance of RTH. In all other cases, the inheritance pattern of RTH is autosomal-dominant, involving a single nucleotide substitution or small deletions. RTH is sporadic in 15% of cases, when a mutation has arisen *de novo*, where neither parent has the abnormal phenotype.

The mutant receptors either cannot bind TH or have reduced affinity for TH. As a result, the corepressor is not released and the gene is not activated. Other receptor mutants have been found that do bind T3 but release the corepressor more slowly than normal or have a weaker interaction with the coactivator.



**Figure 3** The mechanism of TH action. Schematic representation of the elements involved in the mediation of TH action through the regulation of T3-responsive gene expression (A). The unliganded TR homodimerizes with TR, associates with CoR (corepressor), and exerts a suppressive effect on genes positively regulated by TRE (B). Formation of a TR/T3 complex resulting in disassociation of CoR, occupation of TRE by TR/RXR heterodimers, and recruitment of CoA (coactivator) to enhance the transcription of genes positively regulated by TRE (C).

Ten to 15% of families with RTH do not have mutations in the TR $\beta$ . These non-TR RTH subjects have a phenotype similar to those with TR $\beta$  mutations including baseline thyroid function tests as well as metabolic and TRH responsiveness to exogenously administered TH. Since non-TR RTH also has an autosomal-dominant mode of inheritance, it is believed that another transcriptional cofactor involved in TH action could be responsible for the hormone resistance. In support of this notion are studies of the hypothalamic-pituitary-thyroid axis in knockout mice lacking the SRC-1 (coactivator) or the RXR that show that the absence of a cofactor can cause RTH. Further work is required to determine whether there are other forms of hormonal resistance in the non-TR RTH subjects in which the same cofactor may be responsible for mild, but multiple types of hormone resistance.

**Table I** TR $\beta$  Codons with Five or More Families Having the Identical Mutation

Codon mutation	No. of families	Nucleotide substitution
R243W	6	CGG $\rightarrow$ TGG
R243Q	11	CGG $\rightarrow$ CAG
M313T	5	ATG $\rightarrow$ ACG
A317T	15	(C)GCT $\rightarrow$ ACT
R320C	8	CGC $\rightarrow$ TGC
R320H	6	CGC $\rightarrow$ CAC
R338W	23	CGG $\rightarrow$ TGG
R249Q	6	CGG $\rightarrow$ TGG
R348H	12	CGC $\rightarrow$ CAC
P453T	10	CCT $\rightarrow$ ACT
P453S	9	CCT $\rightarrow$ TCT

### Dominant Negative Effect

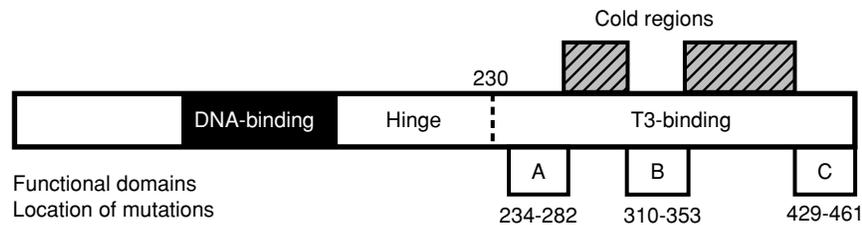
Mutant TR $\beta$ s interfere with the function of the wild-type TRs, a phenomenon termed the dominant negative effect (DNE). Mutant TR $\beta$  receptors are still able to bind to TRE and dimerize with normal TRs or RXRs, interfering with the function of the normal TRs. There is a strong correlation between the DNE measured *in vitro* and decreased release of transcriptional corepressor from TRs.

For a DNE to occur, the TR mutant must bind to TRE, which may explain why no mutations have been identified in the DNA-binding domain and why in a family with a deletion of all coding sequences of the TR $\beta$  gene the heterozygotes did not manifest the phenotype of RTH. Interestingly, one subject with a homozygous deletion of TR $\beta$  Thr-337 (amino acid deletion) with a dominantly inherited RTH pattern manifested the most severe form of RTH.

### PATHOPHYSIOLOGY

The lack of negative feedback of T3 on TSH results in persistent TSH secretion and thyroid gland stimulation, resulting in an increase in TH synthesis and secretion. The normal or slightly increased TSH usually responds by a further increase in the release of TRH, distinguishing this from subjects with a TSH-secreting pituitary adenoma.

The serum TSH of RTH patients is identical immunologically to the serum TSH of normal controls. It does not contain an excess of glycoprotein  $\alpha$ -subunits (TSH $\alpha$ ) that are typically found in the serum of patients with TSH-producing pituitary tumors. However, the TSH has an increased biological



**Figure 4** Location of natural mutations in the *TRβ* gene associated with RTH. Schematic representation of the *TRβ* and its functional domains for interaction with TREs (DNA-binding domain) and with hormone (T3-binding domain). The mutations that have been found in *TRβ* cluster in three areas of the T3-binding domain. The amino acid positions between which the mutations are found are indicated below the boxes (A, B, and C). Note the “cold regions” of 28 and 75 amino acids, devoid of mutations associated with RTH.

potency *in vitro*, perhaps explaining how normal TSH levels produce goiter and hypersecretion of TH by the thyroid gland. Another possible, albeit unproven, explanation of the goiter in RTH subjects with normal serum TSH could be augmented thyrocyte sensitivity to TSH through increased density of TSH receptor units.

Several of the clinical features encountered in some patients with RTH may be the manifestation of selective tissue deprivation of TH during early stages of development. The severity of symptoms is related to the relative expression of the mutant allele (and perhaps cofactors) in the particular tissue. For example, cardiac tissue predominantly expresses *TRα*. Most RTH subjects present with tachycardia: because of RTH subjects' abnormal *TRβ*, they overproduce TH, but since the heart has *TRα*, which is normal, tachycardia results.

Thyroid tissue obtained by biopsy or at surgery revealed various degrees of hyperplasia of the follicular epithelium, supporting the idea of structural heterogeneity of the follicles. Little can be said about the pathological findings in tissues other than the thyroid because of the unavailability of autopsy data from patients with RTH. Metachromasia in fibroblasts is present in patients with RTH as well as in patients with myxedema due to TH deficiency, although treatment with the hormone failed to induce the disappearance of the metachromasia in fibroblasts from patients with RTH.

## CLINICAL FEATURES

### Clinical Classification

The clinical presentation of RTH is highly variable both between families and among family members with identical *TRβ* mutations. The majority of

individuals appear to be euthyroid and therefore they are asymptomatic. Some may manifest symptoms suggestive of TH deprivation, such as growth retardation, impaired cognitive ability, and hypercholesterolemia, whereas others show signs of TH excess, such as tachycardia, advanced bone age, or hyperactivity. Not uncommonly, individuals have symptoms of both TH deficiency and excess.

Subjects with RTH that appear to be eumetabolic and maintain a near normal serum TSH concentration have been classified as having generalized resistance to TH (GRTH). In such individuals, the defect seems to be compensated for by the high endogenous levels of TH. In contrast, patients with equally high levels of TH appear to be hypermetabolic because they are restless or hyperactive or have a rapid heart rate. Such individuals have been classified as having selective pituitary resistance to TH (PRTH). Subdivision of RTH into GRTH and PRTH is an artifact arising from the subjective nature of symptoms and poor specificity of signs. Subjects harboring the same mutations, and even belonging to the same family, have been classified as having GRTH and PRTH. Furthermore, clinical studies have shown that the responses of peripheral tissue markers to TH action were equally attenuated. The classification of RTH into GRTH and PRTH can therefore be viewed as opposite ends of the spectrum of a single disease.

### Clinical Findings

The phenotype of RTH is variable, with most patients presenting with mild to moderate symptoms. In children, investigations leading to diagnosis have been undertaken because of goiter, hyperactive behavior, learning disabilities, or developmental delay.

Most adults seek medical attention because of goiter or rapid heart rate. RTH is being diagnosed in infancy owing to routine neonatal screening programs that measure both TSH and T4 and to hormonal and DNA analyses performed in infants born to parents known to have RTH. In adults, abnormalities found on routine thyroid testing, in the absence of clinical findings, have been responsible for the diagnosis of RTH.

The majority of untreated subjects are eumetabolic at the expense of high levels of TH. The degree of this compensation for the tissue hyposensitivity to the TH is variable among individuals as well as in different tissues. As a consequence, there is clinical and laboratory evidence that TH deficiency and excess often coexist.

Many attempts have been made to demonstrate tissue hyposensitivity *in vitro*. The common finding of paradoxical responses to TH is difficult to explain. They are possibly due to: (1) variable interaction of the mutant TR with the  $\alpha$  and  $\beta$  isoforms of the normal TRs; (2) variable expression of the *TR $\alpha$*  and *TR $\beta$*  genes in different tissues; and (3) differences in the regulation of expression of the *TR $\alpha$*  and *TR $\beta$*  genes by the hormone-activated TR, which in some tissues may be in opposite directions.

On physical examination, goiter is the most common abnormality, occurring in 85% of cases. Tachycardia occurs in 80–90% of subjects with RTH and is caused by excess TH acting on the *TR $\alpha$*  gene. Careful evaluation of subjects with RTH has shown that 50% have learning disabilities, often associated with attention deficit hyperactivity disorder (ADHD) and, on the average, lower intellectual quotients (IQs). However, frank mental retardation (IQ <60) occurs in only 3% of cases. RTH can present with mild to moderate growth retardation and delayed bone maturation in 25% of cases. Hearing defects have been detected in almost 25% of patients (Table II). Other features, such as frequent ear, nose, and throat infections, decreased bone mass, deafness, hypotonia, and seizures, have been recognized.

The course of the disease is as variable as its presentation. Some subjects have normal growth and development and lead a normal life. Others present variable degrees of mental and growth retardation. There is some evidence that the severity of resistance tends to improve as the subject gets older as determined by thyroid function tests. This has been shown in the *TR $\beta$*  knockout mouse model of RTH and is supported from some published studies in humans.

Table II Clinical Features of the RTH

Laboratory findings	Normal <sup>a</sup>	RTH <sup>b</sup>
T4 Levels (pmol/liter)	12.8–24.4	41±2.1
T3 Levels (pmol/liter)	3.8–8.4	11.4±1.5
TSH Level (mU/liter)	0.5–4.5	3.15±0.3
Clinical findings	Frequency (%)	
Goiter	65–95	
Tachycardia	50–80	
Emotional disturbances	70	
Recurrent ear, nose, and throat infections	50	
Attention deficit hyperactivity disorder	40–60	
Hyperkinetic behavior/learning disability	20–50	
Low IQ (IQ <85)	35–50	
Delayed bone age	30–47	
Hearing loss	10–22	
Short stature	18–26	
Mental retardation (IQ <70)	4–16	

<sup>a</sup>Normal denotes normal range in the general population.

<sup>b</sup>RTH denotes mean values of the thyroid function tests in patients with RTH.

## DIAGNOSIS AND DIFFERENTIAL DIAGNOSIS

### Serum Levels of THs and TSH

Increased levels of free T3 and T4 in combination with nonsuppressed TSH in serum are indicative of RTH. One should rule out any condition that can give rise to spuriously elevated T3 and T4 levels with normal or elevated TSH, such as abnormalities in TH transport, drugs, and the presence of antibodies that can interact with the assays.

Serum T3 and T4 values range from just above the upper limit of normal to a severalfold increase in the upper limit of normal. The degree of T3 and T4 elevation is usually congruent, resulting in a normal T3 to T4 ratio. Reverse T3 concentration is also high in patients with RTH. The concentration of serum thyroglobulin reflects the level of TSH-induced thyroid gland activity. The extrathyroidal pool and absolute daily production of T4 and T3 are increased by approximately two- to fourfold. The extrathyroidal conversion of T4 to T3 has been found to be normal.

The fractional uptake of radioiodide by the thyroid gland is high. The iodide appears to be normally organified since no discharge of trapped iodide has

been observed following the administration of perchlorate.

### TRH Testing

A diagnostic feature of the syndrome is the presence of normal or slightly elevated serum TSH levels and preservation of its response to TRH despite elevated TH levels. The TSH response to TRH is either normal or exaggerated; the latter is more common in patients receiving anti-thyroid drug therapy or with a history of prior surgery or radioiodide therapy. The bioactivity of secreted TSH is normal or increased and the concentration of its free  $\alpha$ -subunit is not disproportionately high. In contrast, the  $\alpha$ -subunit to TSH ratio is elevated in TSH-secreting tumors.

### Responses to the Administration of TH and Other Drugs

Because resistance to TH plays a principal role in the pathogenesis of the syndrome, patients have been given exogenous TH in order to observe their responses and thereby establish the presence and degree of insensitivity to the hormone. A standardized diagnostic protocol, using short-term administration of incremental doses of L-T<sub>3</sub>, is recommended. The three doses given in sequence to an adult are a replacement dose of 50  $\mu$ g/day and two supraphysiological doses of 100 and 200  $\mu$ g/day. The hormone is administered in a split dose every 12 h and each incremental dose is given for 3 days. Doses are adjusted in children and in adults of unusual weight to achieve the same level of serum T<sub>3</sub>. L-T<sub>3</sub>, rather than L-T<sub>4</sub>, is used because of its direct effect on tissues, bypassing the potential defects of T<sub>4</sub> transport and metabolism. In addition, the more rapid onset and shorter duration of T<sub>3</sub> action reduce the period required to complete the evaluation and shorten the duration of symptoms that may arise in some individuals.

The administration of supraphysiological doses of TH suppresses TSH secretion, resulting in the decrease and eventually the abolition of the TSH response to TRH. Various responses of peripheral tissues to the administration of TH have been quantitated. Most notable are measurements of the basal metabolic rate, pulse rate, reflex relaxation time, serum cholesterol, lipids, enzymes, sex hormone-binding globulin, and urinary excretion of hydroxyproline, creatinine, and carnitine. Either no significant changes were observed or they were much reduced relative to the amount of TH given.

### Family Studies

Once a diagnosis of RTH is suspected, testing of thyroid function in siblings and parents can be helpful. Confirmation of the same phenotype in other related individuals makes the diagnosis of RTH more likely than a diagnosis of a TSH-secreting adenoma as the latter has not been reported to be familial. Also, identification of an abnormal phenotype in an asymptomatic family member will allow for a correct diagnosis and may prevent unnecessary diagnostic testing and treatment. From a practical standpoint, a physician may find it difficult to arrange for thyroid testing on asymptomatic relatives; however, considering the cost of magnetic resonance imaging (MRI), documentation of an inherited disorder is more cost-effective.

### Imaging Studies

Imaging studies are not necessary for the diagnosis of RTH. Although an MRI of the pituitary is often ordered to satisfy the physician that a TSH-secreting adenoma is absent, it cannot be justified as routine in the work-up for RTH due to the fact that a small microadenoma may be missed by the MRI and 10% of normal individuals may have some abnormality that may not have any clinical significance.

### Genetic Studies

The use of genetic testing in a new family with RTH for mutations in the *TR $\beta$*  genes is limited to research laboratories. In light of the 114 mutations that have been described, "screening" would be useful only if it were to involve sequencing of at least the last five coding exons of the *TR $\beta$* . Once a mutation has been identified, screening of other family members for the mutation can be more sensitive than screening for the phenotype, due to the mild and variable nature of the RTH phenotype. Confirmation of the mutation in a newborn may eliminate the need for treatment. Confirmation of the mutation in a fetus at some point may allow clinicians to prescribe intrauterine treatment, if necessary.

### Summary of Recommended Diagnostic Evaluation

1. The usual presentation is high serum levels of free T<sub>4</sub> and T<sub>3</sub> with nonsuppressed TSH.

2. Confirm the elevated serum levels of free thyroid hormone (T4 and T3) and exclude TH transport defect or antibody interference in the assays.
3. Measure thyroid function tests in other family members.
4. In sporadic cases, exclude the presence of a pituitary adenoma by measurement of the  $\alpha$ -subunit in serum.
5. Demonstrate a blunted TSH suppression and metabolic response to the administration of supra-physiologic doses of TH.
6. Perform genetic studies when indicated.

## TREATMENT

### General Considerations

No specific treatment is available to fully correct the defect. Fortunately, in most cases of RTH, the partial tissue resistance to TH appears to be adequately compensated for by an increase in the endogenous supply of TH and no treatment is necessary in these cases.

In cases where previous erroneous diagnosis has occurred resulting in postsurgical or postradiation hypothyroidism, treatment with TH can be started. The serum TSH level can be used as a guideline for hormone dosage; required T4 doses can be as high as 1000  $\mu\text{g}/\text{day}$ .

Some patients with RTH present with peripheral tissues that are relatively more resistant than the pituitary. Thus, compensation for the defect at the level of peripheral tissue is incomplete. In such instances, administration of supra-physiologic doses of TH is indicated. Since the dose varies among affected subjects, it should be individually determined by assessing tissue responses.

Patients without prior anti-thyroid treatment but with more severe thyrotroph resistance and symptoms of thyrotoxicosis may also require therapy. Common symptoms are hyperactivity, tachycardia, and, less commonly, diarrhea. Usually symptomatic treatment with a  $\beta$ -adrenergic blocking agent would suffice.

Treatment with L-T3 may improve the symptoms of ADHD in a significant proportion of children who also have RTH.

Although RTH is a relatively rare condition, the physician should be aware that common diseases, such as autoimmune-mediated hypothyroidism and hyperthyroidism, can occur in subjects with RTH. Concurrence of Grave's disease and RTH would significantly alter the therapeutic and diagnostic approach to the care of these patients, but the overall

goal would be to obtain clinical euthyroidism with a normal serum TSH.

### Pharmacologic Therapy

TRIAAC has been used successfully to decrease the serum TSH and TH levels, to reduce goiter size, and to alleviate some of the symptoms attributed to the effect of TH on peripheral tissues. However, the concomitant effects of TRIAC on markers that measure TH action on peripheral tissues, as well as on heart rate, are minimal, probably because the decrease in TH levels is offset by the intrinsic specific thyromimetic effect of TRIAC. The ability of TRIAC to suppress TSH without an increase in the thyromimetic effect on peripheral tissues is due to the following properties of this TH analogue: its higher affinity for TR $\beta$  but not TR $\alpha$  as compared with T3 and its more rapid degradation. The mechanism mediating a similar effect attributed to D-T4 is less well understood. TRIAC was used to treat *in utero* a fetus harboring a TR $\beta$  mutation in order to reduce fetal goiter. Although treatment was successful up to a point, some controversy has arisen due to the fact that repeated chordocentesis was necessary and that there is a lack of information about placental transport and fetal metabolism of TRIAC.

Dopaminergic drugs and somatostatin analogues have had limited use because of side effects and a low success rate in maintaining TSH suppression.

General guidelines for treatment with TH, usually L-T4, are as follows: (1) elevated serum TSH levels; (2) failure to thrive that cannot be explained on the basis of another illness or defect; (3) unexplained seizures; (4) developmental delay; and (5) history of growth or mental retardation in affected members of the family.

The future direction for the treatment of RTH will be the use of "designer drugs" that have been formulated to interact with a specific mutant TR $\beta$ . Such a TH analogue would have a strong affinity for the mutant TR $\beta$  and could eliminate the DNE. Such an analogue has already been reported, although no human studies have been reported. The rarity of the syndrome, the fact that RTH is not a life-threatening illness, the expense of the drug, and the cost of the necessary trials to obtain approval, even as an orphan drug, make the widespread availability of these drugs a prospect for the future.

Another set of compounds that are TR $\beta$  isoform-specific agonists has been developed. These drugs, which would not stimulate TR $\alpha$  and therefore be cardioprotective, have also been reported and include medications such as CG-1.

## Management of Pregnancy and RTH

Little is known about the ideal management of a pregnant mother with RTH. Different considerations may need to be made depending on whether the fetus also has RTH. In such an instance, one may predict that an increased level of maternal TH is necessary for the early development of a similarly affected fetus, but could be detrimental to a non-RTH fetus. Conversely, it is unknown whether a normal mother would be able to provide the appropriate amount of TH to an RTH fetus (for example, when the father is affected). Analysis of the limited amount of clinical data available indicates that there is no further morbidity in children and adults with RTH from affected or nonaffected mothers. However, there are no data on the outcome of normal children of RTH mothers. Further investigation is necessary for the evaluation of these important questions.

## See Also the Following Articles

Drug Effects and Thyroid Function • Thyroid Disease, Genetic Factors in • Thyroid Hormone Action • Thyroid Hormone Metabolism • Thyroid Hormone Receptors • Thyroid Hormone-Binding Proteins • TSH (Thyroid-Stimulating Hormone; Thyrotropin)

## Further Reading

- Forrest, D., and Vennstrom, B. (2000). Functions of thyroid hormone receptors in mice. *Thyroid* **10**, 41–52.
- Refetoff, S., Degroot, L. J., Benard, B., and DeWind, L. T. (1972). Studies of a sibship with apparent hereditary resistance to the intracellular action of thyroid hormone. *Metabolism* **21**, 723–756.
- Refetoff, S., DeWind, L. T., and DeGroot, L. J. (1976). Familial syndrome combining deaf-mutism, stippled epiphyses, goiter, and abnormally high PBI: Possible target organ refractoriness to thyroid hormone. *J. Clin. Endocrinol. Metab.* **27**, 279–294.
- Refetoff, S. R., Weiss, R., and Usala, S. (1993). The syndromes of resistance to thyroid hormone. *Endocr. Rev.* **14**, 348–399.
- Sarne, D. H., Refetoff, S., Rosenfield, R. L., and Farfriaux, J. P. (1988). Sex-hormone-binding globulin in the diagnosis of peripheral tissue resistance to thyroid hormone: The value of changes following short-term triiodothyronine administration. *J. Clin. Endocrinol. Metab.* **66**, 740–746.
- Weiss, R. E. (1999). Management of resistance to thyroid hormone. *Thyroid Today* **22**, 1–11.
- Weiss, R. E., Hayashi, Y., Nagaya, T., Petty, K. J., Murata, Y., Tunca, H., Seo, H., and Refetoff, S. (1996). Dominant inheritance of resistance to thyroid hormone not linked to defects in the thyroid hormone receptors  $\alpha$  or  $\beta$  genes may be due to a defective co-factor. *J. Clin. Endocrinol. Metab.* **81**, 4196–4223. [Also see editorial in the same issue.]
- Weiss, R. E., Weinberg, M., and Refetoff, S. (1993). Identical mutations in unrelated families with generalized resistance to thyroid hormone occur in CG-rich areas of the thyroid hormone receptor  $\beta$  gene: Analysis of fifteen families. *J. Clin. Invest.* **91**, 2408–2415.
- Yen, P. M. (2001). Physiological and molecular basis of thyroid hormone action. *Physiol. Rev.* **81**, 1097–1142.

principal forms of premature ovarian failure: premature menopause, with complete follicular depletion of the ovaries, and the resistant ovary syndrome, with the presence of primordial follicles. Resistant ovary syndrome is occasionally associated with autoimmune disease. Possible etiologies of resistant ovary syndrome also include antibodies against the FSH receptor or mutations in the FSH receptor gene.

Ovarian reserve may be evaluated by the following assessment: basal serum FSH levels, basal estradiol levels, basal inhibin B levels, and autoimmune and genetic evaluation; ovarian volume and ovarian biopsy are less important.

### Basal Serum FSH Levels

Once the ovary is more or less exhausted, increased pituitary production of FSH follows. It is accepted that the FSH level on day 2 or 3 of the cycle is the best marker for assessing ovarian reserve and for predicting the response to superovulation, with a good correlation with pregnancy rates. However, lack of a clear cutoff point, large variations among different laboratories, and monthly variations in FSH secretion in the same patient mean that FSH measurement is of only limited value in assessing the prognosis of IVF treatment.

Once the FSH levels start to fluctuate, it is probably too late. A decreased ovarian reserve can already be observed, and it is not clear whether starting stimulation during a later month with "normal" FSH levels will give a better result. It could be useful if each fertility center defined its cutoff point, depending on that center's experience. Furthermore, it could be very important to have another FSH evaluation in a different cycle and to test the pituitary response to the gonadotropin-releasing hormone (GnRH) test. In fact, in women with imminent ovarian failure, the pituitary is more sensitive to GnRH. This leads to higher FSH and LH pulse amplitudes, which underlie the elevated FSH concentrations during the early follicular phase.

### Basal Estradiol Levels

Measurement of basal estradiol levels, in addition to FSH levels, might improve the ability to predict fertility potential compared with measurement of basal FSH and chronological age alone. The gold standard is to evaluate estradiol levels on day 3 of the cycle. A concentration of less than 80 pg/ml with a normal

FSH level in women between 38 and 42 years of age gives a good prognosis of successful treatment.

### Basal Inhibin B Levels

Inhibin B is heterodimeric glycoprotein released by the granulosa cells of the follicle. One study showed that women with a low day 3 inhibin B level (<45 pg/ml) and a poorer response to superovulation for IVF were less likely to conceive a clinical pregnancy. It also showed that a decrease in inhibin B probably precedes the increase in FSH levels.

Inhibin is secreted mostly by granulosa cells and has a role in regulating the pituitary secretion of FSH. Therefore, inhibin is a potential marker for ovarian function and follicular content. The main function of inhibin in women is regulation of pituitary FSH secretion. One study showed that a decrease in serum inhibin concentrations could be observed when the ovarian follicular reservoir begins to diminish.

On the other hand, Petraglia and colleagues found that low levels of circulating inhibin B reflect ovarian failure in women with premature ovarian failure (POF), whereas women with hypogonadotropic hypogonadism have normal levels of inhibin B. The role of inhibin B in the pathogenesis of the resistant ovary syndrome has not been investigated in full, but inhibin B levels may be a helpful diagnostic tool.

### Autoimmunity

To diagnose the resistant ovary syndrome occasionally associated with autoimmune disease, the first important step is to think about it. If a young woman presents an autoimmune disease, the ovarian function must be checked and the antiovary antibody must be tested. Furthermore, if a young woman presents hypergonadotropic amenorrhea, it is very important to evaluate her immunological asset to exclude a concomitant silent immunological disease.

### Genetic

Several findings lean in favor of genetic control of the follicles decreasing during genital life by apoptosis. Ovarian insufficiency can occur by follicular maturation blocking (e.g., modification of genes GDF-9 and GDF-9B, null mutation of the FSH receptor gene, autoimmune polyglandular disease). Even if the etiology of premature ovarian insufficiencies in women remains unknown in more than 90% of the cases,

genetic studies on the family cases should make it possible to identify new genes involved in ovarian control.

### Ovarian Volume

Increased age is associated with a decrease in ovarian volume. Transvaginal measurement of ovarian volume is quick, accurate, and cost-effective. Decreased ovarian volume is a sign of ovarian aging that may be observed earlier than a rise in FSH levels. Small ovaries are associated with poor response to superovulation and a high cancellation rate in IVF.

### Ovarian Biopsy

Ovarian biopsy is more useful when attempting to obtain a correct diagnosis of the ovarian failure than when testing the ovarian response to the stimulation for an IVF protocol. The laparoscopic approach is preferred because it is less invasive.

## THERAPEUTIC STRATEGIES

Genes acting at later stages of oocyte or ovarian follicle function, such as gonadotropin hormones and receptors, are responsible for POF in some women. POF has been found to be a heterogeneous disorder, the dissection of which offers promising insights into mechanisms governing germ cell origination, migration, and proliferation; meiotic mechanisms; and factors governing oocyte maturation and survival. Unfortunately, for these reasons, the genetical resistant ovary syndrome is far from being “solved.”

Identifying patients with autoimmune premature ovarian failure presents the opportunity to restore ovarian function by treating these patients with the proper immune modulation therapy. On the other hand, potent immune modulation therapy can have major complications. Corticosteroid therapy for autoimmune premature ovarian failure should be limited to use in placebo-controlled trials designed to evaluate the safety and efficacy of such treatment.

For the remaining women with an idiopathic resistant ovary syndrome, standard treatment protocols are not available. The most important thing is to have an early diagnosis because only in this way can the assisted procreation procedure be used to attempt to achieve a pregnancy. In fact, if the ovarian resistance become definitive, the only hope is to use donor oocytes for assisted reproduction technology (ART).

Based on our experience starting with evidence-based medicine, we have tried to treat these women with a regimen of low estrogen, using the negative feedback on hypophysis. The gonadotropin ovarian receptors are for some time free and able to respond to new stimuli.

When the ovarian response is restored, even if only for a few cycles, the final decision as to whether to follow these women in their physiological cycle and only monitor their ovulation or to stimulate with human menopausal gonadotropin (hMG) or recombinant FSH (r-FSH) for an insemination or for IVF/intracytoplasmic sperm injection (ICSI) is a decision also related to the age of the patient, the other hormonal parameters, and the quality of seminal fluid.

### See Also the Following Articles

FSH (Follicle-Stimulating Hormone) • Infertility, Overview • In Vitro Fertilization (IVF) • Ovarian Androgen-Producing Tumors • Ovarian Failure Treatment Strategies: Egg Donation • Ovarian-Follicular Apparatus • Ovarian Hyperstimulation Syndrome • Premature Ovarian Failure

### Further Reading

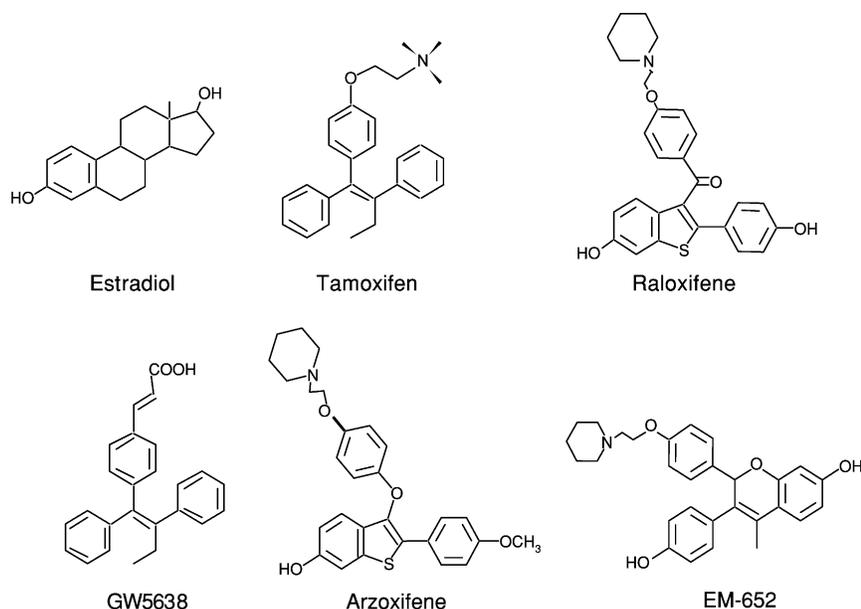
- Arici, A., Matalliotakis, I. M., Koumantakis, G. E., Goumenou, A. G., Neonaki, M. A., and Koumantakis, E. E. (2002). Diagnostic role of inhibin B in resistant ovary syndrome associated with secondary amenorrhea. *Fertil. Steril.* **78**, 1324–1326.
- Bondy, C. A., Nelson, L. M., and Kalantaridou, S. N. (1998). The genetic origins of ovarian failure. *J. Women's Health* **7**, 1225–1229.
- Buyalos, R. P., Daneshmand, S., and Brzechffa, P. R. (1997). Basal estradiol and follicle-stimulating hormone predict fecundity in women of advanced reproductive age undergoing ovulation induction therapy. *Fertil. Steril.* **68**, 272–277.
- Christin-Maitre, S. (2002). Ovarian genes. *Gynecol. Obstetr. Fertil.* **30**, 827–830.
- de Koning, C. H., Popp-Snijders, C., Schoemaker, J., and Lambalk, C. B. (2000). Elevated FSH concentrations in imminent ovarian failure are associated with higher FSH and LH pulse amplitude and response to Gn-RH. *Hum. Reprod.* **15**, 1452–1456.
- Dicker, D., Goldman, J. A., Ashkenazy, J., et al. (1991). Age and pregnancy rates in *in vitro* fertilization. *J. In Vitro Fertil. Embryo Transfer* **8**, 141–144.
- Faddy, M. J., and Gosden, R. G. (1995). A mathematical model of follicle dynamics in the human ovary. *Hum. Reprod.* **10**, 770–775.
- Jequier, A. M. (1990). Conception in the resistant ovary syndrome occurring during hormone replacement therapy: A report of two cases. *Aust. N.Z. J. Obstetr. Gynecol.* **30**, 176–178.
- Kalantaridou, S. N., Braddock, D. T., Patronas, N. J., and Nelson, L. M. (1999). Treatment of autoimmune premature ovarian failure. *Hum. Reprod.* **14**, 1777–1782.
- Kalantaridou, S. N., Davis, S. R., and Nelson, L. M. (1998). Premature ovarian failure. *Endocrinol. Metab. Clinics North America* **27**, 989–1006.

- Lass, A., Skull, J., McVeigh, E., *et al.* (1997). Measurement of ovarian volume by transvaginal sonography before human menopausal gonadotrophin superovulation for *in vitro* fertilization can predict poor response. *Hum. Reprod.* **12**, 294–297.
- Petraglia, F., Hartmann, B., Luisi, S., Florio, P., Kirchengast, S., Santuz, M., Genazzani, A. D., and Genazzani, A. R. (1998). Low levels of serum inhibin A and inhibin B in women with hypergonadotropic amenorrhea and evidence of high levels of activin A in women with hypothalamic amenorrhea. *Fertil. Steril.* **70**, 907–912.
- Scott, R. T., and Hoffmann, G. E. (1995). Prognostic assessment of ovarian reserve. *Fertil. Steril.* **63**, 1–11.
- Scott, R. T., Toner, J. P., Muasher, S. J., *et al.* (1989). Follicle-stimulating hormone levels on cycle day 3 are predictive of *in vitro* fertilization outcome. *Fertil. Steril.* **51**, 651–654.
- Seifer, D. B., Lambert-Masserlian, G., Hogan, J. W., *et al.* (1997). Day 3 serum inhibin-B is predictive of assisted reproductive technologies outcome. *Fertil. Steril.* **67**, 110–114.



## Rickets

see Vitamin D Deficiency, Rickets, and Osteomalacia



**Figure 1** Chemical structures of estradiol and selective estrogen receptor modulators: tamoxifen, raloxifene, GW5638, arzoxifene, and EM-652.

cancer, clinical trials showed that tamoxifen increased the risk of endometrial cancer in the uterus by 0.1%. These results strongly indicated that tamoxifen was not a pure anti-estrogen but had selective functions depending on the target tissue.

## SELECTIVE ESTROGEN RECEPTOR MODULATION

In the 1980s, the concept of selective estrogen receptor modulators (SERMs) was recognized from research studies that investigated the pharmacological properties of tamoxifen at diverse tissue sites. For example, tamoxifen has estrogen-like effects on bone in ovariectomized rats. However, in the uterus, tamoxifen is both a partial estrogen and an anti-estrogen, but in the rat mammary gland, tamoxifen is an anti-estrogen/antitumor agent. Most importantly, observations in the mouse, where tamoxifen is predominantly an estrogen, were used to illustrate the target site specificity rather than the species specificity of tamoxifen. When athymic mice are bi-transplanted with both human breast and endometrial tumors, tamoxifen inhibits the growth of breast tumors while simultaneously stimulating the growth of endometrial tumors. Thus, the concept of the selective actions by nonsteroidal compounds was first recognized in the laboratory and then applied to clinical drug development. Tamoxifen maintains bone

density and lowers circulating cholesterol levels in postmenopausal women. It was known that tamoxifen lowers the recurrence of ER-positive breast cancer and decreases the incidence of contralateral breast cancer by 50% but causes a modest increase in the risk of endometrial cancer by 0.1% in postmenopausal women. The beneficial effects of tamoxifen at various sites in a woman's body were used to justify the evaluation of tamoxifen as a preventive in high-risk pre- and postmenopausal women. Predictably, tamoxifen reduces the incidence of invasive breast cancer and ductal carcinoma *in situ* by 50%.

The discovery that tamoxifen was a SERM has provided an opportunity for the development of new drugs with tissue selectivity. Multifunctional compounds that provide the potential positive effects of estrogen on the skeleton and cardiovascular system and prevent postmenopausal-related hot flashes, but avoid the negative effects on the breast and the uterus, are in high demand, particularly since it has been reported that hormone replacement therapy (HRT) increases a woman's risk of developing breast cancer and coronary heart disease but has only modest beneficial effects on bone density. One of these new SERMs, raloxifene, is similar in chemical structure to tamoxifen (Fig. 1) and has been shown to be an estrogen-like agonist on bone with a reduction of fractures, but with protective effects on the reduction of breast cancer incidence and little detrimental effect

on the uterus. Unfortunately, raloxifene does not decrease the incidence of hot flashes. Raloxifene is being tested as a chemopreventive for breast cancer in the Study of Tamoxifen and Raloxifene trial and because raloxifene lowers circulating cholesterol levels it is also being tested in the study of Raloxifene Use for the Heart trial as a preventive for coronary heart disease.

Development of new SERMs as multifunctional medicines is of growing interest to the clinical community. Thus, discovery of the mechanism of action of SERMs would greatly enhance the development of new agents. Some progress has been made in understanding the function of the SERM-ER complex *in vitro*. However, the exact mechanism of action in various tissues has not been discovered. Therefore, the future of SERM development in the clinic will depend on identifying new target sites for SERM action in the laboratory.

## MECHANISM OF ACTION

### ER- $\alpha$ and ER- $\beta$

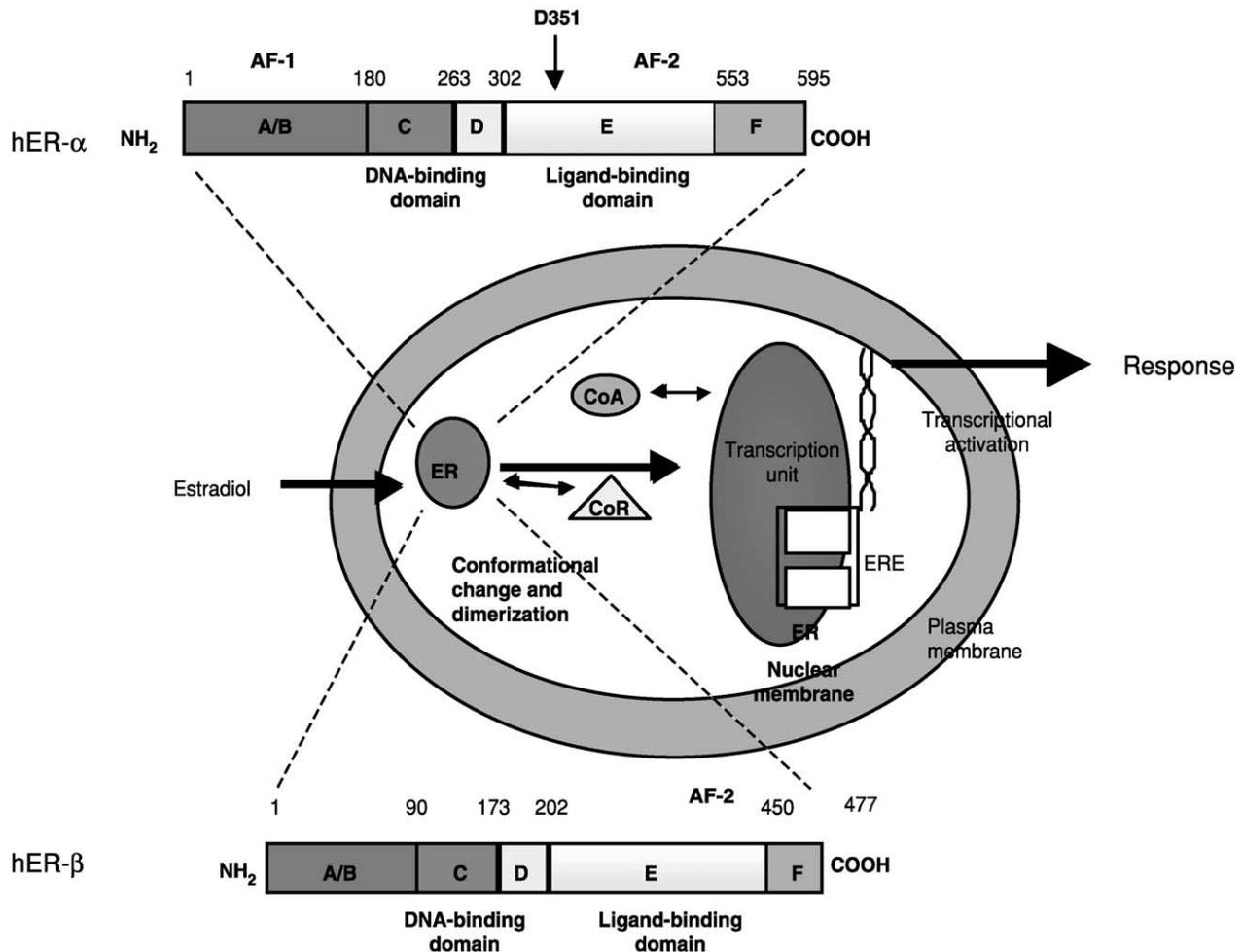
An understanding of the selective nature of SERMs in patients first requires an understanding of the function of the ER at the molecular level. The human ER is a member of the nuclear receptor family of ligand-inducible transcription factors. The classical model of ER function is that estradiol binds to the ligand-binding domain (LBD) of the nuclear ER in target tissues to induce a conformational change in the three-dimensional structure of the complex. The ER dimerizes and subsequently binds to DNA sequences referred to as estrogen-response elements (EREs). These EREs are *cis*-acting enhancer elements located within regulatory regions of target genes. The DNA-bound ER contacts the preinitiation transcriptional machinery either directly or indirectly by coregulatory proteins (Fig. 2). In contrast, the nonclassical or tethered pathway for ER activity is the indirect interaction of the ER with regulatory elements in the promoter of target genes through direct interaction with transcription factors, such as jun/fos, or activator protein-1 (AP-1) and the subsequent activation of AP-1-driven genes. Thus, the ER can regulate the transcription of a wide variety of genes by both classical and nonclassical pathways.

The identification of different receptor isoforms is one recognized pharmacological method for developing tissue-selective drugs. The discovery of two genetically distinct ERs, ER- $\alpha$  and ER- $\beta$ , provides a possible explanation for the tissue-selective nature of SERMs. The two receptors have some degree

of homology as both receptors have a LBD and a DNA-binding domain (Fig. 2). However, there are notable differences in the two activating functions (AFs). First, ER- $\beta$  does not have an AF-1 region but does retain an AF-2 region. These differences in the AF regions alter the activity of the SERM-ER complex, resulting in increased or decreased estrogenic activity. For example, tamoxifen is more anti-estrogenic when it is complexed with ER- $\beta$ . However, little is known about the role of ER- $\beta$  in estrogen and SERM action. What is known from mRNA expression profiles is that ER- $\beta$  is expressed in a wide variety of tissues—including the breast and uterus. However, the existence of significant levels of ER- $\beta$  protein in all mRNA-expressing tissues has been more difficult to interpret and requires further investigation. Studies *in vitro* indicate that both ER- $\alpha$  and ER- $\beta$  activate transcription in response to estradiol at ERE-containing promoters, with ER- $\alpha$  being a more efficient transcriptional activator than ER- $\beta$ . However, when both receptors are expressed in cells, ER- $\beta$  decreases the activity of ER- $\alpha$  and decreases the overall sensitivity of agonists. The mechanism is not known and there are few *in vivo* data showing competition between the two receptors for the same DNA sequence and/or heterodimerization of ER- $\beta$  with ER- $\alpha$ . In cells where tamoxifen activates ER- $\alpha$ -mediated, ERE-driven transcriptional activity, coexpression of ER- $\beta$  completely suppresses this activity. Thus, expression levels of ER- $\beta$  might regulate the magnitude of the tamoxifen-ER- $\alpha$  complex on ERE-driven genes but only if the ratio of receptors was appropriate. In AP-1-containing elements/ER- $\alpha$  systems, where ER- $\alpha$  binds indirectly to AP-1 promoters by interaction with AP-1, estrogens activate transcription and SERMs display a range of activities from partial to full agonists. In contrast, estrogens antagonize and all SERMs tested activate transcription in cells expressing AP-1-ER- $\beta$  systems. However, demonstration of the physiological significance of the AP-1-ER interaction *in vivo* has yet to be shown. What is known from *in vitro* studies is that tamoxifen appears to inhibit estradiol-activated transcription by both ER- $\alpha$  and ER- $\beta$ . However, in some cell and promoter contexts, the tamoxifen-ER- $\alpha$  complex manifests partial agonist activity. Therefore, it remains to be explained how the same ER- $\alpha$ -ligand complex can be recognized differently in different cells.

### ER Coregulators

Another explanation for the selective activities of SERMs in different cell types is the expression pattern



**Figure 2** Function of estrogen receptors (ERs)  $\alpha$  and  $\beta$  in a cellular context. The human estrogen receptor proteins  $\alpha$  and  $\beta$  share identity. The major difference between the two receptors is that ER- $\beta$  is truncated by 118 amino acid residues and contains no activating function 1 (AF-1) region. Both receptors contain DNA- and ligand-binding domains. Both bind one molecule of estradiol, dimerize, and recruit coactivators (CoAs) and/or corepressors (CoRs). If more CoAs are present than CoRs, transcription of target genes driven by estrogen-responsive elements (EREs) is activated, followed by a cellular response, such as cell division.

of coregulatory proteins. Estrogen receptors directly or indirectly activate or repress target genes by binding to hormone-response elements in promoter or enhancer regions (Fig. 2). Modulation of transcription by ERs requires the recruitment of coregulators. Coregulators are transcription modifiers that can either activate or repress the activity of the ER complex. A basic mechanism for the switching of target genes from off to on requires a ligand-dependent exchange of corepressors for coactivators. Members of the family of steroid receptor coactivators (SRC-1,-2, and-3) complement the activity of the ER. Ligand-dependent recruitment of coactivators is dependent on the AF-2 region within the C terminus of

the LBD of both ER- $\alpha$  and ER- $\beta$ . The precise region of the coactivators that interacts with the AF-2 region of the ER is an LXXLL domain, where L represents a leucine and X represents any amino acid. The ER-coactivator complex functions to stabilize the preinitiation transcriptional machinery and promotes chromatin remodeling at the promoter region of the target gene. Thereafter, transcription of the target gene is either activated or repressed depending on the promoter and cellular context. SRC-1 has been shown to recruit CBP/p300, an acetyltransferase that acetylates histones bound to chromatin, facilitating remodeling. More than 30 additional putative coactivators have been identified. Thus, expression levels of

coactivators might explain the actions of SERMs in different cells. The estrogen-like activity of the tamoxifen-ER complex has been shown to be due to higher levels of SRC-1 in endometrial cancer cells versus MCF-7 breast cancer cells, where the tamoxifen-ER complex is mostly anti-estrogenic.

ER- $\alpha$  and ER- $\beta$  also exert critical roles by repressing gene transcription. They function as ligand-independent repressors on some target genes or ligand-dependent repressors on others. Corepressors are coregulators that repress the activity of both ER- $\alpha$  and ER- $\beta$ . The known co-repressors are N-CoR, SMRT, and REA. These coregulators contain multi-independent repressor domains that interact with histone deacetylases to mediate the deacetylation of histones and promote the condensation of chromatin, repressing the transcription of the target gene. The ER-corepressor complex prevents the interaction of the ligand-less ER with the transcriptional machinery, resulting in transcriptional repression. Transcriptional repression of target genes by SERM-ER complexes may depend on the structure of the SERM, the three-dimensional structure of the SERM-ER complex, and the availability of corepressors versus coactivators.

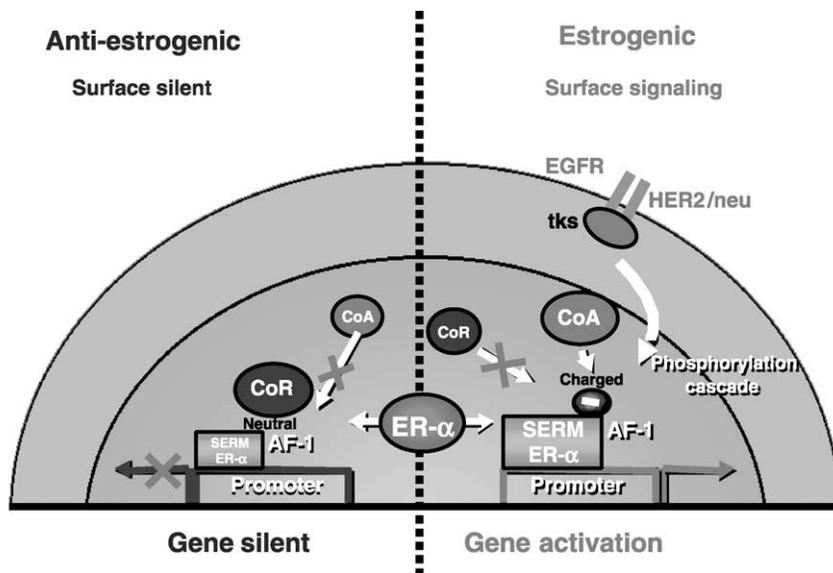
### Structure-Function Relationship

The ligand that occupies the LBD of ER- $\alpha$  programs the external shape of the ER complex. Estradiol is a small, planar, and lipophilic steroid molecule (Fig. 1) that is able to fit into the LBD with high affinity. Estradiol binding to ER- $\alpha$  controls the activity of the AF-2 region, which is located in the LBD. The activity of the AF-1 region of ER- $\alpha$  is regulated by phosphorylation. These two activation domains can function either independently or synergistically, depending on the cell type and the target promoter of a particular ER-responsive gene. Numerous molecular models of estrogen, partial estrogen, and anti-estrogen action were developed in the 1980s based on structure-activity relationships of the ligand-bound ER- $\alpha$ . The model suggests that sealing an estradiol molecule in the LBD controls the intrinsic activity of the ER- $\alpha$  complex. Anti-estrogens block the binding of estradiol to the LBD by competitive inhibition. It has been shown that the length of the anti-estrogenic side chain regulates the anti-estrogenicity of the SERM. Partial agonist activity results from a mixture of diverse-shaped SERM-ER- $\alpha$  complexes, depending on the structure of the SERM. The identification of a single point mutation in the LBD, D351Y, of ER- $\alpha$  in a tamoxifen-stimulated breast tumor model and the finding that the D351YER $\alpha$  converts raloxifene from

an anti-estrogen to an estrogen demonstrated that a specific region in the ER controls the estrogen-like actions of a SERM. The crystal structures of both the tamoxifen-ER- $\alpha$ LBD and the raloxifene-ER- $\alpha$ LBD complexes show that the anti-estrogenic side chains of SERMs directly interact with D351, located on the surface of the LBD of the ER. The data demonstrate that raloxifene interacts tightly with the aspartate residue, resulting in neutralization of the carboxylate anion. However, tamoxifen is not able to shield the aspartate residue completely and thus is unable to neutralize the anion so that the tamoxifen-ER- $\alpha$  complex is more estrogen-like at target genes. Numerous other mutants of D351 were subsequently generated to address the hypothesis that the D351 residue in the LBD regulates the estrogen-like actions of tamoxifen. A D351G (aspartate to glycine) mutation, transfected into ER-negative MDA-MB-231 breast cancer cells, allosterically converts tamoxifen from an estrogen to a complete anti-estrogen, suggesting that the charge at position 351 of LBD is critical for the estrogenic activity. In contrast, expression of a D351E (aspartate to glutamate) or a D351Y (aspartate to tyrosine) mutation restores the estrogenic activity of the raloxifene-ER complex. The same result is seen when the anti-estrogenic side chain of raloxifene is changed from a piperidinyl ring to a cyclohexanyl ring (Fig. 1). Conversely, the D351F (aspartate to phenylalanine) mutation completely negates the agonist activity of raloxifene. The results, taken together, suggest that the interrelationship of the anti-estrogenic side chain of the SERM and residue D351 of the ER is critical for the estrogenic activity of SERM-ER complexes at specific target sites.

### DRUG RESISTANCE

The activity of the SERM-ER- $\alpha$  complex is also regulated by phosphorylation of either the AF-1 domain at the amino terminus of ER- $\alpha$  or coactivators. Numerous *in vitro* studies demonstrate that the p42/44 mitogen-activated protein kinase (MAPK) pathway phosphorylates ER- $\alpha$  or SRC-3 and enhances the estrogenic activity in a ligand-independent and/or-dependent manner. Due to the fact that cell surface receptors activate intracellular phosphorylation cascades such as the MAPK pathway, one hypothesis to explain the differential activities of SERMs in different tissues is crosstalk between the ER- $\alpha$  and the surface receptors (Fig. 3). Evidence to suggest that crosstalk between surface receptors and the ER- $\alpha$  alters the activity of the SERM-ER- $\alpha$  complex came



**Figure 3** Integrated mechanism for the target site-specific action of SERMs in breast or uterine cancer. The two extremes of anti-estrogenic or full estrogenic actions are shown. Estrogen-like actions could occur in cells expressing an excess of coactivators (CoAs) and/or a decrease in corepressors (CoRs). The charged surface of a tamoxifen-ER complex prevents CoR binding. The estrogenic action would be amplified by surface signaling from dimers of epidermal growth factor receptor (EGFR) and HER2/neu activating tyrosine kinases (tk). The phosphorylation cascade can activate AF-1 on ER- $\alpha$  directly or activate the excess of CoAs in a high-ER environment. Reduced levels of ER prevent the signal transduction pathway and promote anti-estrogenic actions in a surface silent cell. Reprinted from Jordan, V. C. (2002). The secrets of selective estrogen receptor modulation: Cell-specific coregulation. *Cancer Cell* 1(3), pp. 215–217. Copyright 2002, with permission from Elsevier Science.

in 1993, when Benz *et al.* showed that overexpression of HER2/neu, a member of the epidermal growth factor receptor family of receptor tyrosine kinases, in MCF-7 cells resulted in resistance to tamoxifen. Clinical evidence also strongly supports a correlation between overexpression of HER2/neu and the lack of efficacy of tamoxifen in ER- $\alpha$ -positive breast cancer. However, there are two models of tamoxifen resistance, observed in the laboratory and the clinic. Intrinsic resistance to tamoxifen occurs in ER- $\alpha$ -negative tumors or might result from the overexpression of HER2/neu in ER- $\alpha$ -positive tumors. Acquired resistance to tamoxifen can result from extended therapy of ER- $\alpha$ -positive breast tumors that initially regress with treatment but recur by tamoxifen-stimulated growth. Thus, the change in activity of the tamoxifen-ER- $\alpha$  complex from anti-estrogenic to estrogenic during breast cancer treatment might occur by crosstalk between surface receptors (HER2/neu) and the SERM-ER- $\alpha$  complex (Fig. 3).

## NOVEL AGENTS

New drug discovery for SERMs is driven by the known side effects of tamoxifen, for example, the

increase in the incidence of endometrial cancer and the report of the negative effects of HRT on coronary heart disease. Several approaches are being pursued by altering the anti-estrogenic side chain or improving the pharmacokinetics of existing molecules. A novel SERM, GW5638, is a tamoxifen derivative containing a carboxylic acid side chain rather than a tertiary amino group (Fig. 1). As demonstrated in the rat, GW5638 is a nonsteroidal SERM with estrogen-like activity in bone without known uterotrophic activity. GW7604, the active metabolite of the prodrug GW5638, is less estrogenic than the active metabolite of tamoxifen (4-OHT) due to the carboxylate side chain of GW7604 causing a strong repulsion of the aspartate anion at residue D351 in the LBD of the ER- $\alpha$ . GW5638 might be a good candidate as a novel SERM for the prevention and treatment of osteoporosis and breast cancer with little risk of developing endometrial cancer because the drug has a longer biological half-life than raloxifene.

Arzoxifene is another novel SERM developed to improve the poor bioavailability of raloxifene (Fig. 1). Arzoxifene is being tested in clinical trials for the treatment of breast cancer. Preliminary clinical data show that arzoxifene is only a modest anti-estrogen for the treatment of breast cancer and is predominantly an

anti-estrogen in the rat uterus. Tamoxifen-stimulated endometrial tumors grown in athymic mice continue to grow with tamoxifen or arzoxifene, indicating that arzoxifene might be a poor candidate as a second-line treatment after tamoxifen failure.

EM-800 was developed in an effort to increase the bioavailability of SERMs for longer pharmacological effects. The active metabolite of EM-800, EM-652, is similar in structure to raloxifene (Fig. 1) and is a potent anti-estrogen in both *in vitro* and *in vivo* biological assays. EM-652 inhibits estrogen-responsive proliferation of a variety of human breast cancer cell lines and is observed to inhibit tumor growth and have little stimulatory effect on human endometrial cancer cells. EM-652 is a potent blocker of both estrogen-activated ER- $\alpha$  and ER- $\beta$ , resulting in repression of estrogen-responsive gene transcription. The activation of both ERs by either H-Ras or SRC-1 is completely abrogated by EM-652. The pharmacology of EM-652 is similar to that of raloxifene due to the similar manner in which they both fit into the LBD and interact with residue D351. Based on the analogy of the prodrug EM-800 with raloxifene, the prediction would be that EM-800 should be effective as a bone-preserving SERM with little uterotrophic activity.

## IDEAL MULTIFUNCTIONAL SERMS

Although the new SERMs, GW5638, arzoxifene, and EM-800, are improvements over tamoxifen and raloxifene with regard to bioavailability and the lack of uterotrophic activity, the need for the ideal multifunctional SERM still exists. The ideal SERM should have full estrogenic activity in the skeleton to maintain bone density, lower circulating levels of cholesterol to prevent coronary heart disease, and prevent hot flashes and should have anti-estrogenic activity in the breast and uterus to prevent breast and endometrial cancers. Most importantly, a SERM that prevents hot flashes will dramatically enhance compliance. Development of the ideal SERM with tissue-selective activity requires elucidating the exact mechanisms of SERM-ER complexes in diverse cellular and molecular contexts. The exploitation of both *in vivo* and *in vitro* SERM-ER models will assist in the identification of signal transduction pathways that interact with the SERM-ER- $\alpha$  or SERM-ER- $\beta$  complex. Furthermore, understanding the role of coactivators and corepressors in modulating the activity of SERM-ER complexes in different tissues will further explain the selective nature of SERMs. Ultimately, this information will provide insight for development of

SERMs that will advance the prevention of breast cancer, endometrial cancer, coronary heart disease, osteoporosis, and postmenopausal symptoms.

## Acknowledgments

The authors thank The Avon Foundation and The Lynn Sage Breast Cancer Research Foundation for their generosity. Funding provided by Breast Cancer Spore Grant CA89018-03 is most gratefully acknowledged.

## See Also the Following Articles

Breast Disease: Impact of Sex Steroid Replacement • Cardiovascular Disease: Impact of Sex Steroid Replacement • Estrogen Replacement, Oral • Estrogen Replacement, Vaginal • Hormone Replacement, Transdermal • Hyperparathyroidism, Primary • Osteoporosis in Older Women • Osteoporosis in Reproductive Age Women: Impact of Sex Steroid Replacement

## Further Reading

- Bentrem, D., Dardes, R., Liu, H., MacGregor-Schafer, J., Zapf, J., and Jordan, V. (2001). Molecular mechanism of action at estrogen receptor  $\alpha$  of a new clinically relevant antiestrogen (GW7604) related to tamoxifen. *Endocrinology* **142**, 838–846.
- Benz, C. C., Scott, G. K., Sarup, J. C., Johnson, R. M., Tripathy, D., Coronado, E., Shepard, H. M., and Osborne, C. K. (1993). Estrogen-dependent, tamoxifen-resistant tumorigenic growth of MCF-7 cells transfected with HER2/neu. *Breast Cancer Res. Treat.* **24**, 85–95.
- Berry, M., Metzger, D., and Chambon, P. (1990). Role of the two activating domains of the estrogen receptor in the cell-type and promoter-context dependent agonistic activity of the anti-estrogen 4-hydroxytamoxifen. *EMBO J.* **9**, 2811–2818.
- Brzozowski, A. M., Pike, A. C., Dauter, Z., Hubbard, R. E., Bonn, T., Engstrom, D., Ohman, L., Greene, G. L., Gustafsson, J. A., and Carlquist, M. (1997). Molecular basis of agonism and antagonism in the estrogen receptor. *Nature* **389**, 753–758.
- Cummings, S. R. (1999). The effect of raloxifene on risk of breast cancer in postmenopausal women: Results from the MORE randomized trial. Multiple Outcomes of Raloxifene Evaluation. *J. Am. Med. Assoc.* **281**, 2189–2197.
- Dardes, R. C., Bentrem, D., O'Regan, R. M., Schafer, J. M., and Jordan, V. C. (2001). Effects of the new selective estrogen receptor modulator LY353381.HCl (arzoxifene) on human endometrial cancer growth in athymic mice. *Clin. Cancer Res.* **7**, 4149–4155.
- Ettinger, B. (1999). Reduction of vertebral fracture risk in postmenopausal women with osteoporosis treated with raloxifene: Results from a 3-year randomized clinical trial. Multiple Outcomes of Raloxifene Evaluation (MORE) Investigators. *J. Am. Med. Assoc.* **282**, 637–645.
- Fisher, B., Costantino, J. P., Redmond, C. K., et al. (1994). Endometrial cancer in tamoxifen-treated breast cancer patients: Findings from the National Surgical Adjuvant Breast and Bowel Project (NSABP) B-14. *J. Natl. Cancer Inst.* **86**, 527–537.

- Fisher, B., Constantino, J. P., Wickerham, D. L., *et al.* (1998). Tamoxifen for prevention of breast cancer: Report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study (NSABP). *J. Natl. Cancer Inst.* **90**, 1371–1388.
- Gottardis, M., Robinson, S. P., Satyaswaroop, P. G., *et al.* (1988). Contrasting actions of tamoxifen on endometrial and breast tumor growth in the athymic mouse. *Cancer Res.* **48**, 812–815.
- Jordan, V. C. (2001). Selective estrogen receptor modulation: A personal perspective. *Cancer Res.* **61**, 5683–5687.
- Jordan, V. C., Lababidi, M. K., and Langan-Fahey, S. (1991). Suppression of mouse mammary tumorigenesis by long-term tamoxifen therapy. *J. Natl. Cancer Inst.* **83**, 492–496.
- Jordan, V. C., Phelps, E., and Lindgren, J. U. (1987). Effects of antiestrogens on bone in castrated and intact female rats. *Breast Cancer Res. Treat.* **10**, 31–35.
- Kato, S., Endoh, H., Masuhiro, Y., *et al.* (1995). Activation of the estrogen receptor through phosphorylation by mitogen-activated protein kinase. *Science* **270**, 1491–1494.
- Lerner, L. J., and Jordan, V. C. (1990). Development of antiestrogens and their use in breast cancer: Eighth Cain Memorial Award Lecture. *Cancer Res.* **50**, 4177–4189.
- Liu, H., Park, W. C., Bentrem, D. J., *et al.* (2002). Structure–function relationships of the raloxifene estrogen receptor- $\alpha$  complex for regulating transforming growth factor- $\alpha$  expression in breast cancer cells. *J. Biol. Chem.* **277**, 9189–9198.
- Onate, S. A., Tsai, S. Y., Tsai, M. J., *et al.* (1995). Sequence and characterization of a coactivator for the steroid hormone receptor superfamily. *Science* **270**, 1354–1357.
- Schafer, J. I., Liu, H., Tonetti, D. A., *et al.* (1999). The interaction of raloxifene and the active metabolite of the antiestrogen EM-800 (SC 5705) with the human estrogen receptor. *Cancer Res.* **59**, 4308–4313.
- Shang, Y., and Brown, M. (2002). Molecular determinants for the tissue specificity of SERMs. *Science* **295**, 2465–2468.
- Willson, T. M., Norris, J. D., Wagner, B. L., *et al.* (1997). Dissection of the molecular mechanism of action of GW5638, a novel estrogen receptor ligand, provides insights into the role of estrogen receptor in bone. *Endocrinology* **138**, 3901–3911.
- Writing Group for the Women’s Health Initiative Investigators (2002). Risks and benefits of estrogen plus progestin in healthy postmenopausal women: Principal results from the Women’s Health Initiative randomized controlled trial. *J. Am. Med. Assoc.* **288**, 321–333.



## Sertoli Cell Pathology

see Androgen Insensitivity Syndrome

due to the upregulation of nitric oxide synthesis. This concept is supported by a positive effect of administration of dopamine agonists such as apomorphine, bromocriptine, and pergolide on spontaneous penile erections, increased libido reported after dopamine precursor L-dopa is administered, and impaired libido associated with antidopaminergic agents. Epidemiological data include evidence that higher serum testosterone levels are associated with greater sexual activity in healthy, older but not younger men. Sexual studies suggest libido is restored in hypogonadal males when testosterone levels are increased from subnormal to the low normal range. Other studies showed that higher testosterone levels may shorten the latency of erection stimulated by the exposure to erotic material, and testosterone replacement in hypogonadal males restores sexual interest, shortens latency, and increases frequency and magnitude of nocturnal penile tumescence (NPT). Furthermore, withdrawal of androgen therapy in hypogonadal males leads to a decline in libido in 3 to 4 weeks and unreplace hypogonadal males have impairment in spontaneity of erections. These observations appear to have a threshold effect, as levels within the normal range seem to be adequate for normal libido and increasing the doses to superphysiologic levels does not further enhance sexual drive. It is also of note that testosterone is supportive of sexual drive and seems to be responsive to external stimuli. This is shown by the work of J. M. Davidson demonstrating that men with hypogonadism are able to achieve erections in response to viewing of erotic films.

### Erectile Function

Penile erections occur when the paired corpora cavernosa and corpus spongiosum of the testes are filled with blood under pressure. Erection is the ultimate response to multiple psychogenic and sensory stimuli from imaginative, visual, auditory, olfactory, tactile, and genital reflexogenic sources that affect several neurologic and vascular cascades that lead to penile tumescence and rigidity sufficient for vaginal penetration. Coincident with penile engorgement sexual arousal is heightened, with testicular ascent and swelling and dilation of the urethral bulb. There is also an increase in the glands and coronal size and vasodilation of blood vessels in the skin of the chest and buttocks.

The mechanisms that increase penile blood flow have been studied extensively, with nitric oxide having been identified as the predominant endothelial-derived relaxation factor. Nitric oxide is synthesized from its precursor arginine under the enzymatic

regulation of nitric oxide synthase (NOS). While a constitutive NOS is produced by the endothelial cells and nerve terminals of the penile vascular tissue, the predominant effect leading to an erection is an increase in neuronal NOS. Nitric oxide is a major vasodilator with its effects mediated by the production of cyclic GMP. Cyclic GMP is in turn inactivated by a relatively penile specific enzyme phosphodiesterase type 5. It was the recognition that phosphodiesterase 5 is a regulator of the persistence of nitric oxide induced cyclic GMP effects on penile blood flow that led to the development of a class of drugs (phosphodiesterase 5 inhibitors; e.g., Sildenafil) that selectively enhance erectile function. While nitric oxide seems to be the predominant stimulus for local penile vasodilatation, other factors, including vasoactive intestinal polypeptide (VIP), bradykinin, peptide, histidine, nithiamide, pituitary adenal cyclase activating polypeptide, helospectin, gallamine, calcitonin-gene-related peptide (CGRP) and prostaglandin E-1, may play a role. Thus, increased sexual drive produced by cognitive and other erotic signals from the brain, sleep-induced CNS signaling, and tactile stimulation of the sensory component of the penis all induce increased vascular dilatation.

Once increased penile blood flow occurs, there is compression of the penile veins, which course on an angle from the corpora cavernosa to the tunica albuginea to the deep dorsal vein. Engorged corpora cavernosa compress the venous outflow maintaining erection until the signals for enhanced blood flow disappear. It should be noted that the nonadrenergic perisymphatic pathway provides the neuronal pathway between the brain, spinal cord, and the penis. The sympathetic pathway probably plays a role in the ultimate detumescence, providing inhibitory signals to the NOS containing perisymphatic neuron and having direct vasocontractile effects on the smooth muscle. The latter appears to be mediated by norepinephrine and neuropeptide Y.

### Ejaculation

The ejaculation phase is controlled by sympathetic innervation of the genital organs. It occurs as a result of a spinal cord reflex arc. There is considerable voluntary inhibitory control over this space of the sexual response, and this has been described as having two sequential components. The first component is called emission and is associated with deposition of seminal fluid into the posterior urethra. Contraction of the ampullae of the vas deferens, the seminal vesicles, and the smooth muscles of the prostate mediate emission.

The second process is the true ejaculation; it results when the seminal fluid is expelled from the posterior urethra through the penile meatus as a result of organized smooth muscle contraction.

### Orgasm

Much less is understood about the regulation of male orgasm than about erectile function. It is generally believed that both physiological and psychogenic elements contribute to the genesis phase. The process seems to be associated with afferent stimuli that transmit via the pudendal nerve, inducing the following physiologic events: smooth muscle contraction of the accessory sex organs; build-up and release of pressure in the posterior urethral; sensation of the ejaculatory inevitability; contraction of the urethral bulb and perineum; rhythmic contractions of the pelvic floor muscles; semen emission and ejaculation; and finally the reversal of the generalized physiologic changes and sexual tension. These processes are sensed by the cortical neuron as pleasurable. Factors that influence the subjective sensation of orgasmic pleasure include the degree of sexual excitement, recency of sexual activity, and the psychosexual makeup of the individual. Although uncommon, it is possible for orgasm to occur without either erection or ejaculation. Conversely, contraction of the pelvic musculus and ejaculation may occur in the absence of the pleasurable component of orgasm.

### Detumescence

The final phase of the erectile process is called detumescence. In this phase the penis returns to the flaccid state. This occurs when vasoconstriction occurs in the arterials, reducing the blood flow to the penis and allowing increase in venous drainage of the corpora cavernosal contents. Local penile  $\alpha$ -adrenergic receptor activation is the most important neuro mediator effecting detumescence. Interference to this function through the  $\alpha$ -1 adrenergic receptor blockade may lead to the development of priapism.

A list of the causes of sexual dysfunction by clinical manifestations is provided in [Table I](#).

## ANDROGENS AND ERECTILE FUNCTION

### Androgen Actions on Libido and Penile Erections

Testosterone has its primary effect on erectile function by enhancing libido with secondary effects on

penile NOS activity. Libido is highly sensitive to testosterone, thus explaining the preservation of erectile capacity in many men with partial androgen deficiency. Young men with marked hypogonadism are usually free of functional and anatomical impairment of the penile vasodilatory apparatus. Thus, severely hypogonadal young men with impaired libido and secondary decrease in erectile activity usually benefit from treatment with testosterone therapy. In contrast, middle-aged and, in particular, older males with androgen deficiency may have a secondary penile neurovascular disorder and may respond poorly to testosterone therapy alone.

### Testosterone Deficiency and Impaired Sexual Function

There has been considerable debate about the usefulness and cost-effectiveness of hormonal evaluation and the extent to which testosterone deficiency should be investigated in men presenting with erectile dysfunction. Of all men with erectile dysfunction, only 8–10% have demonstrated low total testosterone levels. Because serum testosterone concentrations fall progressively with age and sex hormone binding globulin concentrations increase with age, the incidence of chemically low bioavailable or free testosterone is higher in the older age group. Using total testosterone as a chemical marker, studies have reported that only 6–8% of men with erectile dysfunction have an endocrine basis to the condition, particularly in the older age group. These data are contrasted with the high incidence of subnormal serum-free and non-SHBG bound testosterone in the overall elderly male population. These findings are confirmed in the MMAS study, where longitudinal data revealed a 1.6% per year and 2–3% per year decline in total and bioavailable testosterone, respectively, in men between the ages of 40 and 70 years.

Erectile dysfunction and androgen deficiency may be considered two common but often independently distributed disorders. Thus, while testosterone deficiency represents only a small component of the total population of men with erectile dysfunction, it is important to be able to identify testosterone deficiency in this patient population, in part because androgen deficiency is a correctable cause of sexual dysfunction, and some men with erectile dysfunction and low testosterone concentrations respond rapidly to testosterone replacement. In addition, testosterone deficiency is responsible for many other systemic complaints, including muscle wasting, osteoporosis, increased

Table 1 Causes of Sexual Dysfunction in the Male Classified by Clinical Manifestation

Clinical manifestation	Most common causes
Disorders of desire	
Hypoactive sexual desire (HSD)	Psychogenic (e.g., depression, marital discord leading to desire deficiency, performance anxiety leading to excitement inhibition) CNS disease (partial epilepsy, Parkinson's, poststroke, adrenoleukodystrophy) Androgen deficiency (primary or secondary), androgen resistance Drugs (antihypertensives, psychotropics, alcohol, narcotics, dopamine blockers, antiandrogens)
Compulsive sexual behaviors	Psychogenic (obsessive-compulsive sexuality, excessive sex-seeking in association with affective disorders, addictive sexuality, sex impulsivity)
Erectile dysfunction	Psychogenic Drugs (antihypertensives, anticholinergics, psychotropics, cigarette smoking, substance abuse) Systemic diseases (cardiac, hepatic, renal, pulmonary, cancer, metabolic, postorgan transplant, pelvic irradiation) Androgen deficiency (primary or secondary), androgen resistance, other endocrinopathies Vascular insufficiency (atherosclerosis, pelvic steal, penile Raynaud's, venous leakage) Neurological disorder (Parkinson's, Alzheimer's, Shy-Drager, encephalopathy, spinal cord or nerve injury) Penile disease (Peyronie's, priapism, phimosis, smooth muscle dysfunction, trauma)
Disorders of ejaculation	Psychogenic (neurotic personality, anxiety/depression, partner discord or other situational factors)
Premature ejaculation (primary or secondary)	Organic (increased central dopaminergic activity, increased penile sensitivity)
Absent or retarded emission	Sympathetic denervation (diabetes, surgical injury, irradiation) Drugs (sympatholytics, CNS depressants) Androgen deficiency (primary or secondary), androgen resistance
Postejaculation	Psychogenic
Orgasmic dysfunction	Drugs (selective serotonin reuptake inhibitors, tricyclic antidepressants, monoamine oxidase inhibitors, substance abuse) CNS disease (multiple sclerosis, Parkinson's, Huntington's chorea, lumbar sympathectomy) Psychogenic (performance anxiety, conditioning factors, fear of impregnation, hypoactive sexual desire)
Failure of detumescence	
Structural penile disease	Penile structural abnormalities (Peyronie's, Phimosis)
Priapism (primary or secondary)	Primary priapism: Idiopathic Priapism secondary to disease: Hematologic (sickle cell anemia, leukemia, multiple myeloma), infiltrative (Faber's disease, amyloidosis), inflammatory (tularemia, mumps), and neurologic disease, solid tumors, trauma Priapism secondary to drugs: Phenothiazines, trazodone, cocaine, intrapenile vasoactive injections

body fat, fatigue, and mood impairment. Androgen deficiency may be due to multiple causes, which can be generally divided into a central (hypothalamic, pituitary) and primary testicular disease. Only a small fraction of men with erectile dysfunction and low testosterone levels have been found to have space-occupying lesions of the hypothalamic-pituitary region. The vast majority of individuals with structural hypothalamic pituitary lesions have serum testosterone levels below 150 ng/dL and/or significantly elevated prolactin levels.

### Testosterone Treatment for Erectile Dysfunction

There have been numerous studies characterizing the dose-response relationship of testosterone treatment to improve sexual desire, motivation, and sexual performance. A positive response may be seen within 30 days of initiation of testosterone treatment. Positive responses are seen in treatment with a wide range of testosterone delivery systems including injections, oral agents, transdermal patches, and testosterone

gels. As suggested previously, the majority of elderly men have impaired penile vasodilation as the cause of their erectile dysfunction. A majority of these men respond to various penile vasodilatory agents, the most popular of which are phosphodiesterase 5-inhibitors. Studies have demonstrated that individuals who have androgen deficiency and impaired erectile function may benefit by the combined treatment of a phosphodiesterase-5 inhibitor and testosterone.

### See Also the Following Articles

Aging and the Male Reproductive System • Androgen Biosynthesis Inhibitors and Androgen Receptor Antagonists • Androgen Insensitivity Syndrome • Contraception, Male • Erectile Dysfunction • Gonadotropins and Testicular Function in Aging • Hormone Replacement Therapy, Male • Impotence and Aging • Spermatogenesis, Endocrine Control of

### Further Reading

- Cunningham, G. R., Hirshowitz, M., Korenman, S. G., and Karacan, I. (1990). Testosterone replacement therapy and sleep-related erections in hypogonadal men. *J. Clin. Endocrinol. Metab.* **70**, 792–797.
- Feldman, H. A., Longcope, C., Derby, C. A., Johannes, C. B., Arango, A. B., Coviello, A. D., Bremner, W. J., and McKinlay, J. B. (2002). Age trends in the level of serum testosterone and other hormones in middle aged men: Longitudinal results from the Massachusetts Male Aging Study. *J. Clin. Endocrinol. Metab.* **87**, 589–598.
- Goldstein, I., Lue, T. F., Padma-Nathan, H., Rosen, R. C., Steers, W. D., and Wicker, P. A. (1998). Oral sildenafil in the treatment of erectile dysfunction. Sildenafil Study Group. *N. Engl. J. Med.* **338**, 1397–1404.
- Harman, S. M., Mettler, E. J., Tobin, J. D., Pearson, J., and Blackman, M. R. (2001). Longitudinal effects of aging on serum total and free testosterone levels in healthy men. *J. Clin. Endocrinol.* **86**, 724–731.
- Johannes, C. B., Araujo, A. B., Feldman, H. A., Derby, C. A., Leinman, K. P., and McKinlay, J. B. (2000). Incidence of erectile dysfunction in men 40 to 69 years old: Longitudinal results from the Massachusetts Male Aging Study. *J. Urol.* **163**, 460–463.
- Kandeel, F. R., Koussa, V. K. T., and Swerdloff, R. S. (2001). Male sexual function and its disorders: Physiology, pathophysiology, clinical investigation and treatment. *Endocr. Rev.* **22**(3), 342–388.
- Kwan, M., Greenleaf, W. J., Mann, J., Crapo, L., and Davidson, J. M. (1983). The nature of androgen action on male sexuality: A combined laboratory-self report study on hypogonadal men. *J. Clin. Endocrinol. Metab.* **57**, 557–562.
- Laumann, E. O., Paik, A., and Rosen, R. C. (1999). Sexual dysfunction in the United States: Prevalence and predictors. *J. Am. Med. Assoc.* **281**, 537–544.
- Mooradian, A. D., Morley, J. E., and Korenman, S. G. (1987). Biological actions of androgens. *Endocr. Rev.* **8**, 1–28.
- Rajfer, J., Aronson, W. J., Bush, P. A., Dorey, F. J., and Ignarro, L. J. (1987). Nitric oxide as a mediator of relaxation of the corpus cavernosum in response to nonadrenergic, noncholinergic neurotransmission. *N. Engl. J. Med.* **326**, 90–94.
- Shabsigh, R. (2003). Hypogonadism and erectile dysfunction: The role for testosterone therapy. *Intl. J. Potency Res. Suppl.* **4**, S9–S13.
- Wang, C., Swerdloff, R. S., Iranmanesh, A., Dobs, A., Synder, P. J., Cunningham, G., Matsumoto, A. M., Weber, T., and Berman, N. (2000). Transdermal testosterone gel improves sexual function, mood, muscle strength, and body composition parameters in hypogonadal men. Testosterone Gel Study Group. *J. Clin. Endocrinol. Metab.* **85**, 2839–2853.

Table I Stages of Sexual Maturation (P1 to P5) at Puberty in Girls

Stage	Description
P1	Prepubertal stage
P2	Early development of subareolar breast bud; widening of areolae; small amounts of labial, pubic, and/or axillary hair may be present
P3	Increase in size of palpable breast tissue and areolae; increased amount of pubic hair and axillary hair; development of adult body odor
P4	Further increase in breast tissue and areolae that protude above breast level; adult amounts of pubic hair; acne
P5	Menarche; adult breast and areolar size; adult amount and distribution of pubic hair with extension to the upper thigh

In most developed countries, 50% of girls show minimal breast development by the age of 10.5 years (Table II). Very few of them experience breast budding before the age of 9 years and nearly all have some breast development by the age of 14 years. Menarche is present in 50% of girls at 12.9 years of age and is considered early if it occurs before 10 years of age. Almost all girls experience menarche by the age of 15 years. In Europe, precocious puberty is therefore

usually defined by the appearance of pubertal signs before the age of 8 years. In the United States, it has been suggested that precocious puberty should be defined as sexual maturation that is observed before the age of 7 years. Delayed puberty is defined by the absence of breast budding by the age of 14 years. Primary amenorrhea is considered to occur if menarche has not occurred by the age of 15 years. The size of the uterus is another way to examine how pubertal

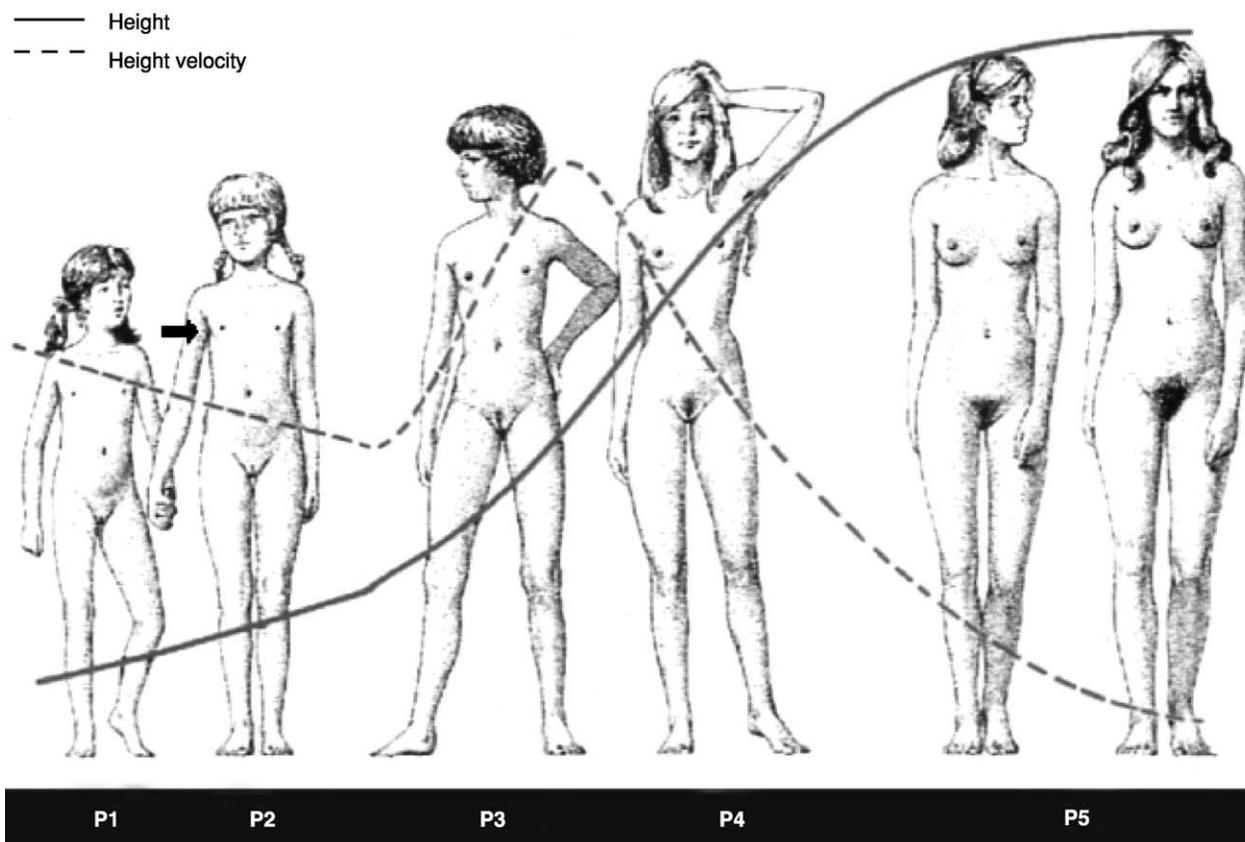


Figure 1 Stages of puberty and height increase and height velocity during sexual maturation. (P1) Absence of sexual maturation. (P2) First signs of sexual maturation, i.e., onset of puberty with the first breast budding (arrow). (P3, P4) Further development of the genitalia and development of secondary sexual characteristics, such as pubic and axillary hair. (P5) Menarche and adult status.

Table II Stages of Development of Breast and Peak Height Velocity in Girls in Relation to Chronological Age and Bone Age<sup>a</sup>

	Chronological age (years)	Bone age (years)
<b>Development of the breast</b>		
B1 Prepubertal		
B2 First budding enlargement of the areola	10.9 (8.5–13.3)	10.5 (8.5–13.2)
B3 Enlargement of the breast palpable mammary gland	12.2 (9.8–14.6)	12.0 (10.2–14.0)
B4 Additional enlargement of the breast and of the areola	13.2 (11.4–15.0)	13.5 (11.5–15.0)
B5 Adult breast	14.0 (11.6–16.4)	15.0 (12.5–16.0)
<b>Menarche</b>	12.9 (9.9–14.9)	12.8 (11.3–13.6)
<b>Peak height velocity</b>	12.2 (10.2–14.2)	12.4 (10.0–14.5)

Adapted from Sempé *et al.* (1996). Croissance et développement. In "Précis de Pédiatrie" (P. C. Sizonenko and C. Griscelli, eds.), pp. 64–65. Editions Payot, Lausanne, Switzerland, with permission.

<sup>a</sup>Chronological age and bone age are expressed as means and confidence limits (given in parentheses).

development is progressing. It requires an ultrasonographic examination of the pelvis, which is costly and perhaps not as precise as the assessment of breast size and other signs of sexual maturity. However, it can contribute significantly as it allows one to measure the ovaries and to observe their morphology. Multifollicular or microscopic cystic morphology is observed in 50% of ovaries in normal adolescents after the age of 9 years. Such morphology can last for several years after menarche, before the ovary becomes homogenous with the typical appearance of the dominant follicle.

## HORMONAL CHANGES

During pubertal development, the secretion of pituitary gonadotropins increases and, in response, the ovary will develop, with increasing secretion of estradiol (Fig. 2). The increase in plasma estradiol occurs between 10 and 12 years of age. A plasma concentration >25 pg/ml suggests pubertal ovarian secretion. However, marked variations in plasma estradiol levels are observed over 24h, even after menarche. Therefore, the measurement of plasma estradiol is not a reliable marker of the onset of puberty. The levels of plasma estrone, testosterone (Fig. 2), and adrenal androgens ( $\Delta^4$ -androstenedione and DHEA) increase during sexual maturation (Fig. 3). Plasma prolactin and inhibin concentrations also increase during pubertal development.

## GROWTH HORMONE AND GROWTH FACTORS

As observed mainly in boys, plasma growth hormone levels as measured over 24h indicate a rise in

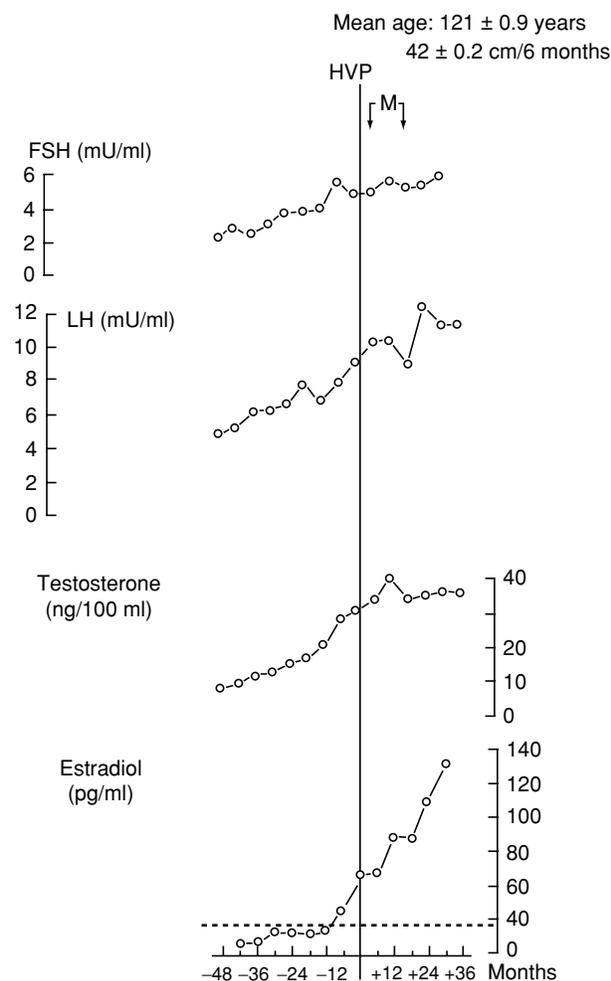
growth hormone (GH) secretion every 24h, mainly in relation to the increased secretion of sex steroids inducing an increase in GH pulse amplitude. Insulin-like growth factor-I (IGF-I) and insulin-like growth factor-binding protein (IGFBP-3) are secreted at an increased rate, in order to promote, together with GH, skeletal growth. Plasma levels of IGF-I and IGFBP-3 increase maximally at the time of the pubertal growth spurt (Figs. 4 and 5).

## NUTRITION AND BODY COMPOSITION

Nutrition is considered to be an important factor in the onset of puberty and menarche. In 1972, Frisch and Revelle showed that in order to experience menarche, the adolescent girl should achieve an ideal critical weight, which is  $47.8 \pm 0.51$  kg (mean  $\pm$  standard error of the mean). Although the concept still remains generally valid, there is a considerable variation within that critical weight. It depends on the adolescent height and should be adjusted for height. Fat, which secretes leptin, is considered to be one of the critical players in the onset of puberty and the occurrence of menarche.

## PATTERNS OF GROWTH DURING PUBERTY

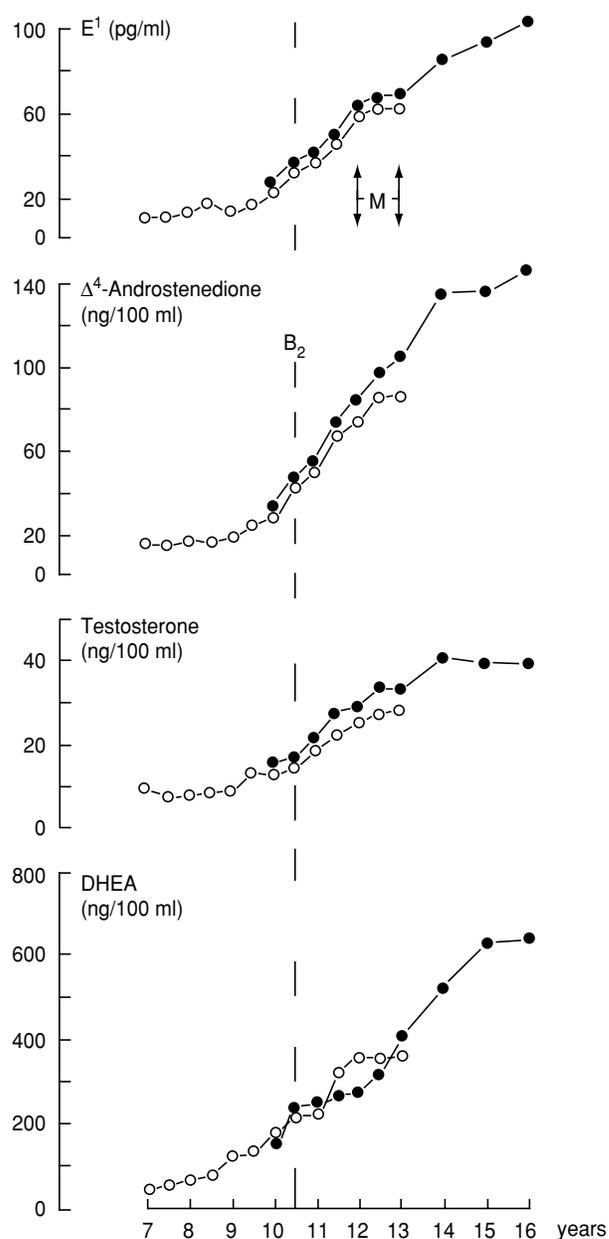
In a normal population of girls, peak height velocity in normal puberty is usually  $8.6 \pm 1.2$  cm/year at a mean chronological age of 12.0 years and at a mean bone age of 12.2 years (Table II). Remaining growth is approximately 4 to 5 cm during the 2 years following the occurrence of menarche.



**Figure 2** Changes in follicle-stimulating hormone (FSH), luteinizing hormone (LH), testosterone, and estradiol during female sexual maturation plotted at 6-month intervals before and after height velocity peak (HVP), which occurred at a mean age of 12 years and was 4.2 cm/6 months. Reprinted from Sizonenko and Aubert (1990), with permission.

## SECULAR TREND

Data obtained since the midnineteenth century confirm that there has been a decrease in the age of onset of puberty, mainly in developed countries, as observed by the age at which menarche occurs, ranging from 16–17 years in 1880 in Norway, Finland, and Sweden to 12–13 years in the past decade (Fig. 6). The median age for the onset of puberty (i.e., changes in the genitalia) was 10.0–10.5 years in the United States in 1960, not different from the median age of 10 years found in surveys performed in the past decade. No significant secular trend is observed in developed countries. Such changes have not been observed at all in European countries, such as the Netherlands, or



**Figure 3** Changes in plasma concentrations of estrone (E1),  $\Delta^4$ -androstenedione, testosterone, and DHEA during female sexual maturation, in girls with normal development [younger group (open circles) and older group (closed circles)]. Vertical interrupted line represents the onset of puberty (P2) for both normal groups. Both groups have menarche (M). The older group was followed until the end of puberty (stage P5). The younger group achieved midpubertal stage P3. Reprinted from Sizonenko and Aubert (1990), with permission.

else only small changes were observed, such as a shift of 0.6 years in Sweden between 1970 and 1980. The appearance of pubic hair was observed at a median age of 10.9 to 11.9 years between 1970 and 1980, decreasing to 11.3 years in the 1990s. Maturity and tempo of

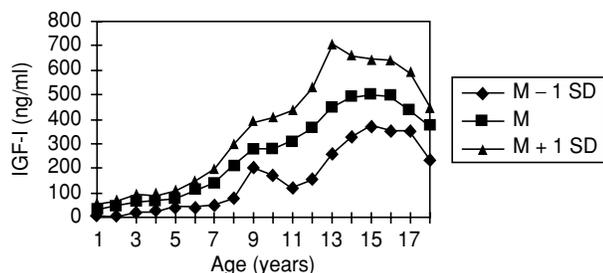


Figure 4 Mean ( $\pm 1$  SD) plasma IGF-I levels during female sexual maturation.

puberty appear to remain the same, suggesting that sexual maturation takes longer to be completed. This secular trend, which is so clearly shown in girls, has been attributed to the improvement in socioeconomic conditions (food, hygiene, better working conditions in lower social classes) in developed countries. As mentioned earlier, such a secular trend seems to disappear and, if present, remains very small and barely discernible.

## BONE MASS DENSITY

During puberty, there is a marked increase in the bone mass and in particular an increase in the bone mineral density, as measured by several methods. Several endocrine and growth factors, in addition to genetic factors and nutrition, in particular calcium intake, are responsible for bone mineral accretion: sex steroids (particularly estradiol), IGF-I, GH, and two growth factors, IGFBP-4 and IGFBP-5, as well as the hormones involved in calcium and bone metabolism: parathyroid hormone and vitamin D and its metabolites. In a group of female adolescents, followed longitudinally, bone mineral density increases markedly at the time of puberty. In girls, the increase in the bone mineral density of the lumbar vertebrae, L2 to L4, already begins at 11–12 years of age and is at a

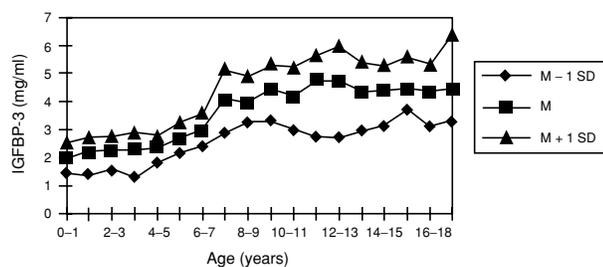


Figure 5 Mean ( $\pm 1$  SD) plasma IGFBP-3 levels during female sexual maturation.

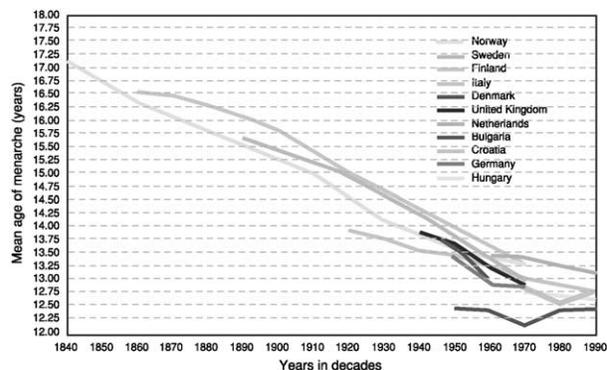


Figure 6 Secular trend of age of menarche in European countries since 1840 till 1990. Reproduced from Lee, P. A. (2002). *Top. Endocrinol.* 20, 7, © Chapterhouse Codex Ltd.

maximum between 12 and 14 years of age, at stage P3–P4. The increase in the bone mineral density of the femoral neck follows the same pattern. Bone mineral accretion decreases after the age of 15 years in girls and stops after the age of 18 years in girls. Therefore, during puberty, there is a considerable increase in the amount of minerals in the bone.

## CONCLUSION

The evaluation of pubertal development is purely clinical, by staging the maturation of the sexual characteristics. The early stage of puberty is defined in girls by the onset of breast budding. Menarche appears 2.5 to 3 years after the onset of puberty. A considerable gain in height and bone mineral density is observed during the same period.

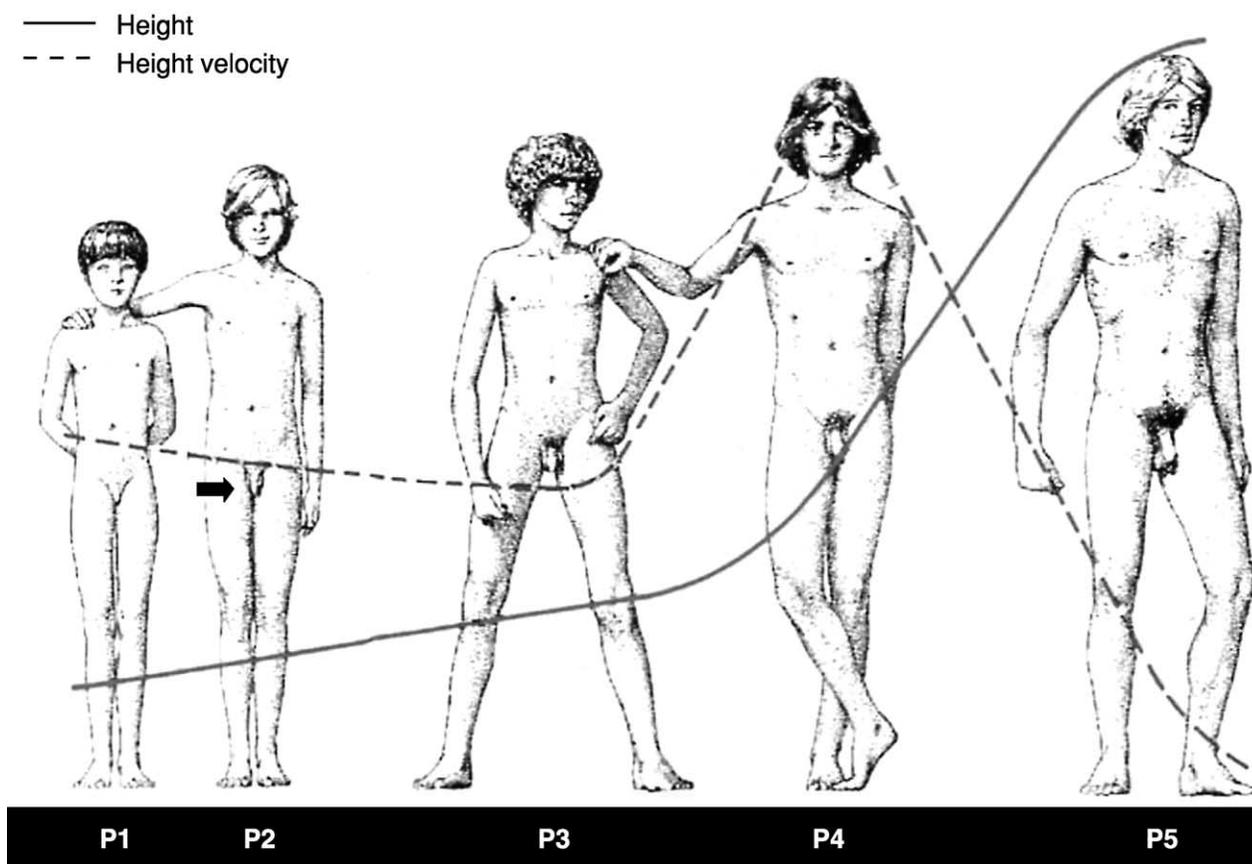
## See Also the Following Articles

Androgens, Gender and Brain Differentiation • Anti-Müllerian Hormone • Body Composition During Growth • Bone Mass Measurement • Eating Disorders and the Reproductive Axis • Growth Hormone (GH) • Insulin-like Growth Factors • Menstrual Cycle: An Integrative View • Prolactin (PRL) • Sexual Maturation, Male • Skeletal Development During Childhood and Adolescence

## Further Reading

- Blum, W. F. (1992). Insulin-like growth factors and their binding proteins. In "Functional Endocrinologic Diagnostics in Children and Adolescents" (M. B. Ranke, ed.), pp. 102–117. J & J Verlag, Mannheim, Germany.
- Bridges, N. A., *et al.* (1993). Standards for ovarian volume in childhood and puberty. *Fertil. Steril.* 60, 456–460.

- Forest, M. G. (1990). Pituitary gonadotrophin and sex steroids secretion during the first two years of life. *In* "Control of the Onset of Puberty" (M. M. Grumbach, P. C. Sizonenko, and M. L. Aubert, eds.), pp. 451–478. Williams & Wilkins, Baltimore, MD.
- Frisch, R. E., and Revelle, R. (1979). Height and weight at menarche and a hypothesis of critical weights and adolescent events. *Science* **169**, 397–399.
- Grumbach, M. M., and Kaplan, S. L. (1990). The neuroendocrinology of human puberty: An ontogenic perspective. *In* "Control of the Onset of Puberty" (M. M. Grumbach, P. C. Sizonenko, and M. L. Aubert, eds.), pp. 1–62. Williams & Wilkins, Baltimore, MD.
- Marshall, W. A., and Tanner, J. M. (1970). Variations in the pattern of pubertal changes in boys. *Arch. Dis. Childhood* **45**, 13–23.
- Mauras, N., Rogol, A. D., Haymond, M. W., and Veldhuis, J. D. (1996). Sex steroids, growth hormone, IGF-I: Neuroendocrine and metabolic regulation in puberty. *Horm. Res.* **45**, 74–80.
- Sizonenko, P. C. (1989). The adrenal gland and sexual maturation. *In* "Control of Onset of Puberty III" (H. A. Delemarre-van de Waal, T. M. Plant, G. P. van Rees, and J. Schoemaker, eds.), pp. 375–387. Elsevier Science B. V., Amsterdam.
- Sizonenko, P. C. (2002). Delayed puberty and hypogonadism. *In* "Oxford Textbook of Endocrinology and Diabetes" (J. A. H. Wass and S. M. Shalet, eds.), pp. 1035–1046. Oxford University Press, Oxford, UK.
- Sizonenko, P. C., and Aubert, M. L. (1990). Pituitary gonadotropins, prolactin and sex steroids: Secretion in prepuberty and puberty. *In* "Control of Onset of Puberty" (M. M. Grumbach, P. C. Sizonenko, and M. L. Aubert, eds.), pp. 479–505. Williams & Wilkins, Baltimore, MD.
- Theintz, G., Buchs, B., Rizzoli, R., *et al.* (1992). Longitudinal monitoring of bone mass accumulation in healthy adolescents: Evidence for a marked reduction after 16 years at the levels of lumbar spine and femoral neck in female subjects. *J. Clin. Endocrinol. Metab.* **75**, 1060–1065.



**Figure 1** Stages of puberty and height increase and height velocity during sexual maturation in boys. (P1) Absence of sexual maturation. (P2) First signs of sexual maturation, i.e., onset of puberty with the increase in testicular volume  $>4$  ml (or  $4$  cm<sup>2</sup>) (arrow). (P3, P4) Further development of the genitalia and development of secondary sexual characteristics, such as pubic and axillary hair and change in voice. (P5) Adult status.

Prader orchidometer or by the testicular volume index, which is expressed by the length  $\times$  width of the right side plus the length  $\times$  width of the left side, divided by 2. A testicular size of less than 4 ml ( $4$  cm<sup>2</sup>)

is prepubertal. A testicular volume of  $>4$  ml signifies pubertal development. The subsequent stages are marked by further increases in the size of the testes, by an increase in the size of the penis, and by the

**Table II** Stages of Development of Testicular Volume (Stages of Genitalia G1 to G5) and Peak Height Velocity in Boys in Relation to Chronological Age and Bone Age<sup>a</sup>

	Chronological age (years)	Bone age (years)
<i>Testicular volume<sup>b</sup></i>		
G1 prepubertal testes (TVI $<4$ ml)		
G2 pubertal testes (TVI between 4 and 5 ml)	11.2 (9.2–14.2)	11.5 (9.0–14.2)
G3 TVI between 7 and 11 ml	12.9 (10.5–15.4)	13.2 (10.5–15.0)
G4 TVI between 9 and 17 ml	13.8 (11.6–16.0)	14.5 (12.5–16.0)
G5 adult testes	14.7 (12.5–16.9)	
<i>Peak height velocity</i>	13.9 (12.3–15.5)	14.5 (12.5–16.0)

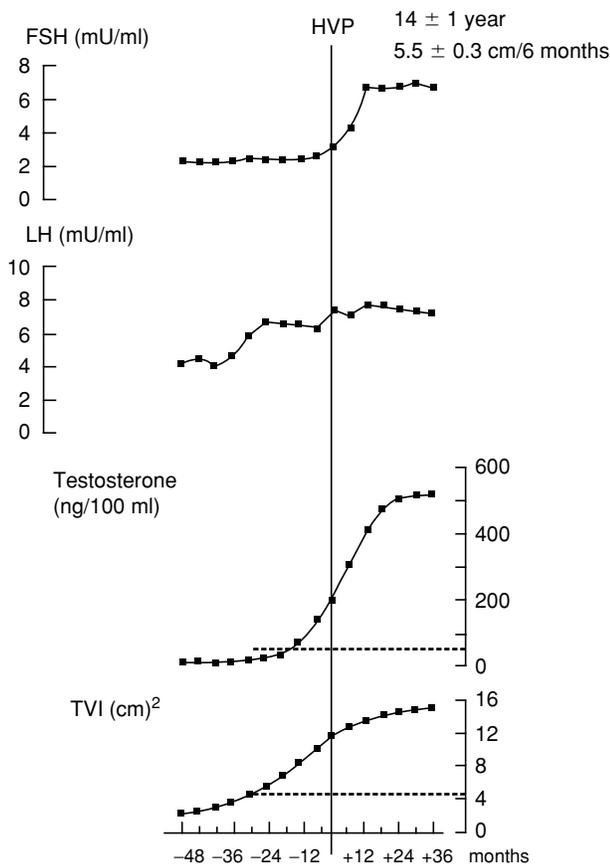
Adapted from Sempé *et al.* (1996). Croissance et développement. In "Précis de Pédiatrie" (P. C. Sizonenko and C. Griscelli, eds.), pp. 64–65. Editions Payot, Lausanne, Switzerland, with permission.

<sup>a</sup>Chronological age and bone age are expressed as means and 95% confidence limits (given in parentheses).

<sup>b</sup>Testicular volume index (TVI) is expressed by the length  $\times$  width of the right side plus the length  $\times$  width of the left side, divided by 2.

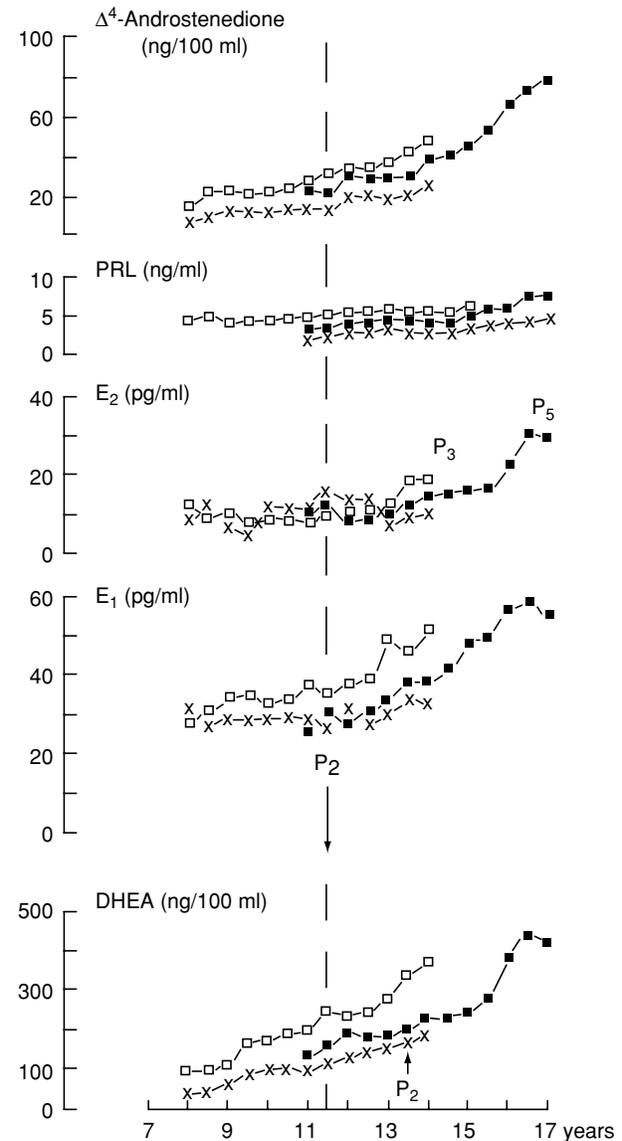
appearance of the pubic and axillary hair. Stage P3 is also characterized by an increase in height and growth velocity—the pubertal growth spurt, which is crucial to achieving normal final height (Fig. 1).

Fifty percent of boys experience the first increase in the size of the testes, to a volume >4ml, at the age of 11.5 years. Almost none of them have a testicular volume >4ml before the age of 9 years and almost all of them have passed stage P2 at age 14 (Table II). Precocious puberty can be defined as puberty occurring before the age of 9 years in boys and delayed puberty can be defined as puberty occurring after the age of 14 years. It is evident from the range in ages observed for each stage of pubertal maturation that individual variability in pubertal development is important. In a few boys, sperm were found in the urine by the age of 12, and in 50% of boys, sperm were found by 14.5 years of age.

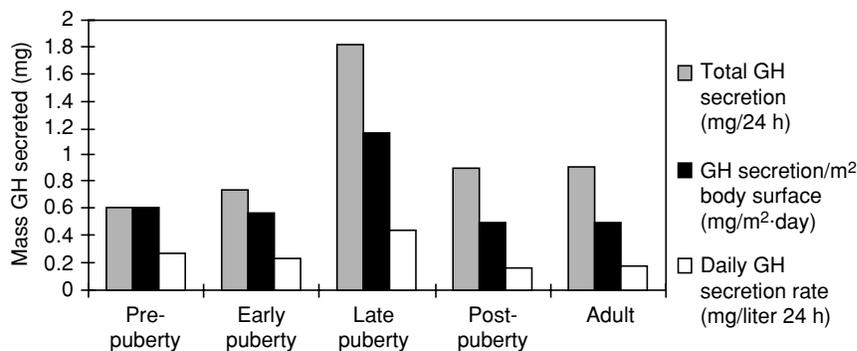


**Figure 2** Changes in follicle-stimulating hormone (FSH), luteinizing hormone (LH), and testosterone during male sexual maturation and testicular volume index (TVI; in cm<sup>2</sup>) plotted at 6-month intervals before and after height velocity peak (HVP), which occurred at a mean age of 14 years and was 5.5 cm/6 months. Reprinted from Sizonenko and Aubert (1990), with permission.

Testicular volume was approximately 8–15 ml. Both adrenal androgens and testosterone are responsible for skeletal growth, growth of the penis, scrotum, prostate, and seminal vesicles, the development of axillary and pubic hair, the change in voice, and increase in muscle strength.



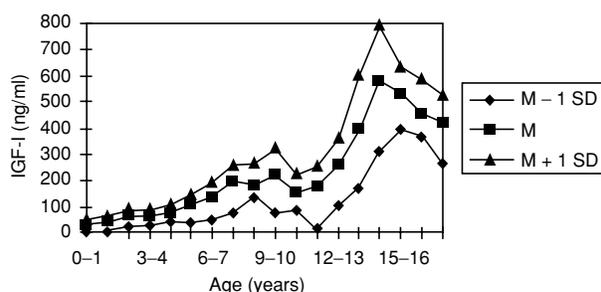
**Figure 3** Changes in plasma concentrations of Δ<sup>4</sup>-androstenedione, prolactin (PRL), estradiol (E<sub>2</sub>), estrone (E<sub>1</sub>), and dehydroepiandrosterone (DHEA) during male sexual maturation, in boys with normal [younger group (open squares) and older group (closed squares)] and delayed sexual maturation (×symbols). Vertical dashed line represents the onset of puberty (P2) for both normal groups. The older group was followed until the end of puberty (stage P5). The younger group achieved midpubertal stage P3. The onset of puberty of the delayed-maturation group was retarded by 2 years. Reprinted from Sizonenko and Aubert (1990), with permission.



**Figure 4** Growth hormone (GH) secretion during male sexual maturation. Reprinted from Sizonenko and Aubert (1990), with permission.

## HORMONAL CHANGES

The hormonal changes that occur at the onset of puberty are characterized by an increase in both of the gonadotropins, follicle-stimulating hormone and luteinizing hormone, as well as testosterone. Testosterone is an excellent parameter along with testicular size for the determination of the onset of puberty. Most boys who have started puberty have testosterone levels  $>0.25$  ng/ml (Fig. 2). Plasma adrenal androgens, prolactin, and estrogens increase during sexual maturation (Fig. 3). Plasma inhibin concentrations also increase during pubertal development. In addition, plasma growth hormone (GH) levels as measured over 24 h indicate an increase in GH secretion, mainly in relation to increased secretion of sex steroids, inducing an increase in GH pulse amplitude (Fig. 4). Secretions of insulin-like growth factor-I (IGF-I) and insulin-like growth factor-binding protein (IGFBP-3) increase markedly, in order to promote, together with GH, skeletal growth. Plasma levels of IGF-I and IGFBP-3 increase maximally at the time of the pubertal growth spurt (Figs. 5 and 6).



**Figure 5** Mean ( $\pm 1$  SD) plasma IGF-I levels during male sexual maturation.

Nutrition is considered to be an important factor for the onset of puberty. However, this factor has not been as well studied in boys as it has been in girls.

Peak height velocity occurs at approximately 14 years of chronological age and at 14.5 years of bone age. Mean peak height velocity is  $10 \pm 1.2$  cm (Table I).

## SECULAR TREND

Data obtained since the midnineteenth century confirm that there has been a decrease in the age of onset of puberty, mainly in developed countries, as observed by the age at which the change in voice in choirboys occurs. The median age of onset of puberty (i.e., changes in genitalia) was 11.5–12 years in the United States in 1960, which is different from the median age of 10 years that was observed in surveys performed in the past decade. Such changes have not been observed at all in European countries, such as the Netherlands, or else only small changes were observed, such as a shift of 0.6 years in Sweden between 1970 and 1980. The appearance of pubic hair was observed at a median age of 12.5 years between 1970 and 1980, decreasing to 11.3 years in the 1990s. The maturity and tempo of puberty appear to remain the same, suggesting that sexual maturation takes longer to be completed.

## BONE MASS DENSITY

During this period of rapid growth and sexual maturation, a marked increase in bone mass and in particular an increase in the bone mineral density, as measured by several methods, are observed. In addition to genetic factors and nutrition, in particular calcium intake, several endocrine and growth factors

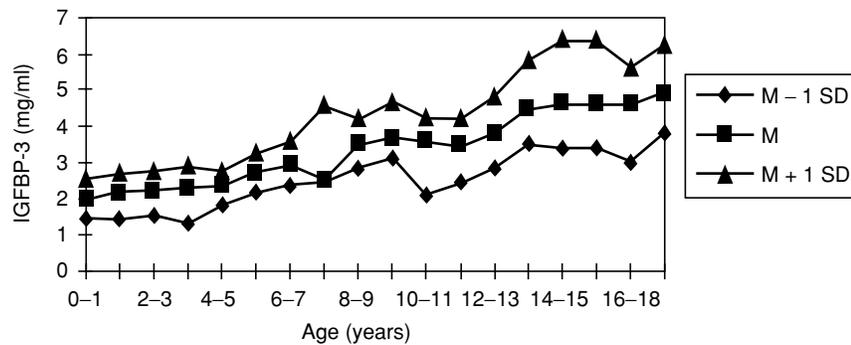


Figure 6 Mean ( $\pm 1$  SD) plasma IGFBP-3 levels during male sexual maturation.

are responsible for bone mineral accretion: sex steroids (testosterone and estradiol), IGF-I, GH, and two growth factors, IGFBP-4 and IGFBP-5, as well as the hormones involved in calcium and bone metabolism: parathyroid hormone and vitamin D and its metabolites. In boys, the increase in the bone mineral density of the lumbar vertebrae, L2 to L4, starts at 13–14 years of age and is at a maximum between 14 and 16 years of age, at stage P3–P4. The increase in the bone mineral density of the femoral neck follows the same pattern in both boys and girls. Bone mineral accretion decreases after the age of 16 years and stops after the age of 20 years. Therefore, during puberty, there is a considerable increase in the amount of minerals in the bone.

## CONCLUSION

The evaluation of pubertal development is purely clinical, by staging the maturation of the sexual characteristics. The early stage of puberty is defined by the increase in the size of the testes and also by plasma testosterone levels. A considerable gain in height and bone mineral density is observed during the same period. Finally, data fail to show a clear secular trend for an earlier onset of puberty during the past decades in developed countries.

## See Also the Following Articles

Androgens, Gender and Brain Differentiation • Anti-Müllerian Hormone • Bone Mass Measurement • Endocrine Disrupters and Male Sexual Differentiation • Germ Cell Differentiation Signaling Events, Male • Growth Hormone (GH) • Prolactin (PRL) • Sexual Maturation, Female • Skeletal Development During Childhood and Adolescence • Testes, Embryology of

## Further Reading

- Blum, W. F. (1992). Insulin-like growth factors and their binding proteins. *In* "Functional Endocrinologic Diagnostics in Children and Adolescents" (M. B. Ranke, ed.), pp. 102–117. J & J Verlag, Mannheim, Germany.
- Forest, M. G. (1990). Pituitary gonadotrophin and sex steroids secretion during the first two years of life. *In* "Control of the Onset of Puberty" (M. M. Grumbach, P. C. Sizonenko, and M. L. Aubert, eds.), pp. 451–478. Williams & Wilkins, Baltimore, MD.
- Grumbach, M. M., and Kaplan, S. L. (1990). The neuroendocrinology of human puberty: An ontogenic perspective. *In* "Control of the Onset of Puberty" (M. M. Grumbach, P. C. Sizonenko, and M. L. Aubert, eds.), pp. 1–62. Williams & Wilkins, Baltimore, MD.
- Marshall, W. A., and Tanner, J. M. (1970). Variations in the pattern of pubertal changes in boys. *Arch. Dis. Childhood* **45**, 13–23.
- Mauras, N., Rogol, A. D., Haymond, M. W., and Veldhuis, J. D. (1996). Sex steroids, growth hormone, IGF-I: Neuroendocrine and metabolic regulation in puberty. *Horm. Res.* **45**, 74–80.
- Schaefer, F., Marr, J., Seidel, C., *et al.* (1990). Assessment of gonadal maturation by evaluation of spermaturia. *Arch. Dis. Childhood* **65**, 1205–1207.
- Sizonenko, P. C. (1989). The adrenal gland and sexual maturation. *In* "Control of Onset of Puberty III" (H. A. Delemarre-van de Waal, T. M. Plant, G. P. van Rees, and J. Schoemaker, eds.), pp. 375–387. Elsevier Science B. V., Amsterdam.
- Sizonenko, P. C. (2002). Delayed puberty and hypogonadism. *In* "Oxford Textbook of Endocrinology and Diabetes" (J. A. H. Wass and S. M. Shalet, eds.), pp. 1035–1046. Oxford University Press, Oxford, UK.
- Sizonenko, P. C., and Aubert, M. L. (1990). Pituitary gonadotropins, prolactin and sex steroids: Secretion in prepuberty and puberty. *In* "Control of Onset of Puberty" (M. M. Grumbach, P. C. Sizonenko, and M. L. Aubert, eds.), pp. 479–505. Williams & Wilkins, Baltimore, MD.
- Theintz, G., Buchs, B., Rizzoli, R., *et al.* (1992). Longitudinal monitoring of bone mass accumulation in healthy adolescents: Evidence for a marked reduction after 16 years at the levels of lumbar spine and femoral neck in female subjects. *J. Clin. Endocrinol. Metab.* **75**, 1060–1065.
- Zachmann, M., Prader, A., Kind, H. P., *et al.* (1974). Testicular volume during adolescence. Cross-sectional and longitudinal studies. *Helv. Paediatr. Acta* **29**, 61–72.

**Table 1 Hemodynamic Patterns Determined by Pulmonary Artery Catheter Measurements in Specific Forms of Shock**

Type of shock	CVP and PCWP	CO	SVR	SvO <sub>2</sub>
Hypovolemic	↓	↓	↑	↓
Cardiogenic	↑	↓	↑	↓
Septic (early)	↑ or ↓	↑	↓	↑
Neurogenic	↓	↓	↓	↓

*Note.* CO, cardiac output; CVP, central venous pressure; PCWP, pulmonary capillary wedge pressure; SvO<sub>2</sub>, mixed venous saturation; SVR, systemic vascular resistance.

Inflammation plays an important role in the progression of shock. Shock and tissue injury lead to the activation of pro-inflammatory cells, such as neutrophils, macrophages, and platelets. Activation of these inflammatory cells leads to the release of inflammatory mediators, which govern virtually all manifestations of shock. These inflammatory mediators include chemokines/cytokines, eicosanoids, activated complement components, and coagulation cascades.

The primary pro-inflammatory cytokines include tumor necrosis factor  $\alpha$  (TNF $\alpha$ ) and interleukin-1 $\beta$  (IL-1 $\beta$ ) (Table 1). These molecules have potent biologic effects, inducing fever, anorexia, hypercortisolemia, neutrophilia, hypotension, and activation of the coagulation cascades and they greatly augment the entire immunoinflammatory process. Activation of complement components produces chemokines and activators of neutrophils and monocytes (C3a and C5a) and may also lead to the C5–C9 “attack complex,” causing further cytolysis. Activation of the coagulation cascade produces microvascular thrombosis, worsening cellular ischemia. Eicosanoids are vasoactive and immunomodulatory products of the arachidonic acid pathway released from the cell membrane. These include prostaglandins (PG) and thromboxane A<sub>2</sub>. PGI<sub>2</sub> and PGE<sub>2</sub> cause vasodilation, enhance capillary leakage, and contribute to edema. Thromboxane A<sub>2</sub> causes vasoconstriction, contributing to pulmonary hypertension and renal acute tubular necrosis. Nitric oxide produced by inducible nitric oxide synthase in endothelial cells and macrophages is released in excess, leading to further vascular dilation and permeability and secondary oxidant-induced bystander tissue injury.

## END-ORGAN EFFECTS

Shock has profound effects on every organ system in the body. However, this discussion is limited to

four specific areas that illustrate the interdependent physiology of shock.

## Neuroendocrine

Decreased blood volume is sensed by carotid and aortic arch baroreceptors and mechanoreceptors in the right atrium. This directly stimulates the sympathetic nervous system (SNS), which increases the heart rate, causes peripheral vasoconstriction, and increases pituitary release of adrenocorticotropic hormone (ACTH) and antidiuretic hormone (ADH). ACTH stimulates cortisol release from the adrenal cortex and ADH cause the kidneys to retain water. The SNS stimulates catecholamine release from the adrenal medulla, which increases heart rate and causes peripheral and mesenteric vasoconstriction to maintain mean arterial pressure (MAP) and conserve blood flow to the heart and brain. Hypotension also causes activation of the renin–angiotensin arc, the net effect of which is to maintain blood pressure through vasoconstriction and retain sodium and water by the kidneys. Increased cortisol leads to decreased glucose uptake in the tissues and enhanced lipolysis, increasing serum osmolality. These hormonal effects all work synergistically to maintain blood pressure.

## Cardiovascular

Cardiac function may be depressed in all forms of shock. Maintenance of blood pressure is dependent on three cardiac factors: preload, afterload, and contractility. Preload is the volume seen by the heart during ventricular filling, afterload is the resistance to ventricular ejection, and contractility is the ability of the myocardium to eject blood during systole. The combination of stroke volume and heart rate determines cardiac output, which is the major determinant of tissue perfusion. With decreased blood volume, preload and therefore stroke volume are decreased. The heart rate increases to maintain cardiac output but has a limited ability to compensate. Shock, through ischemia and reperfusion injury, often leads to increased myocardial interstitial fluid and decreased myocardial compliance, and a higher than normal filling pressure is required to attain an adequate end-diastolic volume to generate a normal stroke volume. In addition, the shock-related problems of acidemia, hypothermia, trauma, and sepsis may also depress myocardial contractility due to mitochondrial dysfunction and uncoupling of oxidative phosphorylation.

## Pulmonary

The pulmonary vasculature, like the systemic vasculature, generally vasoconstricts in shock. This increased pulmonary vascular resistance creates a relative hypoxemia and the patient becomes tachypneic to compensate. However, the rapid, shallow breathing increases both dead space and work of breathing through the increased minute ventilation. In addition, shock is a major cause of the acute respiratory distress syndrome or noncardiogenic pulmonary edema, which is likely secondary to an ischemia–reperfusion-induced systemic inflammatory injury. Histologically, there is diffuse alveolar epithelial and capillary endothelial damage, influx of neutrophils, and loss of surfactant. This leads to alveolar flooding, decreased lung compliance, alveolar collapse, and resistant hypoxemia.

## Renal

Oliguria is a frequent manifestation of shock. The decreased urine output is not simply due to hypoperfusion and dysfunction, but also a renal compensatory response to decreased blood flow (i.e., “renal success”). Increased afferent arteriolar vasoconstriction, stimulated by the renin–angiotensin arc, causes decreased glomerular filtration. Increased ADH and aldosterone increase urine concentration. The retention of sodium and distal tubular reabsorption of water produce concentrated urine with low sodium content and high osmolarity. However, if hypotension goes uncorrected, renal ischemic toxicity may occur. Necrosis of tubular epithelium, tubular obstruction

by debris, and depletion of renal ATP all lead to impaired renal function, manifesting as acute tubular necrosis (ATN). In addition, ATN can be seen with skeletal muscle trauma, with precipitation of muscle breakdown products, such as myoglobin, in the kidneys. Radiocontrast dye, aminoglycoside antibiotics, and other nephrotoxic agents may also lead to ATN in patients with shock.

## SPECIFIC TYPES OF SHOCK

Shock has classically been arbitrarily divided into four basic subgroups with somewhat unique hemodynamic profiles, recognizing that there is frequent overlap. Although debate continues on the therapeutic utility of pulmonary arterial catheterization, most critical care physicians use the data generated from these catheters to help both diagnose and treat the various types of shock (Table II).

### Hypovolemic

Hypovolemic shock occurs due to decreased intravascular volume. Most commonly, this is due to external or internal hemorrhage. Other sources of massive fluid loss include diarrhea, cholera, burns, or diffuse “capillary leakage” that occurs with pancreatitis or following severe trauma with ischemia–reperfusion-induced injury. Hypovolemia produces the classic hemodynamic picture of shock syndrome. It is characterized by decreased cardiac filling pressures and stroke volume with a compensatory increased heart rate and peripheral vascular resistance. Mild hypovolemia ( $\leq 20\%$  of the circulating blood volume) may be

Table II Sources and Effects of Selected Cytokines

Cytokine	Cellular source	Biologic effects
TNF $\alpha$	Macrophages	Fever, anorexia, increased ACTH, hypotension, hypercoagulability, others
Interleukin-1	Macrophages, others	Similar to TNF $\alpha$
Interleukin-6	Macrophages, others	Fever, promotes B cell maturation, stimulates adrenals
Interleukin-8	Macrophages, endothelial/epithelial cells	Chemokine for neutrophils
Prostanoids (PG, TX)	Macrophages, platelets, endothelial cells, others	Vasoactive agent (dilators and constrictors), permeability, thrombosis
Nitric oxide	Endothelial cells, macrophages, others	Vasodilator, excessive production causes oxidant-induced damage
Interferon- $\gamma$	TH1 cells	Activates macrophages, promotes TH1 differentiation
Granulocyte colony-stimulating factor	Macrophages, fibroblasts	Enhances bone marrow production of granulocytes

Note. ACTH, adrenocorticotropic hormone; PG, prostaglandin; TX, thromboxane; TH1 cells, T-helper 1 cells; TNF $\alpha$ , tumor necrosis factor  $\alpha$ .

well compensated by cellular and interstitial fluid shifts and remain virtually asymptomatic, particularly in a young, supine patient. Moderate hypovolemia (20–40% decrease in blood volume) is accompanied by increased tachycardia, postural hypotension, and anxiety, as homeostatic mechanisms begin to fail to maintain perfusion. Classic severe shock, including hypotension and other signs of end-organ hypoperfusion, appears with loss of 40% or more of the blood volume. Treatment includes rapid replacement of intravascular volume with isotonic crystalloid solutions and/or blood with simultaneous treatment of the underlying disease process.

### Cardiogenic

Patients in cardiogenic shock have a primary failure of myocardial contractility (i.e., “pump failure”). This most commonly is due to acute myocardial infarctions or chronic congestive heart failure and may manifest as right, left, or complete cardiac dysfunction. In these individuals, the filling pressures, as well as the systemic vascular resistance, will be high in an attempt to maintain cardiac output and MAP, which are both characteristically very low. The patient may show signs of intrinsic cardiac dysfunction, such as murmurs or arrhythmias, pulmonary edema (with left heart failure), and distended neck veins (with right heart failure). In contrast to all other forms of shock, neither absolute nor relative hypovolemia is present, but frequently, even in the presence of total body fluid overload, myocardial function will transiently improve with a small bolus of fluid until the diagnosis is confirmed. Treatment includes coronary revascularization, if possible, and cardiac support with inotropes, vasodilators, or intra-aortic balloon pump, as necessary.

### Septic

Septic shock is caused by the systemic response to severe infections. Although gram-negative rods are typically responsible, gram-positive cocci or rods, fungi, viruses, and other microorganisms may all cause shock. Bacterial endo- or exotoxins lead to the release of cytokines, including IL-1 $\beta$  and TNF $\alpha$ . Fibrin deposition increases, which, if unchecked, may lead to disseminated intravascular coagulation. Inducible nitric oxide synthase is stimulated, releasing nitric oxide, a powerful vasodilator and, in excess, a source of oxidant-induced injury. In addition, a massive capillary leak occurs with the progressive endothelial cell damage.

Sepsis classically creates a hyperdynamic picture. Even though there is splanchnic vasoconstriction and intestinal hypoperfusion, there is peripheral arterial vasodilation and cardiac output becomes supranormal. In contrast to other forms of shock, oxygen extraction by the tissues is reduced due to impairment of microcirculatory blood flow and mitochondrial oxidative metabolism. Diffuse microvascular injury and thrombosis lead to massive capillary leak and severe intravascular hypovolemia. In end-stage sepsis, however, patients become hypodynamic from organ failure, primarily cardiac. Abnormalities in myocardial contractility have been linked to endotoxins, which cause myocardial depression mediated by TNF and other potent inflammatory mediators. The treatment includes massive fluid repletion and hemodynamic support, along with prompt treatment of the underlying infection and adequate source control.

### Neurogenic

After trauma to the spinal cord, hypotension occurs due to the interruption of sympathetic vasomotor input with profound loss of venous and arterial vascular tone and pooling of blood peripherally. Because of the vasodilation, the extremities are typically warm and well perfused. There is no reflex-induced tachycardia. Treatment is directed toward repletion of the relative hypovolemia due to dilation of venous capacitance vessels and augmentation of vascular tone with pressors, including norepinephrine, if necessary, once hemorrhage is ruled out.

### Hypo-adrenal

Sudden loss of glucocorticoid production (which may be seen with adrenal hemorrhage associated with sepsis, malignancy, or anticoagulation) or a sudden, severe stress in individuals with chronic (due to pharmacologic corticosteroid therapy) or acute relative adrenal insufficiency may precipitate fulminant shock or hemodynamic instability. This is manifested by decreased systemic vascular resistance, refractory hypotension, and reduced cardiac output. Hypoadrenal shock often occurs concomitantly with other shock states, such as hypovolemia and sepsis, and diagnosis requires a high index of suspicion. The diagnosis is made with a short synthetic ACTH stimulation test. Treatment consists of fluid resuscitation, pressor support as needed, and intravenous steroids, 100 mg hydrocortisone every 6–8 h.

Several other types of shock have been described, including anaphylactic shock (similar to hyperdynamic sepsis with massive “capillary leak”) and extracardiac compressive shock (similar to cardiogenic shock and due primarily to pericardial tamponade, particularly acute, and tension pneumothorax from trauma). Each, however, has subtle, unique hemodynamic differences that require specific therapies. Interested readers should consult the reference materials at the end of this article.

## PROMISING THERAPIES AND THE FUTURE

Shock represents the clinical syndrome due to loss of homeostasis. Until the advent of critical care medicine with ventilator support, recognition and emergent surgical interventions, modern antimicrobial therapy, and intravenous resuscitation, shock was not survivable. The mortality of shock from all causes has been gradually decreasing, although it still approaches 50%. Advances in shock research have uncovered promising new therapies that have demonstrated significant survival advantages. For patients in shock, early goal-directed resuscitation targeting specific oxygen delivery end-points has been shown to improve survival. In addition, treatment with activated protein C, a molecule with anticoagulant and anti-inflammatory properties, has been shown to reduce mortality among patients in septic shock, although its applicability to all patients is still being tested. Among patients with cardiogenic shock, early angioplasty with revascularization and anti-thrombotic agents has been shown to significantly reduce mortality. The careful use of insulin to tightly control glucose levels and the avoidance of unnecessary blood transfusions have both led to a decrease in infectious

complications, organ dysfunction, and mortality. These and many other advances in shock research are continuing to improve outcomes for this once unsurvivable “unhinging of the basic machinery of life.”

## See Also the Following Articles

Cytokine Actions, Cellular Mechanism of • Cytokines, Extracellular Transport and Processing

## Further Reading

- Annane, D., *et al.* (2002). Effect of treatment with low doses of hydrocortisone and fludrocortisone on mortality in patients with septic shock. *J. Am. Med. Assoc.* **288**, 862–871.
- Bernard, G. R., *et al.* (2001). Efficacy and safety of recombinant human activated protein C for severe sepsis. *N. Engl. J. Med.* **344**, 699–762.
- Fink, M. P. (1997). The role of cytokines as mediators of the inflammatory response. In “Sabiston Textbook of Surgery” (C. M. Townsend, *et al.*, eds.), 15th ed., pp. 28–43. W. B. Saunders Co., Philadelphia, PA.
- Gattinoni, L., *et al.* (1995). A trial of goal-oriented hemodynamic therapy in critically ill patients. SvO<sub>2</sub> Collaborative Group. *N. Engl. J. Med.* **333**, 1025–1032.
- Jimenez, E. J. (1997). Shock. In “Critical Care” (J. M. Civetta, R. W. Taylor, and R. R. Kirby, eds.), 3rd ed., pp. 359–387. Lippincott-Raven, Philadelphia, PA.
- Maier, R. V. (2001). Shock. In “Harrison’s Principles of Internal Medicine” (Braunwald, Fauci, Kasper, Hauser, Longo, and Jameson, eds.), 15th ed., pp. 222–228. McGraw-Hill, New York.
- Menon, V., and Hochman, J. S. (2002). Management of cardiogenic shock complicating acute myocardial infarction. *Heart* **88**, 531–537.
- Rhodes, A., Cusack, R. J., Newman, P. J., Grounds, R. M., and Bennett, E. D. (2002). A randomized, controlled trial of the pulmonary artery catheter in critically ill patients. *Intensive Care Med.* **28**, 256–264.
- Rivers, E., *et al.* (2001). Early goal-directed therapy in the treatment of severe sepsis and septic shock. *N. Engl. J. Med.* **345**, 1368–1377.

are frequent and can be of great diversity by nature; they can also involve both the number (loss, multiplication) and the structure of chromosomes (duplication, inversion, translocation, ring formation). Such aberrations either occur during the development of an individual or are inherited from a parent. If the chromosomal aberration occurs at the level of oocyte-sperm fusion (zygote), all cells of the body will subsequently be affected (constitutional aberrations). If such an aberration occurs at a later stage of development, the aberration is passed on only to the cells that derive from the affected cell. Thus, the patient will have cells containing normal as well as abnormal chromosomes, an occurrence that is referred to as “mosaicism”; this indicates that not all chromosomal abnormalities are detectable if chromosomes from blood leukocytes are investigated. It is important to establish whether an aberration of the chromosomal constitution of an individual is causally related to a disorder, since some deviations are functionally irrelevant (e.g., balanced translocations). On the other hand, there are specific chromosomal abnormalities that lead to typical structural and functional disorders in an individual, the most prominent of which is impaired growth. Traditionally, evidence of chromosomal abnormalities is sought in children whose physical features are indicative of a congenital disorder or in whom abnormal structures or abnormal functioning of internal organs has been established. Chromosomal abnormalities occur more frequently as parental age increases; thus, if the expectant parents are older, fetal cells are harvested during early pregnancy for chromosomal analysis by means of amniocentesis or chorionic villous sampling. Genetic counseling for an abnormality discovered at the prenatal stage necessitates detailed knowledge on the lifelong development of affected individuals. The aim of this article is to provide a brief discussion of a few chromosomal abnormalities that present with a characteristic clinical pattern and short stature.

### Distinct Characteristics

Disorders caused by chromosomal abnormalities are commonly associated with four key features: (1) pre- and postnatal growth retardation, (2) mental retardation, (3) organ malformations, and (4) abnormal facial and body features. Chromosomal abnormalities occur when there is either an excess or a loss of a total chromosome or when a portion of a chromosome is absent (macrodeletions and microdeletions). Only a few of these chromosomal aberrations are compatible with pre- and postnatal life. For example, only the loss of a whole sex chromosome is compatible with life

(Turner’s syndrome), whereas the total loss of any other chromosome is not. Therefore, no information is available on long-term growth development in most disorders.

## DISORDERS INVOLVING AN EXCESS OF ONE CHROMOSOME

There are only a few disorders involving a complete trisomy of an autosome compatible with postnatal life: Down’s syndrome (trisomy 21), Patau’s syndrome (trisomy 13), and Edwards’ syndrome (trisomy 18). Information on the long-term development of children with trisomy 13 and 18 (see [Table I](#)) is limited because of their reduced life expectancy. No disease-specific growth curves exist yet and the genetic foundation of the growth disorder is unknown. The development of long-term survivors tends to follow an individual path and thus requires special attention to individual needs. Children with the rare case of trisomy 8 [deep set eyes, thick lips, inability to stretch fingers and toes (camptodactyly)] survive only in the case of mosaicism of the chromosomal abnormality; they may vary in stature from short to small. The same holds true for children with trisomy 9 (hearing defect, joint contractures, mental retardation, malformed ears).

### DOWN’S SYNDROME (TRISOMY 21)

The clinical appearance of Down’s syndrome was first described by the physician Langdon Down, for whom this chromosomal abnormality was named (see [Table II](#) and [Fig. 1](#)). After the technique of chromosomal analysis became available, it was observed that Down’s syndrome is caused by an additional chromosome 21 (three instead of the normal two). The extra chromosome is either in the cell in free form (95%) or is attached to another chromosome (translocation). Down’s syndrome is also the most familiar of chromosomal abnormalities, not least because of its relatively high frequency and the characteristic (mongoloid) appearance of the affected individual. Down’s syndrome is characterized mainly by a consistent clinical appearance, mental retardation, and short stature. The affected individuals are basically affectionate by nature and have a special sense for music: their life expectancy, however, is limited, particularly among those with a complex cardiac failure.

Chromosome 21 is a rather small chromosome, but the genetic basis of the disorder is not yet known in detail. Information is still being acquired on some critical regions of the chromosome that relate to

Table I Characteristics of Edwards' Syndrome and Patau's Syndrome

<b>Edwards' Syndrome</b>	
Chromosomal aberration	Trisomy of chromosome 18 (47,XY + 18)
Incidence (cases per number of newborns)	1 in 3000; male to female ratio 1:3
Characteristic clinical features	Polihydramnios, severe mental deficiency (seizures), small oral opening, clenched hand, index finger overlapping third, hypoplasia of nails, small nipples, herniation of abdominal wall, redundancy of skin, heart failure (VSD)
Growth pattern	Death is common during first year (90%) Small placenta, birth weight $\approx$ 2300 g at term, progressive short stature with an adult height of <150 cm
<b>Patau's Syndrome</b>	
Chromosomal aberration	Trisomy of chromosome 13 (47,XY + 13) Trisomy 13 mosaicism and partial trisomy (13pter to q14) are rare
Incidence (cases per number of newborns)	1 in 5000
Characteristic clinical features	Incomplete development of forebrain, severe mental retardation, seizures, microcephaly, hearing defect, cleft lip, polydactyly, cardiac failure (e.g., VSD [ventricular septal defect]), genital abnormalities in both sexes
Growth pattern	Death common before 6 months of age Small size at birth ( $\approx$ 2500 g), short stature in childhood, restricted information on adult height

specific features. In particular, there is no information on the cause of the growth disorder. Thus far, there is circumstantial evidence indicating that the growth hormone system plays no major role in the development of short stature, which suggests that the

principal defect is connected to bone growth. The typical growth pattern in children and adolescents with Down's syndrome is, however, well-described. The birth size of newborns is normal. From infancy onward, there is a progressive decrease in height

Table II Characteristics of Down's Syndrome

Chromosomal aberration	Trisomy of chromosome 21 (47,XY + 21); in rare cases (5%) translocations (partly inherited) or partial trisomy
Incidence (cases per number of newborns)	Depending on the age of mother (<35 years 1:1000; > 45 years 1:50)
Characteristic clinical features	"Mongoloid" position of the eyes Speckling of iris (Brushfield spots) Open mouth and protruding tongue Short hands with simian crease Wide gap between first and second toes Mental retardation (moderate to severe) Muscular hypotonia
Major associated abnormalities	Cardiac failure (AV canal, VSD, ASD) Hypogonadism Hypothyroidism
Growth pattern	Normal size at birth Progressive short stature; adult height $\approx$ 153 cm in males and $\approx$ 145 cm in females Microcephaly and obesity
Life expectancy	Reduced; with cardiac failure <30 years, without cardiac failure <50 years
Treatment (general)	Functional abnormalities (e.g., cardiac abnormalities and hypothyroidism are treated accordingly) -Patients need lifelong special care -Patients are sociable and musically inclined
Treatment (growth disorder)	None

Note: AV, atrioventricular; VSD, ventricular septal defect; ASD, atrioventricular septal defect.



**Figure 1** Girl with Down's syndrome as an infant (A) and at school age (B).

increment; the onset of puberty is delayed and a reduced pubertal growth spurt occurs. Female patients generally reach an adult height of approximately 140 cm and males attain an adult height of approximately 150 cm (Fig. 2). There is a tendency toward obesity. The discussion about improving height with growth hormone treatment—as for other chromosomally caused growth disorders—centers around the issues of the disease-specific efficacy, safety, and psychosocial implications. Since patients with Down's syndrome have a high risk of developing leukemia, there have been no conclusive long-term studies conducted on the efficacy and safety of growth hormone treatment for this diagnosis. In addition, the question needs to be asked whether a therapy-induced increase in height would improve the quality of life for the affected individuals.

#### DISORDERS INVOLVING A LOSS OF A PORTION OF THE CHROMOSOME

The loss of a total chromosome is not compatible with life. The loss of a sex chromosome (X or Y) is an exception. The most frequent and characteristic disorder is Turner's syndrome (TS), which involves the total or partial loss of an X chromosome (typically 45,X). This disorder occurs in approximately 1:2000 newborn females. The individuals are physically and emotionally female in gender and have a characteristic appearance; there is a malformation of internal organs (e.g., heart failure), absent ovaries (gonadal dysgenesis), and short stature. TS girls are characteristically short during childhood and their spontaneous final height is approximately 20 cm below the mean female height for a given population (equivalent to  $\approx 145$  cm). Spontaneous growth in Turner's syndrome has been well described in several populations. Effective treatment with growth hormone is possible. The growth disorder is caused by the loss of one SHOX gene (also known as "haploinsufficiency"), which is located at the tip of the short arm of the X (and Y) chromosome (Xp23) (the presence of two SHOX genes is required for normal growth). The SHOX genes are thought to play a major role in the prenatal development of bones. The selective loss of a SHOX gene can lead to disproportionate short stature (Léri-Weill syndrome). It is remarkable that Noonan's syndrome (incidence 1:1000), which greatly resembles Turner's syndrome (and can affect both sexes), is caused by a mutation of a gene on chromosome 12 (PTPN11 on 12q24) encoding a protein phosphatase. Growth in Noonan's

Table III Characteristics of Syndromes with a Partial Loss of Chromosomes

Name	Chromosomal defect	Symptoms
Wolf-Hirschhorn syndrome	Deletion at 4p (~20 Mb)	Microcephaly with frontal hemangioma, prominent nose, cleft lip/palate, coloboma of eye, heart failure, kidney malformations, severe mental retardation, severe intrauterine growth retardation
Cri-du-chat syndrome	Deletion at 5p (~20 Mb); incidence 1:20,000	Cat-like screaming during infancy, microcephaly, flat nasal bridge, downward slant of the palpebral fissures, short hands and feet, mental retardation, low birth weight, short stature
William's syndrome	Microdeletion 7q11 (~2 Mb); incidence 1:10,000	Supravalvular stenosis of the aortic stenosis, peripheral pulmonary arterial stenoses, perinatal hypercalcemia, typical facies (elfin face), mild mental retardation, short stature <sup>a</sup>
Prader-Willi syndrome	Microdeletion 15q11-q13 (paternal chromosome) (~4 Mb) in ≈70%	Muscular hypotonia in infancy, hyperphagia and obesity in later life, hypogenitalism, mental retardation, short stature
CATCH22 (cardiac defect, abnormal facies, zymic hypoplasia, cleft palate, hypocalcemia, 22q11 deletion)	Microdeletion 22q11 (~2 Mb)	Extremely varied picture; most frequent is the di George syndrome—immune deficiency due to thymic aplasia, hypocalcemia, cardiac failure, facial dysmorphia, stature varies

<sup>a</sup>See Fig. 3.

syndrome is also well described and females with this diagnosis are slightly taller than TS patients.

The partial loss or augmentation of chromosomal material may relate to a whole arm (p denotes “petite,” referring to the short arm; “q” is the next letter after “p” and refers to the long arm) or even smaller regions. The visibility of a microdeletion in chromosomal

material depends both on the size and on the methodology. It is possible to visualize a loss of chromosomal material equivalent to approximately 5 Mb. This is the magnitude available with FISH, a relatively new diagnostic tool. Using this method, the DNA that carries information on a specific chromosomal region is marked with a dye and is mixed with the patient's

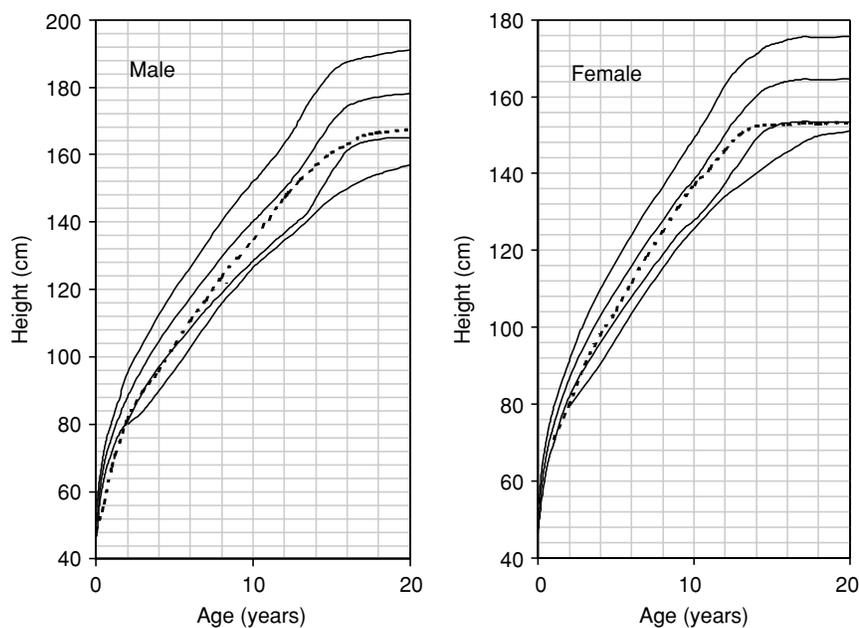


Figure 2 Mean height development of children with Down's syndrome (thick boldface line) and Williams' syndrome (dotted line).

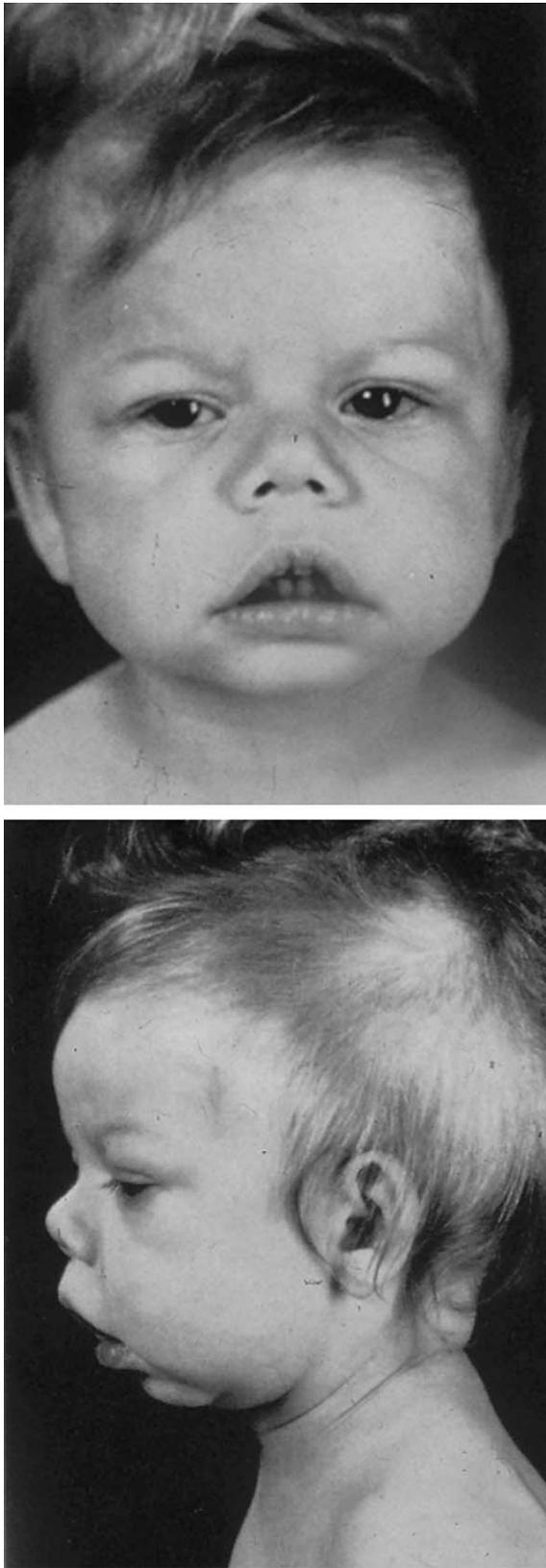


Figure 3 Infant with Williams' syndrome.

chromosomes on glass laboratory slides. If corresponding chromosomal material exists in the patient's DNA, a visible signal appears through the fluorescence microscope; if the signal does not appear, it can be assumed that the material is not present in the patient's DNA. Conversely, more than two signals from the normal two chromosomes are indicative of an excess of genetic material. The number of disorders in this category is extremely large, although the condition itself is rare. Some disorders are listed in Table III; see also Fig. 3.

Despite the fact that the human genome has been analyzed in totality, there is incomplete knowledge pertaining to the genetic code and its implications for the development and maintenance of structures and the functioning of the various parts of the human body. It is known that even in well-defined abnormalities of a single gene, the resulting functional deviation from normal may vary widely. This is particularly so in cases in which several genes are affected, such as in cases involving chromosomal abnormalities. Therefore, researchers need to refrain from regarding the affected individuals only from the perspective of chromosomes and, instead, study them as unique expressions of the variability of human creation.

### See Also the Following Articles

Growth, Normal Patterns and Constitutional Delay • Intrauterine Growth Retardation • Noonan Syndrome • Postnatal Normal Growth and Its Endocrine Regulation • Prader-Willi Syndrome • Puberty: Physical Activity and Growth • *SHOX* Disorders • Skeletal Development During Childhood and Adolescence • Turner Syndrome

### Further Reading

- Cronk, C., Crocker, A. C., Pueschel, S. M., Shea, A. M., Zackai, E., Pickens, G., and Reed, R. B. (1988). Growth charts for children with Down syndrome: 1 month to 18 years of age. *Pediatrics* **81**, 102–110.
- Gardner, R. J. M., and Sutherland, R. G. (1996). "Chromosome Abnormalities and Genetic Counselling." Oxford University Press, New York.
- Hall, J. G., Roster-Iskenius, U. G., and Allanson, J. E. (1989). "Handbook of Normal Physical Measurement." Oxford University Press, Oxford, UK/New York.
- Jones, K. L. (1997). "Recognizable Patterns of Human Malformation." W. B. Saunders, Philadelphia, PA.
- Partsch, C. J., Dreyer, G., Gosch, A., Winter, M., Schneppenheim, R., Wessel, A., and Pankau, R. (1999). Longitudinal evaluation of growth, puberty and bone maturation in children with Williams syndrome. *J. Pediatr.* **134**, 82–89.
- Ranke, M. B. (1996). Disease-specific standards in congenital syndromes. *Horm. Res.* **45**, 35–41.

chromosomes) play a key role in height determination. It is well established that complete or partial loss of the human X chromosome is directly implicated in the development of short stature together with a number of other skeletal abnormalities and somatic stigmata observed in Turner syndrome. Genotype–phenotype correlations suggested the major pseudo-autosomal region (PAR1), located at the tip of the short arm of the X and Y chromosomes, as a candidate region for the presence of a growth-determining genetic locus. Since genes residing within PAR1 escape X inactivation, two active copies of these loci are required for normal function. Consequently, the loss of a single copy of such genes leads to a lower dose of functional gene product, a condition known as haploinsufficiency.

In 1997, two independent research teams identified a candidate gene for short stature within PAR1 from patients with short stature who had rearrangements encompassing the Xp22 and Yq11 regions of the sex chromosomes. As a consequence of its location within the pseudo-autosomal region, this gene—designated short stature homeobox-containing gene or *SHOX*—was proposed to also be responsible for the short stature phenotype associated with Turner syndrome. According to its chromosomal localization, the *SHOX* gene has two active copies in normal individuals, and a defect in one copy caused by point mutations, deletions, or chromosomal rearrangements results in *SHOX* haploinsufficiency. A number of studies have shown that this *SHOX* haploinsufficiency not only leads to short stature, but is also involved in a number of skeletal abnormalities and other congenital problems that are seen in a variety of clinical conditions that include idiopathic short stature, Léri-Weill dyschondrosteosis [LWD, or Léri-Weill syndrome (LWS)], Langer mesomelic dwarfism, and Turner syndrome.

The underlying and consistent phenotypic feature of each of these disorders is impaired growth. However, the extent of short stature in these different entities is variable and may also vary between affected individuals within the same family. Léri-Weill dyschondrosteosis, Langer mesomelic dysplasia, and Turner syndrome are also characterized by a range of skeletal abnormalities, again of variable expression, which start to develop during embryogenesis and continue to progress during adolescence. From this correlation, and from the fact that the expression of *SHOX* is primarily found in the midportion of the limbs (i.e., the radius/ulna and the tibia/fibula) as well as the first and second pharyngeal arches, it has

been proposed that the *SHOX* gene acts as a key regulator of skeletal development within these regions.

Accurate identification of defects in the *SHOX* gene is possible thanks to the growing understanding of the molecular biology of this gene. It is possible to screen large numbers of patients for deletions and base pair mutations that result in clinical phenotypes. This will allow the physician to offer patients a more accurate diagnosis and will inform both the patient and physician about the likely abnormalities that may occur over time in these syndromes. As a consequence, appropriate therapeutic interventions that may enhance the quality of life of individuals with these disorders can be undertaken.

## THE *SHOX* GENE

The existence of a critical gene involved in growth determination, which resides in the pseudo-autosomal region, was proposed after deletions of the short arm of the X chromosome and small terminal deletions of the short arm of the Y chromosome were shown to be consistently associated with short stature. Genes residing in the PAR escape X inactivation, and so, for normal function, both copies of the gene need to be expressed. Therefore, it was proposed that these short stature phenotypes arose as a consequence of haploinsufficiency of the critical short stature gene.

Significant efforts were made to identify and characterize the critical gene. Mapping studies narrowed the short stature critical interval to a 170 kb region residing 500 kb from the Xp telomere and allowed the molecular cloning of a single gene, designated *SHOX*. Shortly thereafter, another team independently reported the isolation of the identical gene from a YAC contig of this chromosomal region.

*SHOX* encodes a homeodomain protein that has been shown to act as a transcription factor. A variety of studies have indicated that the *SHOX* product is likely to be a key regulator of growth. The gene spans approximately 40 kb of the PAR and consists of seven exons that encode two alternatively spliced transcripts, *SHOXa* and *SHOXb*. These are translated to produce two isoforms of 292 and 225 amino acids residues, respectively. The gene was shown to be expressed from both the active and the inactive X chromosomes, as is expected based on its pseudo-autosomal location. Expression studies show that the two isoforms differ in their expression patterns.

*SHOXa* is widely expressed in different tissues, whereas the *SHOXb* isoform has a more restricted expression pattern and is found at its highest levels in bone marrow fibroblasts, providing additional evidence for its role in linear growth. Rao *et al.* have suggested that *SHOXb* may act as a transcriptional modulator of *SHOXa*.

Furthermore, *in situ* hybridization studies of *SHOX* gene expression during human embryogenesis have shown that expression occurs in the midlimb regions (i.e., the radius, ulna, elbow, wrist, tibia, fibula, knee, and ankle), which is in agreement with both the short stature phenotype and the typical skeletal dysmorphisms that are observed in patients with *SHOX* gene defects and disorders such as Turner syndrome.

### SHOX GENE DEFICIENCY—CLINICAL CONSEQUENCES

The phenotypes associated with *SHOX* gene deficiency vary widely. Although short stature is the underlying phenotypic feature, *SHOX* haploinsufficiency is also associated with significant skeletal abnormalities as seen in Léri-Weill dyschondrosteosis and Turner syndrome. The complete absence of any *SHOX* gene product is the basis for the severe short stature and dysmorphic features that are seen in Langer mesomelic dysplasia. In contrast, females with an extra X chromosome (47,XXX) have three copies of the *SHOX* gene and are taller than normal 46,XX females. Thus, there exists a continuum of growth defects that relates directly to the dose of functional *SHOX* gene product.

Therefore, *SHOX* deficiencies are correlated with a variety of different skeletal phenotypes and clinical conditions with a significant degree of clinical and molecular overlap complicating the differential diagnosis.

#### Léri-Weill Dyschondrosteosis

LWD, or LWS, is characterized by disproportionate short stature with shortening of the middle segments of the long bones of the arm and leg, more usually referred to as mesomelic short stature, and was first recognized by Léri and Weill in 1929. Like the other *SHOX* disorders, short stature in LWS is variable, with affected individuals ranging in adult height from 135 cm to normal height for their age and gender. The primary skeletal abnormality, the sign that permits the clinical diagnosis of LWS, is Madelung deformity.

This is a characteristic deformity of the forearm, consisting of shortening and bowing of the radius and distal dislocation of the ulna, leading to limited mobility of the wrist and the elbow. Madelung wrist deformity likely originates with disorganized growth of part of the radial epiphysis, leading to radial bowing, premature fusion of that epiphysis, dorsal dislocation of the ulna, and wedged carpal bones.

Another feature of LWS is the striking preponderance of affected females, with four times as many females affected as males. However, this predominance may be the result of a bias of ascertainment, as females appear to be more severely affected than males and thus may come to a clinician's attention more readily. However, since the location of the *SHOX* gene has been unambiguously determined, the inheritance can be classified as pseudo-autosomal dominant.

#### Langer Mesomelic Dysplasia

Langer mesomelic dysplasia was described by Langer in 1967. This disorder is thought to be the homozygous form of Léri-Weill dyschondrosteosis. Evidence in favor of this proposal came from an observation of a male newborn diagnosed with Langer mesomelic dysplasia, whose parents and maternal family exhibited Madelung deformity and mesomelic dwarfism. Complete absence of the *SHOX* gene has been demonstrated in a fetus with Langer mesomelia.

The clinical features of this disorder are severe. Most patients exhibit a much more extreme reduction in stature than in LWS, with severe mesomelic shortening of the limbs. The limb malformation is severe hypoplasia of the ulna and/or fibula with a severely shortened radius and tibia. Other features include Madelung deformity and hypoplasia of the mandible, or micrognathia. The main radiologic findings observed in Langer patients include severely shortened long bones of the limbs, Madelung deformity, varus deformity of the humeral head, angulation of the radial shaft, distortion of the carpals, short femoral neck, early fusion of tibial epiphyses, hypoplastic or absent proximal half of fibula, and lumbar lordosis.

#### Turner Syndrome

Turner syndrome (monosomy of the X chromosome, 45,X/partial monosomy of the X) is the most common sex chromosome abnormality known in females. It is associated with a high degree of embryonic lethality and thus fewer cases than expected are observed clinically. In fact, approximately 3% of all female conceptions have X chromosomal abnormalities typical of

Turner syndrome but, as a consequence of this high rate of embryonic lethality, only 1:2500 of live female births are diagnosed with the disorder. All affected individuals are female and the most common, invariable physical features that are observed are short stature and ovarian failure. The syndrome is also characterized by a variety of major and minor skeletal malformations as well as heart and renal abnormalities.

In patients with Turner syndrome, the incidence of short stature and ovarian failure is nearly 100%. Skeletal anomalies appear at a lower frequency—approximately 35–60% for abnormalities such as short fourth metacarpals, cubitus valgus, genu valgum, high-arched palate, scoliosis, micrognathia, and short neck. Madelung deformity and mesomelic limb shortening, which characterize LWS, also occasionally occur in Turner syndrome.

### Idiopathic Short Stature

The term “idiopathic short stature” represents a heterogeneous group of patients who have short stature for different unknown reasons, but not due to any of the syndromes described above. Patients are generally physically normal and routine laboratory tests also appear normal.

Diagnostic endocrinological tests do not normally reveal any abnormalities relating to short stature and there is, e.g., no evidence of growth hormone deficiency. However, patients’ heights are below the normal adult range. By definition, 3% of the general population fall below the third percentile and are therefore described as short. The majority of these patients are defined as having idiopathic short stature.

Careful review of a child’s history and physical exam may reveal clues to the presence of an unsuspected condition or disease, but the underlying basis for retarded growth remains unknown—hence the use of the term “idiopathic” to describe this condition. Rappold *et al.* have described *SHOX* gene mutations in 2.4% of children with idiopathic short stature.

### CLINICAL BENEFITS OF IDENTIFYING *SHOX* DEFICIENCY

There are a number of benefits that can be realized from early identification of defects in the *SHOX* gene. With the advent of molecular techniques, screening for *SHOX* defects has become fairly routine and large numbers of samples can be tested to permit early identification of mutations. The main benefits are outlined below.

### Identification of *SHOX* Defects

Identification of *SHOX* defects can be helpful, informative, and educational in helping patients understand the basis of their condition. This may resolve patients’ anxiety regarding the etiology of their condition and may provide a greater understanding of the development of their disease over time. In addition, genetic counseling can help individuals understand the probability of having affected children in the future.

### Identification of Contiguous Gene Disorders

Large-scale *SHOX* deletions that extend beyond the boundary of the pseudo-autosomal region on the X chromosome may express themselves as contiguous gene deletion syndromes. Therefore, it is important for the physician to be aware of the extent of the deletion in order to assess the likelihood that such disorders will develop. Depending on the size of the deletion, the X recessive form of chondrodysplasia punctata, mental retardation, ichthyosis, and Kallmann syndrome can be expected in boys with a complete Xp22.3 deletion.

### Genetic Counseling

Although Léri-Weill syndrome may be a debilitating condition, patients can, with appropriate therapy, live normal lives. Genetic counseling may provide useful advice in terms of the probability of a child having LWS. For example, if either parent has LWS, there is a 50% likelihood that any given male or female offspring will have LWS. Both the knowledge of the inheritance of *SHOX* mutations and the use of techniques to identify such defects are extremely important in cases in which both parents have *SHOX* defects. In these cases, offspring have a 25% chance of developing the more severe disorder, Langer mesomelic dysplasia. The ability to identify such defects prenatally together with ultrasound techniques to identify early skeletal abnormalities will allow parents the opportunity to make informed decisions.

### Surgical/Clinical Implications

Identification of a *SHOX* defect should alert the physician to monitor the development of skeletal abnormalities such as Madelung deformity. In general, skeletal abnormalities tend to be more severe in females and increase in severity with increasing age, particularly during puberty. Therefore, a physician

may intervene at an early stage with appropriate surgery to minimize the effects of the expected skeletal abnormality. Disorders such as Turner syndrome are associated with a variety of somatic stigmata that also require effective monitoring and clinical management.

### Differential Diagnosis

Molecular analysis of *SHOX* is a potentially important and effective tool in differentiating between dyschondrosteosis and other types of skeletal dysplasia associated with disproportionate short stature, including hypochondroplasia. This is of particular value when used in less severe cases and in early childhood, where clinical signs may not be readily apparent.

### Opportunities for Therapy

Patients with Turner syndrome are indicated for growth hormone therapy. Anecdotal reports have demonstrated beneficial effects of treating *SHOX*-deficient patients with growth hormone.

### Future Therapeutic Opportunities

In the future, identification of *SHOX* mutations and other defects, together with identification of target genes and proteins that *SHOX* acts on, may permit the development of novel therapeutic approaches that minimize the effects of such defects. For example, in haploinsufficiency of *SHOX*, a potential therapeutic objective would be to up-regulate the remaining copy of the gene at an early stage in development.

### See Also the Following Articles

Growth Hormone (GH) • Pituitary Gland: Growth and Growth Failure • Short Stature and Chromosomal Abnormalities • Turner Syndrome

### Further Reading

- Belin, V., Cusin, V., Viot, G., *et al.* (1998). *SHOX* mutations in dyschondrosteosis (Léri-Weill syndrome). *Nat. Genet.* **19**, 67–69.
- Binder, G., Schwarze, C. P., and Ranke, M. B. (2000). Identification of short stature caused by *SHOX* defects and therapeutic effect of recombinant human growth hormone. *J. Clin. Endocrinol. Metab.* **85**, 245–249.
- Blaschke, R. J., Monaghan, A. P., Schiller, S., *et al.* (1998). *SHOT*, a *SHOX*-related homeobox gene, is implicated in craniofacial, brain, heart, and limb development. *Proc. Natl. Acad. Sci. USA* **95**, 2406–2411.
- Clement-Jones, M., Schiller, S., Rao, E., *et al.* (2000). The short stature homeobox gene *SHOX* is involved in skeletal abnormalities in Turner syndrome. *Hum. Mol. Genet.* **9**, 695–702.
- Ellison, J. W., Wardak, Z., Young, M. F., Robey, P. G., Webster, M., and Chiong, W. (1997). *PHOG*, a candidate gene for involvement in the short stature of Turner syndrome. *Hum. Mol. Genet.* **6**, 1341–1347.
- Fryns, J. P., and Van den Berghe, H. (1979). Langer type of mesomelic dwarfism as the possible homozygous expression of dyschondrosteosis. *Hum. Genet.* **46**, 21–27.
- Kosho, T., Muroya, K., Nagai, T., *et al.* (1999). Skeletal features and growth patterns in 14 patients with haploinsufficiency of *SHOX*: Implications for the development of Turner syndrome. *J. Clin. Endocrinol. Metab.* **84**, 4613–4621.
- Langer, L. O., Jr. (1967). Mesomelic dwarfism of the hypoplastic ulna, fibula, mandible type. *Radiology* **89**, 654–660.
- Rao, E., Blaschke, R. J., Marchini, A., Niesler, B., Burnett, M., and Rappold, G. A. (2001). The Léri-Weill and Turner syndrome homeobox gene *SHOX* encodes a cell-type specific transcriptional activator. *Hum. Mol. Genet.* **10**, 3083–3091.
- Rao, E., Weiss, B., Fukami, M., *et al.* (1997). Pseudoautosomal deletions encompassing a novel homeobox gene cause growth failure in idiopathic short stature and Turner syndrome. *Nat. Genet.* **16**, 54–63.
- Rappold, G. A., Fukami, M., Niesler, B., *et al.* (2002). Deletions of the homeobox gene *SHOX* are an important cause of growth failure in children with short stature. *J. Clin. Endocrinol. Metab.* **87**, 1384–1388.
- Roth, P., Agnani, G., Arbez-Gindre, F., Maillet, R., and Colette, C. (1996). Langer mesomelic dwarfism: Ultrasonographic diagnosis of two cases in early mid-trimester. *Prenat. Diagn.* **16**, 247–251.
- Shears, D. J., Vassal, H. J., Goodman, F. R., Palmer, R. W., Reardon, W., Superti-Furga, A., Scambler, P. J., and Winter, R. M. (1998). Mutation and deletion of the pseudoautosomal gene *SHOX* cause Léri-Weill dyschondrosteosis. *Nat. Genet.* **19**, 70–73.



model in position and topography the shape of the bone. Depending on the location of these condensations, the cells differentiate into osteogenic cells to form bone directly (intramembranous, or membrane bones) or these mesenchymal condensations differentiate into chondrocytes to form bone indirectly (endochondral bone).

Mesenchymal condensations in the region of the future clavicle, most of the mandible, and most of the skull form bone directly. The other bones form indirectly via replacement of a cartilage model or anlage. Membrane bones enlarge only by appositional growth. Endochondral bones, however, can also grow interstitially by taking advantage of this property of cartilage in their growth plates. Synchondroses mediate growth in the skull base and are essentially double-ended growth plates. Connections between bones depend on their location and their functions. Those between two membrane bones occur at sutures, which permit little movement, but isolated membrane bones (clavicle and mandible) connect to other bones via other types of joints that allow considerable movement.

## FORMING A SKELETON: MECHANISMS AND REGULATION

The skeleton is the result of a series of inductive interactions between cell populations that elaborate matrices that control their differentiation. The initiation and coordination of skeletal growth are accomplished by regulating the temporospatial expression of many genes. Among those known to participate are transcription factors, extracellular matrix molecules, growth factors, cell adhesion molecules, and membrane receptors. Many of these genes were first discovered in *Drosophila melanogaster* and their unusual names reflect this heritage. What is known about each stage of skeletal development, including the events themselves, the genes known to be involved, and their interactions, is discussed next.

### Pattern Formation

The position of future skeletal elements is first indicated in the mesenchyme of the embryo by aggregates of cells at specific sites. This mesenchymal pattern thus establishes the location of the bone to be formed. Pattern formation proceeds differentially in a cranio-caudal direction and has a distinct dorso-ventral axis, but generally exhibits bilateral symmetry. The skeleton consists of axial and appendicular components. The axial components are derived from mesenchyme in the somites (sclerotome) postcranially and from

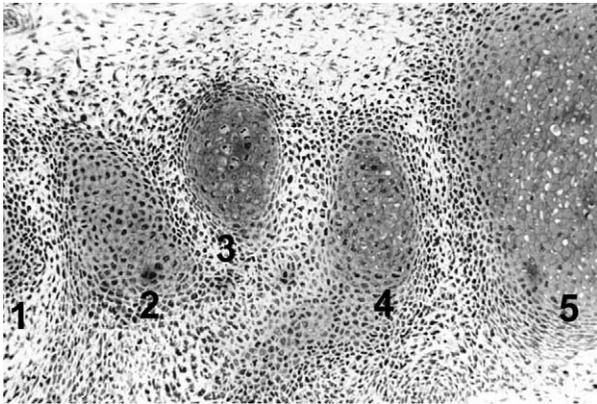
neural crest cells in the craniofacial skeleton. The appendicular skeleton is derived from mesenchyme in the lateral plates of the body wall.

*Hox* gene expression determines developmental events in the cranio-caudal axis. These transcription factors are found in four gene clusters, for a total of 39 genes in vertebrates, and they specify regional skeletal development by regulating the local expression of other genes. Similarly, members of the *Pax* gene family regulate development in the dorso-ventral (anterior-posterior) axis. Ventral signaling is derived from the notochord, especially cranially, and involves sonic hedgehog (*Shh*) and *Noggin*, an antagonist of bone morphogenetic protein (BMP). Dorsal signaling is derived from the dorsal neural tube and surface ectoderm and involves the *Wnt* (*Wingless*) family. The appendicular skeleton arises as limb buds from the lateral plate mesenchyme and develops in a complex pattern along three axes. The proximo-distal axis is established by orchestrating the participation of a variety of transcription factors, growth factors, and adhesion molecules. The initial outgrowth, the apical ectodermal ridge (AER), is dependent on the fibroblast growth factor (FGF) family and their receptors (FGFRs), whose expression patterns dictate the position and eventual size of the limb. The cranio-caudal limb axis is specified by *Shh* gene expression in a small patch of mesoderm at the caudal junction of the AER and the body wall. Dorso-ventral patterning is under the control of *Wnt* genes, specifically *Wnt7a*.

Developmental abnormalities in skeletal patterning result in a variety of skeletal aplasias and dysplasias.

### Mesenchymal Condensations

Mesenchymal condensations establish the position and shape of the eventual skeletal component. They are initiated by epithelial-mesenchymal interactions that up-regulate the production of cell adhesion molecules, which, together with extracellular matrix proteins, cause cell aggregations. Proliferation, adherence, and differentiation in these mesenchymal condensations set them apart from adjacent cell populations. Cells in these condensations eventually differentiate into either bone cells (osteoblasts) or cartilage cells (chondrocytes). Where a series of bones will appear, such as in the limbs, less condensed areas demarcate the positions of joints. This pattern of mesenchymal condensations for the metatarsal bones can be seen in Fig. 2. These bones form by the endochondral pathway and the central cells have already advanced to the next stage, differentiating into chondrocytes and secreting cartilage matrix, which is visualized by its



**Figure 2** Mesenchymal condensations as illustrated by the developing metatarsals of a rat fetus. Metatarsals (numbered) show the progressive development of mesenchymal condensations, including transformation into chondrocytes (darkly staining matrix-secreting cells) and hypertrophy (visible in the center of metatarsal 3).

metachromatic staining in this preparation. Note that the intervening areas between metatarsals are less differentiated and occupy the sites of the eventual attachments between these bones.

Genes regulating mesenchymal condensations include *Shh*, FGFs, BMP2–5 and 7, and other members of the transforming growth factor- $\beta$  (TGF- $\beta$ ) superfamily. They program the site-specific expression of adhesion molecules and extracellular matrices and regulate the proliferation and differentiation in these condensations. Expression of *Sox9* and *Cbfa1* occurs in all mesenchymal condensations as a prelude to their differentiation along the endochondral or intramembranous pathways.

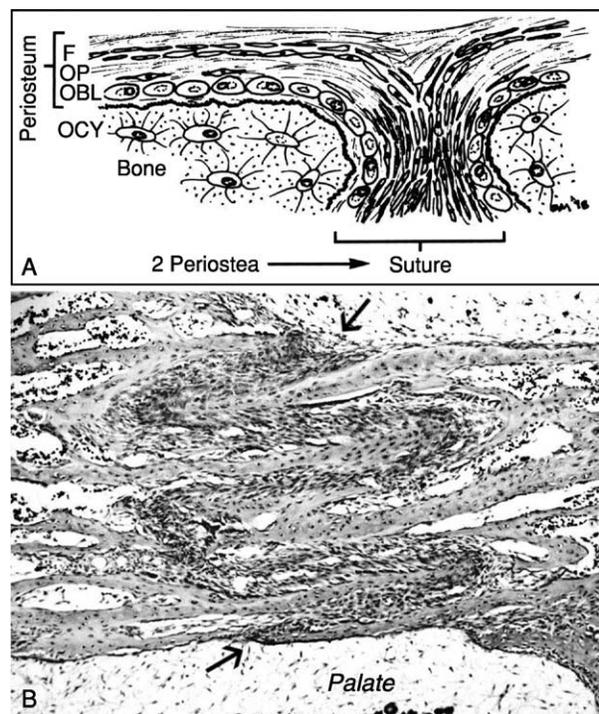
Developmental disturbances of mesenchymal condensations lead to a variety of skeletal dysplasias, including craniofacial defects, chondrodysplasias, and osteodystrophies.

### Direct Bone—The Intramembranous Pathway

Once the cells in the initial condensation of mesenchymal cells differentiate into osteoblasts, they form bone as flat plates. Some of the osteoblasts become surrounded by matrix and become osteocytes. The osteocyte/osteoblast network is connected by canaliculi that maintain metabolic communications between all cells. When these metabolic units reach the limits of diffusion, vascularized Haversian systems (osteons) form to allow them to remain connected metabolically (this process also occurs in endochondral bones). All membrane bones eventually develop marrow spaces by resorption internally. In the flat bones of the

skull, the marrow spaces are small, discrete areas between the internal and external cortical plates. In the mandible and clavicle, these spaces are larger. Adjacent membrane bones form a specialized interface called a suture, a site where the periosteae of two membrane bones come in contact. This is shown in [Fig. 3](#). Note that cells in the center of sutures contain the progenitors of osteoblasts for adjacent bone surfaces and that proliferation and differentiation among these cells is the mechanism by which membrane bones grow at their edges. As bones grow in thickness, sutures become more complex with numerous, complex interdigitations. These are most obvious in the membrane bones of the skull that must grow rapidly to accommodate the early expansion of the brain.

Bone formation is dependent on expression of *Cbfa1* (*Am13* or *Runx2*), a member of the Runt-domain family of transcription factors. *Cbfa1* is expressed by all osteoblasts and its expression is necessary for their



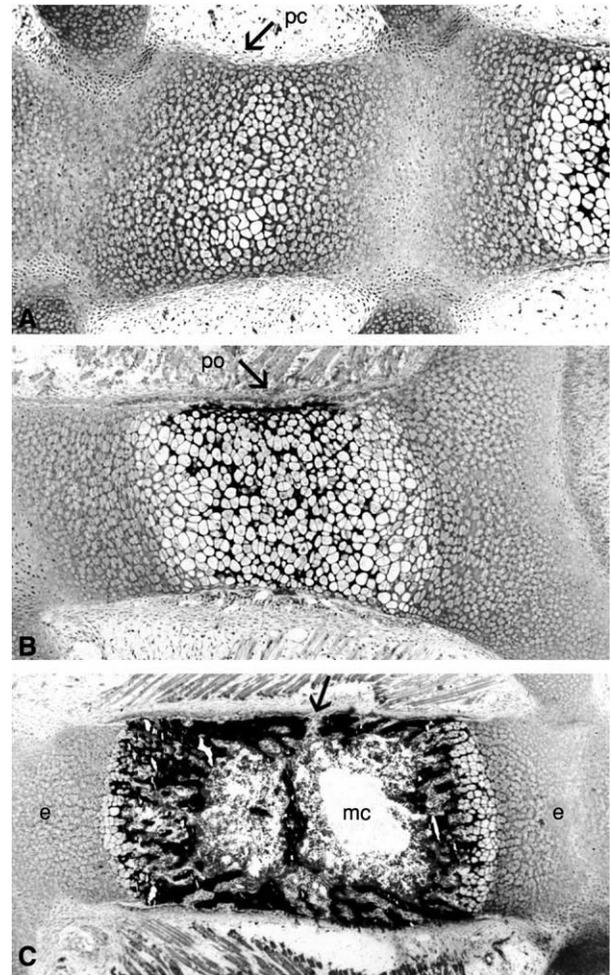
**Figure 3** Cellular organization in sutures. (A) Cellular relationships illustrate that sutures are composed of the apposed periosteae of adjacent bones with a central proliferative region blending into the osteoprogenitors and osteoblasts on bone surfaces. F, fibroblast; OP, osteoprogenitor cell; OBL, osteoblast; OCY, osteocyte. Reprinted from Marks *et al.* (1999). *Dev. Dyn.* **215**, 117–125, with permission. (B) Histological section of the suture between the maxilla (left) and the premaxilla (right) in a 10-day-old rat. The densely cellular suture (arrows) oscillates between the two bone surfaces, which are highly interdigitated during this period of maximal growth.

differentiation. Its absence is lethal—no osteoblasts develop, no bone is formed, and the rib cage collapses. Haploinsufficiency of *Cbfa1* results in cleidocranial dysplasia.

Growth at sutures is regulated by growth factors and transcription factors. Signaling through fibroblast growth factor receptor 1 (FGFR1) induces osteogenesis by up-regulating *Cbfa1* and type I collagen formation, whereas signaling through FGFR2 promotes stem cell proliferation in the center of the suture. During growth, these are balanced. The cessation of sutural growth and closure of sutures is promoted by BMP4, BMP7, FGF9, TWIST (a basic helix-loop-helix transcription factor), MSX1 (a *Hox* family member), and MSX2, which upset this balance in favor of osteogenesis, leading to the obliteration of most sutures. Premature closure leads to a variety of congenital craniosynostoses.

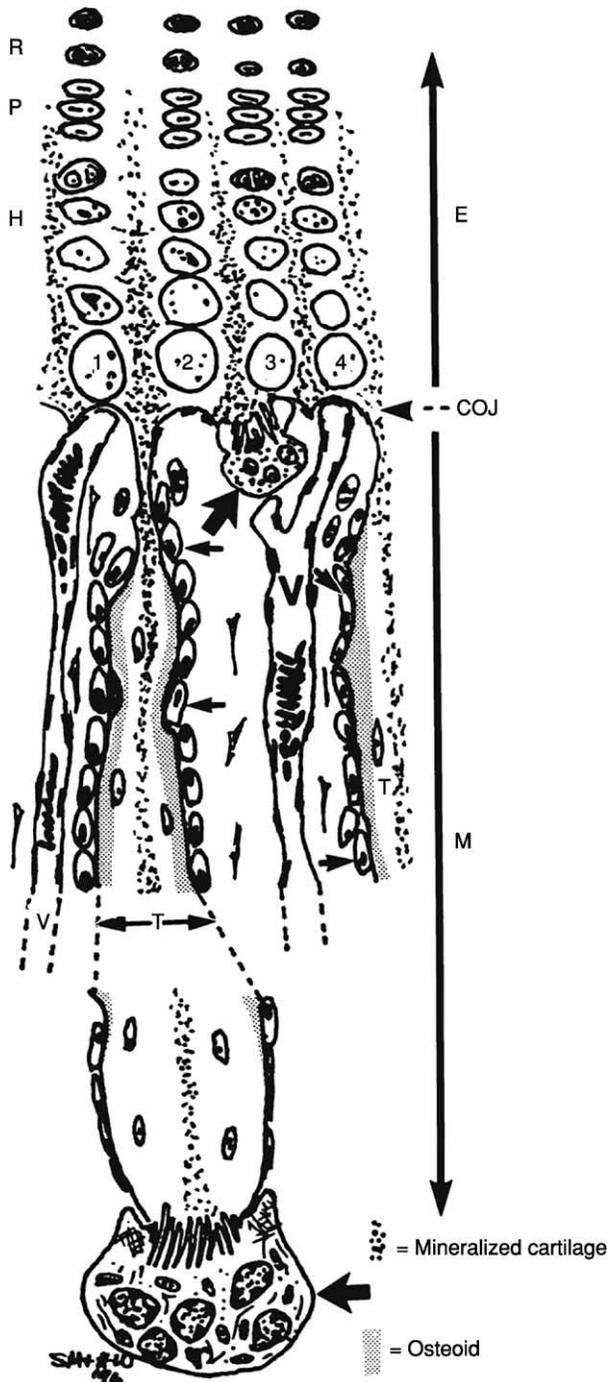
### Indirect Bone—The Endochondral Pathway

The differentiation of cells in mesenchymal condensations into chondrocytes marks the beginning of the endochondral (“inside cartilage”) pathway, which gives rise to the axial and appendicular skeletal elements, and can be seen in Fig. 2. Subsequent stages in the endochondral pathway are shown in Fig. 4. After formation of the miniature cartilage model, or “anlage,” of the future bone, endochondral bones can grow both externally, by appositional growth via the surrounding sheath of cells called the perichondrium, and—unique to endochondral bones—internally via proliferation and hypertrophy of the chondrocytes and the expansion of their secreted matrix, which consists mainly of type II collagen. Cells in the center of these chondrogenic islands enlarge (hypertrophy) and begin to mineralize the adjacent matrix (dark central areas of Fig. 4B). Soon thereafter, the cells surrounding this mineralized central region differentiate into osteoblasts, thus forming a periosteum, and they secrete type I collagen-containing bone matrix that becomes mineralized, forming a “bony collar” that surrounds the central portion of what will be the bone shaft. The periosteum becomes vascularized and its vessels invade the mineralized cartilage interior through an opening that eventually becomes the nutrient foramen of the bone (Fig. 4C). The oldest (central) hypertrophied chondrocytes begin to die as vessels invade, bringing with them osteoclast precursors that resorb approximately two-thirds of the mineralized matrix, producing the earliest marrow spaces. Marrow stromal



**Figure 4** Stages of endochondral bone formation as illustrated by segments of the rat sternum at birth. Hypertrophy of chondrocytes in the center of the cartilage model (A) is followed by mineralization of cartilage matrix and transformation of the central perichondrium (pc in A) into a vascularized periosteum (po in B), which forms a collar of mineralized bone (B). Vessels in the periosteum invade (arrow in C) the mineralized cartilage model, bringing osteoclast precursors, which differentiate and resorb the mineral, and hematopoietic stem cells, which populate the newly formed marrow cavity (mc). Osteoblasts also invade and form bone on longitudinally oriented rods of mineralized cartilage. This produces a central marrow space surrounded by a mineralized cylinder (diaphysis) with mineralized cartilage and bone at each end (dark areas in C). This leaves cartilage islands, epiphyses (e), with regions of proliferation and hypertrophy, in their bony ends.

cells occupy this space and hematopoietic stem cells then migrate from the spleen. Osteoblasts begin to secrete bone matrix onto the mineralized cartilage septa that remain. Thus, a disc of bone is formed and divides the cartilage model in half between its proximal and distal ends. The transition point between cartilage and bone is called the chondroosseous junction.



**Figure 5** Schematic of cellular location and activities at the chondroosseous junction (COJ) of a growing bone. The epiphyseal growth plate (E) consists of resting (R), proliferating (P), and hypertrophied (H) chondrocytes. In the metaphysis (M), trabeculae (T) alternate longitudinally with vascular channels (V). Osteoblasts (small arrows) line trabecular surfaces beginning just below the COJ and osteoclasts (large arrows) are found in two locations: at the COJ and at the marrow cavity ends of trabeculae. Chondrocytes are aligned in columns (numbered) and their alignments are maintained by mineralization of the longitudinal matrix between columns

Growth in length in endochondral bones is accomplished by continuing proliferation and hypertrophy of the chondrocytes on either end of the central bony region. The process of hypertrophy results in a four- to sixfold increase in cell volume and drives the lengthening of the bone. The chondrocytes align themselves along the (weight-bearing) long axis of the bone and differentiate directionally to form columns of proliferating and hypertrophied cells. In the zone closest to the chondroosseous junction, the longitudinal septa of cartilage (i.e., between columns) become mineralized, whereas the transverse septa (i.e., between cells in the same column) do not. On reaching hypertrophy, chondrocytes' energy metabolism switches to glycolysis, and when their abundant stores of glycogen are exhausted, they undergo programmed cell death, the final metabolic state of the last cell in each column. The last transverse septum is resorbed by specialized perivascular cells, clearing the way for vascular invasion and transition to bone.

Approximately two-thirds of the longitudinal septa are resorbed by osteoclasts and the remainder are covered by bone matrix to form bony trabeculae. These trabeculae inhabit the region where the bone narrows, or flares, from the rounded ends to the shaft. The trabeculae are in turn resorbed from their ends as the bone grows, leaving an open marrow space surrounded by the dense bone of the shaft. The flared area containing the trabeculae is called the metaphysis and the central shaft is the diaphysis.

The rounded ends of the cartilage bone model, the epiphyses, enlarge for a time as cartilaginous structures, but eventually they too undergo a wave of vascular invasion and replacement by bone, forming a secondary ossification center at each end of the bone. Between the bony interior of the epiphysis and the metaphysis lies a discoid layer of growing cartilage called the physis, or the epiphyseal growth plate. The growth plate is the site at which the continued lengthening of the bone occurs, again via chondrocyte proliferation and hypertrophy. As the bone grows,

beginning in the zone of proliferation. Osteoclasts resorb two-thirds of these mineralized cartilage struts at the COJ to make room for vascular channels (V). The vascular channels and adjacent cells resorb the horizontal septum between the last hypertrophied chondrocyte and the metaphysis (illustrated for columns 1 and 4). Osteoblasts form bone on the remaining mineralized cartilage struts at the COJ and the bone content of trabeculae increases with time. At the marrow ends of trabeculae, osteoclasts resorb both cartilage and bone. Reprinted from Marks, S. C. (1998). In "Diagnostic Imaging in Child Abuse" (P. K. Kleinman, ed.), pp. 2-7, Mosby, St. Louis, MO, with permission.

chondrocytes at the chondroosseous junction die and are replaced by bone and the lengthening trabecula are resorbed at their ends in a closely regulated manner that is proportional to the length of the bone. The result is that the metaphysis occupies a nearly constant fraction of a bone's length. These events are shown diagrammatically in Fig. 5 and in a photomicrograph of a rat chondroosseous junction in Fig. 6.

The conversion of the epiphysis from cartilage to bone is initiated by the formation of canals from the surface, which are invasions by vascular structures that bring skeletal cells and/or their precursors with them. The beginning of this process is shown in Figs. 7 and 8.

It should be noted that the cartilaginous epiphysis consists of chondrocytes that have divided only once, producing lacunae occupied by two cells. The canals enter the epiphysis and expand and their interiors become ossified. The enzymes involved have been identified and are shown in Fig. 9, where they are compared and contrasted with similar resorptive events at the chondroosseous junction.

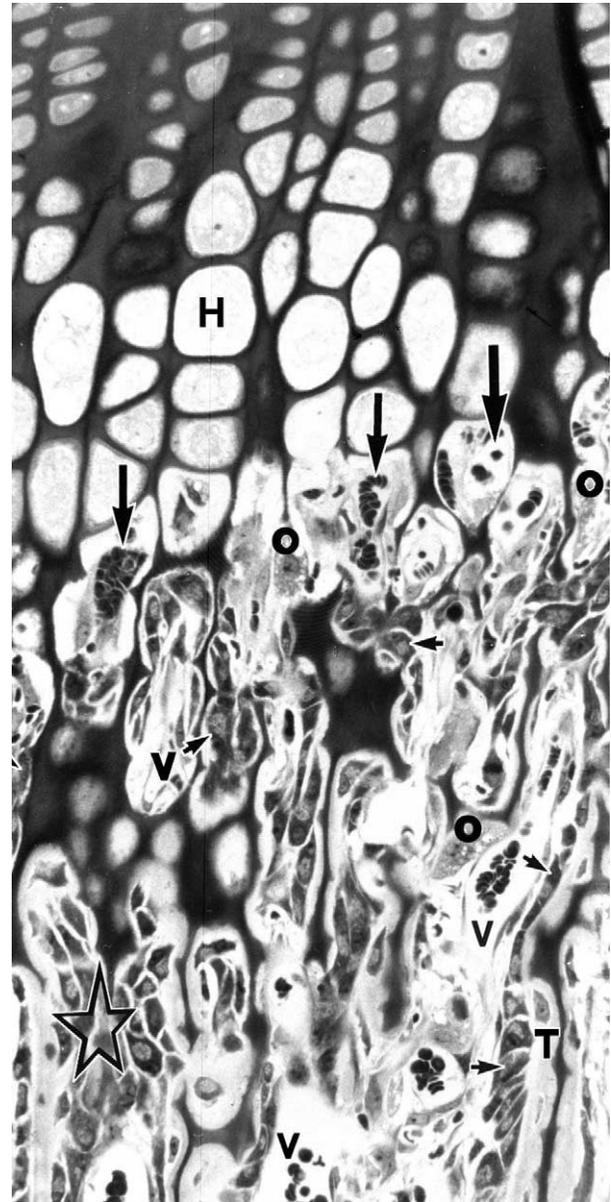
Long bones grow in width along the central shaft by formation on the periosteal surface and resorption on the endosteal surface. This is coordinated with each bone's growth in length to maintain its characteristic shape and may not be bilaterally symmetrical. Thus, the flared epiphyseal ends of long bones, largely cancellous bone (i.e., occupied by trabecula), are converted to the denser bone of the diaphysis by sequential waves of formation and resorption on endosteal and periosteal surfaces. These regions and their activities are summarized in Fig. 10.

When bone growth is complete, the mechanisms for growth in length and girth disappear. The cartilage of the growth plate disappears when chondrocyte proliferation ceases, hypertrophy is exhausted, and bones in the epiphysis and metaphysis are joined. Similarly, bone formation and resorption no longer occur on opposite surfaces in the shaft.

Each bone has a characteristic time at which the primary (diaphyseal) and secondary (epiphyseal) ossification centers appear. Similarly, for each bone there is a time at which its growth plates disappear by fusion of the primary and secondary centers. These events can be monitored radiographically and form the basis for determinations of skeletal or metabolic age, as opposed to the chronological age of an individual.

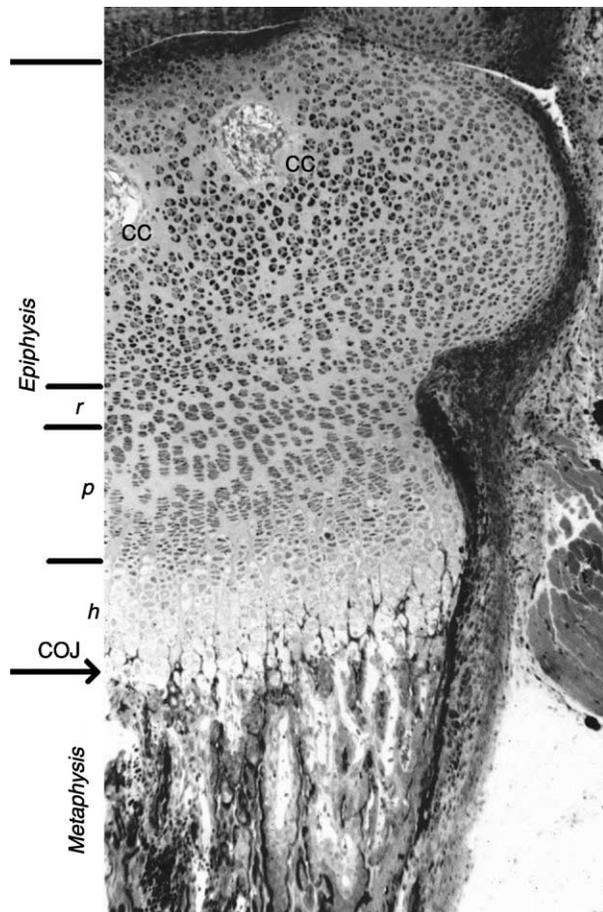
### Regulating Chondrogenesis and Endochondral Growth

Chondrocytes clearly must integrate a complex variety of signals from both local and circulating sources



**Figure 6** Photomicrograph of the chondroosseous junction (COJ) in a young rat. Hypertrophied chondrocytes (H) occupy most of the growth plate visible in this field. Vascular invasion of chondrocyte lacunae is occurring at several sites (vertical arrows) along the COJ, and vascular channels (V) are common between trabeculae (T). Osteoblasts (small arrows) can be identified on most metaphyseal surfaces and a large group (star) appears where trabeculae converge out of the plane of this section. Several osteoclasts (O) can be seen near the COJ. Toluidine blue stain. Reprinted from Marks, S. C. (1998). In "Diagnostic Imaging in Child Abuse" (P. K. Kleinman, ed.), pp. 2–7, Mosby, St. Louis, MO, with permission.

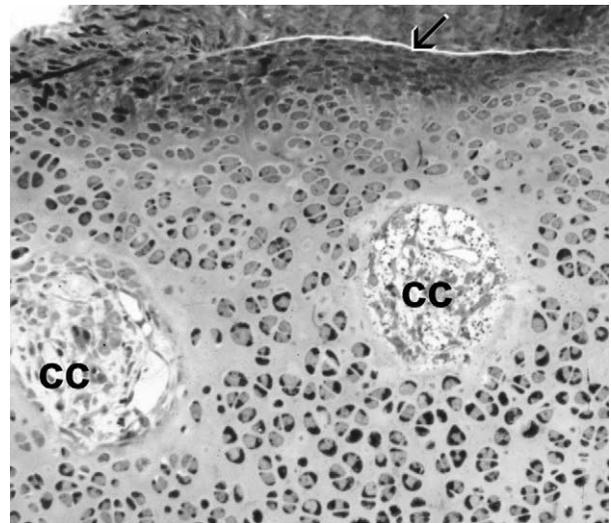
during their proliferation and differentiation in the growth plate. Gene expression in the endochondral pathway begins with the *Sox9* transcription factor



**Figure 7** Parts of a growing long bone as illustrated in the proximal tibia of a newborn mouse. The knee joint cavity can be seen as a thin white space at the top of the figure. The epiphysis stretches from the knee joint to the chondroosseous junction (COJ). At the distal end of the epiphysis, the growth plate consists of regions of resting (r), proliferating (p), and hypertrophied (h) chondrocytes. In the proximal epiphysis, cartilage canals (cc) have invaded to establish the secondary (epiphyseal) ossification center.

and with type II collagen, aggrecan, and other characteristic proteoglycans. These can be found in the earliest cartilage condensations (Figs. 2 and 4). Chondrocyte hypertrophy is essential for endochondral bone formation and *Cbfa1* is required for this step. This may be related to the concurrent expression of this gene in the adjacent periosteum. Chondrocyte proliferation is promoted by *Sox9* and Indian hedgehog (*Ihh*) and inhibited by FGF signaling. Chondrocyte differentiation and hypertrophy are promoted by *Ihh* and *Cbfa1* and inhibited by parathyroid hormone-related peptide (PTHrP) and *Sox9*.

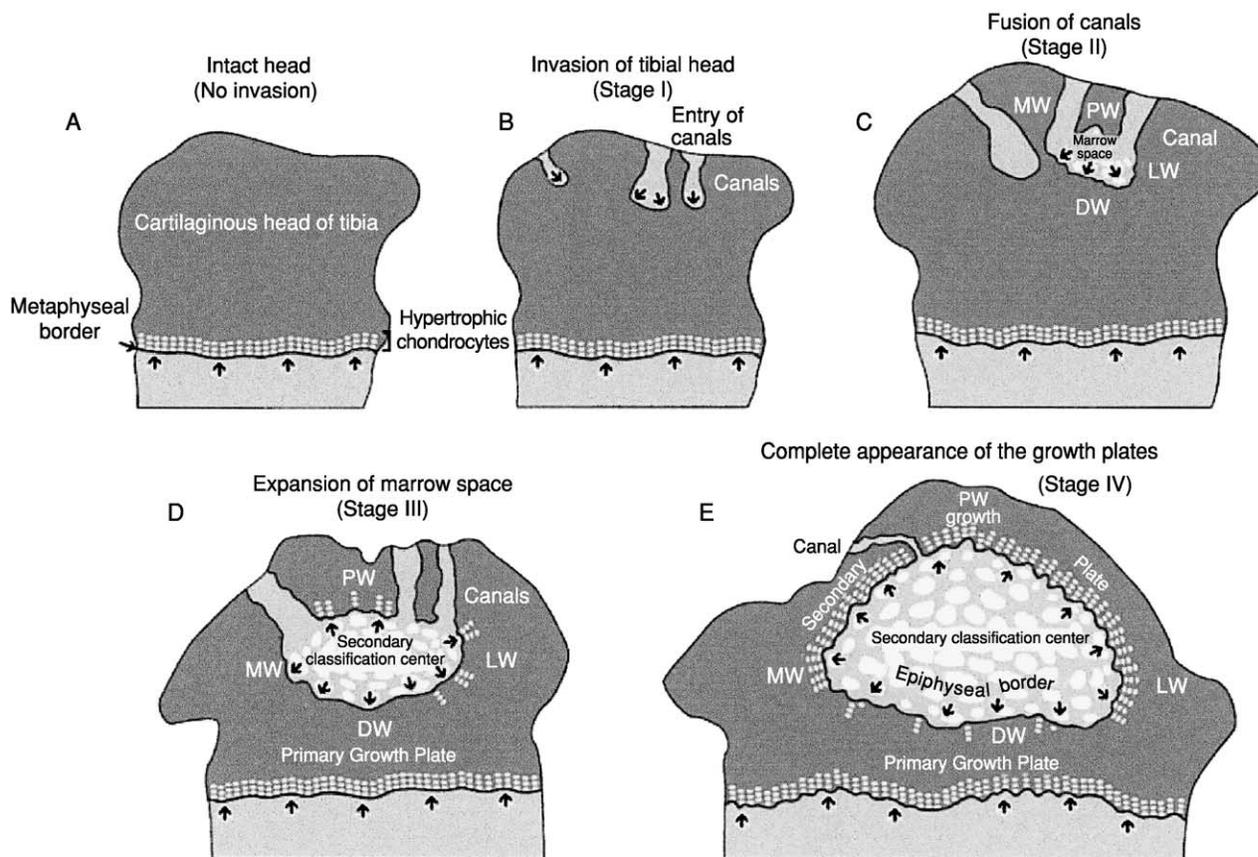
The dynamics of chondrocyte biology in the growth plate are produced by the integrated interactions of



**Figure 8** Higher magnification of the proximal epiphysis in Fig. 7, showing the cartilage canals and their cellular composition, which is distinct from that of the surrounding cartilage. The tibial articular chondrocytes (arrow) are flatter and longer than the other chondrocytes in the epiphysis.

these local influences as modified by systemic influences. PTHrP is produced by periarticular chondrocytes and its receptor is expressed at high levels in prehypertrophic chondrocytes. *Ihh* is expressed at the prehypertrophic–hypertrophic chondrocyte boundary. Cells that escape the inhibitions of PTHrP signaling therefore express *Ihh*, which feeds back to increase PTHrP expression. Thus, PTHrP and *Ihh* are major players in a negative feedback loop that regulates the proliferation/differentiation balance in the growth plate. *Ihh* signaling promotes the proliferation and columnar orientation of chondrocytes, activates PTHrP to inhibit hypertrophy, and is necessary for osteoblast differentiation in endochondral bones. Furthermore, *Ihh* signals the conversion of the perichondrium to the periosteum, another overlapping function with *Cbfa1*. Developmental disturbances in chondrogenesis are the major causes of chondrodysplasias.

Adding to the complexities of endochondral growth are the events at the chondroosseous junction (and its precursor, the initial vascular invasion of the lacunae of hypertrophied chondrocytes and resorption of mineralized cartilage in the center of the developing bone). Matrix metalloproteinases (MMPs), particularly MMP-9, and vascular endothelial growth factor (VEGF) are required for osteoclast recruitment and differentiation and for vascular invasion. MMP-9 is produced by mesenchymal cells in the



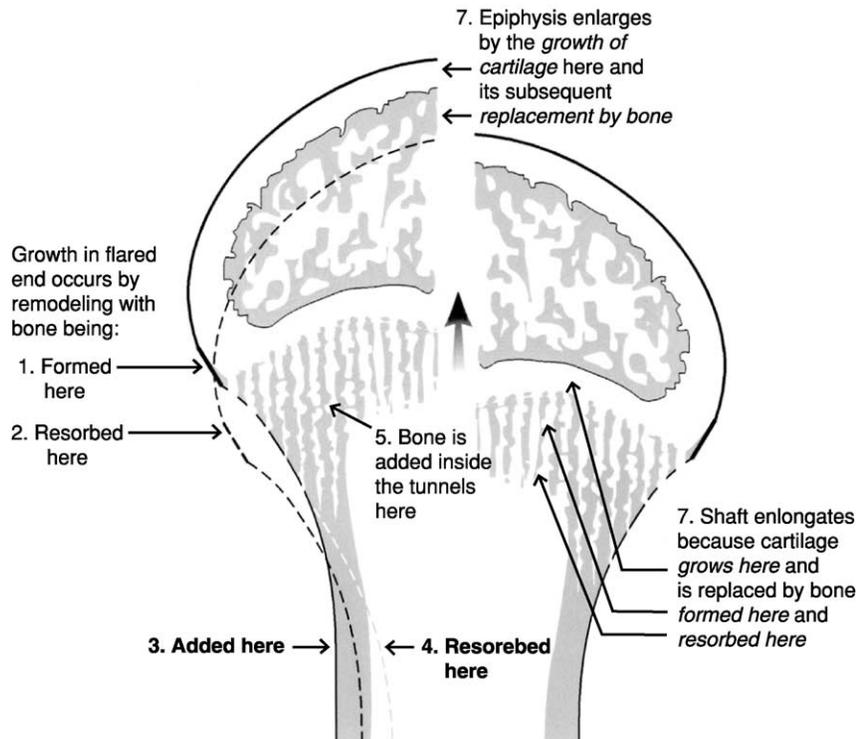
**Figure 9** Epiphyseal transformation from cartilage into bone, illustrated diagrammatically. Secondary ossification centers are initiated by vascular penetration of epiphyses, producing canals. These erosions depend on the local production of a specific enzyme, MMP-14, eventually coalescing to form a central space where bone is formed on trabeculae. As this space expands laterally and proximally, a secondary growth plate is established to provide for epiphyseal growth. Modified from Lee *et al.* (2001). Enzymes active in the areas undergoing cartilage resorption during development of the secondary ossification center in the tibiae of rats ages 0–21 days. I. Two groups of proteinases cleave the core protein of aggrecan. *Dev. Dyn.* 222, 52–70, with permission.

periosteum and by osteoclasts. VEGF is made by hypertrophied chondrocytes, is stored in the adjacent matrix, and is released during hypertrophy. Resorption of the transverse septum between the last hypertrophied chondrocyte and the metaphysis releases additional VEGF, which promotes angiogenesis at this site. Thus, VEGF-mediated angiogenesis appears to be a key event at the chondroosseous junction, where chondrogenesis is coupled to osteogenesis.

MMPs are also involved in ossification of the epiphyses, in this case MMP-14. Elaboration of MMP-14 by the invading cartilage canals is essential for vascular penetration of epiphyses. In its absence, ossification and skeletal age are markedly delayed.

In addition to these local regulations of skeletal development, there are many systemic regulations that are likely mediated through local processes. Among them are growth hormone and insulin-like growth factor-I (IGF-I), which promote cell proliferation in growth plates. Thyroid hormones support metabolism generally and have an indirect effect on growth hormone release and IGF-I action. Estrogens are significant promoters of prepubertal growth and the general maintenance of bone mass. Vitamin D metabolites are important for mineralization. Glucocorticoids, on the other hand, stunt growth by shortening the life span of skeletal cells.

A variety of endochondral skeletal malformations ensue when these processes do not occur in the proper sequence, time, or place.



**Figure 10** Diagram of regional changes in cartilage and bone that produce growth in the length (large arrow) and width of long bones. Reprinted from Marks, S. C. (1998). In "Diagnostic Imaging in Child Abuse" (P. K. Kleinman, ed.), pp. 2-7, Mosby, St. Louis, MO, with permission.

## Joints

Connections between skeletal elements differ depending on whether there will be significant movement between the bones. Joints with little or no motion are fibrous or cartilaginous. Growth plates and synchondroses are temporary cartilaginous joints between the diaphysis and epiphyses or between the bones of the craniofacial base, respectively, and disappear when growth is complete. Symphyses are fibrocartilaginous articulations that permit movement that varies directly with the amount of intervening fibrocartilage. The pubic symphysis is a wide joint with considerable movement, whereas the mandibular symphysis is extremely narrow, has no movement, and is temporary. Sutures are fibrous joints with little or no movement. Synovial joints are designed to allow significant movement and mechanical load transfer between bones. The articulating surfaces of both bones throughout the range of movement are covered by a thin layer of cartilage overlying bone. A fibrous capsule encloses the joint and most of its internal surface consists of a synovial membrane that extends over all internal surfaces except those actually in

contact during movement. The synovial fluid, secreted by this membrane, is a highly viscous material that participates in joint lubrication and cellular maintenance. Depending on the shapes of the articulating surfaces, joints may be strengthened internally and/or externally by ligaments and tendons and the congruence of articulating surfaces enhanced by fibrocartilaginous discs. Various classifications of joints characterize mechanisms and/or movements.

Articular cartilage is derived from the original epiphyseal cartilage but does not undergo hypertrophy. Resisting differentiation, they produce abundant matrix, including tenascin-C, an extracellular matrix protein unique to articular chondrocytes. Endogenous TGF- $\beta$  appears to maintain the appropriate undifferentiated state of articular chondrocytes.

## SKELETAL TURNOVER AND MAINTENANCE

Once formed, the skeleton continues to renew itself throughout life. This is accomplished by a precise, localized resorption that is followed by formation,

Development, maintenance, and pathology of the skeleton			
	<u>Development</u>	<u>Maintenance</u>	<u>The osteoporoses</u>
<u>Age</u>	0–20	20–50	50+
<u>Sequence</u>	BF → BR	BR → BF	BR → BF
<u>Activity</u>	Modeling	Remodelling	
	BF > BR	BF = BR	BF < BR

**Figure 11** Development, maintenance, and acquired pathology of the skeleton. Summary of the relative levels of bone formation (BF) and bone resorption (BR) during the human life cycle. Reprinted from Marks and Odgren (2002), with permission.

which restores the original volume of bone at that site. This is summarized in [Fig. 11](#).

Skeletal development takes place during the first 2 decades of life. During this period, formation of bone precedes and exceeds resorption. This phase is termed modeling. On skeletal maturation, skeletal mass is maintained and microdamage is repaired for the next 3 decades or longer by localized remodeling in which approximately 10% of the skeleton is replaced each year. This begins by resorption followed by an equal amount of formation. Remodeling leaves a trail in cortical bone by producing overlapping Haversian systems of different configurations. With increasing age, the osteoblastic progenitor pool is depleted sooner than the osteolytic pool, leaving the formative

phase of maintenance progressively reduced compared to the resorptive phase. This reduces skeletal mass, the hallmark of osteoporosis.

## DEVELOPMENTAL PATHOLOGY OF THE SKELETON

In each phase of skeletal development, the appropriate, timely integration of the complex expression of many genes is responsible for normal development. Many skeletal disorders result from the inappropriate expression or nonexpression of genes and these gene sequences and genetic disorders are available at the Online Mendelian Inheritance in Man web site. This database includes the latest tabulation of more than 11,000 congenital human malformations, first published as 1487 entries in book form by Victor McCusick, in 1968. Many disorders affecting the skeleton are summarized in [Table I](#).

## Acknowledgments

This work was supported in parts by Grants DE13961 (P.R.O.) and DE07444 (S.C.M.) from the National Institute of Dental and Craniofacial Research (NIDCR), USA. Its contents are solely the responsibility of the authors and do not necessarily represent

**Table I** The Emergent Molecular Biology and Pathology of Skeletal Development

Event	Genes	Developmental pathologies
Patterning/mesenchymal condensations	FGF, FGFR families  <i>HoxD13</i> <i>Hox</i> (MSX1)	Achondroplasia, thanatophoric dysplasia, Crouzon syndrome, Jackson-Weiss syndrome, Pfeiffer syndrome  Synpolydactyly Witkop syndrome
Chondrogenesis	PTHrP, <i>Ibb</i> , BMP7, <i>Sox9</i> , Col II and X	Chondrodystrophies
Osteogenesis	<i>Cbfa1</i> ( <i>Runx2</i> ) FGF, FGFR families Col I  <i>Hox</i> (MSX2)	Homozygous = lethal, heterozygous = cleidocranial dysplasia Craniosynostosis Osteogenesis imperfecta Craniofacial malformations Craniosynostosis
Mineralization	Vitamin D	Rickets
Chondroosseous coordination	MMP-9 MMP-14 VEGF	Chondrodystrophies

*Note.* FGF, fibroblast growth factor; FGFR, FGF receptor; *HoxD13*, *Hox* family of patterning genes; MSX1, -2, homeobox genes; PTHrP, parathyroid hormone-related peptide; *Ibb*, Indian hedgehog; BMP7, bone morphogenic protein family member, *Sox9*, transcription factor; Col I, II, X, collagens type I, II, and X; *Cbfa1* (*Runx2*), mammalian homologue of *Drosophila* Runt protein; MMP-9, -13, matrix metalloproteinases 9 and 13; VEGF, vascular endothelial growth factor.

the official views of the NIDCR or the National Institutes of Health.

### See Also the Following Articles

Bone Mass Measurement • Bone Structure • Bone Turnover Markers • Growth, Normal Patterns and Constitutional Delay • Hypercalcemia and Hypercalcemia Treatment • Osteoporosis, Overview • Skeletal Development During Childhood and Adolescence

### Further Reading

Boyadjiev, S. A., and Jabs, E. W. (2000). Online Mendelian Inheritance in Man (OMIM) as a knowledge base for human developmental disorders. *Clin. Genet.* **57**, 253–266.

de Crombrughe, B., Lefebvre, V., and Nakashima, K. (2001). Regulatory mechanisms in the pathways of cartilage and bone formation. *Curr. Opin. Cell Biol.* **13**, 721–727.

Hickok, N. J., Haas, A. R., and Tuan, R. S. (1998). Regulation of chondrocyte differentiation and maturation. *Microsc. Res. Tech.* **43**, 174–190.

Karaplis, A. C. (2002). Embryonic development of bone and the molecular regulation of intramembranous and endochondral bone formation. In “Principles of Bone Biology” (J. P. Bilezikian, L. G. Raisz, and G. A. Rodan, eds.), 2nd ed., Part 1, Chap. 3, pp. 33–58. Academic Press, San Diego, CA.

Marks, S. C., Jr., and Odgren, P. R. (2002). Structure and development of the skeleton. In “Principles of Bone Biology” (J. P. Bilezikian, L. G. Raisz, and G. A. Rodan, eds.), 2nd ed., Part 1, Chap. 1, pp. 3–15. Academic Press, San Diego, CA.

Yang, X., and Karsenty, G. (2002). Transcription factors in bone: Developmental and pathological aspects. *Trends Mol. Med.* **8**, 340–345.

more descriptive term material BMD ( $BMD_{\text{material}}$ ) may be preferable, because a variety of densities have been labeled “true.”

$BMD_{\text{material}}$  reflects the degree of mineralization of organic bone matrix. The degree of matrix mineralization varies widely in any single piece of bone and therefore  $BMD_{\text{material}}$  is always the average of a continuum of density values. Bone matrix has a mineral density of zero when it is released from the osteoblast and mineralization starts only  $\approx 2$  weeks later at a typical remodeling site. Within a few days after the start of mineralization, inorganic material has filled 75% of the matrix volume originally occupied by water molecules (primary mineralization, not to be confused with primary bone). During the following 6 months, mineral continues to be slowly incorporated into the matrix (secondary mineralization). Because of this time-dependent increase, recently deposited bone matrix has a lower mineral density than “old” matrix. From this fact, it follows that  $BMD_{\text{material}}$  is inversely related to bone remodeling activity: when remodeling activity is high, there will be more unmineralized osteoid and there will be more “young” bone matrix, which has not yet completed secondary mineralization. These considerations refer to secondary (remodeled) bone. Primary (unremodeled) bone, which osteoblasts add to the periosteal surface during growth, is denser than secondary bone, even though it is younger. Therefore, when there is rapid periosteal expansion and little intracortical remodeling, cortical  $BMD_{\text{material}}$  will be high. The functional mechanisms influencing  $BMD_{\text{material}}$  and the structural components reflected by  $BMD_{\text{material}}$  are summarized in [Table I](#).

**Table I** Factors Determining  $BMD_{\text{material}}$

Factor	Effect on $BMD_{\text{material}}$ when increased
Functional mechanisms	
Periosteal expansion (modeling)	↑
Remodeling	↓
Structural features	
Ratio between primary and secondary bone	↑
Relative volume of unmineralized osteoid	↓
Average age of bone matrix	↑

Reproduced from *J. Bone Miner. Res.* 2001; **16**: 597–604, with permission of the American Society for Bone and Mineral Research.

## Age Dependency

In adults, the relative ash weight of trabecular bone is a few percent lower than that of cortical bone. This probably reflects the fact that the turnover of trabecular bone is higher than that of cortical bone. As trabecular bone turnover decreases from childhood to adulthood,  $BMD_{\text{material}}$  should increase during bone development. In fact,  $BMD_{\text{material}}$  of trabecular iliac bone shows such a trend between the first and the fourth decades of life.

## THE DEVELOPMENT OF TRABECULAR AND CORTICAL TISSUE—COMPARTMENT BONE MINERAL DENSITY

The trabecular compartment is defined as the space within the endocortical surface. The cortical compartment is delimited by the periosteal and endocortical surfaces. Both compartments contain not only bone matrix, but also nonbone tissue. In trabecular bone, this is mainly hematopoietic and fat marrow. In cortical bone, the nonbone composites are blood vessels and the connective tissue within osteonal (Haversian) and Volkmann’s canals. In this article, the term compartment BMD ( $BMD_{\text{compartment}}$ ) is used to denote the mass of mineral per unit volume of the trabecular or cortical compartment. In the densitometric literature, trabecular  $BMD_{\text{compartment}}$  has been called cancellous bone density or volumetric trabecular apparent BMD. The cortical  $BMD_{\text{compartment}}$  has been referred to as true cortical bone density or material density of cortical bone.  $BMD_{\text{compartment}}$  depends on  $BMD_{\text{material}}$  and relative bone volume, which is the relative amount of space occupied by bone matrix.

Structural features and functional mechanisms reflected by  $BMD_{\text{compartment}}$  are listed in [Table II](#). In trabecular bone,  $BMD_{\text{compartment}}$  depends on trabecular number (defined as the number of trabeculae that an imaginary line through the bone would hit per millimeter of its length) and mean trabecular thickness. The mechanisms determining trabecular number during growth have not been well characterized. Trabecular thickness depends on remodeling activity and on remodeling balance.

## Age Dependency

$BMD_{\text{compartment}}$  of the femoral midshaft decreases in the first months after birth and then increases until adulthood. The increase in  $BMD_{\text{compartment}}$  between

Table II Factors (Other than  $BMD_{\text{material}}$ ) Determining  $BMD_{\text{compartment}}$ 

Factor	Effect on $BMD_{\text{compartment}}$ when increased
Functional mechanisms	
Remodeling	
Cortical bone	↓
Trabecular bone	↓↑
Structural features	
Trabecular bone	
Trabecular number	↑
Trabecular thickness	↑
Cortical bone	
Number of osteonal canals	↓
Size of osteonal canals	↓

Reproduced from *J. Bone Miner. Res.* 2001; **16**: 597–604, with permission of the American Society for Bone and Mineral Research.

the age of 3 and 5 months and adulthood amounts to 36%, but most of this increase occurs in early childhood. The difference between the age group of 4–13 years and young adults is only  $\approx 6\%$ . These variations in cortical  $BMD_{\text{compartment}}$  appear to be more pronounced than the corresponding changes in relative ash weight. The changes in cortical  $BMD_{\text{material}}$  probably are caused by variations in both  $BMD_{\text{material}}$  and relative bone volume.

## THE DEVELOPMENT OF TOTAL BONE DENSITY

The volume to determine total BMD ( $BMD_{\text{total}}$ ) is enclosed by the bone's periosteal envelope and articular surfaces. Terms used in the literature to denote  $BMD_{\text{total}}$  include bone mineral apparent density, volumetric bone density, and true bone density.  $BMD_{\text{total}}$  is determined by trabecular and cortical  $BMD_{\text{compartment}}$  and by the relative volumes of the two compartments. Table III lists the functional mechanisms and structural features that influence  $BMD_{\text{total}}$ .

### Age Dependency

The dry total density of long bones (i.e., mineral plus organic mass divided by external bone volume) decreases by approximately 30% in the first few months after birth. This is followed by a rapid increase until approximately 2 years of age and a slower increase thereafter. The postnatal drop in  $BMD_{\text{total}}$  is partly

Table III Factors (Other Than  $BMD_{\text{compartment}}$ ) Determining  $BMD_{\text{total}}$ 

Factor	Effect on $BMD_{\text{total}}$ when increased
Functional mechanisms	
Modeling	
Endocortical resorption	↓
Endocortical apposition	↑
Periosteal apposition	↑
Structural features	
Relative volume of the cortical compartment	↑
Relative volume of the trabecular compartment	↓

Reproduced from *J. Bone Miner. Res.* 2001; **16**: 597–604, with permission of the American Society for Bone and Mineral Research.

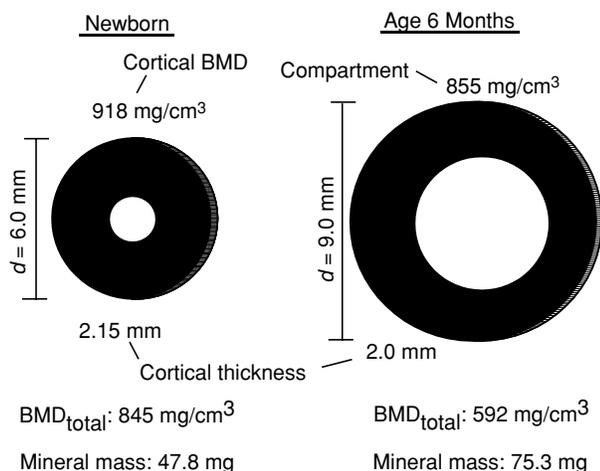
caused by decreasing relative cortical area. The simultaneous decline in cortical  $BMD_{\text{material}}$  and  $BMD_{\text{compartment}}$ , as outlined previously, also will contribute to lower  $BMD_{\text{total}}$  in early infancy.

The decrease in cortical thickness during the first months of life was dubbed “physiological osteoporosis of infancy” a century ago. However, in growing children, a decrease in cortical thickness and  $BMD_{\text{total}}$  is not necessarily a sign of bone loss. This is exemplified in Fig. 1, which summarizes studies on femoral midshaft development during the first 6 months of postnatal life. In healthy babies, the decline in  $BMD_{\text{total}}$  reflects a redistribution of bone tissue from the endocortical surface to the periosteal surface rather than bone loss. Therefore, bone mineral mass increases considerably during the first months of life, despite the precipitous drop in  $BMD_{\text{total}}$  (Fig. 1).

## THE REGULATION OF BONE DEVELOPMENT

Interpreting bone density in terms of the underlying biological organization not only provides a better understanding of bone development, but is helpful in reconciling seemingly contradictory results of densitometric studies in children with bone diseases.

A viable model of functionally controlled bone development must take into account a general principle of regulation, which applies to biology, engineering, and even human organizations: Control requires not only the ability to act, but also information on the current state of affairs. Thus, a controlled

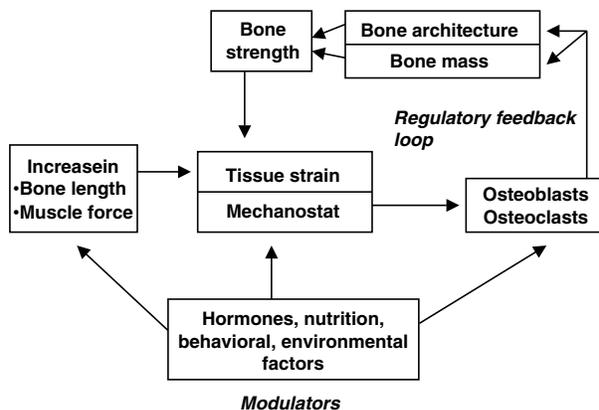


**Figure 1** A model of bone development at the femoral midshaft from birth to 6 months of age. At birth, the external bone diameter ( $d$ ) is approximately 6.0 mm and cortical thickness is 2.15 mm. Thus, 92% of the cross section is cortical bone. At 6 months of age,  $BMD_{total}$  has dropped by 30% and cortical thickness has decreased. However, the external bone diameter has increased to 9.0 mm. Therefore, the mineral mass in a 2 mm thick slice of bone has increased by 58%. Reproduced from *J. Bone Miner. Res.* 2001; **16**: 597–604, with permission of the American Society for Bone and Mineral Research.

system needs to have both effector and sensor mechanisms. The former performs an action and the latter generates feedback signals, which indicate whether or not the desired effect has been achieved. This automatically indicates that there must be information within the system about what the “desired effect” is.

These requirements for a controlled system are met by Frost’s mechanostat model and related approaches (Fig. 2). It is proposed that the desired effect of bone homeostasis is to keep the mechanically induced deformation of bone (in biomechanical terminology this is called strain) close to a preset level, called the setpoint. The deformation of a bone is a surrogate measure of its strength, because a strong bone will deform less than a weak bone when a mechanical challenge is applied. Bone deformation generates canalicular fluid flow, which could be monitored by osteocytes. When bone deformation exceeds a certain acceptable limit, osteocytes might sense this and send out signals, which could lead to adaptations in bone mass and architecture. These adaptations increase bone strength and the mechanical strain returns to the acceptable level.

According to this model, changes in bone mass and architecture occur when bone stability is challenged and bone deformation exceeds an acceptable level. To put it differently, the required mechanical strength of



**Figure 2** A functional model of bone development based on mechanostat theory. The central feature of bone regulation is the feedback loop between bone deformation (tissue strain) and bone strength. During growth, this homeostatic system is continually forced to adapt to external challenges. Factors shown at the bottom modulate various aspects of the central regulatory system.

a bone determines its mass and architecture, not vice versa. During growth, bone stability is continually threatened by two processes: the increase in bone length and the increase in muscle force. Longitudinal growth increases lever arms and bending moments and therefore leads to greater bone deformation. Greater muscle force will also increase bone deformation during muscle contraction. Body weight alone puts relatively small loads on bones, but the effect of weight is amplified by muscle action. These challenges create the need for adaptational changes in bone mass and architecture.

## BONE DEVELOPMENT IN NEWBORNS AND PREMATURE INFANTS

The role of estrogen in the mechanostat system is to lower the setpoint on endosteal bone surfaces, thereby increasing the amount of endocortical bone. This could not only explain certain aspects of pubertal bone development in girls, but is also of relevance for early postnatal events. During intrauterine life, the fetus is exposed to high placental estrogen levels. Accordingly, endosteal surfaces are very sensitive to mechanical strain, leading to small marrow cavities and high organ-level bone density. Cutting off the placental estrogen supply at birth should increase the mechanostat setpoint on endosteal surfaces. Therefore, a substantial amount of bone next to the marrow should then be interpreted as mechanically unnecessary by

the mechanostat. This leads to endocortical resorption and expansion of the marrow cavity, which in densitometric terms corresponds to a decrease in volumetric bone mineral density. There is a postnatal increase in the size of the marrow cavity, which leads to a drop of approximately 30% in directly determined bone density within the first 6 months of life.

At the same time, bones rapidly grow in length. From the perspective of the mechanostat theory, the destabilizing effect of longitudinal growth leads to the addition of bone tissue on periosteal surfaces, where the effect on stability is largest. The combined effect of decreased estrogen levels and increasing mechanical strain during the first postnatal months is a redistribution of bone tissue from the endocortical surface to the periosteal surface. This is a useful mechanism, because it optimizes the distribution of bone mass with regard to mechanical stability and decreases the amount of calcium that must be added from nutritional sources.

When birth occurs prematurely, the placental estrogen supply is cut off earlier. According to mechanostat theory, decreasing estrogen levels should lead to marrow cavity expansion and a decrease in organ-level bone density, similar to mature newborns. Compared to mature newborns at birth, premature babies have larger marrow cavities and lower bone density when they have reached expected term, because they have already started to adapt their bones to postnatal conditions. The differences in bone density between term and preterm babies are only transient. Premature babies have a larger marrow cavity and lower densitometric results than term newborns have at birth. Although prematurity may have a deleterious effect on longitudinal bone growth, there is ample evidence from follow-up studies that the mass of the (shorter) bones is normal for size. From the perspective of mechanostat theory, the so-called “osteopenia of prematurity” (not to be confused with rickets of prematurity) is the expression of physiological postnatal adaptations.

## BONE DEVELOPMENT DURING PUBERTY

When estrogen levels rise again in female puberty, endosteal bone surfaces are resensitized to mechanical strain, leading to endocortical apposition at many skeletal sites. Adding bone to endocortical surfaces has a smaller effect on bone stability than adding the same amount of bone to the periosteal surface.

Consequently, postpubertal, premenopausal girls and women have more bone relative to their mechanical needs than males. The purpose of this estrogen-dependent “excess bone” on endocortical surfaces could be to create a calcium reservoir that can be tapped during pregnancy and lactation.

## THE FRAGILITY OF GROWING BONES

According to mechanostat theory, bone reacts to challenges to its stability. Since the response can only follow but not precede the challenge, bones are “overloaded” and thus fragile for as long as growth continues. The lag between growth in length and growth in strength should be exaggerated when longitudinal growth accelerates. In fact, the dissociation between bone’s growth in length and in mass is a well-documented phenomenon during the pubertal growth spurt in both genders and could explain the increased fracture rate during that period in life. Since the timing of the maximal increase in bone length and muscle mass differs between musculoskeletal regions, the increase in bone mass follows a region-specific pattern during puberty.

## See Also the Following Articles

Body Composition During Growth • Body Proportions • Bone Mass Measurement • Bone Structure • Bone Turnover Markers • Growth, Normal Patterns and Constitutional Delay • Puberty: Physical Activity and Growth • Skeletal Development

## Further Reading

- Dickerson, J. W. (1962). Changes in the composition of the human femur during growth. *Biochem. J.* **82**, 56–61.
- Frost, H. M. (1990). Skeletal structural adaptations to mechanical usage (SATMU). 1. Redefining Wolff’s law: The bone modeling problem. *Anat. Rec.* **226**, 403–413.
- Frost, H. M. (1990). Skeletal structural adaptations to mechanical usage (SATMU). 2. Redefining Wolff’s law: The remodeling problem. *Anat. Rec.* **226**, 414–422.
- Trotter, M. (1971). The density of bones in the young skeleton. *Growth* **35**, 221–231.
- Trotter, M., and Hixon, B. B. (1974). Sequential changes in weight, density, and percentage ash weight of human skeletons from an early fetal period through old age. *Anat. Rec.* **179**, 1–18.
- Trotter, M., and Peterson, R. R. (1970). The density of bones in the fetal skeleton. *Growth* **34**, 283–292.
- Vinz, H. (1970). Studies on the density, water and mineral contents of compact human bone tissue depending on age. *Morphol. Jahrb.* **115**, 273–283.

at the expense of rT3. Finally, to reconcile the finding of decreased thyroid hormone levels and normal/decreased TSH values reported by some authors, an impairment of TSH secretion at the hypothalamic and/or pituitary level might concomitantly occur. No evidence for this effect of smoking has been thus far provided and it must be mentioned that, at least under acute conditions (20 min of continuous smoking), cigarette smoking does not cause any significant variation in serum TSH levels.

In summary, cigarette smoking seems to be associated with minor and controversial changes in thyroid function tests, whose pathophysiological relevance seems to be marginal (Table I).

## SMOKING AND GOITER

Several reports have documented that smoking is associated with an increased prevalence of goiter (Table I). In addition, it was shown that heavy smokers have an increased frequency of nodular goiter. These findings may be related to iodine intake, because many of these studies were performed in iodine-deficient areas. At variance, no differences in the prevalence of goiter among smokers and nonsmokers were found in iodine-sufficient areas. It is likely that cigarette smoking represents only a cofactor and its goitrogenic effect becomes more apparent when other goitrogenic factors, and particularly iodine deficiency, are also present.

It is not completely clear how cigarette smoking can contribute to the development of goiter, but a candidate goitrogen in smoke is thiocyanate, produced by detoxification of hydrogen cyanide. The role of thiocyanate in the etiology of endemic goiter has been clearly shown in the presence of iodine deficiency. Body fluids of smokers contain increased thiocyanate concentrations. The increase in the thyroid volume/birth weight ratio in newborns parallels the increase in cord serum thiocyanate levels, taken as an index of

maternal smoking habits, suggesting that smoking during pregnancy may be a relevant cause of thyroid gland enlargement in the newborn. The effects of thiocyanate and other cigarette smoking products (nicotine, cotinine) have also been studied *in vitro* using porcine thyroid follicles in culture; whereas nicotine and cotinine did not inhibit iodide transport or thyroid hormone synthesis, thiocyanate, at concentrations equivalent to those reached in the serum of smokers, inhibited iodide transport and iodine organification while increasing iodide efflux. In addition to underscoring the role of thiocyanate, these findings may explain the interaction of cigarette smoking and iodine deficiency in the development of goiter.

In summary, it appears that smoking plays a role in the development of goiter and most reports agree on a higher prevalence of goiter in smokers. Although it is conceivable that many (and yet unidentified) smoke products may contribute, thiocyanate appears to be the most likely culprit, in view of the high circulating levels found in smokers and in view of its complex effects on thyroid function. Smoking-related goitrogenesis appears to be frequent, particularly in iodine-deficient areas, where smoke may also represent a relevant cause of neonatal thyroid enlargement.

## RELATIONSHIP BETWEEN SMOKING AND THYROID AUTOIMMUNE DISORDERS

Cigarette smoking has a number of immunological effects involving both humoral and cellular components of the immune system. These include a depression of natural killer cell activity, an increase in the total number of T lymphocytes, with a relative decrease in CD4<sup>+</sup> (helper) and increase in CD8<sup>+</sup> (suppressor) subpopulations in heavy smokers. Serum immunoglobulin (Ig) G, IgM, and IgA levels are decreased by 10–20% in the serum of smokers, whereas IgE levels are increased

Table I Effects of Smoking on the Thyroid

<b>Thyroid function tests</b>	Minor and controversial changes in serum thyroid hormone and TSH concentrations
<b>Goiter</b>	Goitrogenic effect, more evident in iodine-deficient areas
<b>Graves' disease</b>	Increased prevalence of smokers among Graves' disease patients; increased risk of occurrence of Graves' disease in smokers, higher recurrence rate after anti-thyroid drug withdrawal
<b>Graves' ophthalmopathy</b>	Increased prevalence of smokers among Graves' patients with ophthalmopathy; relationship between smoking and severity of eye disease; lower effectiveness of treatment for ophthalmopathy in smokers
<b>Hashimoto's thyroiditis</b>	Association of smoking with thyroiditis, but not with subsequent progression to hypothyroidism
<b>Postpartum thyroid dysfunction</b>	Conflicting results on the role of smoking in the occurrence of postpartum thyroid disorders

in light-to-moderate smokers and decreased in heavy smokers. A lower level of immunosuppression has been observed in mice as an effect of smoking. The numerous immunological effects of smoking may have some relevance in human pathophysiology. For example, it has been shown that, although smoking is associated with Crohn's disease, ulcerative colitis is strongly associated with not smoking. The underlying mechanism might not necessarily be related to immunological effects of smoking, but to substances contained in smoke that might be somehow beneficial to ulcerative colitis patients. An association between rheumatoid arthritis and cigarette smoking, as well as an increased prevalence of antinuclear antibodies in smokers, has been reported.

### Graves' Disease

Graves' disease is an autoimmune disease to which both genetic and environmental factors contribute. The latter may include stressful life events and infections. Smoking represents an additional and important environmental factor. An increased prevalence of smokers has been observed in patients with Graves' disease and smoking increases the risk of developing this disease. In addition, smoking has been reported to be associated with an increased incidence of recurrence of Graves' hyperthyroidism following anti-thyroid drug treatment withdrawal.

Why should cigarette smoking be associated with an increased risk of developing Graves' disease? Smoking might simply be a coincidental, unrelated factor and Graves' patients might smoke more because they are nervous and stressed. Stress is indeed associated with an increased desire to smoke. It should be mentioned that, in most cases, Graves' patients smoked prior to the development of hyperthyroidism. Alternatively, smoking might enhance other activities (e.g., drinking alcohol or coffee) or other factors that bear the true responsibility for the occurrence of the disease. However, the possibility cannot be excluded that smoking may be effectively involved in the pathogenesis of the disease. This might occur through different smoke-related mechanisms, such as an alteration of the TSH receptor structure, an enhancement of responsiveness to a factor(s) responsible for the initiation of the disease, or an impairment of restoration of tolerance to thyroid autoantigens. Evidence that these hypothetical mechanisms participate in the development of Graves' disease is lacking.

### Graves' Ophthalmopathy

Graves' ophthalmopathy is the most frequent extrathyroidal manifestation of Graves' disease and represents a complex therapeutic problem. A possible relationship between cigarette smoking and Graves' ophthalmopathy was initially suggested in a small series of patients and subsequently confirmed in a large number of patients in a different series (Table II). Another series reported a 64% prevalence of smokers among 307 patients with Graves' ophthalmopathy, a percentage much higher than that in Graves' patients without ophthalmopathy (48%) or controls (29%). Furthermore, there seems to be a relationship between smoking and disease severity. Interestingly, the current number of daily cigarettes smoked, but not lifetime tobacco consumption, appears to be an independent risk factor, and ex-smokers have a lower risk of developing severe Graves' ophthalmopathy than current smokers with a comparable lifetime tobacco consumption. In a survey by the European Thyroid Association, 43% of respondents indicated an apparent decrease in the incidence of clinically relevant Graves' ophthalmopathy in the past decades, whereas 12% reported an increasing incidence of the disease. Interestingly, most respondents of the former group belonged to countries (e.g., the United Kingdom, France, the Netherlands) where tobacco consumption has decreased; conversely, respondents in the second group belonged to countries (e.g., Hungary, Poland) where tobacco consumption has increased.

**Table II** Prevalence of Smokers in Patients with Graves' Ophthalmopathy

Author	Year	No. of patients	No. of smokers	Percentage
Hagg <i>et al.</i>	1987	12	10	83
Bartalena <i>et al.</i>	1989	307	196	64
Shine <i>et al.</i>	1990	85	53	62
Balazs <i>et al.</i>	1990	38	36	95
Tellez <i>et al.</i>	1992	52	23	44
Winsa <i>et al.</i>	1993	63	30	48
Tallstedt <i>et al.</i>	1993	24	19	79
Prummel <i>et al.</i>	1993	100	81	81
Pfeilschifter <i>et al.</i>	1996	52	23	44
Hofbauer <i>et al.</i>	1997	27	18	67
Mann <i>et al.</i>	1999	50	35	70
Salvi <i>et al.</i>	2000	72	31	43
Allahabadia <i>et al.</i>	2000	216	102	47
Marcocci <i>et al.</i>	2001	82	54	66
Total		1179	711	60

This association between cigarette smoking and Graves' ophthalmopathy is unexplained. In addition to direct irritative effects, cigarette smoking might influence ongoing immune reactions in the orbit. It is well accepted that cytokines play an important role in the pathogenesis of Graves' ophthalmopathy. Cigarette smoking may affect the process, because smoking-induced hypoxia in the retrobulbar tissue increases the release of cytokines from orbital fibroblasts *in vitro*. Furthermore, it has been reported, but not unequivocally confirmed, that smokers, compared to nonsmokers, have lower circulating levels of soluble interleukin-1 (IL-1) receptor antagonist levels, an anti-cytokine antagonizing the effects of IL-1.

Cigarette smoking may also affect treatment outcomes of Graves' ophthalmopathy (Table I). A more favorable response to orbital radiotherapy and high-dose systemic glucocorticoid treatment for severe Graves' ophthalmopathy is more frequent in smokers than in nonsmokers. In patients with nonsevere Graves' ophthalmopathy, radioiodine treatment may cause a progression of eye disease in approximately 15% of cases; postradioiodine progression is more likely to occur in smokers, whereas improvement of preradioiodine ophthalmopathy with the concomitant glucocorticoid treatment is more frequent in nonsmokers.

In summary, although the mechanisms whereby cigarette smoking operates remain to be elucidated, the relationship between smoking and eye disease is well established, in terms of the epidemiology of the disease and its influence on the severity of the disease and the efficacy of its management. Accordingly, Graves' ophthalmopathy patients should be vigorously encouraged to refrain from smoking, even though it remains to be proven that smoking withdrawal beneficially affects the course of the disease.

### Hashimoto's Thyroiditis

Data on the relationship between smoking and the occurrence or exacerbation of Hashimoto's thyroiditis are not completely clear. Several studies failed to show a sound relationship. However, a meta-analysis of several published papers seems to indicate that a relationship exists between smoking and autoimmune thyroiditis, but not between smoking and the subsequent progression of thyroiditis to hypothyroidism (Table I). In this regard, it is worth mentioning that anthracene derivatives cause experimental autoimmune thyroiditis in susceptible rats.

### Postpartum Thyroid Dysfunction

Conflicting results have been reported concerning a possible association between cigarette smoking and the development of postpartum thyroid dysfunction. However, in most studies, smoking apparently does not play a relevant role in the occurrence of postpartum thyroid disorders (Table I).

### CONCLUDING REMARKS

The relationship between cigarette smoking and the thyroid is complex and not yet completely understood. Marginal changes in thyroid function have been described in smokers, but an overall evaluation of reported series shows that these effects are controversial and, therefore, unlikely to have a relevant pathophysiological significance. More clearly established is the goitrogenic effect of smoking, which is probably related to the action of thiocyanate (and possibly of other compounds liberated in smoke). This effect might be particularly important in iodine-deficient areas. An interesting observation is the high prevalence of smokers among patients with Graves' disease, but it is unclear whether this has a pathogenic importance or merely represents an epiphenomenon of behavioral changes related to thyroid hyperfunction. Even more striking is the association between smoking and Graves' ophthalmopathy. Although the underlying mechanisms need to be elucidated, smoking might play a role in the pathogenesis of Graves' ophthalmopathy and unfavorably affect its course.

### See Also the Following Articles

Goitrogens, Environmental • Graves' Disease • Graves' Ophthalmopathy • Hashimoto's Disease • Iodine Deficiency • Thyroid Autoimmunity • Thyroid Hormone Action • TSH (Thyroid-Stimulating Hormone; Thyrotropin)

### Further Reading

- Bartalena, L., Bogazzi, F., Tanda, M. L., Manetti, L., Dell'Unto, E., and Martino, E. (1995). Cigarette smoking and the thyroid. *Eur. J. Endocrinol.* **133**, 507–512.
- Bartalena, L., Marcocci, C., and Pinchera, A. (2002). Cigarette smoking and Graves' ophthalmopathy. In "Thyroid Eye Disease: Diagnosis and Treatment" (J. J. Dutton and B. J. Haik, eds.), pp. 251–259. Dekker, New York.
- Bartalena, L., Marcocci, C., Tanda, M. L., Manetti, L., Dell'Unto, E., Bartolomei, M. P., Nardi, M., Martino, E., and Pinchera, A. (1998). Cigarette smoking and treatment outcomes in Graves' ophthalmopathy. *Ann. Intern. Med.* **129**, 632–635.

- Bartalena, L., Pinchera, A., and Marcocci, C. (2000). Management of Graves' ophthalmopathy: Reality and perspectives. *Endocr. Rev.* **21**, 168–199.
- Berthelsen, J. B., and Hegedus, L. (1994). Cigarette smoking and the thyroid. *Thyroid* **4**, 327–331.
- Glinoe, D., De Nayer, P., Bex, M., and the Belgian Collaborative Study Group on Graves' Disease (2001). Effects of L-thyroxine, TSH-receptor antibodies and smoking on the risk of recurrence in Graves' hyperthyroidism treated with antithyroid drugs: A double-blind prospective randomized study. *Eur. J. Endocrinol.* **144**, 475–483.
- Prummel, M. F., and Wiersinga, W. M. (1993). Smoking and risk of Graves' disease. *J. Am. Med. Assoc.* **269**, 479–482.
- Vestergaard, P. (2002). Smoking and thyroid disorders—A meta-analysis. *Eur. J. Endocrinol.* **146**, 153–161.
- Weetman, A. P., and Wiersinga, W. M. (1998). Current management of thyroid-associated ophthalmopathy in Europe: Results of an international survey. *Clin. Endocrinol.* **49**, 21–28.

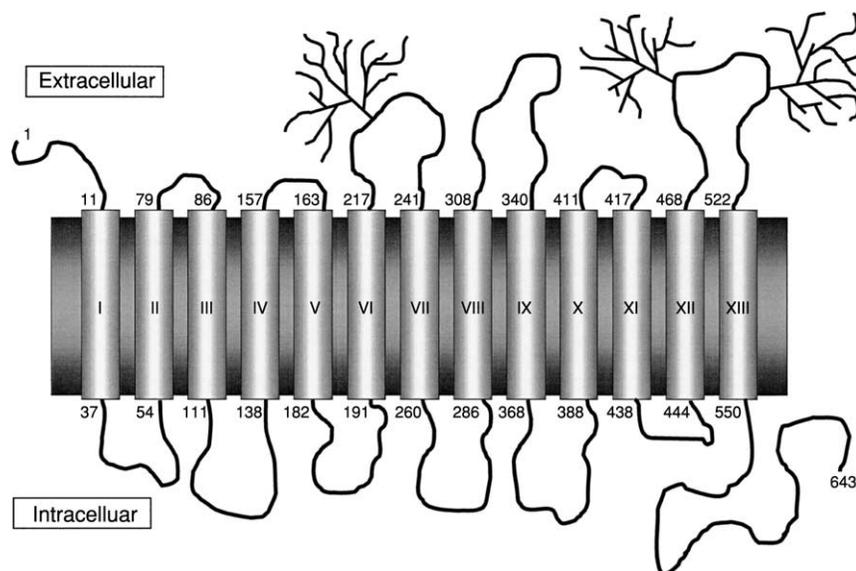


Figure 1 Schematic representation of the human sodium iodide symporter.

cyclase–cAMP signal transduction pathway. Forskolin and  $(\text{Bu})_2\text{cAMP}$  mimic this stimulatory effect on both iodide transport activity and NIS expression. Furthermore, adenosine, which acts as an autocrine factor in cultured thyrocytes, has been shown to stimulate NIS mRNA and protein expression, as well as iodide transport *in vitro*. In addition to various agents that stimulate NIS expression and function, several inhibitors of NIS expression and/or iodide uptake activity have been identified *in vitro* and *in vivo*. Thyroglobulin, a suppressor of thyroid-specific gene expression, ceramide and sphingomyelinase, which play important roles in the nonthyroidal illness syndrome, in age-associated hypothyroidism, and in intrathyroidal apoptosis, and excessive iodide (Wolff-Chaikoff effect) all act to decrease NIS mRNA and protein expression in FRTL5 cells or *in vivo*. Whereas estradiol increases the proliferation of FRTL5 cells, it down-regulates NIS mRNA expression. Moreover, various cytokines, such as transforming growth factor- $\beta$ 1, tumor necrosis factor  $\alpha$ , interferon- $\gamma$ , interleukin-1 $\alpha$  (IL-1 $\alpha$ ), IL-1 $\beta$ , and IL-6 act to down-regulate NIS expression and iodide uptake.

In the lactating mammary gland, iodide accumulates in breast milk by NIS-mediated transport in order to supply adequate amounts of iodide to the newborn. Several lines of evidence suggest that iodide transport activity and NIS expression in mammary gland are stimulated by prolactin, representing an important regulatory mechanism for iodide accumulation in breast milk. Further studies in rats showed that NIS protein expression in the lactating mammary

gland is up-regulated during nursing and is significantly decreased within 24 h following separation of pups from their mothers. These data indicate that NIS expression in the mammary gland is regulated in a reversible manner by suckling, suggesting a regulatory mechanism involving the hormone oxytocin.

#### EXPRESSION OF SODIUM IODIDE SYMPORTER IN BENIGN AND MALIGNANT THYROID TISSUE

Several studies have demonstrated that benign toxic multinodular goiters reveal a heterogeneous pattern of NIS protein expression, with overall expression being stronger than in normal thyroid gland. Furthermore, NIS mRNA levels have been shown to be low in nontoxic multinodular goiter and diffuse iodine deficiency goiter. Whereas NIS mRNA and protein expression levels are increased in autonomously functioning thyroid nodules with increased radioiodine uptake, they are decreased in cold thyroid nodules with low or absent radioiodine uptake.

In contrast to benign thyroid tissue, malignant thyroid tumors reveal decreased, unchanged, or even increased levels of NIS mRNA and decreased levels of cell membrane-associated NIS protein expression with less pronounced basolateral orientation, indicating a defect in intracellular processing and trafficking. Correct NIS protein expression at the cell membrane appears to correlate inversely with the state of cellular differentiation, with more advanced tumor stages

being associated with lower levels of NIS expression. A lack of cell membrane-associated NIS protein expression has been observed in anaplastic and Hurthle cell carcinomas. These data suggest that oncogene activation may be involved in the suppression of NIS expression in thyroid cancer. In a limited number of cases, comparison of radioiodine scans and NIS expression patterns in thyroid carcinomas and metastases revealed a positive correlation between iodide uptake activity and NIS expression levels, suggesting that NIS expression levels might predict radioiodine uptake behavior and therapeutic effectiveness of radioiodine therapy in thyroid cancer. However, NIS mRNA expression has been detected in lymph node metastases with negative iodine uptake, indicating that NIS mRNA expression is not an adequate predictor of NIS functional activity.

### NIS IN THYROID AUTOIMMUNE DISEASES AND IN DISORDERS OF IODINE TRANSPORT

There is general agreement that NIS expression is markedly stimulated in active Graves' disease, whereas it is low or suppressed in Hashimoto's thyroiditis. In contrast, studies exploring the role of NIS autoantibodies in autoimmune thyroid diseases have generated inconclusive and often contradictory results, mainly due to methodological difficulties in identifying and differentiating linear and conformational NIS epitopes. Therefore, it has remained difficult to determine the existence, true prevalence, functional activities, and possible pathological implications of NIS-directed autoantibodies in autoimmune thyroid diseases. Using stringent cutoff criteria, NIS autoantibodies are detected in less than 6% of patients with Graves' disease and in less than 7% of patients with Hashimoto's thyroiditis. Thus, it appears that there is no clinical rationale for measuring NIS autoantibodies in patients with autoimmune thyroid diseases.

Congenital iodine transport defect (ITD) is a rare autosomal recessive condition caused by mutations in NIS. The clinical picture comprises hypothyroidism, the presence of a goiter, low thyroid iodine uptake, and a low saliva/plasma iodine ratio. Fifty-eight cases of ITD from 33 families have been reported. Among 27 cases from 13 families, nine NIS mutations have been identified. Only three mutations (T354P, Q267E, and G395R) have been characterized in greater detail, revealing some clues as to the molecular mechanisms that underlie the effects of these mutations. Continued study of naturally occurring NIS mutations is likely to

advance the understanding of functionally significant residues or segments of NIS.

### EXPRESSION OF SODIUM IODIDE SYMPORTER IN EXTRATHYROIDAL TISSUES

The thyroid gland shares its capacity to actively accumulate iodide with several other tissues, such as salivary and lacrimal gland, gastric mucosa, lactating mammary gland, choroid plexus, ciliary body of the eye, skin, nasopharynx, thymus, and placenta. Iodide transport in these nonthyroidal tissues is TSH-independent, but reveals several functional similarities to that in the thyroid gland, such as inhibition by thiocyanate and perchlorate and generation of iodide concentration gradients of similar magnitude. In addition, numerous data suggest that iodide is organified by certain extrathyroidal tissues. Detection of NIS mRNA and protein expression in salivary and lacrimal glands, gastric mucosa, and mammary gland suggests that iodide transport in these tissues is conferred by the expression of functional NIS protein. Inorganic iodide secreted by exocrine glands, such as salivary and lacrimal glands, and gastric mucosa followed by oxidation to hypoiodite may act as an antimicrobial agent, offering mucosal protection against environmental microorganisms. Consequently, salivary and lacrimal gland dysfunction observed in thyroid cancer patients after treatment with radioiodine may result from radioiodine accumulation in glandular tissue by NIS. Moreover, lactating breast tissue is capable of concentrating iodide in milk via NIS, thereby providing an adequate supply of iodide to the newborn.

### NIS-BASED THERAPEUTIC APPROACHES IN THYROIDAL AND NONTHYROIDAL CANCERS

As decreased NIS expression and iodide uptake in dedifferentiated thyroid carcinomas reduce the efficiency of radioiodine therapy, several novel therapeutic strategies to stimulate NIS expression levels in dedifferentiated thyroid cancer cells have been explored. A promising candidate seems to be retinoic acid, as it has been shown to suppress iodide uptake and NIS mRNA levels in normal, nontransformed FRTL-5 thyroid cells, whereas NIS mRNA expression levels were up-regulated in human follicular thyroid carcinoma cell lines *in vitro*. Treatment with retinoic acid

may therefore act to up-regulate iodide transport in thyroid cancer cells while down-regulating iodide accumulation in surrounding normal tissue, thereby reestablishing or enhancing the therapeutic effect of radioiodine therapy on malignant cells while sparing normal thyroid tissue. Furthermore, data suggest that DNA methylation may be involved in the loss of NIS expression in thyroid cancer, which might offer the possibility of restoring responsiveness to radioiodine treatment by chemical demethylation. A therapeutic effect was also achieved in mice xenotransplanted with transfected and NIS-expressing thyroid tumor cells following radioiodine treatment.

Tissue-specific expression of functionally active NIS in cancer cells without endogenous NIS may confer onto these cells the capacity to concentrate iodine from plasma and thereby establish the possibility of radioiodine therapy for nonthyroidal cancer cells. This therapeutic approach is theoretically feasible in all cancers that specifically express a particular gene, such as calcitonin (medullary thyroid cancer), prostate-specific antigen (prostate tumors), and carcinoembryonic antigen (colorectal cancer), by transferring an expression vector coding for the full-length human NIS under the control of this tissue-specific promoter. Several studies have demonstrated both radioiodine uptake and a selective cytotoxic effect of trapped radioiodine *in vitro* and *in vivo* in malignantly transformed cells without iodide transport activity, following transfection with a NIS-coding expression vector. Using retroviral or adenoviral vectors for NIS gene delivery, expression of functionally active NIS under the control of a nonspecific viral promoter was demonstrable in melanoma, liver, lung, breast, cervical, prostate, ovarian, and colon carcinoma cells, as well as in human follicular thyroidal, cervical, and breast cancer xenotransplants *in vitro* and *in vivo*. Although these data underscore the potential of NIS gene transfer to induce iodide accumulation activity in tumor cells, the therapeutic efficacy of accumulated radioiodine remains to be confirmed in an *in vivo* setting that resembles more closely the clinical situation.

Prostate cancer cells were shown to be selectively killed by accumulated  $^{131}\text{I}$  following induction of tissue-specific iodide uptake activity by prostate-specific antigen promoter-directed NIS expression *in vitro*. The magnitude of iodide uptake observed *in vivo* suggested that a therapeutic effect of accumulated radioiodine is likely to be achieved *in vivo*, which has been confirmed. Thus, prostate cell-specific NIS gene transfer appears to be a promising approach to prostate cancer therapy using radioiodine, particularly once the disease has

reached the metastatic stage. Whether and how this approach can be applied to its main clinical target, metastatic cancer of the prostate, remains to be seen. Furthermore, strong evidence of NIS expression in breast cancer has been reported. From these studies, it appears that active iodide transport in mammary gland tissue is mediated by NIS in healthy lactating mammary gland as well as in mammary tumors, but not in non-lactating mammary gland. In contrast to thyroidal NIS, expression of NIS by mammary gland tissue is not regulated by transcription factors such as Pax-8, TTF-1, and TTF-2, but is induced by a combination of estrogen, oxytocin, and prolactin. Over 80% of human breast cancer samples examined by immunohistochemistry expressed NIS immunoreactivity, compared with none of the normal, nonlactating breast tissue samples obtained from reductive mammoplasties. In contrast to normal thyroid or lactating mammary cells, expression of NIS protein was not restricted to the cell membrane of tumor cells, but was also detected intracellularly. Demonstration of NIS expression in mammary tumors may explain why these tumors are sometimes visualized by radioiodine and  $^{99\text{m}}\text{Tc}$ -pertechnetate scintigraphy. Knowing the level of functional NIS protein expression in breast cancer cells may be informative both therapeutically and prognostically. Furthermore, similar to expression of HER2/neu or the estrogen/progesterone receptor, NIS expression may play a role in the prognosis of patients. In NIS-positive breast cancer, additional radioiodine therapy might be considered, similar to anti-estrogen therapy in estrogen receptor-positive tumors, or chemotherapy combined with anti-HER2/neu antibody therapy in HER2/neu-positive breast cancers.

### See Also the Following Articles

Iodine • Nontoxic Goiter • Thyroid Hormone Metabolism • TSH (Thyroid-Stimulating Hormone; Thyrotropin)

### Further Reading

- Dai, G., Levy, O., and Carrasco, N. (1996). Cloning and characterization of the thyroid iodide transporter. *Nature* **379**, 458–460.
- Dohán, O., De la Vieja, A., Paroder, V., Riedel, C., Artani, M., Reed, M., Ginter, C. S., and Carrasco, N. (2003). The sodium/iodide symporter (NIS): Characterization, regulation and medical significance. *Endocr. Rev.* **24**, 48–77.
- Mandell, R. B., Mandell, L. Z., and Link, C. J. (1999). Radioisotope concentrator gene therapy using the sodium/iodide symporter gene. *Cancer Res.* **59**, 661–668.
- Schmutzler, C., Winzer, R., Meissner-Weigl, J., and Koehrle, J. (1997). Retinoid acid increases sodium/iodide symporter mRNA levels in human thyroid cancer cell lines and suppresses expression of functional symporter in nontransformed FRTL-5 rat thyroid cells. *Biochem. Biophys. Res. Commun.* **240**, 832–838.

- Smanik, P. A., Liu, Q., Furminger, T. L., Ryu, K., Xing, S., Mazzaferri, E. L., and Jhiang, S. M. (1996). Cloning of the human sodium iodide symporter. *Biochem. Biophys. Res. Commun.* **226**, 339–345.
- Smanik, P. A., Ryu, K.-Y., Thiel, K. S., Mazzaferri, E. L., and Jhiang, S. M. (1997). Expression, exon-intron organization, and chromosome mapping of the human sodium iodide symporter. *Endocrinology* **138**, 555–558.
- Spitzweg, C., Joba, W., Eisenmenger, W., and Heufelder, A. E. (1998). Analysis of human sodium iodide symporter gene expression in extrathyroidal tissues and cloning of its complementary deoxyribonucleic acids from salivary gland, mammary gland, and gastric mucosa. *J. Clin. Endocrinol. Metab.* **83**, 1746–1751.
- Spitzweg, C., Joba, W., Schriever, K., Goellner, J. R., Morris, J. C., and Heufelder, A. E. (1999). Analysis of human sodium iodide symporter immunoreactivity in human exocrine glands. *J. Clin. Endocrinol. Metab.* **84**, 4178–4184.
- Spitzweg, C., O'Connor, M. K., Bergert, E. R., Tindall, D. J., Young, C. Y., and Morris, J. C. (2000). Treatment of prostate cancer by radioiodine therapy after tissue-specific expression of the sodium iodide symporter. *Cancer Res.* **60**, 6526–6530.
- Tazebay, U. H., Wapnir, I. L., Levy, O., Dohan, O., Zuckier, L. S., Hhua Zhao, Q., Fu Deng, H., Amenta, P. S., Fineberg, S., Pestell, R. G., and Carrasco, N. (2000). The mammary gland iodide transporter is expressed during lactation and in breast cancer. *Nat. Med.* **6**, 871–878.

California, the presence of a factor with GH-release inhibiting activity (somatotropin-release inhibitory factor, SRIF) was confirmed. Only about 8.5 mg of a product with SRIF activity could be extracted from about 500,000 sheep hypothalamic fragments. The chemical structure of SRIF was determined and the tetradecapeptide was synthesized. The synthetic peptide had the biological activity of the native SRIF on pituitary GH release. The cyclic peptide was called somatostatin. P. Brazeau and colleagues reported the findings of this group in January 1973 in a *Science* article.

It took nearly another 10 years before two groups, those of R. Guillemin and of W. Vale, working independently at the Salk Institute, reported in 1982 the isolation and sequencing of GH-RF peptides, the target of research which led to the discovery of somatostatin.

The simple concept of somatostatin as a 14 amino acid-containing peptide whose main function is the inhibition of growth hormone secretion has been substantially widened through further research in the 1970s and 1980s. Somatostatin-related peptides constitute a family that includes the originally identified peptide (somatostatin-14), an amino-terminal extended somatostatin (somatostatin 28), several species-specific variants, and larger prohormone forms. Somatostatin is distributed widely in cells that have nothing to do with growth hormone regulation, including an extensive distribution within the nervous system, and is found in many tissues outside of the brain, including the gut and endocrine and exocrine glands, the retina, the immune system, and walls of blood vessels.

### Biological Effects of Somatostatin

In addition to being a potent inhibitor of pituitary GH secretion, somatostatin also inhibits the secretion of other hormones such as thyrotropin (TSH), insulin, glucagon, pancreatic polypeptide (PP), and vasoactive intestinal polypeptide (VIP). Furthermore, a variety of physiological functions such as gastrointestinal motility, gastric acid production, pancreatic exocrine secretion, bile, and colonic fluid secretion are also inhibited. In the light of this additional knowledge acquired in the 10 years following its discovery, it is evident that the name somatostatin for this peptide does not take into account the multitude of its activities that reach far beyond GH inhibition. Somatostatin acts as a neurohormone, a neurotransmitter, or as a local factor acting via autocrine or paracrine mechanisms.

### SOMATOSTATIN RECEPTORS

Actions of somatostatin are mediated through specific membrane receptors. They have been shown in the anterior pituitary, the endocrine and exocrine pancreas, the mucosa of the gastrointestinal tract, and in various regions of the brain, as well as in cells of the immune system. Five subtypes of the human somatostatin receptor (sst<sub>1</sub> to sst<sub>5</sub>) have been cloned and characterized. These subtypes are identical in 42 to 60% of their amino acid sequences. They belong to the superfamily of G protein-coupled receptors with seven membrane-spanning domains. The genes for these subtypes are located on different chromosomes, suggesting different functions in different organs.

All five receptor subtypes are able to mediate the inhibition of adenylate cyclase activity. Signal transduction mechanisms not coupled with specific somatostatin receptor subtypes are the activation of potassium channels and inhibition of voltage-dependent calcium channels that are reversed by pertussis toxin. In the brain, somatostatin activates phospholipase C and A and mobilizes calcium. Stimulation of sst<sub>1</sub> and sst<sub>2</sub> results in the activation of tyrosine phosphatase, which is related to the antimetabolic effects of somatostatin in some types of cells. In addition, inhibition of cell proliferation is mediated through sst<sub>5</sub>, probably involving changes in intracellular calcium mobilization.

Table I shows a summary of the biological effects mediated through the different subtypes of somatostatin receptors.

### SOMATOSTATIN ANALOGS

The ability of somatostatin to inhibit a variety of physiological processes created expectations that this peptide could be of therapeutic value in clinical conditions involving hyperfunction of several organ systems. However, native somatostatin has several properties that limit its therapeutic use. First, the elimination half-life is only about 3 min, making continuous intravenous infusions mandatory. Somatostatin exerts at pharmacological concentrations multiple simultaneous effects in different organs. This lack of selectivity could lead to adverse events. After stopping a somatostatin infusion, a rebound effect with post-infusion hypersecretion of hormones is observed. There was therefore a clear need to synthesize analogs that are more stable to enzymatic degradation and have a more selective activity than native somatostatin.

Several groups of peptide chemists were involved in the design of analogs with improved properties

**Table I** Summary of Selected Biological Effects Mediated through the Different Somatostatin Receptor Subtypes (sst)

Biological effect	Sst <sub>1</sub>	Sst <sub>2</sub>	Sst <sub>3</sub>	Sst <sub>4</sub>	Sst <sub>5</sub>
<i>Inhibition of secretion</i>					
GH	(+)	+			+
Insulin		+			+
Glucagon		+			
Prolactin		(+)			+
Serotonin		+			
Gastric acid		+			
Amylase					+
<i>GI motility</i>					
Gastric and intestinal relaxation			+		
Colonic contraction					+
<i>Cell proliferation</i>					
Cell cycle arrest	+	+		+	+
Induction of apoptosis		+	+		
Inhibition of angiogenesis	+	+			

compared to natural somatostatin. However, the first analogs tested in humans showed no significant advantages. It was an increased understanding of structure–activity relationships of the somatostatin molecule that allowed the synthesis of smaller, more stable, and more potent molecules.

The same methodology that was used for the discovery of somatostatin was also applied for the characterization of analogs. It was based on their ability to inhibit the release of GH in monolayer cultures of rat pituitary cells. This approach led to the selection of analogs with high potency for GH inhibition, or as it is now understood, a high affinity to somatostatin receptor subtype 2.

In 1982, W. Bauer *et al.*, working in the laboratories of the former Sandoz Company (now Novartis) in Basel, Switzerland, introduced octreotide to the scientific community. Octreotide represented the first

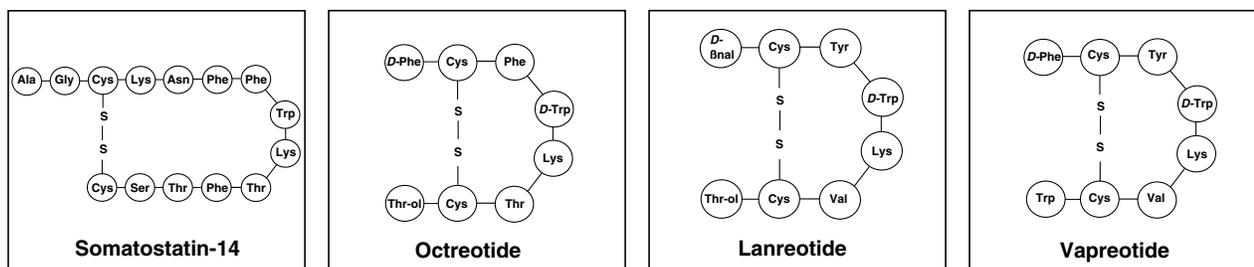
analog that obtained marketing authorization. Its duration of action after subcutaneous injection is between 6 and 12 hours for GH inhibition and the elimination half-life is 90 min. The potency of octreotide for GH, glucagon, and insulin inhibition in monkeys is 45 times, 11 times, and 1.3 times higher, respectively, than that of native somatostatin 14. No rebound hypersecretion of hormones occurs after cessation of octreotide administration.

Other cyclic analogs with similar activity profiles, lanreotide and vapreotide, were clinically tested and in the case of lanreotide also marketed in some countries outside the United States.

Figure 1 shows the amino acid structure of somatostatin and somatostatin analogs and Table II compares their affinity for the five somatostatin receptor subtypes. Somatostatin-14 and somatostatin-28 bind with high affinity to all five somatostatin receptor subtypes, whereas the analogs octreotide, lanreotide, and vapreotide bind predominantly to sst<sub>2</sub> and to a lesser degree also to sst<sub>3</sub> and sst<sub>5</sub>. They have very low affinities for sst<sub>1</sub> and sst<sub>4</sub>.

### Long-Acting Depot Formulations

The first formulation of octreotide for clinical use was an immediate release aqueous solution for subcutaneous injection. It has to be administered two or three times daily in doses of 100 or 150 µg. Individual patients may need daily doses up to 1500 µg. The observation that a given dose of octreotide administered as a continuous subcutaneous infusion was more efficacious than three daily subcutaneous injections, along with the inconvenience of three-times daily injections, led to the development of depot formulations for intramuscular injection. The active molecules are embedded in microspheres of DL-lactide-co-glycolide polymer. The duration of action of octreotide LAR is 28–42 days and that of lanreotide SR 7–14 days. Octreotide LAR is supplied as 10 mg, 20 mg, and 30 mg and lanreotide SR as 30 mg. Recently, a



**Figure 1** Amino acid structure of somatostatin-14 and somatostatin analogs.

Table II Binding Properties of Human Somatostatin Receptor Subtypes  $sst_1$ – $sst_5$  Expressed in Chinese Hamster Ovary Cells ( $sst_1$ – $sst_4$ ) or African Green-Monkey COS-1 Cells ( $sst_5$ )

Peptide	pK <sub>i</sub> [–log M]				
	Sst <sub>1</sub>	Sst <sub>2</sub>	Sst <sub>3</sub>	Sst <sub>4</sub>	Sst <sub>5</sub>
Somatostatin-14	9.2	9.7	9.6	8.9	9.4
Somatostatin-28	9.2	9.7	9.5	9.1	9.4
Octreotide	6.6	9.5	8.3	< 6.0	8.3
Lanreotide	6.3	9.3	7.2	< 6.0	8.4
Vapreotide	6.8	10.1	7.7	6.8	8.3

new once-monthly aqueous preparation of lanreotide, lanreotide autogel, has become available in some European countries. For chronic treatment with somatostatin analogs, long-acting depot preparations are the formulations of choice.

## CLINICAL APPLICATIONS OF SOMATOSTATIN ANALOGS

### Pituitary Adenomas

#### ***GH-Secreting Pituitary Adenomas***

In more than 95% of patients, acromegaly is caused by a GH-secreting pituitary adenoma. Therapeutic options include surgery, radiotherapy and drug treatment. Transsphenoidal surgery can achieve biochemical disease control, defined as circulating concentration of  $GH \leq 2.5 \mu\text{g/liter}$  and insulin-like growth factor-1 (IGF-1) levels within the age- and sex-adjusted normal range, in up to 90% of patients harboring tumors with a diameter of  $\leq 10 \text{ mm}$ , a so-called microadenoma. However, larger tumors, especially those with extrasellar extension, reach postoperative biochemical disease control in less than 50%. Radiotherapy takes up to 10 years before being fully effective, and hypopituitarism is a frequent side effect. Dopamine agonists have been used for the treatment of acromegaly since the mid 1970-s. High doses of these drugs are needed, and the proportion of patients who can reach a satisfactory response is maximally 40%.

The addition of somatostatin analogs to the armamentarium of physicians in the late 1980s revolutionized the therapy of acromegalic patients. Chronic treatment with octreotide LAR can achieve biochemical control in 60–70% of patients. Symptoms of acromegaly improve significantly within days of the start of therapy. Several publications have reported a reduction in the size of the pituitary adenoma in the majority

of patients during treatment with somatostatin analogs. There are no reports of an increase in size of GH-secreting adenomas in patients who responded to somatostatin analogs with a decrease in the circulating GH concentration as long as therapy is continued. It has been suggested that somatostatin analogs may be used as primary medical therapy in acromegalic patients who have a small chance of surgical cure or who are unwilling to undergo an operation. Presurgical treatment for up to 3 months before surgery is also employed by some centers; the purpose is to improve the patient's condition and thereby reduce the perioperative morbidity as well as change the consistency of the adenoma to ease extirpation. The evidence of the usefulness of preoperative somatostatin analog treatment is not strong.

#### ***TSH-Secreting Pituitary Adenomas***

TSH-secreting pituitary adenomas causing hyperthyroidism are very rare tumors. They respond to somatostatin analog treatment with a reduction in TSH levels and a normalization of serum thyroxin concentrations in about 75% of patients.

### Functioning Endocrine Tumors of the Gastro-Entero-Pancreatic System

Endocrine tumors of the gastro-entero-pancreatic (GEP) system are rare. They are divided into functionally active and inactive tumors. The clinical symptoms of the functionally active tumors are mainly caused by the secretory products released into the circulation. They usually present with metastatic disease, not allowing curative surgical resection. Table III summarizes the leading clinical symptoms/syndromes, the main secretory tumor product, the main primary site of these tumors, and the percentage of malignancy of these tumors. Nearly all of the functionally active tumors express somatostatin receptors. Sst<sub>2</sub> are found in 80 to 90% of them. The therapeutic response to the available somatostatin analogs is correlated with the density of the sst<sub>2</sub> expression of these tumors. Symptoms of most patients are controlled with daily doses of 300 to 450  $\mu\text{g}$  of octreotide divided into three doses, 20 mg of octreotide LAR once monthly, or 30 mg of lanreotide SR every 10 or 14 days.

#### ***Types of GEP Endocrine Tumors***

##### ***Carcinoid Tumors and Carcinoid Syndrome***

Carcinoid tumors behave differently depending on the localization of the primary. For example, appendiceal carcinoids are practically never malignant,

Table III Main Characteristics and Leading Symptoms of GEP Endocrine Tumors

Tumor	Leading clinical symptoms	Main tumor product	Location of primary tumor	Percentage malignant
<b>Carcinoid tumor</b>	Carcinoid syndrome (flushing, diarrhea); dependent on location of primary tumor	Serotonin, tachykinins, histamine, (ACTH, GH-RH, etc.)	Small intestine (20%), bronchial system (15%), appendix (40%), rectum (20%), other locations (5%)	2–80% (dependent on location of primary tumor)
<b>VIPoma</b> (Verner–Morrison Syndrome)	Watery diarrhea, hypokalemia, achlorhydria	VIP	Pancreas	80–90%
<b>Glucagonoma</b>	Necrolytic migratory erythema, diabetes, cachexia	Glucagon	Pancreas	60%
<b>Insulinoma</b>	Hypoglycemia	Insulin	Pancreas	5–10%
<b>Gastrinoma</b> (Zollinger–Ellison Syndrome)	Peptic ulcers, diarrhea	Gastrin	Pancreas (30–60%), duodenum (30–40%), other locations (10–20%)	60–90%

whereas midgut carcinoid tumors are more than 80% malignant, and secrete high amounts of serotonin, histamine, tachykinins, and peptides that can cause the carcinoid syndrome. The carcinoid syndrome is a systemic manifestation of the release of tumoral secretion into the systemic circulation, comprising flushing of the face and upper part of the body, diarrhea, cardiac valvular lesions, and wheezing. Somatostatin analogs are indicated only for the treatment of the carcinoid syndrome. Their usefulness in the therapy of functionally inactive carcinoid tumors has not been established. Flushing episodes and diarrhea are improved or resolved in up to 80% of patients during treatment with somatostatin analogs. A significant reduction of the main biochemical marker, urinary excretion of 5-hydroxy-indolic acetic acid (5-HIAA), is observed in up to 70%. The median duration of biochemical response is about 12 months.

Somatostatin analogs have been successfully used to prevent or treat the life-threatening carcinoid crisis, which can be triggered by laparotomy or embolization of liver metastases.

#### *VIPoma (Verner-Morrison Syndrome)*

Control of watery, high-volume diarrhea and decrease of elevated levels of circulating VIP can be achieved in 80 to 90% of patients with inoperable VIPoma. Somatostatin analogs are the treatment of choice for symptoms due to the VIP hypersecretion of this tumor type.

#### *Glucagonoma*

Necrolytic migratory erythema, diarrhea, mild diabetes mellitus, and protein hypercatabolism are the leading symptoms in patients with the glucagonoma syndrome. Treatment with somatostatin analogs can significantly improve these symptoms.

#### *Insulinoma*

Insulinomas express in only about 50% of sst<sub>2</sub>. Somatostatin analog therapy will reduce tumoral insulin secretion only in the presence of sst<sub>2</sub>. However, in their absence, somatostatin analogs suppress GH and glucagon more pronounced than tumoral insulin, resulting in a worsening of hypoglycemia. The presence or absence of sst<sub>2</sub> can be determined *in vivo* with OctreoScan.

#### *Gastrinoma (Zollinger-Ellison Syndrome)*

Somatostatin analogs inhibit gastrin release from tumor cells and suppress gastric acid secretion by a direct effect on the parietal cell. However, with the introduction of proton pump inhibitors, the importance of

somatostatin analog treatment has been reduced. There may be subgroups of malignant gastrinoma patients that could benefit from this treatment.

### **General Remarks on the Treatment of GEP Endocrine Tumors with Somatostatin Analogs**

#### *Tachyphylaxis*

Loss of therapeutic effect to somatostatin analog treatment after initial response is a well-recognized phenomenon in patients harboring GEP endocrine tumors. The median duration of response to somatostatin analog therapy is about 12 months, but the beneficial effects can decrease, in a small number of patients, just a few days after initiation of therapy. The loss of response may be overcome by an increase of the dose or a drug-free interval. The mechanism for tachyphylaxis is largely unknown. Interestingly, this phenomenon does not exist in acromegalic patients.

#### *Antiproliferative Effects of Somatostatin Analogs in GEP Endocrine Tumors*

*In vitro* and *in vivo* studies have demonstrated antiproliferative effects of somatostatin analogs. The growth rate of endocrine tumors of the GEP system is in general slow but variable: in addition to long phases of spontaneous tumor standstill or even tumor regression without any treatment, rapid increase in tumor burden have been observed. Using the classical oncological criteria for tumor growth assessment can be misleading. During chronic somatostatin analog treatment, regression of the tumor size is observed in about 10% and stabilization was reported in up to 70%. There are, however, still many open questions such as whether somatostatin analog therapy prolongs survival in patients harboring GEP endocrine tumors, duration of antiproliferative effect, or optimal dose of somatostatin analogs.

### **Prevention of Complications Following Pancreatic Surgery**

Patients undergoing major pancreatic surgery are at high risk of developing postoperative complications related to exocrine secretion such as fistula, fluid collection, and leakage of the anastomosis. Somatostatin analogs temporarily inhibit the exocrine pancreatic secretion. Several double-blind, randomized, placebo-controlled studies have shown that octreotide administered subcutaneously started at least 1 hour before laparotomy and continued for 5–7 days is able to reduce the postoperative complication rate significantly.

## Esophageal Variceal Bleeding

More than 50% of patients with liver cirrhosis will develop esophageal varices. About one-third of them will experience variceal bleeding within 2 years of diagnosis. Thirty percent of those who bleed will eventually die as a consequence of the bleeding episode. Somatostatin and its analogs have been shown to decrease the hepatic venous pressure gradient and azygos blood flow in patients with liver cirrhosis. Intravenous infusion of somatostatin or octreotide has been helpful in the management of patients with bleeding varices. Administration shortly after admission to the hospital was able to stop bleeding and to facilitate the endoscopic intervention. Continuation of the infusion for 2–5 days reduced the number of rebleeding episodes and the need for blood transfusions. However, somatostatin or octreotide infusions did not influence the long-term mortality.

## Diarrhea

Severe refractory diarrhea may constitute a serious and challenging clinical issue. Diarrhea occurring in conditions such as AIDS, after antitumor chemotherapy or radiotherapy applied to the abdomen or pelvis, or in diabetes mellitus or intestinal graft-versus-host disease may fail to respond to specific antimicrobial treatment. The rationale for the use of somatostatin analogs in diarrhea is based on their effects on the gastrointestinal tract: inhibition of gastrointestinal motility, inhibition of intestinal and exocrine pancreatic secretion, inhibition of gastrointestinal peptides, and induction of net water and electrolyte absorption. The quality of most of the clinical studies performed in these conditions is not optimal and the response rates are highly variable. Well-designed controlled trials with a sufficient sample size are needed to define the value of somatostatin analogs in the treatment of patients with refractory diarrhea.

## Diabetic Complications

Endocrine and growth factors that may play a role in the pathogenesis of later stages of diabetic retinopathy and nephropathy include GH, IGF-1, and vascular endothelial growth factor (VEGF). The role of GH for the progression of diabetic retinopathy is especially supported by experimental data and clinical observations. A small controlled trial demonstrated that treatment with octreotide was able to stop or delay the progression of preproliferative diabetic retinopathy. Another study in patients who underwent pan-retinal

laser photocoagulation for proliferative diabetic retinopathy showed that octreotide could decrease neovascularization and prevent vitreous hemorrhages. However, larger controlled studies are needed to confirm these findings.

There were only short-term experimental studies performed for the investigation of the effects of somatostatin analogs in diabetic nephropathy. They demonstrated that octreotide treatment can reduce albuminuria.

## Oncological Indications

The rationale for the use of somatostatin analogs for solid nonsecreting tumors is mainly based on three potential mechanisms of action: (1) Many malignant tumors originating from tissue that is free of somatostatin receptors under physiological condition start to express somatostatin receptors at advanced stages. The somatostatin receptor expression is not homogeneously distributed in the nonendocrine tumoral tissues. (2) Growth factors such as IGF-1, GH, or gastrin influence the proliferation rate of several tumors. Somatostatin analogs are potent inhibitors of some of these factors. (3) Octreotide has been shown to possess angiogenesis inhibitory properties.

Cell line experiments but also a series of experiments with transplanted tumors in animals supported the antiproliferative activity of somatostatin analogs. Clinical trials in different adenocarcinomas, including breast and pancreas, were not able to confirm the postulated antitumor effect of monotherapy with the presently available sst<sub>2</sub>-predominant somatostatin analogs.

## Other Clinical Applications

In the mid 1980s, it was hoped that somatostatin analogs would be beneficial for improving blood glucose control in type 1 diabetic patients. The hypothesis was that inhibition of nightly surges of GH secretion would decrease the early morning hyperglycemia known as the dawn phenomenon. Another potential application was the reduction of acute bleeding from gastric or duodenal ulcers, the rationale being the somatostatin-induced decrease of splanchnic blood flow. For both of these two indications, large clinical programs did not demonstrate a benefit of treatment with somatostatin analogs.

The effects of somatostatin analogs were studied in a multitude of other applications. Some of the pilot studies gave positive results, e.g., external pancreatic fistula, hyperinsulinemic hypoglycemia syndromes

of infancy, and hypoglycemia due to overdosage of sulfonylureas. For various reasons, none of them was proposed for a full clinical development.

## SAFETY AND TOLERABILITY OF SOMATOSTATIN ANALOGS

Chronic therapy with somatostatin analogs is generally well tolerated. Transitory adverse events frequently seen after initiating somatostatin analog treatment include abdominal discomfort, diarrhea, steatorrhea, nausea, and bloating. They can be explained by the inhibition of the exocrine pancreatic secretion and effects on the motility of the gastrointestinal tract. These adverse drug reactions usually improve without treatment within 10 days, despite continued therapy with somatostatin analogs.

Somatostatin and its analogs inhibit the secretion of hormones involved in carbohydrate metabolism, such as insulin, glucagon, and growth hormone. Furthermore, the absorption of carbohydrates is delayed during somatostatin analog therapy. Long-term treatment with somatostatin analogs seldom leads to overt diabetes mellitus, and the occurrence of hypoglycemia is even rarer.

Formation of cholesterol-rich gallstones occurs in 20–30% of patients on chronic therapy with somatostatin analogs. Only about 1% of them cause symptoms. Their prevalence varies geographically and may be influenced by dietary or racial factors. Reduced gallbladder contractility through inhibition of cholecystokinin and decreased gastrointestinal motility resulting in increased intestinal and biliary production of deoxycholic acid are contributing to the increased prevalence of gallstones.

## RADIOLABELED SOMATOSTATIN ANALOGS

In the late 1980s, E. Krenning and S. W. J. Lamberts, in Rotterdam, The Netherlands, together with researchers from the former Sandoz company (now Novartis) developed the concept of targeting  $ss_{t_2}$ -expressing tissues with an octreotide derivative coupled with a  $\gamma$ -rays-emitting isotope for diagnostic purposes. The product of their efforts is  $^{111}\text{In}$ -linked with a chelating agent to D-Phe-octreotide, which has been marketed under the name OctreoScan since 1994. Somatostatin receptor scintigraphy has proved especially helpful for staging of GEP endocrine tumors. An extension of the use of radioactive isotopes bound to

somatostatin analogs for therapy of tumors with high expression of  $ss_{t_2}$  was proposed by the same group. One of the preferred radionuclides for therapeutic purposes is the  $\beta$ -particle emitter  $^{90}\text{Y}$ . Compounds using this principle of targeted radiotherapy are being clinically studied.

## FUTURE OF SOMATOSTATIN ANALOGS

Somatostatin analogs available for clinical use have been selected from a large number of analogs based on their ability to inhibit pituitary GH secretion, long before somatostatin receptor subtypes were cloned. The therapeutic potential of somatostatin analogs with high affinity for receptor subtypes other than  $ss_{t_2}$  is largely unknown. I. Shimon, from S. Melmed's group in Los Angeles, and A. Saveanu, from P. Jaquet's laboratories in Marseille, France, have shown that a combination of an analog with high affinity for  $ss_{t_2}$  and one for  $ss_{t_5}$  is able to inhibit GH secretion of cell cultures of GH-secreting pituitary adenomas stronger than octreotide or lanreotide alone. An  $ss_{t_5}$  agonist—in contrast to  $ss_{t_2}$  agonists—also decreased prolactin secretion of prolactinoma cells. These examples, together with the increase in knowledge on specific physiological regulatory roles of the individual somatostatin receptor subtypes and their expression in tumoral and nontumoral tissues, could significantly expand the potential for new applications of  $ss_{t}$  selective somatostatin analogs.

Y. C. Patel's group in Montreal, Canada, published data that generated new insights into somatostatin receptor physiology. They demonstrated ligand-induced somatostatin receptor homo- or heterodimerization of  $ss_{t_1}$  and  $ss_{t_5}$ . In another experiment, they showed that the dopamine receptor D2R and  $ss_{t_5}$  can also interact through hetero-oligomerisation. This creates a new receptor with enhanced functional activity.

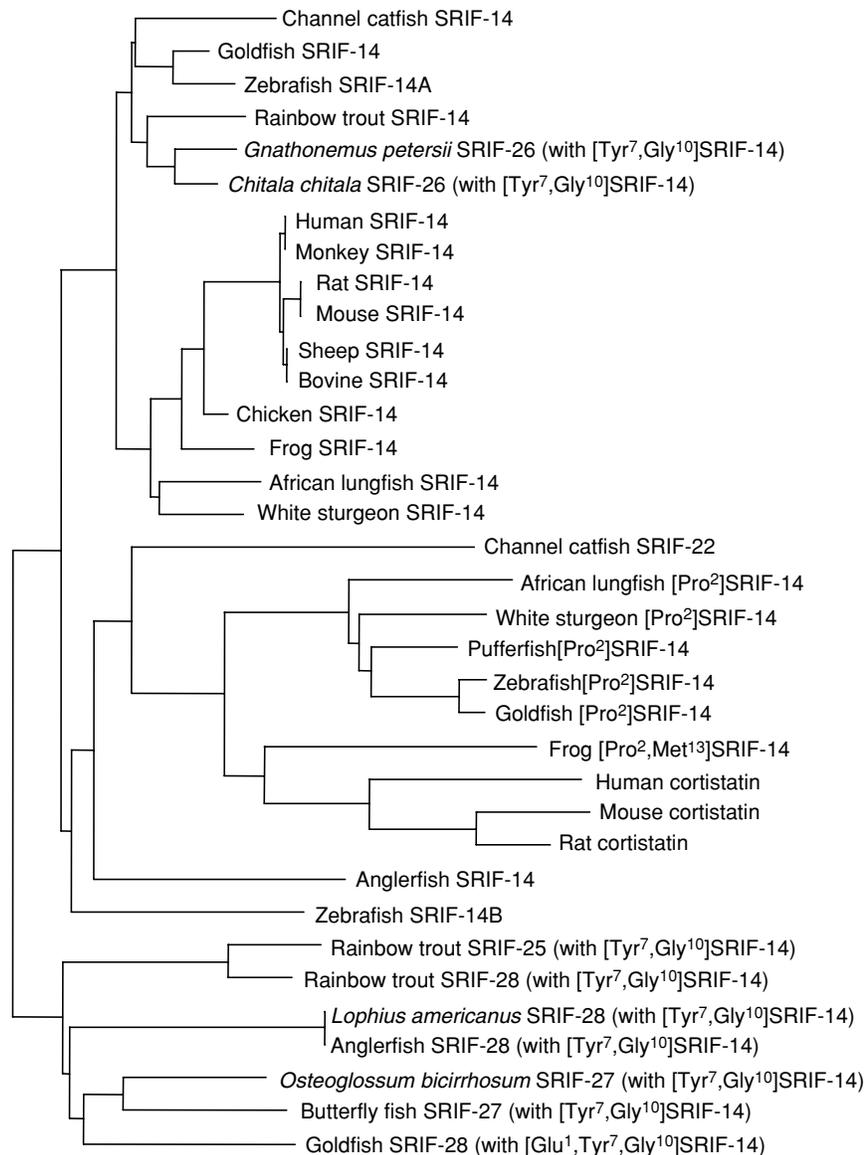
The findings that there are interactions between somatostatin receptor subtypes and that there is a cross-talk with other G protein-coupled systems creates new opportunities for medicinal chemists. Somatostatin analogs that will be designed considering this knowledge could result in drugs with an increase in the responder rate in the classical somatostatin analog indications, but they could also represent new therapeutic approaches for autoimmune and endocrine diseases. Oncological applications may also become targets for the next generation of somatostatin analogs.

### See Also the Following Articles

Acromegaly, Clinical Features of • GI Hormones in Cancer • G Protein-Coupled Receptors • Growth Hormone (GH) • Peptide Neurotransmitters and Smooth Muscle in the Gut • Pituitary Adenomas, TSH-Secreting • Somatostatin, Evolution of

### Further Reading

- Arnold, R., Wied, M., and Behr, T. H. (2002). Somatostatin analogues in the treatment of endocrine tumours of the gastrointestinal tract. *Expert Opin. Pharmacother.* **3**, 643–656.
- Hejna, M., Schmidinger, M., and Raderer, M. (2002). The clinical role of somatostatin analogues as antineoplastic agents: Much ado about nothing? *Ann. Oncol.* **13**, 653–668.
- Lamberts, S. W. J., van der Lely, A. J., de Herder, W. W., and Hofland, L. J. (1996). Octreotide. *N. Engl. J. Med.* **334**, 246–254.
- Lamberts, S. W. J., van der Lely, A. J., and Hofland, L. J. (2002). New somatostatin analogs: Will they fulfil old promises? *Eur. J. Endocrinol.* **146**, 701–705.
- Patel, Y. C., Greenwood, M. T., Panetta, R., Demchyshyn, L., Niznik, H., and Srikant, C. B. (1995). Mini review: The somatostatin receptor family. *Life Sci.* **57**, 1249–1265.
- Sheppard, M. C. (2003). Primary medical therapy for acromegaly. *Clin. Endocrinol.* **58**, 387–399.



**Figure 1** Phylogenetic tree based on the alignment of amino acid sequences of the known preprosomatostatins in vertebrates. The multiple sequence alignment and phylogenetic tree were derived using the public domain ClustalX program with default parameters. The phylogenetic tree in ClustalX uses the neighbor-joining method on a matrix of “distances” between all sequences. The branch lengths are proportional to the estimated divergence along each branch.

products. The PSS-II gene identified in a number of teleost has not been found in tetrapods.

In goldfish brain, three SRIF genes have been identified. Two of them code for PSS-I and PSS-II, whereas the third gene from goldfish brain encodes a precursor (PSS-III) with [Pro<sup>2</sup>]SRIF-14 at its C terminus. [Pro<sup>2</sup>]SRIF-14 peptide was first identified in Russian sturgeon and then its cDNA was cloned from African lungfish, pufferfish, zebrafish, and sturgeon. In frog, SRIF-14 and [Pro<sup>2</sup>, Met<sup>13</sup>]SRIF-14 peptides have been isolated and their cDNAs

cloned from brain. In addition, a SRIF-related gene, termed cortistatin (CST), has been described in rat, mouse, and human. The cortistatin precursor contains a tetradecapeptide at its C terminus sharing an 11-amino-acid homologous sequence with SRIF-14. A second SRIF cDNA coding for [Pro<sup>2</sup>]SRIF-14 was isolated from chicken. These extend the existence of a multigene family for SRIF to tetrapods.

A 22-amino-acid peptide (SRIF-22), which differs from SRIF-14 at 7 of 14 residues, and the cDNA for SRIF-22 have been identified in catfish. The cDNA

sequence for SRIF-22 shows less than 30% similarity to those of the cDNAs for teleost PSS-I and PSS-II.

## PHYLOGENETIC RELATIONSHIP OF SOMATOSTATIN GENES

According to phylogenetic analysis based on the amino acid sequence alignment of the known SRIF precursors deduced from the known cDNA or gene sequences (Fig. 1), the vertebrate SRIF genes can be grouped into three major clades. PSS-I genes, which code for the SRIF-14 precursor, are highly conserved from fish to mammals. Anglerfish SRIF-14 precursor and the second zebrafish SRIF-14 precursor show a relatively lower degree of homology to other SRIF-14 precursors in vertebrates. PSS-II from several teleost species are grouped together in a separate clade, except that the PSS-II from *G. petersii* and *C. chitala* share a high level of homology with fish PSS-Is. This suggests that the teleost PSS-II gene is phylogenetically close to the PSS-I gene and the PSS-II from *G. petersii* and *C. chitala* could serve as intermediate forms between the two genes. The precursors of [Pro<sup>2</sup>]SRIF-14 from several teleost species (goldfish, African lungfish, sturgeon, zebrafish, and pufferfish) are highly homologous to that from chicken (not shown in Fig. 1) and these [Pro<sup>2</sup>]SRIF-14 precursors also show homology to frog [Pro<sup>2</sup>,Met<sup>13</sup>]SRIF-14 precursor and mammalian CST precursors. Generally, the PSS-I gene product is widely distributed in brain and peripheral tissues including the gastrointestinal (GI) tract. However, the mammalian CST gene is expressed in a restricted fashion in the brain with functions in neuronal depression and sleep modulation. In frog, [Pro<sup>2</sup>,Met<sup>13</sup>] gene expression is restricted to the brain. Similarly, the [Pro<sup>2</sup>]SRIF-14 gene is expressed only in brain of African lungfish. In goldfish, the PSS-I and PSS-II genes are expressed in brain and peripheral tissues, such as the GI tract, whereas the PSS-III ([Pro<sup>2</sup>]SRIF-14) gene is expressed in brain, but not in the GI tract. The restricted expression of the third SRIF gene in brain supports a close phylogenetic relationship among the members of this gene group.

Multigene families for neuropeptides arise mainly by duplication of individual genes or duplication of the entire genome (e.g., tetraploidization) and subsequent accumulation of mutations. Extreme conservation of the primary structure of SRIF-14 in vertebrates suggests that the SRIF gene likely existed earlier than the appearance of cyclostomes. Notably, SRIF-14 immunoreactivity has been detected in invertebrate species, plant, and *Escherichia coli* extracts. Based on the phylogenetic relationship of SRIF

precursors and the phylogenetic distribution of SRIF forms, a scheme for SRIF evolution has been hypothesized. Duplication of an ancestral SRIF-14-like gene occurred at an early stage of vertebrate evolution. One SRIF-14 gene is conserved throughout vertebrate evolution from cyclostomes to mammals, whereas the other SRIF-14 gene gave rise to the [Ser<sup>12</sup>]SRIF14 gene in cyclostomes, the [Ser<sup>5</sup>]SRIF-14 gene in holocephalians, and the [Pro<sup>2</sup>]SRIF-14 gene in dipnoans and chondrosteans by accumulation of mutations. In some primitive fish species closely related to dipnoans, from which ancestral amphibians were derived, another duplication of the [Pro<sup>2</sup>]SRIF-14 gene may have taken place, giving rise to the amphibian [Pro<sup>2</sup>,Met<sup>13</sup>]SRIF-14 and mammalian CST genes. The [Pro<sup>2</sup>]SRIF-14 gene is maintained in some primitive fish species (e.g., lungfish, sturgeon), in some teleost species (e.g., goldfish), and in avian species (e.g., chicken). On the other hand, duplication of the SRIF-14 gene in fish gave rise to a gene coding for PSS-II containing [Tyr<sup>7</sup>,Gly<sup>10</sup>]SRIF-14 in teleosts.

## See Also the Following Articles

ACTH,  $\alpha$ -MSH, and POMC, Evolution of • Angiotensin, Evolution of • GI Hormone Development (Families and Phylogeny) • Insulin and Insulin-like Growth Factors, Evolution of • Natriuretic Peptide System, Evolution of • Neuropeptide Y, Evolution of • Opioid/Orphanin Gene Family, Evolution of • Prolactin, Evolution of • Steroid Receptors, Evolution of

## Further Reading

- Conlon, J. M., Tostivint, H., and Vaudry, H. (1997). Somatostatin and urotensin II-related peptides: Molecular diversity and evolutionary perspectives. *Regul. Peptides* **69**, 95–103.
- Lin, X. W., Otto, C. J., and Peter, R. E. (1998). Evolution of neuroendocrine peptide systems: Gonadotropin-releasing hormone and somatostatin. *Comp. Biochem. Physiol.* **119C**, 375–388.
- Lin, X., Otto, C. J., Cardenas, R., and Peter, R. E. (2000). Somatostatin family of peptides and its receptors in fish. *Can. J. Physiol. Pharmacol.* **78**, 1053–1066.
- Patel, Y. C. (1999). Somatostatin and its receptor family. *Front. Neuroendocrinol.* **20**, 157–198.
- Spier, A. D., and de Lecea, L. (2000). Cortistatin: A member of the somatostatin neuropeptide family with distinct physiological functions. *Brain Res. Rev.* **33**, 228–241.
- Su, C.-J., White, J. W., Li, W.-H., Luo, C.-C., Frazier, M. L., Saunders, G. F., and Chan, L. (1988). Structure and evolution of somatostatin genes. *Mol. Endocrinol.* **2**, 109–216.
- Tannenbaum, G. S., and Epelbaum, J. (1999). Somatostatin. In "Handbook of Physiology, Section 7, The Endocrine System, Vol. V, Hormonal Control of Growth" (J. L. Kostyo and H. M. Goodman, eds.), pp. 221–265. Oxford University Press, New York.

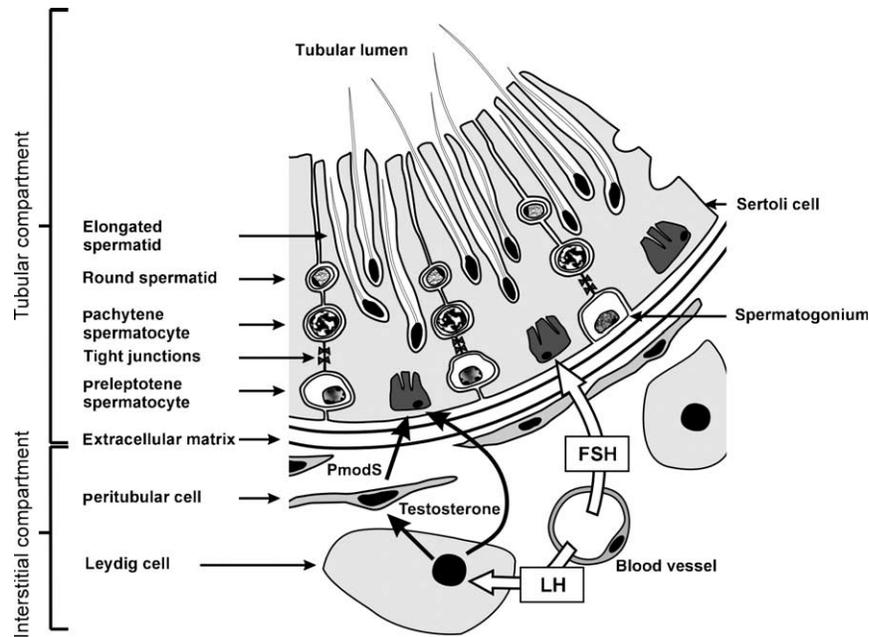


Figure 1 Organization of spermatogenesis.

in an adluminal compartment containing all later germ cell types. The blood–testis barrier is very important since the milieu in the adluminal compartment differs considerably from serum or interstitial fluid and is stringently controlled by the Sertoli cells. Complex mechanisms allow the germ cells to enter the adluminal compartment without disrupting the blood–testis barrier.

Temporally, the successive changes in developing germ cells follow a very strict time schedule and, as a result, only particular associations of cells are observed in transverse sections of testicular tubules. Such associations are indicated as stages of spermatogenesis. In human, six stages are routinely described. A spermatogenic cycle is the time required to pass through these successive stages (16 days in human). The complete development from spermatogonium to spermatozoon requires approximately 74 days (more than four cycles). Successive stages of spermatogenesis are observed not only on transverse sections but also along the length of testicular tubules (a wave of spermatogenesis). In rats and mice, transverse sections reveal only one particular stage. In human, the situation is more complicated since stages are arranged in an intertwining helical pattern. As a result, single-tubule cross sections may display up to six different stages. Interestingly, studies in rats have revealed that Sertoli cells also undergo cyclic functional changes during progression of the spermatogenic cycle.

## FACTORS CONTROLLING QUANTITY AND QUALITY OF SPERM OUTPUT

The final efficiency and quality of sperm production depend on at least three determinants: a genetic program inherent to the germ cells themselves, a number of variables defining maximal spermatogenic capacity, and a series of regulatory factors deciding whether the system is used at maximal capacity or not. This article focuses only on the second and third of these determinants in which hormones play a definite role. The role of factors inherent to the germ cells themselves is probably illustrated most clearly by germ cell transplantation experiments showing that essential characteristics, such as the duration of the spermatogenic cycle, are maintained when rat germ cells are transplanted into a mouse testis.

## DETERMINANTS OF MAXIMAL SPERMATOGENIC CAPACITY

The capacity of the testis to produce sperm cells differs considerably from species to species. In human, for instance, total daily sperm production expressed per gram of testis is five times lower than that in rat, rabbit, or rhesus monkey. Maximal spermatogenic capacity depends on the number of germ cells that can be supported by an individual Sertoli cell, the number of Sertoli cells available, the number

of germ cells entering the spermatogenic cycle, and the number of germ cells that become lost during the entire process.

The number of germ cells that can be nurtured by an individual Sertoli cell varies considerably from species to species. It is three times lower in human than in rabbit or rat.

The number of Sertoli cells determines the ultimate ceiling of sperm output and the size of the testis. Sertoli cell proliferation is maximal in the perinatal period and ceases at the time of puberty, concordant with Sertoli cell maturation and the formation of tight inter-Sertoli cell junctions. This explains why, after unilateral orchidectomy or other testicular damage, compensatory hypertrophy is observed only prepubertally. In human, some proliferation may continue throughout the prepubertal period. The most important regulator of Sertoli cell replication is FSH and, accordingly, absent or reduced levels of FSH will limit Sertoli cell numbers, testis size, and sperm output. Research findings indicate that, apart from FSH, thyroid hormone may also affect testicular size. Hypothyroidism in the prepubertal period delays Sertoli cell maturation, prolongs the period of FSH-dependent proliferation, and increases testicular size as well as sperm output.

The number of germ cells entering the spermatogenic cycle depends on species-specific characteristics, such as the number of mitotic divisions at the level of type A spermatogonia and hormonal factors affecting spermatogonial proliferation.

Finally, it is important to know that large numbers of germ cells undergo apoptosis during germ cell development. In rat, cell losses are observed mainly during spermatogonial stages and may exceed 75%. In human, losses seem to be most pronounced during meiosis and may reach as high as 40%. This reduction in germ cells may serve to adapt the number of germ cells to the germ cell supportive capacity of Sertoli cells and may also prevent further progression of (chromosomally) aberrant germ cells through spermatogenesis. Disturbances in the endocrine and local control mechanisms affecting spermatogenesis, however, may markedly increase germ cell apoptosis.

## ENDOCRINE CONTROL OF SPERMATOGENESIS

An overwhelming amount of data indicates that gonadotropins are the major hormones controlling spermatogenesis (Fig. 2). Partial or complete inhibition of germ cell production is observed after procedures that

suppress gonadotropins [e.g., hypophysectomy, gonadotropic hormone-releasing hormone (GnRH) agonists or antagonists, immunization against GnRH] or selectively reduce LH (e.g., supraphysiological doses of testosterone in the rat) or FSH (FSH antibodies). None of these procedures is both completely effective and selective, however. Moreover, the effects of a particular intervention may differ depending on the species studied and whether the selected endpoint is initiation, maintenance, or restoration of spermatogenesis. It is safe to say that quantitatively and qualitatively normal spermatogenesis requires both LH (androgens) and FSH, but that the individual actions and interactions of these hormones remain only partially understood. Reports of rare mutations as well as the development of knockout animals have offered additional clues to the relative role of FSH and androgens in spermatogenesis.

## FSH and Spermatogenesis

FSH does not act on the germ cells directly but activates specific receptors on the Sertoli cells. The FSH receptor belongs to a large family of membrane receptors that are coupled to G proteins and that are characterized by seven hydrophobic helices forming the transmembrane domain. This family also comprises the receptors for LH and TSH. Ligand binding results in activation of the adenylate cyclase signaling pathway and other signaling pathways and induces a cascade of protein phosphorylations that ultimately mediates the intracellular effects.

As discussed above, FSH is the major factor responsible for the proliferation of Sertoli cells during prepubertal life and accordingly its presence during this critical window is of major importance in determining the ultimate size and spermatogenic capacity of the testis. However, FSH action may not be compulsory for pubertal onset of spermatogenesis and for adult fertility. The most convincing evidence for this contention comes from a report of five Finnish males who are homozygous for a mutation in the FSH receptor gene that severely impairs signal transduction. As expected, these men have reduced testicular volumes but two of them fathered two children each. Semen examination revealed a variable degree of oligozoospermia and teratozoospermia, but none of the five men displayed azoospermia. Very similar observations were made in FSH receptor knockout mice. In these animals also, a lack of FSH action resulted in small testis size and partial spermatogenic failure. Fertility was reduced in one study and normal in

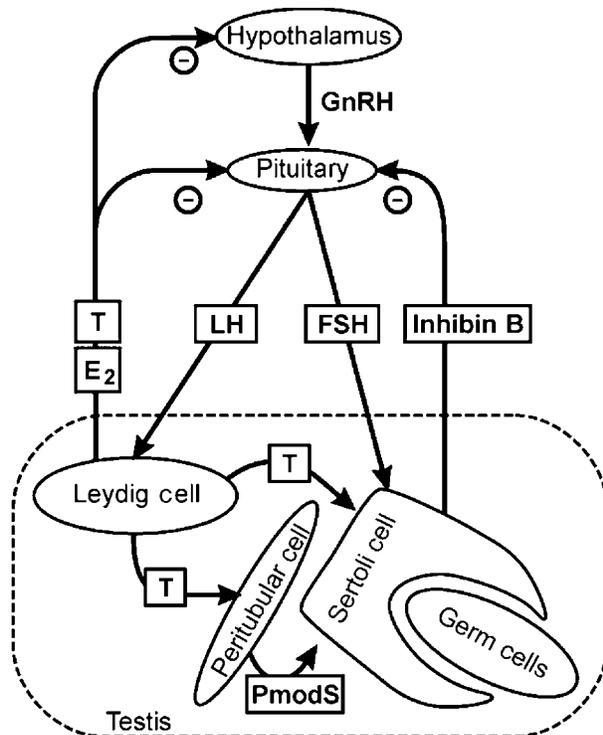


Figure 2 Endocrine control of testicular function.

another. Taken together, these observations suggest that FSH is required for quantitatively and qualitatively normal spermatogenesis, but that, in the presence of normal testicular androgen production, fertility is possible in the absence of FSH action.

Conversely and surprisingly, under some conditions, FSH may be able to maintain spermatogenesis in the presence of low concentrations of androgens. Relevant to this issue is the anecdotal description of a hypophysectomized patient undergoing treatment with testosterone who had normal spermatogenesis despite undetectable levels of gonadotropins and in which spermatogenesis appeared to be maintained by an activating mutation of the FSH receptor.

Stereological studies in rodents as well as in monkeys suggest a major and specific role for FSH in the proliferation of spermatogonia. In addition, FSH seems to favor spermatocyte and round spermatid development, most probably by supporting cell survival. Finally, and in cooperation with testosterone, FSH may promote the release of mature spermatozoa from Sertoli cells. The last two effects might be related to the actions of FSH on cytoskeletal elements and on the specialized Sertoli cell junctional apparatus.

## Androgens and Spermatogenesis

Androgens play a key role in the control of spermatogenesis. This key role is probably best illustrated by the fact that androgens alone can restore spermatogenesis and fertility in a number of test systems where FSH is extremely low or absent, including severely hypogonadotropic mice deficient in GnRH (*hpg* mice). Many questions remain, however, regarding the ultimate target cells for androgen action, the concentration of androgens required to maintain spermatogenesis, the exact effects of androgens, and the role played by active androgen metabolites.

Similar to FSH, androgens do not act directly on the germ cells. The effects of androgens are thought to be mediated by the androgen receptor (AR), a member of the nuclear receptor superfamily, and most studies have failed to demonstrate this AR in germ cells. Moreover, transplantation experiments of AR-deficient germ cells ( $T^{fm}X/Y$  mice) into the seminiferous tubules of azoospermic mice expressing a functional AR resulted in complete and qualitatively normal donor-derived spermatogenesis. Accordingly, the effects of androgens on spermatogenesis are probably indirect and mediated by somatic testicular cells. Sertoli cells are considered the prime candidates. These cells express the AR, they interact intimately with the developing germ cells, and they control the intratubular environment. Nonetheless, attempts to identify specific genes or molecular events controlled by androgens in Sertoli cells have been rather disappointing and the possibility that androgens have effects on other somatic components of the testis cannot be excluded. Peritubular myoid cells surrounding the seminiferous tubules, for instance, also express AR and some data suggest that under the influence of androgens these cells produce one or more local mediators referred to as "PmodS," modulating Sertoli cell function in an androgen-controlled way.

The maintenance of quantitatively and qualitatively normal sperm production apparently requires intratesticular levels of testosterone that are 20–50 times higher than those observed in the circulation. These high concentrations are normally present in the testis as a result of local production by LH-stimulated Leydig cells. Administration of physiological or slightly suprphysiological amounts of testosterone, however, suppresses LH, reduces these high intratesticular concentrations of testosterone, and suppresses spermatogenesis. It is unclear why spermatogenesis requires such high levels of androgens. Given the low level of 5 $\alpha$ -reductase in testicular target cells, testosterone rather than its active metabolite

5 $\alpha$ -dihydrotestosterone (DHT) is considered to be the active androgen in the testis. The absence of DHT formation might explain the need for 3- to 10-fold higher concentrations of testosterone. It is unlikely that this is the only mechanism involved, however. The possibility of other mechanisms reducing androgen sensitivity or alternative pathways of androgen action cannot be excluded.

Stereological approaches to the study of spermatogenesis and particularly the so-called optical disector model have contributed considerably to a more quantitative evaluation of the effects of androgens on germ cell development. In rats, testosterone seems to be essential for the adhesion of round spermatids to Sertoli cells. In its absence, round spermatids are sloughed from the epithelium and spermatid elongation fails. Spermiation also is under the control of testosterone (and FSH) in rat as well as in monkey and human. Studies on the contraceptive effects of testosterone indicate that impairment of spermiation may be a key to achieving azoospermia. Rodent models provide little evidence that androgens promote spermatogonial proliferation. Rather, studies on the restoration of spermatogenesis in irradiated rats suggest that high concentrations of androgens are detrimental to spermatogonial development.

In many target organs, androgen action is mediated by so-called active metabolites. Testosterone can be 5 $\alpha$ -reduced to form DHT, a more potent androgen with a higher affinity for the AR. Alternatively, it can be aromatized to 17 $\beta$ -estradiol, which can act via one of the estrogen receptors (ER- $\alpha$  or ER- $\beta$ ). As explained above, testosterone is considered to be directly responsible for the maintenance of spermatogenesis under normal conditions. However, under conditions in which intratesticular testosterone levels are reduced, for instance, by the administration of testosterone-based contraceptives, an up-regulation of 5 $\alpha$ -reductase activity has been observed. Similarly, in rats treated with suboptimal concentrations of testosterone, coadministration of a 5 $\alpha$ -reductase inhibitor further impairs the production of mature spermatids. These findings suggest that suppression of intratesticular testosterone may induce a compensatory reaction involving enhanced conversion to DHT. Differences in the ability to activate this reaction could contribute to the observed interracial and interindividual differences in the efficiency of testosterone-based contraceptives.

Along the same lines, there is increasing evidence that aromatization of testosterone to 17 $\beta$ -estradiol is required for normal spermatogenesis. Aromatase activity and estrogen receptors are widely distributed in the somatic and germinal elements of the testis and in

the efferent ducts. Moreover, fertility defects have been described in mice with a knockout of the ER- $\alpha$  or the aromatase. Some of these defects may be related to defective fluid resorption at the level of the efferent tubules and retrograde accumulation of fluid, causing Sertoli cell dysfunction. Additional sites of action for estradiol, however, seem very likely. Only two men with mutations resulting in aromatase deficiency and one man with a mutated ER- $\alpha$  have been described. Data available are too limited to allow definitive conclusions, but there are indications that here too germ cell development may be impaired.

### Interactions of Testosterone and FSH

The data discussed above indicate that both FSH and testosterone affect multiple steps in the spermatogenic process, though to a variable degree. Although under some conditions one of these hormones may maintain spermatogenesis in the absence of the other, qualitatively and quantitatively normal spermatogenesis requires both hormones. Multiple sites of interaction have been identified. Some parameters of Sertoli cell function require both androgens and FSH for maximal expression (e.g., secretion of androgen-binding protein [ABP], expression of N-cadherin) and others (e.g., aromatase activity) are regulated in an opposite fashion by FSH (stimulation) and testosterone (suppression). Different stages of the cycle may display a variable need for either androgens or FSH. In rat Sertoli cells, AR concentration and androgen sensitivity are maximal at stages VII and VIII, whereas FSH effects are more pronounced at stages XII to II. Finally, AR concentration in Sertoli cells is up-regulated by FSH, whereas maturation of Leydig cells and production of androgen are promoted by Sertoli cell-derived paracrine regulators produced under the influence of FSH.

The mechanisms by which androgen- and FSH-induced changes in Sertoli cell function ultimately affect germ cell development also remain far from understood. As already mentioned, both hormones contribute to the development of specific cell-cell contacts that may allow adhesion, migration, and release of germ cells as well as exchange of information and material between Sertoli and germ cells. Furthermore, Sertoli cells control the concentrations of ions, nutrients, and transport molecules responsible for the specific environmental conditions in the adluminal compartment. Finally, FSH and androgens undoubtedly affect the complex network of paracrine communications responsible for bilateral communications between germ cells and Sertoli cells.

## CONCLUSIONS

Quantitatively and qualitatively normal spermatogenesis requires both FSH and high local concentrations of androgens. FSH plays a key role in determining the ultimate number of Sertoli cells and the spermatogenic capacity of the testis, but in rodents and in human some degree of sperm production and fertility may be maintained by high concentrations of androgens in the virtual absence of FSH. The reasons for the need for high local concentrations of androgens and the exact role played by active androgen metabolites remain unknown. Androgen-based contraceptives may act both by suppressing intratesticular androgen levels and by lowering FSH. The data available suggest that both FSH and androgens affect germ cell development indirectly by modulating the function of somatic testicular cells and particularly of Sertoli cells. The processes and genes involved are only partially understood but it may be anticipated that approaches such as DNA microarray technology will provide powerful tools to clarify these issues in the near future.

### See Also the Following Articles

Androgen Biosynthesis Inhibitors and Androgen Receptor Antagonists • Contraception, Male • Fertility in Men with Spermatogenesis Abnormalities • Fertilization • FSH

(Follicle-Stimulating Hormone) • Germ Cell Differentiation Signaling Events, Male • Gonadotropin-Releasing Hormone (GnRH) Actions • Gonadotropins and Testicular Function in Aging • Sexual Function and Androgens

### Further Reading

- Amory, J. K., and Bremner, W. (2001). Endocrine regulation of testicular function in men: Implications for contraceptive development. *Mol. Cell. Endocrinol.* **182**, 175–179.
- McLachlan, R. I., O'Donnell, L., Meachem, S. J., Stanton, P. G., de Kretser, D. M., Pratis, K., and Robertson, D. M. (2002). Identification of specific sites of hormonal regulation in spermatogenesis in rats, monkeys, and man. *Rec. Prog. Horm. Res.* **57**, 149–179.
- McLean, D. J., Friel, P. J., Pouchnik, D., and Griswold, M. D. (2002). Oligonucleotide microarray analysis of gene expression in follicle-stimulating hormone-treated rat Sertoli cells. *Mol. Endocrinol.* **16**, 2780–2792.
- Nieschlag, E., Simoni, M., Gromoll, J., and Weinbauer, G. F. (1999). Role of FSH in the regulation of spermatogenesis: Clinical aspects. *Clin. Endocrinol.* **51**, 139–146.
- O'Donnell, L., Robertson, K. M., Margaret, E. J., and Simpson, E. R. (2001). Estrogen and spermatogenesis. *Endocr. Rev.* **22**, 289–318.
- Sharpe, R. M. (1994). Regulation of spermatogenesis. In "The Physiology of Reproduction" (E. Knobil and J. D. Neill, eds.), Chap. 22, pp. 1363–1434. Raven Press, New York.
- Themmen, A. P. N., and Huhtaniemi, I. T. (2000). Mutations of gonadotropins and gonadotropin receptors: Elucidating the physiology and pathology of pituitary–gonadal function. *Endocr. Rev.* **21**, 551–583.

Table I Definitions of Metabolic Syndrome in Men

	National Cholesterol Evaluation Program 2001	World Health Organization 1998
	<i>Three or more from</i>	<i>One of the two major criteria<sup>a</sup> AND at least two minor criteria<sup>b</sup></i>
Insulin resistance		<sup>a</sup> Insulin resistance in top 25% by euglycemic hyperinsulinemic clamp
Hyperglycemia	Fasting plasma glucose $\geq 6.1$ mM	<sup>a</sup> Impaired glucose tolerance or type 2 diabetes mellitus
Hypertriglyceridemia	Serum triglycerides $\geq 1.7$ mM; HDL cholesterol $< 1.0$ mM	<sup>b</sup> Serum triglycerides $\geq 1.7$ mM or HDL cholesterol $< 0.9$ mM
Hypertension	BP $\geq 130/85$ mm Hg	<sup>b</sup> BP $\geq 160/90$ mm Hg
Obesity	Waist circumference $> 102$ cm	<sup>b</sup> WHR $> 0.90$ or BMI $\geq 30$ kg/m <sup>2</sup>
Other		<sup>b</sup> Microalbuminuria

Note. BMI, body mass index; BP, blood pressure. HDL, high-density lipoprotein; WHR, waist/height ratio.

popularly dubbed the “pear” distribution. Once the metabolic syndrome was defined and its boundaries were ratified, it became the subject of intense investigation, particularly investigation of potential mechanisms that might explain the links between hitherto discrete disorders within the syndrome.

The metabolic syndrome is common, affecting up to 20% of Western populations and increasingly so. Each component is prevalent and the key effector processes, such as insulin resistance, are becoming much more frequent, driven by the rapidly rising prevalence of obesity. The changes in the incidence of obesity are too fast to be accounted for by genetic alterations. Instead, they are commonly attributed to the dietary excess and relative sedentary nature of modern urban life, acting on a basic genotype primed by natural selection over millennia of human evolution to avidly retain calories in times of plenty for the inevitable times of famine to come (the thrifty genotype hypothesis). It is estimated that more than 30% of Western populations will have abdominal obesity (waist circumference greater than 40 in. in men and greater than 37 in. in women) within the next 20 years. This is associated with very substantially increased morbidity (type 2 diabetes mellitus, dyslipidemia, hypertension, and heart disease) and mortality. The disorder is more common among men, but women after menopause have an increasing incidence of the metabolic syndrome, which has implicated sex steroid deficiency in its pathogenesis.

## THE METABOLIC SYNDROME AND GLUCOCORTICOID

The metabolic syndrome strikingly resembles Cushing’s syndrome, which is caused by circulating

glucocorticoid excess (either due to elevated levels of endogenous cortisol or to exogenous pharmacotherapy). A series of studies suggests that plasma cortisol is subtly elevated in subjects with hypertension, glucose intolerance, and dyslipidemia. However, it has not been possible to reliably document plasma glucocorticoid excess in obesity. Plasma cortisol has a pronounced diurnal rhythm, with concentrations being highest on waking and lowest on retiring to bed. Although in obesity, plasma cortisol levels are sometimes slightly elevated at the diurnal nadir in the afternoon and evening, they are typically low at the morning peak. Thus, although elevated circulating cortisol has been invoked as a cause of the association between some aspects of the metabolic syndrome, it does not explain amplification of the features of the syndrome by obesity. A series of studies has taken an entirely different tack and examined tissue sensitivity to glucocorticoids in obesity and the metabolic syndrome.

## TISSUE GLUCOCORTICOID SENSITIVITY

Tissue sensitivity to glucocorticoids is classically determined by the relative density and affinity of intracellular receptors for cortisol. There are two receptors: the lower affinity glucocorticoid receptor (GR) and the higher affinity mineralocorticoid receptor (MR). Both are members of the nuclear hormone receptor superfamily of ligand-activated transcription factors. As far as the metabolic syndrome is concerned, the primary target tissues (liver, adipose tissue, vascular tissue) typically express high levels of GR, although MRs are reported in the heart and vessels too.

The effects of alterations in GR or MR function depend on the tissue concerned. GR mediates the adverse effects of glucocorticoids on glucose and lipid metabolism in liver, adipose tissue, and skeletal muscle. MR may contribute to adverse effects in more selected sites including blood vessels, heart, and kidney. However, both receptors are also involved in glucocorticoid negative feedback control of the hypothalamic–pituitary–adrenal (HPA) axis, which suppresses endogenous glucocorticoid production in the presence of glucocorticoid excess.

“Global” changes in GR function, affecting both peripheral tissues and central HPA feedback sites, may not influence net GR activation because of a compensatory change in glucocorticoid circulating levels, although there may be other consequences. For example, a number of families (fewer than 10) have glucocorticoid resistance due to deleterious mutations of the GR gene. Affected individuals show HPA axis activation, which compensates for the glucocorticoid resistance, but the concomitant elevated adrenocorticotropic hormone levels drive adrenal androgen excess, producing masculinization in females, and elevated deoxycorticosterone levels, which are assumed to be responsible for the hypertension observed. The patients do not have obesity or the metabolic syndrome.

A key finding, however, is that control of glucocorticoid sensitivity is tissue-specific rather than generalized. Thus, there is no correlation between sensitivity to glucocorticoids in tests of vascular responses and HPA feedback. This is perhaps not surprising given the numerous alternative promoters/first exons of the GR gene that show some tissue-associated independence of expression.

A series of studies of patients with components of the metabolic syndrome have suggested increased glucocorticoid sensitivity in some peripheral tissues, notably the vasculature and skeletal muscle. At least in muscle, the change occurs at the level of mRNA expression. Such increased tissue sensitivity in the absence of compensatory decreased cortisol levels, since HPA feedback sensitivity is unaltered, might be anticipated to increase glucocorticoid signaling in the affected organs, producing disease. The overall importance of this change remains to be fully determined.

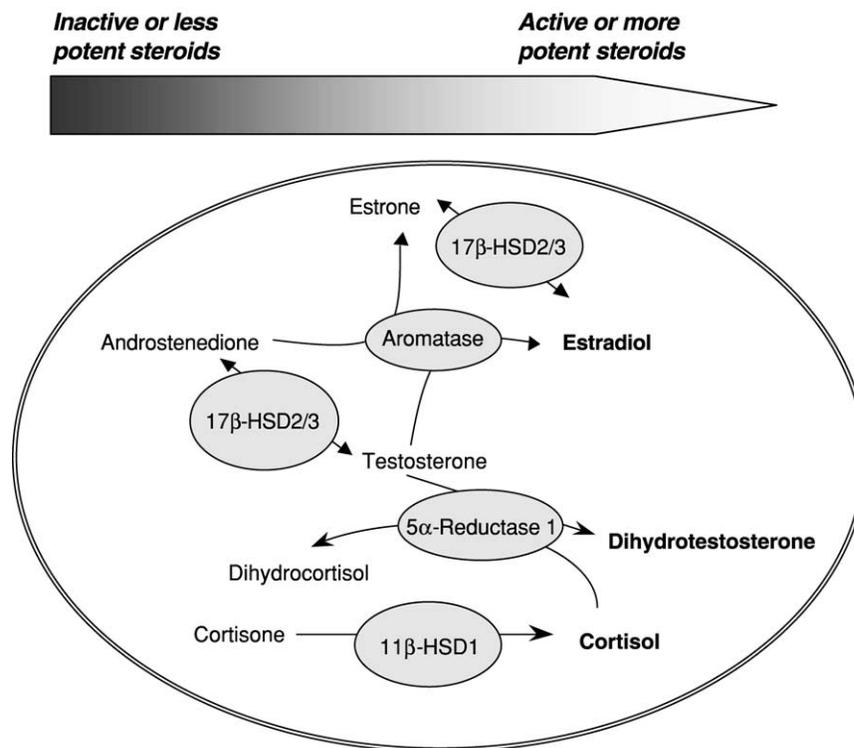
The possibility of selective changes in HPA feedback sensitivity has been researched in great depth in patients with obesity and the metabolic syndrome. In general, little conclusive evidence has emerged, although some data in obese rats and humans suggest a possible role for MR dysfunction in producing

feedback insensitivity in some such individuals. This is difficult biology to dissect in humans, because of the overriding effects of stress.

## 11 $\beta$ -HYDROXYSTEROID DEHYDROGENASE TYPE 1

Recent studies have suggested that the actions of glucocorticoids are determined not only by the concentration of active steroids in the circulation and the tissue density of intracellular GR and MR, but also by prereceptor metabolism by enzymes (Fig. 1). The best understood are the 11 $\beta$ -hydroxysteroid dehydrogenases (11 $\beta$ -HSDs). 11 $\beta$ -HSDs were first identified over 50 years ago. They are microsomal enzymes that catalyze the interconversion of active 11-hydroxy glucocorticoids (cortisol in most mammals including humans and corticosterone in rats and mice) and their inert 11-keto forms (cortisone and 11-dehydrocorticosterone), which do not bind to GR or MR. There are two isozymes of 11 $\beta$ -HSD that are the products of distinct genes and differ considerably in tissue distribution and function. 11 $\beta$ -HSD type 2 (11 $\beta$ -HSD2) has a well-recognized function as a potent dehydrogenase that rapidly inactivates glucocorticoids, thus allowing aldosterone selective access to the otherwise nonselective MR in the distal nephron. Humans or mice homozygous for deleterious mutations of the 11 $\beta$ -HSD2 gene exhibit the syndrome of apparent mineralocorticoid excess in which cortisol and (or) corticosterone illicitly occupy renal MR, causing sodium retention, severe hypertension, and hypokalemia.

The function of the original 11 $\beta$ -HSD type 1 (11 $\beta$ -HSD1) isozyme, though purified and characterized from rat liver almost two decades ago by Monder and his colleagues, has been much less readily discerned. 11 $\beta$ -HSD1 is highly expressed in liver, adipose tissue, and the central nervous system, but until the 1990s, its function was entirely obscure. The key to understanding the role of 11 $\beta$ -HSD1 came from studies of its reaction direction in intact cells cultured *in vitro*. Although 11 $\beta$ -HSD1 shows bidirectional activity in tissue homogenates and purified microsomal preparations, it acts predominantly as a reductase in intact cells transfected with 11 $\beta$ -HSD1 cDNA. Such cells convert 11-dehydrocorticosterone in the medium to corticosterone and not vice versa. This is not merely due to differential access of 11-hydroxy and 11-keto steroids through the cell membrane, since both steroids can activate reporter genes driven by GR or MR in cells transfected with 11 $\beta$ -HSD1



**Figure 1** Intracrine conversions of relatively inactive steroids to potent steroids within the adipocyte.

cDNA but only corticosterone activates transcription in the same cell line lacking 11 $\beta$ -HSD1. This is also not merely an effect confined to clonal cells, since intact primary cultures of hepatocytes, lung cells, neurons, and adipocytes also show predominant 11 $\beta$ -reduction of inert 11-keto steroids to their active forms. This suggests that it is the subcellular context that is likely to determine 11 $\beta$ -HSD1 reaction direction in intact cells; presumably the environment in the inner leaflet of the endoplasmic reticulum where the enzyme is thought to lie favors 11 $\beta$ -reduction. This 11 $\beta$ -reductase reaction direction, far from inactivating glucocorticoids, regenerates active glucocorticoids in target cells from the substantial levels of inert 11-keto steroids in the blood, themselves produced largely by the actions of renal 11 $\beta$ -HSD2.

The 11 $\beta$ -reductase activity of 11 $\beta$ -HSD1 also predominates in intact organs and *in vivo*. Thus, the intact liver perfused *in situ* converts ~50% of 11-dehydrocorticosterone to active corticosterone on a single pass, whereas dehydrogenation of corticosterone is four- to fivefold lower. Moreover, the liver can activate large amounts of 11-dehydrocorticosterone, suggesting a large turnover of this reaction direction *in vivo*. In humans, reduction also predominates across the liver.

## 11 $\beta$ -HSD1 AND THE METABOLIC SYNDROME

Of course, recognizing that key metabolic tissues such as liver and adipose tissue express a glucocorticoid-regenerating enzyme adds a novel dimension to hypotheses of the etiology of the metabolic syndrome–Cushing’s syndrome paradox. A series of studies in genetically obese rat and mouse strains has shown two- to threefold increased activity of 11 $\beta$ -HSD1 selectively in adipose tissue, and particularly in abdominal adipose tissue, whereas the enzyme is down-regulated in the liver in the same animals. In humans, most but not all studies suggest the same pattern in obese subjects of two- to threefold increased 11 $\beta$ -HSD1 expression in adipose tissue (in this case, using superficial adipose tissue biopsies, not intra-abdominal fat) and yet decreased enzyme activity in liver.

To examine the pathogenic potential of such increased 11 $\beta$ -HSD1 in adipose tissue alone, mice that overexpress (two- to threefold) the enzyme in adipose tissue under the fat-specific aP-2 promoter have been generated. These mice are viable but have elevated intra-adipose corticosterone levels, whereas plasma levels of glucocorticoids are unaltered. The

consequences are striking. The mice develop selective abdominal obesity due to marked mesenteric fat hypertrophy and also exhibit glucose intolerance, insulin resistance, dyslipidemia, and hypertension. They also show hyperphagia, though expression of the transgene is confined to adipose tissues. Thus, modest overexpression of 11 $\beta$ -HSD1 to levels mimicking those seen in obese rodents and humans can faithfully reproduce all key features of the metabolic syndrome.

### 11 $\beta$ -HSD1 DEFICIENCY AND PROTECTION FROM METABOLIC DISORDERS

A series of studies with (nonselective) 11 $\beta$ -HSD inhibitors has suggested that reduced 11 $\beta$ -HSD1 activity increases insulin sensitivity and/or reduces plasma glucose and perhaps lipid levels in a variety of animal models and humans *in vivo*. Such studies, though interesting, are complicated by the lack of selectivity of the existing licorice-derived inhibitors, which act on both 11 $\beta$ -HSD isozymes and related steroid dehydrogenases, albeit with much lower affinity. More convincing have been studies in mice homozygous for targeted disruption of the 11 $\beta$ -HSD1 gene (11 $\beta$ -HSD1 knockout mice). These animals have clearly shown that 11 $\beta$ -HSD1 is the sole 11 $\beta$ -reductase, at least in mice, since adrenalectomized 11 $\beta$ -HSD1 knockout animals cannot regenerate active corticosterone from inert 11-dehydrocorticosterone. Moreover, the reaction appears to be important since the animals, despite normal or modestly elevated plasma corticosterone levels, resist hyperglycemia associated with stress or a high-fat diet. This appears to be due to attenuated responses of glucocorticoid-sensitive gluconeogenic enzymes in the liver, indicating a reduction in glucocorticoid levels within the hepatocyte. This contention is supported by lower levels of other glucocorticoid-sensitive enzymes in the liver, including those catalyzing the metabolism of lipids. This associates with reduced plasma triglycerides and increased HDL cholesterol levels, producing an apparently “cardioprotective” lipid profile. As predicted, 11 $\beta$ -HSD1 knockout mice are insulin sensitized, both overall in plasma and in specific cells and tissues, notably liver and adipose tissue. Strikingly, when bred onto an obesity-prone background, 11 $\beta$ -HSD1 knockout mice are modestly resistant to weight gain on high-fat feeding, an effect that does not appear to be due to altered calorie intake. Overall, the findings suggest that deficiency of this amplifier of intracellular glucocorticoids leads to a beneficial cardio-metabolic

risk factor profile, though whether it actually attenuates cardiovascular outcomes remains an unresolved question.

### OTHER GLUCOCORTICOID-METABOLIZING ENZYMES AND THE METABOLIC SYNDROME

There are other glucocorticoid-metabolizing enzymes, the function of which is altered in obesity and metabolic syndrome (Fig. 1). These include the A-ring reductase, 5 $\alpha$ -reductase, and 5 $\beta$ -reductase, which convert glucocorticoids to dihydro- and tetrahydro-metabolites, which are excreted mainly as conjugates in urine. Increased A-ring reductase activity in obesity may underlie the tendency to lower plasma cortisol levels and compensatory activation of the HPA axis in obesity. The consequences do not necessarily affect GR activation in peripheral tissues, but may include an increased drive to adrenal androgen production. In women, this may underlie the complaint of excessive hair growth in a male pattern (hirsutism), which, in polycystic ovary syndrome, coassociates with other features of the metabolic syndrome and is aggravated by obesity.

### AROMATASE

Another key determinant of body fat distribution is the balance between estrogens and androgens in adipose tissues. Estrogen deficiency, for example, after menopause, is associated with a change in fat distribution from a gynecoid (pear-shaped) to an android (apple-shaped) pattern and with increased risk of cardiovascular disease.

Estrogens are produced from androgens by aromatase (Cyp19), which is expressed in the ovaries and also in target tissues including muscle and fat, but not liver. Estrogen action within adipose tissue, thought largely to be mediated via the  $\alpha$ -subtype of the estrogen receptor (ER $\alpha$ ) in the adipocyte nucleus, is therefore dependent not only on circulating estrogens but also on local production from circulating androgens (Fig. 1). Aromatase is expressed predominantly in subcutaneous adipose beds as opposed to visceral fat. In addition, aromatase influences androgen receptor activation by “consuming” androgen substrates.

The importance of aromatase in determining ER activation is illustrated in a series of clinical paradigms. Premenopausal women generate abundant circulating estrogens from their ovaries. By contrast, men throughout life rely on adipose production of

estrogen from circulating testosterone (from the testes) and androstenedione (from dehydroepiandrosterone produced by the adrenals). These steroids enter adipose tissue, yielding estradiol and estrone. In postmenopausal women, however, ovarian estrogen (and testosterone) production is negligible and androstenedione from the adrenal gland is the major substrate, yielding estrone, which relies on further transformation by 17 $\beta$ -HSD to estradiol for activation. Paradoxically then, postmenopausal women are more estrogen-deficient than men of a similar age, which is thought to explain the relative protection of men from disorders such as osteoporosis and may explain the “masculinization” of body fat distribution after menopause. A role for aromatase in determining body fat distribution has been further highlighted by the generation of aromatase knockout mice, which have striking abdominal obesity in both genders. There are rare cases of humans with mutations in the aromatase gene and these subjects too are viscerally obese and insulin-resistant. In contrast, a male patient reported with overactivity of aromatase exhibited a feminized (pear-shaped) distribution of body fat. Early studies showed that conversion rates of androstenedione to estrone are increased with increasing body weight but did not characterize subjects according to the distribution of adipose tissue. The use of aromatase inhibitors for breast cancer has been associated with weight gain, although again the specific sites of adipose accumulation have not been defined. Overall, the observations support a role for the generation of active estrogens within adipose tissue in obesity and adipose distribution, but the fine details, notably in human pathology, still remain to be ironed out.

## CONCLUSIONS

The metabolic syndrome, its component cardiovascular risk factors, and visceral obesity are all common and highly complex disorders that are undoubtedly caused by the interplay of multiple genetic and environmental influences. Nevertheless, the notion that steroid hormones play central pathogenic roles is attractive and reflects the striking parallels between the phenotype of the metabolic syndrome and Cushing’s syndrome or, less clearly, estrogen/aromatase deficiency. Moreover,

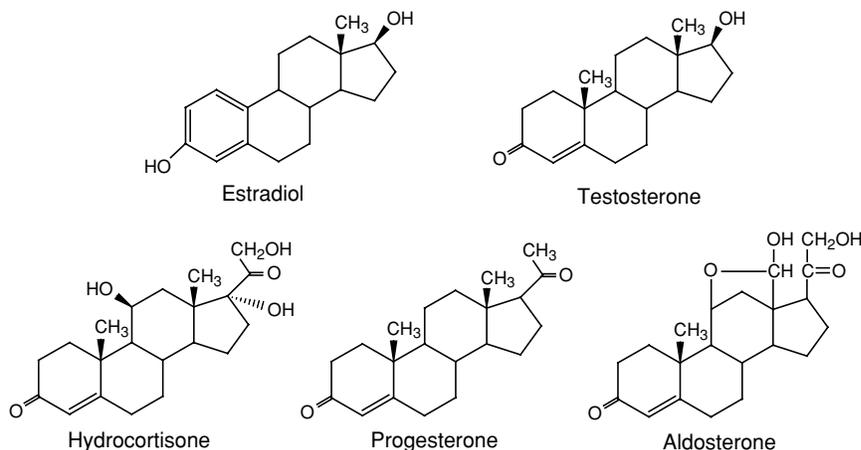
excess steroid production within adipose tissue and perhaps other tissues (liver, muscle) is anticipated to generate a host of downstream effects that contribute to the phenotype. These range from alterations to other local nuclear receptors/transcription factors, such as PPAR $\alpha$  and PPAR $\gamma$  (the target receptors for fibrates and thiazolidinediones, respectively), insulin sensitivity, and more novel adipose gene products including “adipokines” (leptin, adiponectin, tumor necrosis factor  $\alpha$ , and resistin). All are altered in mice with transgenic manipulations of 11 $\beta$ -HSD1. Exactly how these various effectors interact and the relative role of aromatase and other steroid enzymes remain intriguing future topics for study. The possibility of exploiting drugs that manipulate such intracrine systems is an attractive prospect, heralding more tissue-selective targeting of steroid manipulations than the available generalized treatments with their legion of benefits and detriments.

## See Also the Following Articles

Glucocorticoid Receptor • Hyperandrogenism, Hyperinsulinemic • Multiple Autoimmune Endocrinopathy

## Further Reading

- Andrews, R. C., and Walker, B. R. (1999). Glucocorticoids and insulin resistance: Old hormones, new targets. *Clin. Sci.* **96**, 513–523.
- Björntorp, P. (1997). Body fat distribution, insulin resistance, and metabolic diseases. *Nutrition* **13**, 795–803.
- Kotelevtsev, Y., Holmes, M. C., et al. (1997). 11 $\beta$ -Hydroxysteroid dehydrogenase type 1 knockout mice show attenuated glucocorticoid inducible responses and resist hyperglycaemia on obesity or stress. *Proc. Natl. Acad. Sci. USA* **94**, 14924–14929.
- Morton, N. M., Holmes, M. C., et al. (2001). Improved lipid and lipoprotein profile, hepatic insulin sensitivity and glucose tolerance in 11 $\beta$ -hydroxysteroid dehydrogenase type 1 null mice. *J. Biol. Chem.* **276**, 41293–41300.
- National Cholesterol Education Program Expert Panel. (2001). Executive Summary of The Third Report of The National Cholesterol Education Program Expert Panel (Adult Treatment Panel III). *J. Am. Med. Assoc.* **285**, 2486–2497.
- Seckl, J. R., and Walker, B. R. (2001). 11 $\beta$ -Hydroxysteroid dehydrogenase type 1: A tissue-specific amplifier of glucocorticoid action. *Endocrinology* **142**, 1371–1376.
- Stewart, P. M., and Krozowski, Z. S. (1999). 11 $\beta$ -Hydroxysteroid dehydrogenase. *Vitam. Horm.* **57**, 249–324.
- Walker, B. R., and Seckl, J. R. (2001). Cortisol metabolism. In “International Textbook of Obesity” (P. Björntorp, ed.), pp. 241–268. Wiley, Chichester, UK.



**Figure 1** Structures of the adrenal and sex steroids. The major circulating steroids in humans that bind to receptors are estradiol, cortisol, aldosterone, progesterone, and testosterone. In many androgen-responsive tissues, the metabolism of testosterone by 5 $\alpha$ -reductase yields 5 $\alpha$ -dihydrotestosterone, a more potent androgen.

phylogenetic trees are listed, with the hope that the reader will be tempted to use them in the future.

Central in the evolution of a family of related proteins that bind steroids is duplication of an ancestral steroid receptor gene followed by mutations in each gene, leading to two divergent proteins that can have different steroid-binding properties. One needs to be aware that not all mutations result in noticeable functional changes in a protein. Some mutations are “neutral”; that is, they do not change properties of the protein, as far as can be determined. Deciphering which sequence changes are functionally important is one of the applications of bioinformatics.

Two proteins with a common ancestor are homologues. Genes are either homologous to each other or not; there is no in-between state. Sometimes, mistakenly, percentage homology is used instead of percentage identity in a comparison of two proteins, for example, when two proteins are characterized as being 42% homologous, when they really are 42% identical and 100% homologous. Homologues can be either orthologues or paralogues. Analysis of orthologues and paralogues yields important information on the evolution of structure and function in steroid receptors and other proteins. Orthologues are the same gene in different organisms, for example, the estrogen receptor in humans, rats, and frogs. Analysis of orthologues illuminates how a protein evolved in response to different environmental conditions in the host organism. Thus, in fish, which are cold-blooded animals, the estrogen receptor is optimized to bind estradiol at a lower temperature than the estrogen receptor in mammals. This difference is reflected in the sequence in the estrogen-binding domain. Paralogues are genes that arise from

gene duplication and divergence. The paralogous androgen and progesterone receptors arose by a duplication of an ancestral gene. It is after this duplication that more complex regulation of androgen and progestin action evolved.

How can one determine whether two proteins are homologous? This is accomplished with computer programs that align the two protein's amino acid sequences and determine the percentage of identical amino acids and of conservative replacements. The latter are amino acids with similar properties, such as arginine/lysine or aspartic acid/glutamic acid. A variety of statistical algorithms are available to convert these data into a quantitative measure of similarity. Although either amino acid or nucleic acid sequences can be used, amino acid sequences, which have a 20-letter alphabet, are more sensitive to distant relationships than nucleic acid alignments, which have only a 4-letter alphabet. Various alignment programs are available on the Internet. One of the most popular is BLAST, which takes the sequences of two proteins (either the amino acid sequence or the DNA sequence), aligns them, and then calculates the probability of finding two proteins with such similarity by chance based on the number of proteins in the database (GenBank contains sequence data for approximately 900,000 proteins).

To analyze the evolution of steroid receptors, one needs a phylogenetic tree that depicts the relationships between the different steroid receptors from different animals. Just as in constructing a family tree, where one tries to obtain records of births and marriages that cover many generations, the most useful phylogenetic tree is one that contains steroid receptors from animals

that arose from the earliest times to the most recent. In this regard, it is fortunate that in the past few years, the sequences of steroid receptors from fish, amphibians, reptiles, birds, and mammals have been determined. This allows steroid receptor evolution close to the origins of vertebrates approximately 530 million years ago to be examined. Such a phylogenetic tree will be used to investigate the diversification of the steroid response in vertebrates.

The Internet contains a variety of computer programs, such as ClustalW and PHYLIP, that can construct a phylogenetic tree from a collection of sequences. However, analysis of steroid receptor evolution is not that simple. Steroid receptors are mosaics of functional domains, each of which is changing at a different rate, as was discovered in 1985 and 1986, when the glucocorticoid receptor (GR) and human estrogen receptor (ER) were cloned and sequenced. In the middle of each sequence is an ~65-amino-acid segment containing two sets of four cysteine residues complexed with zinc in a characteristic zinc finger motif. These two “zinc-fingers” bind DNA in the GR, ER, and other steroid receptors. At the C terminus is an ~250-amino-acid segment that binds the steroid. The DNA-binding domains on the GR and ER are well conserved; their amino acid sequences are approximately 60% identical. As expected, there is much less amino acid sequence identity—approximately 25%—in the steroid-binding domain on the ER and GR because estrogens and glucocorticoids have important structural differences. Subsequent sequencing of the progesterone receptor (PR), androgen receptor (AR), and mineralocorticoid receptor (MR) revealed that all adrenal and sex steroid receptors have a similar overall structure and there is much more sequence variation in the steroid-binding domain than in the DNA-binding domain. The domains at the amino terminus and the sequence between the DNA- and steroid-binding domains are changing even more rapidly than is the steroid-binding domain. An analysis of the entire steroid receptor sequence will be an average of the evolutionary rates of the various domains. This can obscure the evolution of the steroid-binding domain and the understanding of how the response to different steroids evolved. Another reason for using the steroid-binding domain is that it provides better resolution of the branches compared to the DNA-binding domain, which has a shorter sequence and is changing very slowly.

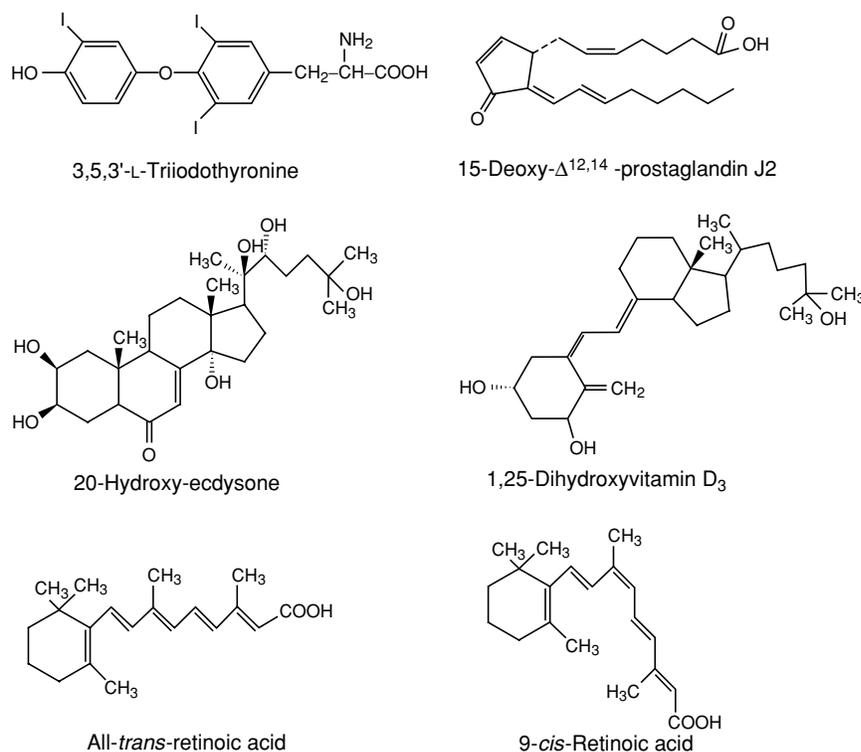
## NUCLEAR RECEPTORS

To obtain a more complete understanding of steroid receptor evolution, steroid receptor homologues that

bind ligands with structures that differ from that of adrenal and sex steroids are included in the analysis. The idea that steroid receptors are a subset of a larger family of ligand-regulated transcription factors arose in 1986 with the surprising discovery that the thyroid hormone receptor is homologous to steroid receptors. Soon other receptors were found with a DNA-binding domain containing a zinc-finger motif and a ligand-binding domain at the C terminus. This includes receptors for retinoids, 15-deoxy- $\Delta^{12,14}$ -prostaglandin J<sub>2</sub>, ecdysone, and 1,25-dihydroxyvitamin D<sub>3</sub>. Since all of these receptors act in the cell nucleus, the term nuclear receptors was used to describe this protein superfamily, which contains hundreds of different proteins. Nuclear receptors are an ancient family that is found in invertebrates and vertebrates. However, sequence searches of the complete genomes of *S. cerevisiae* and *A. thaliana* show that neither genome has a protein with the zinc-finger motif characteristic of nuclear receptors. Thus, nuclear receptors arose in multicellular animals, in which they regulate a diverse set of physiological processes. Indeed, the nuclear receptor family is an excellent example of how gene duplication and divergence can lead to a family of proteins that respond to ligands with very different structures, providing flexibility in the regulation of gene transcription (see Fig. 2). In this article, some ligand-binding nuclear receptors are included in the phylogenetic analysis of adrenal and sex steroid evolution to explore their position in the nuclear receptor family.

## WHEN DID STEROID RECEPTORS EVOLVE?

In 1997, Escriva *et al.* used polymerase chain reaction (PCR) to investigate the presence in vertebrates and invertebrates of genes with sequences that are homologous to DNA-binding domains of steroid receptors as well as other nuclear receptors. They used the DNA-binding domain on steroid receptors because it is changing very slowly, which makes it a good probe for finding homologues in organisms that diverged from a common ancestor over 600 million years ago. PCR with the zinc-finger domain clearly showed that nuclear receptors are present in invertebrates, including jellyfish, which are very simple invertebrates. However, genes with similarity to the DNA-binding domain of steroid receptors were restricted to vertebrates. In particular, Escriva *et al.* found evidence for the AR, ER, and GR in sharks, a cartilaginous fish, and the PR in hagfish, which diverged from the vertebrate line earlier than lamprey, the other jawless fish. The presence of the PR in hagfish



**Figure 2** Other ligands for nuclear receptors. Thyroid hormone, 1,25-hydroxyvitamin D<sub>3</sub>, all-*trans*-retinoic acid, 9-*cis*-retinoic acid, 15-deoxy- $\Delta^{12,14}$ -prostaglandin J<sub>2</sub>, and ecdysone bind with high affinity to nuclear receptors that are distantly related to the adrenal and sex steroid receptors.

places the origin of adrenal and sex steroid receptors in the Early Cambrian, approximately 540 million years ago, when ~35 phyla of multicellular animals appear in the fossil record. Steroid receptors are a “recent” evolutionary innovation, when one considers that unicellular eukaryotes arose at least 2 billion year ago and that plants and animals diverged from a common ancestor approximately 1.2 billion years ago. This means that the steroid response was integrated into what already was a sophisticated and robust signaling network for regulating differentiation, development, homeostasis, and reproduction. Steroid receptors assumed some activities that were regulated by other transcription factors, for example, the regulation by androgen and estrogen of spermatogenesis and oogenesis in vertebrates.

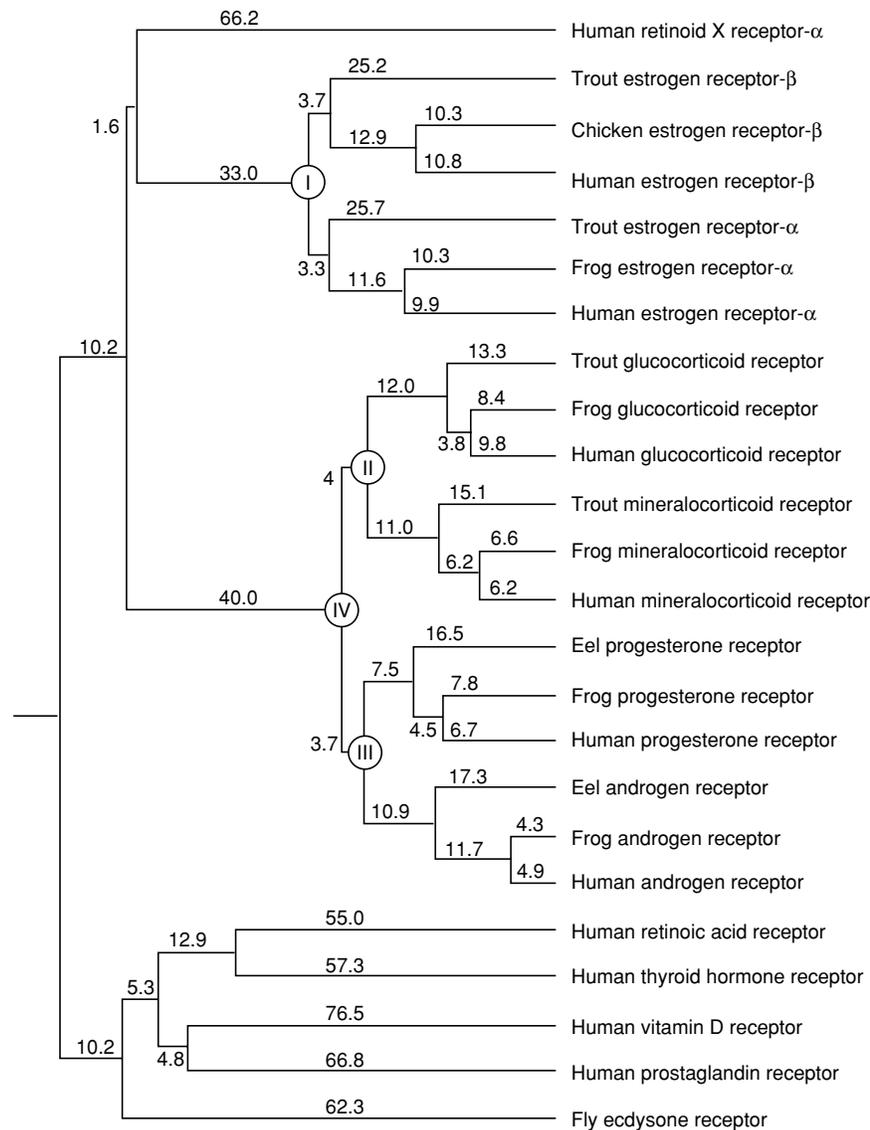
## PHYLOGENETIC ANALYSIS OF ADRENAL AND SEX STEROID RECEPTORS

Figure 3 shows the phylogeny of the ligand-binding domain of adrenal and sex steroid receptors and that of some other nuclear receptors. Sites of gene

duplications are labeled. The lengths of the branches after each duplication are proportional to the changes in the receptor sequence. By adding the numbers between two branches, one can determine how fast a given receptor type is changing over time or how close two receptors are to each other. For example, the AR in frog and the AR in human are 9.2 units distant from each other. In contrast, the GR in frog and the GR in human are 18.2 units from each other, indicating that the GR is changing more rapidly in land animals than is the AR, as was shown by Thornton and Kelley in 1998.

A overview of the steroid receptor tree reveals that the adrenal and sex steroid receptors are a distinct clade in the nuclear receptor family, well separated from receptors for 1,25-dihydroxyvitamin D<sub>3</sub>, ecdysone, thyroid hormone, and retinoids. The adrenal and sex steroid receptors appear to have evolved in jawless fish or a protochordate, which is consistent with the PCR data of Escriva *et al.*

Steroid receptors can be further subdivided. The ER is on one branch that subdivides at node I into ER- $\alpha$  and ER- $\beta$ . In a separate group are the AR, PR, GR, and MR. The PR and AR cluster on one branch and the GR and



**Figure 3** Evolution of the adrenal and sex steroid receptors and other nuclear receptors. The adrenal and sex steroid receptors form a separate clade in the nuclear receptor family. The estrogen receptor is closest to the duplication that led to the separation of estrogen from the  $\Delta^4$  steroids. This suggests that the estrogen receptor is most ancient of the adrenal and sex steroid receptors. Further duplications and sequence divergence led to ER- $\alpha$  and ER- $\beta$  and the formation of the GR, MR, PR, and AR.

MR cluster on another branch. The phylogenetic grouping of the AR, PR, GR, and MR and their subdivision correlates nicely with their transcriptional and steroid-binding properties. The AR, PR, GR, and MR can bind to the same DNA sequence in the nucleus to regulate gene transcription. The AR and PR have similar recognition for some steroids; that is, some progestins have androgenic activity. Similarly, the MR binds glucocorticoids, such as cortisol and corticosterone, with an affinity similar to that of aldosterone, the main mineralocorticoid in humans.

## STEROID RECEPTORS AND THE ORIGINS OF VERTEBRATES

The distances between the branches at the duplications in the ER, AR, PR, GR, and MR are short, which indicates diversification by a series of closely spaced (on a geological timescale) gene duplications. It is thought that two genome-size duplications occurred early in the evolution of vertebrates, before the evolution of fishes with jaws. Each genome-size duplication substantially increased the raw material

for diversification of protein function and complex regulatory networks that are characteristic of vertebrates. Analysis of the chromosomal location of steroid receptors indicates that it is likely that they formed during these genome-size duplications. Indeed, the diversification of steroid responses mediated by the ER, AR, PR, GR, and MR probably contributed a selective advantage to ancestral vertebrates during and after the Cambrian explosion.

### WHAT WAS THE ANCESTRAL STEROID RECEPTOR?

The tree shows that the branch leading to the ER from the ancestral steroid receptor is shorter than the branch leading to the AR, PR, GR, and MR, indicating a slower change in the ER sequence since their separation. This suggests that the ER was under functional constraints during this time, which is consistent with the estrogen response being the most ancient of the adrenal and sex steroid responses. Thornton cloned the ER, PR, and GR from lamprey, a jawless fish, which has ancestors in the Cambrian period from 550 to 535 million years ago. Thornton clearly showed that the ER is the most ancient of the adrenal and sex steroid receptors. However, the identity of the steroid(s) that regulated the action of the primitive ER is not known. It could have been estradiol or a  $\Delta^5$  androgen, both of which bind with high affinity to human ER, or it could have been another steroid or a ligand with a nonsteroidal structure. This question of the identity of ancestral steroids will be explored further in the analysis of other steroid receptors.

### STEROID RECEPTOR EVOLUTION IN FISH AND LAND VERTEBRATES

Figure 3 shows that the rate of change in the sequence, or “molecular clock,” of the hormone-binding domain is not uniform among the various vertebrate steroid receptors. Considering that humans and amphibians diverged from a common ancestor approximately 350 million years ago and that teleosts, fishes with jaws, arose approximately 425 million years ago, the branch lengths between the sequences of the hormone-binding domain in steroid receptors among various land animals are much shorter than said branch lengths between fish and amphibians. This indicates that there was a reduction in the rate of change in the steroid-binding domain in land animals compared to fish. This sequence divergence correlates with differences in ligand specificity between fish and land

vertebrate steroid receptors. Among the adrenal and sex steroid receptors, the hormone-binding domain of the AR has been most conserved and the hormone-binding domain of the GR has been least conserved since the separation of amphibians and humans from a common ancestor.

### RECEPTOR DIVERSIFICATION AND PHYSIOLOGICAL RESPONSES

#### Glucocorticoid and Mineralocorticoid Receptors

The length of the branches in the GR/MR clade indicates that the duplication at node II leading to the MR and GR arose in jawless fish or earlier. In land vertebrates, the MR regulates electrolyte balance under the influence of aldosterone. Aldosterone is not found in most fish. It is in amphibians that aldosterone assumes the role of a mineralocorticoid, regulating electrolyte transport. In fish, cortisol regulates electrolyte transport. It is not known whether cortisol acts exclusively in fish through the MR to regulate electrolyte transport or whether this is accomplished by binding of cortisol to either the GR or the MR.

The evolution of the MR from a GR ancestor is consistent with the steroid specificity of mammalian MR. When recombinant human MR became available for analysis of steroid specificity, it was unexpectedly found to bind cortisol, corticosterone, and aldosterone with similar affinity, clearly showing the similarity of the MR to the GR. The high affinity of the MR for glucocorticoids has led to an alternative name, GR type 1, which also reflects the role of this receptor in mediating responses to glucocorticoids in some tissues, where aldosterone is not likely to regulate electrolyte balance, such as the brain. GR type 2 is the classical glucocorticoid receptor.

Evolutionary analysis of the GR type 1/type 2 can explain the presence of the MR in tissues that are not involved in electrolyte balance; it predicts that in a primitive vertebrate, GR type 1 regulated responses in the brain and these remain under its regulation. Later, in land animals, GR type 1 assumed exclusive regulation of the mineralocorticoid response. It also predicts that the GR type 1 sequence will change more slowly in land animals than the GR type 2 sequence. This is clearly seen in the shorter branches for GR type 1 (MR) in amphibians and mammals compared to GR type 2 in these animals as shown in Fig. 3. The requirement that the MR respond to aldosterone in kidney and gut, and to corticosterone and cortisol in brain and other tissues with GR type I/type II

physiology, puts constraints on changes in the MR sequence.

## Androgen and Progestin Receptors

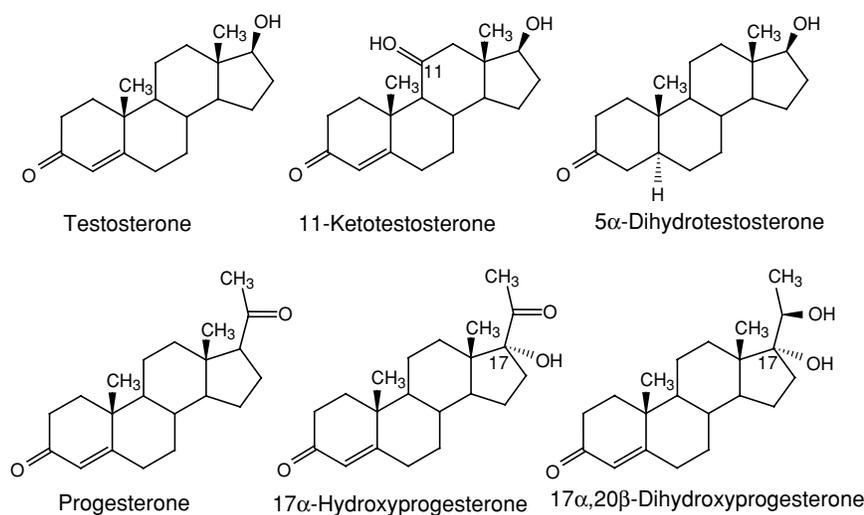
The AR/PR clade also has an origin in a jawless fish or earlier, such as a cephalochordate (modern example: amphioxus) or urochordate (modern example: tunicate). The lengths of the branches indicate a substantial sequence change in the fish AR, in contrast to the slow change of the AR in land animals. Similarly, the fish PR has mutated more than the PR in land animals. The identity of the steroid(s) that regulated the action of the ancestor at node III is unknown. It may have been an androgen or a progestin or both. It is likely that the proto-AR/PR had lower binding affinity and less specificity for its ligand(s) than is found in current AR and PR. Over time, after the gene duplication, these receptors evolved toward higher affinity and greater specificity for their cognate steroids, leading to a division of function between androgens and progestins. A similar division of function occurred in the MR/GR clade. Evolution leading to a division of function provides increased flexibility for specific regulation of gene expression.

It is likely that the ancestral receptor recognized steroids that differ from the active androgens or progestins in mammals because fish and mammals have different circulating androgens and progestins. In most fish, the major circulating androgen is

11-ketotestosterone (see Fig. 4). In land animals, the main circulating androgen is testosterone, which is biologically active. In addition, in some tissues, testosterone is metabolized to 5 $\alpha$ -dihydrotestosterone, a more potent androgen. In fish, progestins with 17 $\alpha$ -hydroxyl substituents have high affinity for the PR. In contrast, in birds and mammals, steroids with 17 $\alpha$ -hydroxyl substituents (e.g., 17 $\alpha$ -hydroxyprogesterone, cortisol) have an affinity for the PR that is less than 1% of that of progesterone. Thus, in the evolution from fish to land animals, there was a change in steroid specificity for androgens and progestins.

## Estrogen Receptors $\alpha$ and $\beta$

The ER- $\alpha$  and ER- $\beta$  branches appear to converge in a jawless fish. The duplication of ER- $\alpha$  and ER- $\beta$  is interesting because it is ancient, yet the two genes have been conserved in an active form despite the fact that there are no major differences in steroid binding. In mammals, there are differences in the tissue-specific expression of these receptors. For example, ER- $\beta$  is found in prostate. Clearly this function and certain other estrogen-dependent activities in tissues such as the placenta were not the function of either ER- $\alpha$  or ER- $\beta$  in fish. Indeed, the ancestral GR, MR, and AR also are likely to lack functions that their orthologues have in land animals. This illustrates an important practical insight from evolutionary analyses. Modern



**Figure 4** Steroids that are active in fish. Testosterone, 11-ketotestosterone, and 5 $\alpha$ -dihydrotestosterone have high affinity for the fish androgen receptor. However, the main androgen in fish is 11-ketotestosterone. In mammals, the biologically active androgens are testosterone and 5 $\alpha$ -dihydrotestosterone. In fish, 17 $\alpha$ -hydroxyprogesterone, 17 $\alpha$ ,20 $\beta$ -hydroxyprogesterone, and progesterone bind to the progesterone receptor. However, it is 17 $\alpha$ ,20 $\beta$ -hydroxyprogesterone that induces the final stages of spermatogenesis in fish. In contrast to fish, in mammals, 17 $\alpha$ -hydroxyprogesterone and other steroids with 17 $\alpha$ -hydroxy substituents have less than 1% of the affinity for the progesterone receptor than that of progesterone.

proteins could not have had some of their modern-day functions early in their evolution. Are some of these ancient functions still important in humans? That is, the reproductive functions of estrogens, androgens, and progestins may be only part of their actions in mammals. In the past decade, it has become clear that the “canonical” functions of steroids are not their only activities. Just as the MR or GR type 1 is important in the hippocampus, so the ER has important actions in the brain. Indeed, estrogens have been implicated in maintaining cognitive function and prevention of Alzheimer’s disease. The ER may have been important in the brain in a primitive vertebrate with a simple body plan, lacking the organs found in sharks, bony fish, and land animals.

Indeed, the cloning of receptors for adrenal and sex steroids in animals such as lamprey, hagfish, and amphioxus and the study of where these receptors are expressed and how expression is regulated will provide clues to their ancient functions. This information is likely to be relevant to less-well-studied functions and even unknown functions of adrenal and sex steroids in humans and other modern organisms.

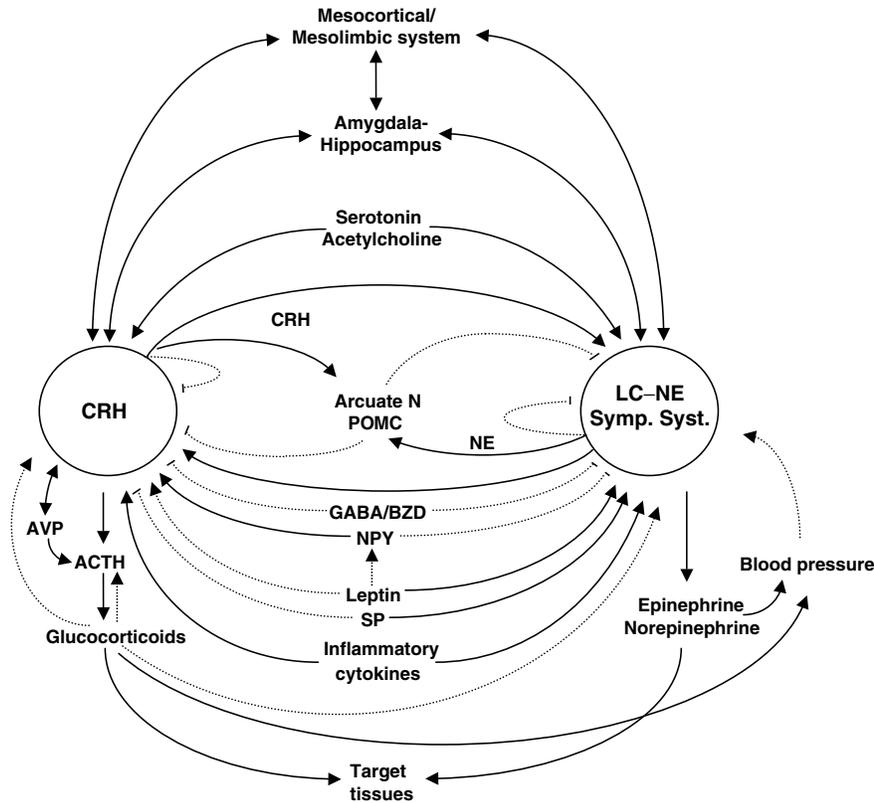
### See Also the Following Articles

ACTH,  $\alpha$ -MSH, and POMC, Evolution of • Angiotensin, Evolution of • GI Hormone Development (Families and Phylogeny) • Insulin and Insulin-like Growth Factors, Evolution of • Natriuretic Peptide System, Evolution of • Neuropeptide Y, Evolution of • Opioid/Orphanin Gene Family,

Evolution of • Prolactin, Evolution of • Somatostatin, Evolution of

### Further Reading

- Baker, M. E. (1997). Steroid receptor phylogeny and vertebrate origins. *Mol. Cell. Endocrinol.* **135**, 101–107.
- Baker, M. E. (2001). Adrenal and sex steroid evolution: Environmental implications. *J. Mol. Endocrinol.* **26**, 119–125.
- Baker, M. E. (2003). Evolution of adrenal and sex steroid action in vertebrates: A ligand-based mechanism for complexity. *BioEssays* **25**, 396–400.
- Escriva, H., Delaunay, F., and Laudet, V. (2000). Ligand binding and nuclear receptor evolution. *BioEssays* **22**, 717–727.
- Escriva, H., Safi, R., Hanni, C., Langlois, M.-C., Saumitou-Laprade, P., Stehelin, D., Capron, A., Pierce, R., and Laudet, V. (1997). Ligand binding was acquired during evolution of nuclear receptors. *Proc. Natl. Acad. Sci. USA* **94**, 6803–6808.
- Giguere, V. (1999). Orphan nuclear receptors: From gene to function. *Endocr. Rev.* **20**, 689–725.
- Laudet, V. (1997). Evolution of the nuclear receptor superfamily: Early diversification from an ancestral orphan receptor. *J. Mol. Endocrinol.* **19**, 207–226.
- Mangelsdorf, D. J., Thummel, C., Beato, M., Herrlich, P., Schutz, G., Umesono, K., Blumberg, B., Kastner, P., Mark, M., Chambon, P., and Evans, R. M. (1995). The nuclear receptor superfamily: The second decade. *Cell* **83**, 835–839.
- Shimeld, S. M., and Holland, P. W. (2000). Vertebrate innovations. *Proc. Natl. Acad. Sci. USA* **97**, 4449–4452.
- Thornton, J. W. (2001). Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions. *Proc. Natl. Acad. Sci. USA* **98**, 5671–5676.
- Thornton, J. W., and DeSalle, R. (2000). Gene family evolution and homology: Genomics meets phylogenetics. *Annu. Rev. Genomics Hum. Genet.* **1**, 41–73.



**Figure 1** A simplified schematic representation of the central and peripheral components of the stress system, their functional interrelations, and their relationships to other central systems involved in the stress response. The CRH–AVP neurons and central catecholaminergic neurons of the LC–NE system reciprocally innervate and activate one another. The HPA axis is controlled by several feedback loops that tend to normalize the time-integrated secretion of cortisol, yet glucocorticoids stimulate the fear centers in the amygdala. Activation of the HPA axis leads to suppression of the GH/insulin-like growth factor-I, luteinizing hormone/testosterone/estradiol, and TSH/triiodothyronine axes; activation of the sympathetic system increases interleukin-6 secretion. Solid lines indicate stimulation; dashed lines indicate inhibition. BZD, benzodiazepine; GABA,  $\gamma$ -aminobutyric acid; NPY, neuropeptide Y; SP, substance P. Modified from Chrousos (1992).

## The Hypothalamic–Pituitary–Adrenal Axis

The hypothalamus controls the secretion of adrenocorticotropic hormone (ACTH) from the anterior pituitary, which, in turn, stimulates the secretion by the adrenal cortex of glucocorticoid hormones, mainly cortisol in humans. The principal hypothalamic stimulus to the pituitary–adrenal axis is CRH, a 41-amino-acid peptide first isolated in 1981 by Wylie Vale. AVP is a potent synergistic factor with CRH in stimulating ACTH secretion; however, AVP has little ACTH secretagogue activity by itself. Furthermore, it appears that there is a reciprocal positive interaction between CRH and AVP at the level of the hypothalamus, with each neuropeptide stimulating the secretion of the other.

In nonstressful situations, both CRH and AVP are secreted in the portal system in a circadian, pulsatile

fashion, with a frequency of approximately two to three secretory episodes per hour. Under resting conditions, the amplitude of the CRH and AVP pulses increases in the early morning hours, resulting finally in ACTH and cortisol secretory bursts in the general circulation. These diurnal variations are perturbed by changes in lighting, feeding schedules, and activity and are disrupted by stress.

During acute stress, the amplitude and synchronization of the CRH and AVP pulsations in the hypothyseal portal system markedly increase, resulting in increases in ACTH and cortisol secretory episodes. Depending on the type of stress, other factors such as AVP of magnocellular neuron origin, angiotensin II, and various cytokines and lipid mediators of inflammation are secreted and act on hypothalamic, pituitary, or adrenal components of the HPA axis, potentiating its activity.

Circulating ACTH is the key regulator of glucocorticoid secretion by the adrenal cortex. Other hormones or cytokines, either originating from the adrenal medulla or coming from the systemic circulation, as well as neuronal information from the autonomic innervation of the adrenal cortex may also participate in the regulation of cortisol secretion.

Glucocorticoids are the final effectors of the HPA axis and participate in the control of whole-body homeostasis and the organism's response to stress. They play a key regulatory role on the basal activity of the HPA axis and on the termination of the stress response by acting at extrahypothalamic centers, the hypothalamus, and the pituitary gland. The inhibitory glucocorticoid feedback on the ACTH secretory response acts to limit the duration of the total tissue exposure to glucocorticoids, thus minimizing the catabolic, anti-reproductive, and immunosuppressive effects of these hormones.

Glucocorticoids exert their effects through their ubiquitous cytoplasmic receptors. On ligand binding, the glucocorticoid receptors translocate into the nucleus, where they interact as homodimers with specific glucocorticoid-responsive elements within the DNA to activate appropriate hormone-responsive genes. The activated receptors also inhibit, through protein-protein interactions, other transcription factors, such as c-jun/c-fos, and nuclear factor  $\kappa$ B, which are positive regulators of the transcription of several genes involved in the activation and growth of immune and other cells. Furthermore, glucocorticoids change the stability of mRNAs and hence the translation of several glucocorticoid-responsive proteins, as well as the electrical potential of neuronal cells.

### The Autonomic Axes

The autonomic nervous system provides a rapidly responding mechanism to control a wide range of functions. Cardiovascular, respiratory, gastrointestinal, renal, endocrine, and other systems are regulated by the sympathetic nervous system or the parasympathetic system or both. Interestingly, the parasympathetic system may assist sympathetic functions by withdrawing and can antagonize them by increasing its activity.

Sympathetic innervation of peripheral organs is derived from the efferent preganglionic fibers, whose cell bodies lie in the intermediolateral column of the spinal cord. These nerves synapse in the bilateral chains of sympathetic ganglia with postganglionic sympathetic neurons that richly innervate the smooth

muscle of the vasculature, the heart, the skeletal muscles, the kidney, the gut, fat, and many other organs. The preganglionic neurons are cholinergic, whereas the postganglionic neurons are mostly noradrenergic. The sympathetic system also has a humoral contribution, providing most of the circulating epinephrine and some of the norepinephrine from the adrenal medulla. In addition to the classic neurotransmitters acetylcholine and norepinephrine, both sympathetic and parasympathetic subdivisions of the autonomic nervous system contain several subpopulations of target-selective and neurochemically coded neurons that express a variety of neuropeptides and, in some cases, ATP, nitric oxide, or lipid mediators of inflammation. Interestingly, CRH, neuropeptide Y, and somatostatin are colocalized in noradrenergic vasoconstrictive neurons. Transmission in sympathetic ganglia is also modulated by neuropeptides released from preganglionic fibers and short interneurons (e.g., enkephalin, neurotensin), as well as from primary afferent collaterals (e.g., substance P).

### REGULATION OF THE STRESS RESPONSE

The orchestrated interplay of several neurotransmitter systems in the brain underlies the characteristic phenomenology of behavioral, endocrine, autonomic, and immune responses to stress. These transmitters include CRH, AVP, opioid peptides, dopamine, and norepinephrine.

#### CRH-AVP

Shortly after its isolation, it became apparent that CRH was implicated in other components of the stress response, such as arousal and autonomic activity. Supportive evidence was derived from intracerebroventricular or selective brain administration of CRH in rodents and nonhuman primates, which precipitated several coordinated responses characteristic of stress. Moreover, administration of CRH peptide antagonists into selected areas of the brain suppresses many aspects of the stress response. Finally, CRH type 1 receptor knockout mice were shown to have a markedly deficient ability to mount an effective stress response.

CRH and CRH receptors were found in many sites in the brain outside of the hypothalamus, including parts of the limbic system and the central arousal sympathetic systems (LC-sympathetic systems) in the brainstem and spinal cord. Stress is a potent general

activator of CRH release from the hypothalamus and extrahypothalamic sites. The mechanisms via which stress stimulates CRH neurons are unclear. Whether CRH or another transmitter (e.g., NE) is upstream in eliciting the neurocircuitry of stress remains to be determined.

CRH-binding sites are also found in various peripheral tissues, such as the adrenal medulla, heart, prostate, gut, liver, kidney, and testes. CRH receptors belong to the G protein-coupled receptor superfamily and CRH binding stimulates the intracellular accumulation of cyclic AMP. Two distinct CRH receptor subtypes, designated CRH-R1 and CRH-R2, have been characterized, encoded by distinct genes that are differentially expressed. CRH-R1 is the most abundant subtype found in the anterior pituitary and is also widely distributed in the brain. CRH-R2 receptors are expressed mainly in the peripheral vasculature and the heart, as well as in subcortical structures in the brain.

### Locus Ceruleus–NE System

The locus ceruleus and other noradrenergic cell groups of the medulla and pons are collectively known as the LC–NE system. Brain epinephrine serves globally as an alarm system that decreases neurovegetative functions, such as eating and sleeping, and that contributes to accompanying increases in autonomic and neuroendocrine responses to stress, including HPA axis activation. NE also activates the amygdala, the principal brain locus for fear-related behaviors, and enhances the long-term storage of aversively charged emotional memories in sites such as the hippocampus and striatum.

Reciprocal neural connections exist between the CRH and catecholaminergic neurons (LC–NE neurons) of the central stress system, with CRH and norepinephrine stimulating each other, the latter primarily through  $\beta_1$ -noradrenergic receptors. There is an ultrashort autoregulatory negative feedback loop on the CRH neurons exerted by CRH itself, just as there is a similar loop in the LC–NE neurons, by way of presynaptic CRH and  $\alpha_2$ -noradrenergic receptors, respectively. There is also parallel regulation of both central components of the stress system by other stimulatory and inhibitory neuronal pathways. Several neurotransmitters, including serotonin and acetylcholine, excite CRH and the LC–NE neurons. The negative feedback controls include glucocorticoids,  $\gamma$ -aminobutyric acid, corticotropin, and several opioid peptides, which inhibit both CRH and LC–NE neurons.

## BODY SYSTEMS' RESPONSES TO STRESS

### HPA Axis–Immune System Interactions

It has been known for several decades that stress, whether inflammatory, traumatic, or psychological, is associated with concurrent activation of the HPA axis. In the early 1990s, it also became apparent that cytokines and other humoral mediators of inflammation are potent activators of the central stress response, constituting the afferent limb of a feedback loop through which the immune–inflammatory system and the CNS communicate.

All three inflammatory cytokines, tumor necrosis factor  $\alpha$ , interleukin-1 $\beta$  (IL- $\beta$ ), and interleukin-6, can cause stimulation of the HPA axis alone or in synergy with one another. There is evidence to suggest that IL-6, the main endocrine cytokine, plays the major role in the immune stimulation of the axis, especially in chronic inflammatory stress.

Some of the activating effects of cytokines on the HPA axis may be exerted indirectly by stimulation of the central catecholaminergic pathways. Also, activation of peripheral nociceptive, somatosensory, and visceral afferent fibers would lead to stimulation of both the catecholaminergic and CRH neuronal systems via ascending spinal pathways. Other inflammatory mediators, such as eicosanoids, platelet-activating factor, and serotonin, may also participate in the activation of the HPA axis, via autocrine/paracrine and/or endocrine effects.

Conversely, activation of the HPA axis has profound inhibitory effects on the inflammatory immune response because virtually all the components of the immune response are inhibited by cortisol. Alterations of leukocyte traffic and function, decreases in production of cytokines and mediators of inflammation, and inhibition of the latter's effects on target tissues are among the main immunosuppressive effects of glucocorticoids.

The efferent sympathetic/adrenomedullary system apparently participates in a major fashion in the interactions of the HPA axis and the immune/inflammatory reaction by being reciprocally connected with the CRH system, by receiving and transmitting humoral and nervous immune signals from the periphery, by densely innervating both primary and secondary lymphoid organs, and by reaching all sites of inflammation via the postganglionic sympathetic neurons. When activated during stress, the autonomic system exerts its own direct effects on immune organs, which can be immunosuppressive or both immunopotentiating and

anti-inflammatory. CRH secreted by postganglionic sympathetic neurons at inflammatory sites has pro-inflammatory properties (immune CRH); one of its key actions is to degranulate mast cells.

## HPA Axis–Other Endocrine Axes Interactions

### Gonadal and Growth Axes

The systems responsible for reproduction and growth are directly linked to the stress system and both are profoundly inhibited by various components of the HPA axis, the effector of the stress response. Either directly or through arcuate proopiomelanocortin (POMC) neuron  $\beta$ -endorphin, CRH suppresses the gonadotropin-releasing hormone (GnRH) neurons of the arcuate nucleus of the hypothalamus. Glucocorticoids exert inhibitory effects at the levels of the GnRH neuron, the pituitary gonadotroph, and the gonads themselves and render target tissues of sex steroids resistant to these hormones. Suppression of gonadal function caused by chronic HPA axis activation has been demonstrated in highly trained athletes of both sexes, in ballet dancers, and in individuals sustaining anorexia nervosa or starvation.

During inflammatory stress, cytokines suppress reproductive function directly and indirectly by activating hypothalamic secretion of CRH- and POMC-derived peptides, as well as by peripheral elevations of glucocorticoids and inhibition of steroidogenesis at both the ovaries and the testes.

The interaction between CRH and the gonadal axis appears to be bidirectional. Thus, the presence of estrogen-responsive elements has been demonstrated in the promoter area of the CRH gene, as well as direct stimulatory estrogen effects on CRH gene expression. This finding indicates that the CRH gene is a potentially important target of ovarian steroids and a potential mediator of gender-related differences in the stress response and HPA axis activity.

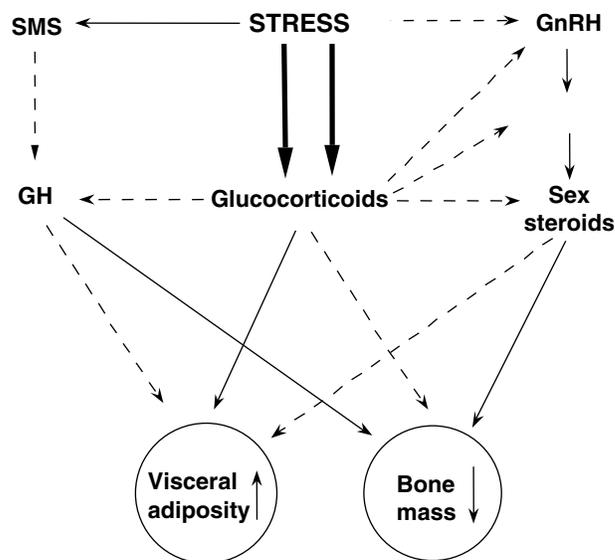
The growth axis is also inhibited at many levels during stress. Prolonged activation of the HPA axis with elevations in glucocorticoids leads to suppression of growth hormone (GH) secretion, inhibition of somatomedin C, and other growth hormone effects on their target tissues. However, acute elevations of growth hormone concentration in plasma may occur at the onset of the stress response or after acute administration of glucocorticoids, presumably through stimulation of the GH gene by glucocorticoids through glucocorticoid-responsive elements in its promoter region. In addition to the direct effects of

glucocorticoids, CRH-induced stimulation in hypothalamic somatostatin secretion may also result in GH suppression, providing another potential mechanism for stress-related suppression of GH secretion.

Similarly, activation of the HPA axis is associated with decreased production of thyroid-stimulating hormone (TSH) and inhibition of the conversion of the relatively inactive thyroxine to the more biologically active triiodothyronine in peripheral tissues (the “euthyroid sick” syndrome). Both phenomena may be caused by the increased levels of glucocorticoids and may serve to conserve energy during stress. In the case of inflammatory stress, inhibition of TSH secretion may occur in part through the action of cytokines on both the hypothalamus and the pituitary.

### Metabolism

Glucocorticoids directly inhibit pituitary growth hormone, gonadotropin, and thyrotropin secretion and make the target tissues of sex steroids and growth factors resistant to these hormones. Thus, glucocorticoids antagonize the beneficial actions of GH and sex steroids on fat tissue (lipolysis) and on muscle and bone anabolism. Chronic activation of the stress system would be expected to increase visceral adiposity, decrease lean body (muscle and bone) mass, and suppress osteoblastic activity (Fig. 2). Interestingly, the phenotype of central obesity and decreased lean body mass is shared by patients with Cushing’s



**Figure 2** Detrimental effects of chronic stress on adipose tissue metabolism and bone mass. Solid lines indicate stimulation; dashed lines indicate inhibition. Modified from Chrousos (1992).

syndrome and some patients with the combined diagnosis of melancholic depression or chronic anxiety disorder and the metabolic syndrome (visceral adiposity, insulin resistance, dyslipidemia, hypertension) or “pseudo-Cushing’s syndrome.”

Because increased hepatic gluconeogenesis is a characteristic feature of the stress response and because glucocorticoids induce insulin resistance, activation of the HPA axis may contribute to the poor control of diabetic patients during periods of emotional stress or concurrently with inflammatory and other diseases.

Obese subjects with psychiatric manifestations ranging from those of melancholic depression to anxiety with perception of “uncontrollable” stress frequently have mild hypercortisolism, whereas carefully screened obese subjects without such manifestations are eucortisolemic. The former may have stress-induced glucocorticoid-mediated visceral obesity and metabolic syndrome manifestations, which in the extreme may be called a pseudo-Cushing’s syndrome state that needs to be differentiated from frank Cushing’s syndrome. Stress-induced hypercortisolism and visceral obesity and their cardiovascular and other sequelae increase the all-cause mortality risk of affected subjects two- to threefold and curtail their life expectancy by several years.

## HPA AXIS: PATHOPHYSIOLOGY

Generally, the stress response with the resultant activation of the HPA axis is meant to be acute or at least of a limited duration. The time-limited nature of this process renders its accompanying anti-reproductive, anti-growth, catabolic, and immunosuppressive effects temporarily beneficial rather than damaging. In contrast, the chronicity of stress system activation would lead to the syndromal state that Selye described in 1936. Because CRH coordinates behavioral, neuroendocrine and autonomic adaptation during stressful situations, increased and prolonged production of CRH could explain the pathogenesis of the syndrome.

The prototypical example of chronic hyperactivation of the stress system (both the HPA axis and the LC-NE system) is manifested in melancholic depression, with dysphoric hyperarousal and relative immunosuppression. Indeed, cortisol excretion is increased and the plasma ACTH response to exogenous CRH is decreased. Hypersecretion of CRH has been observed in depression and suggests that CRH may participate in the initiation or perpetuation of a vicious cycle. Owing to chronically hyperactive stress,

patients with melancholic depression may sustain several severe somatic sequelae, such as osteoporosis, features of the metabolic syndrome, varying degrees of atherosclerosis, innate and T helper 1-directed immunosuppression, and certain infectious and neoplastic diseases. When not treated, these patients have a compromised life expectancy curtailed by 15–20 years after suicides are excluded.

In addition to melancholic depression, a spectrum of other conditions may be associated with increased and prolonged activation of the HPA axis (Table I), including anorexia nervosa with or without malnutrition, obsessive-compulsive disorder, panic anxiety, chronic active alcoholism, alcohol and narcotic withdrawal, excessive exercising, poorly controlled diabetes mellitus, childhood sexual abuse, and hyperthyroidism.

Another group of states is characterized by hypoactivation of the stress system, rather than sustained activation, in which chronically reduced secretion of CRH may result in pathological hypoarousal (Table I). Patients with atypical, seasonal depression and chronic fatigue syndrome fall into this category. Similarly, patients with fibromyalgia have decreased urinary free-cortisol excretion and frequently complain of fatigue. Hypothyroid patients also have clear evidence of CRH hyposecretion.

Withdrawal from smoking has been associated with decreased cortisol and catecholamine secretion. Decreased CRH secretion in the early period of nicotine abstinence could explain the increased appetite and weight gain frequently observed in these patients. In Cushing’s syndrome, the clinical picture of atypical depression, hyperphagia, weight gain, fatigue, and anergia is consistent with suppression of the CRH neuron by the associated hypercortisolism. The periods after cure of hypercortisolism or following cessation of chronic stress and the postpartum period are also associated with atypical depression, suppressed CRH secretion, and decreased HPA axis activity.

It is believed that an excessive HPA axis response to inflammatory stimuli would mimic the stress or hypercortisolemic state and would lead to increased susceptibility of the individual to a host of infectious agents or tumors as a result of T helper-1 suppression, but enhanced resistance to autoimmune/inflammatory disease. In contrast, a defective HPA axis response to such stimuli would reproduce the glucocorticoid-deficient state and would lead to relative resistance to infections and neoplastic disease, but increased susceptibility to autoimmune/inflammatory disease, such as Hashimoto’s thyroiditis or rheumatoid arthritis.

Table I States Associated with Hyperactivation or Hypoactivation of the HPA Axis

Increased HPA axis activity	Decreased HPA axis activity	Disrupted HPA axis activity
Severe chronic disease		
Melancholic depression	Atypical depression	Cushing's syndrome
Anorexia nervosa	Seasonal depression	Glucocorticoid deficiency
Obsessive-compulsive disorder	Chronic fatigue syndrome	Glucocorticoid resistance
	Fibromyalgia	
Panic disorder	Hypothyroidism	
Chronic excessive exercise	Adrenal suppression	
Malnutrition	Post-glucocorticoid therapy	
Diabetes mellitus	Poststress	
	Nicotine withdrawal	
	Postpartum	
Hyperthyroidism	Menopause	
Central obesity	Rheumatoid arthritis	
Childhood sexual abuse		
Pregnancy		

Modified from Chrousos (1992).

Thus, an increasing body of evidence suggests that patients with rheumatoid arthritis have a mild form of central hypocortisolism. Dysfunction of the HPA axis may actually play a role in the development or perpetuation of autoimmune disease, rather than being an epiphenomenon. The same rationale may explain the high incidence of autoimmune disease in the period after cure of hypercortisolism, as well as in glucocorticoid underreplaced adrenal insufficiency.

## CONCLUDING REMARKS

Antalarmin, a novel CRH receptor type 1 antagonist, decreases the activity of the HPA axis and LC-NE system, suppresses neurogenic inflammation, and blocks CRH-induced skin mast cell degranulation, in addition to blocking the development and expression of conditioned fear and stress-induced colonic hyperfunction. Chronic administration of antalarmin is not associated with glucocorticoid deficiency and permits HPA axis and LC-NE responses to severe stress. These data suggest that such CRH antagonists may be useful in human pathologic states, such as melancholic depression and chronic anxiety, associated with chronic hyperactivity of the stress system, along with predictable behavioral, neuroendocrine, metabolic, and immune changes, based on the interrelations outlined above. Conversely, potentiators of CRH secretion and action are needed to treat atypical depression, postpartum depression, and the fibromyalgia/

chronic fatigue syndromes, all characterized by low hypothalamic-pituitary-adrenal axis and LC-NE system activity, fatigue, depressive symptomatology, hyperalgesia, and increased immune/inflammatory responses to stimuli.

## See Also the Following Articles

Adrenergic Mechanisms • Antiadrenergic Agents • Corticotropin-Releasing Hormone, Family of • Pediatric HIV Infection and Hypothalamic-Pituitary-Adrenal Axis • Stress, Aging, and Central Nervous System Interactions

## Further Reading

- Bamberger, C. M., Schulte, H. M., and Chrousos, G. P. (1996). Molecular determinants of glucocorticoid receptor function and tissue sensitivity to glucocorticoids. *Endocr. Rev.* **17**, 245-261.
- Chrousos, G. P. (1992). Regulation and dysregulation of the hypothalamic-pituitary-adrenal axis: The corticotropin releasing hormone perspective. *Endocrinol. Metab. Clin. N. Am.* **21**, 833-858.
- Chrousos, G. P. (1995). The hypothalamic-pituitary-adrenal axis and immune-mediated inflammation. *N. Engl. J. Med.* **332**, 1351-1362.
- Chrousos, G. P. (1996). Organization and integration of the endocrine system. In "Pediatric Endocrinology" (M. Sperling, ed.), pp. 1-14. W. B. Saunders, Philadelphia, PA.
- Chrousos, G. P. (1998). Stressors, stress, and neuroendocrine integration of the adaptive response. The 1997 Hans Selye Memorial Lecture. *Ann. N. Y. Acad. Sci.* **851**, 311-335.
- Chrousos, G. P. (2000). The role of stress and the hypothalamic-pituitary-adrenal axis in the pathogenesis of the metabolic

- syndrome: Neuroendocrine and target tissue-related causes. *Int. J. Obes. Relat. Metab. Disord.* **24**, S50–S55.
- Chrousos, G. P., and Gold, P. W. (1992). The concepts of stress system disorders: Overview of behavioral and physical homeostasis. *J. Am. Med. Assoc.* **267**, 1244–1252.
- Habib, K. E., Gold, P. W., and Chrousos, G. P. (2001). Neuroendocrinology of stress. *Endocrinol. Metab. Clin. N. Am.* **30**, 695–728.
- Tsigos, C., and Chrousos, G. P. (1994). Physiology of the hypothalamic–pituitary–adrenal axis in health and dysregulation in psychiatric and autoimmune disorders. *Endocrinol. Metab. Clin. N. Am.* **23**, 451–466.
- Vale, W. W., Spiess, S., Rivier, C., and Rivier, J. (1981). Characterization of a 41-residue ovine hypothalamic peptide that stimulates secretion of corticotropin and  $\beta$ -endorphin. *Science* **213**, 1394–1397.

hormones, corticotropin-releasing hormone (CRH) and vasopressin. CRH is the main stimulator of both the synthesis and secretion of adrenocorticotropic hormone (ACTH) from the pituitary gland; such secretion is augmented in the presence of vasopressin, which acts as a synergistic factor. ACTH, in turn, stimulates glucocorticoid synthesis and, if secreted in an excessive amount, may induce adrenal cortical hypertrophy and hyperplasia. HPA axis activity follows a diurnal rhythm, characterized in humans by an early morning peak (acrophase), declining over the daytime, a minimal secretory activity at approximately midnight (quiescent period), and an abrupt elevation during late sleep, resulting in an early morning maximum.

In the periphery, glucocorticoids are involved in many important functions including electrolyte homeostasis, glucose utilization, fat and protein metabolism, bone and cartilage formation and resorption, immune and inflammatory reaction regulation, and cardiovascular reactivity. Centrally, glucocorticoids affect the behavior, mood, excitability, and electrical activity of the neurons in a general manner. In addition, glucocorticoids have a regulatory feedback effect on the HPA axis, which might be overrun during stressful conditions. Glucocorticoids inhibit ACTH secretion directly by suppressing proopiomelanocortin expression in pituitary corticotrophs and indirectly by inhibiting the secretion of CRH and vasopressin at the hypothalamic level. A higher nervous system structure, the hippocampus, mediates part of the negative feedback effect of glucocorticoids on the HPA axis. The hippocampus is involved in the regulation of ACTH and glucocorticoid secretion under basal conditions, as well as in limiting the stress-induced activation of the HPA axis.

There are two types of receptors for glucocorticoids; the mineralocorticoid receptors (MRs, also called type I receptors) have approximately 10-fold higher affinity than the glucocorticoid receptors (GRs, also called type II receptors). GRs are widely expressed in the brain, including the hippocampus, and are also highly expressed in the paraventricular nucleus of the hypothalamus, whereas MRs are present mainly in the hippocampus and lateral septum. Since aldosterone, the natural ligand for MRs, is not present in brain, glucocorticoid is the preferential ligand for both receptors. In rats, inhibition of basal secretion of ACTH by corticosterone at the low point of diurnal HPA axis activity appears to be mediated through the MR and inhibition at peak activity occurs through the GR potentiated by the MR. Studies in humans have shown that spironolactone, a mineralocorticoid receptor antagonist, increases HPA axis activity.

## SNS

The autonomic nervous system innervates vascular and visceral smooth muscles, exocrine and endocrine glands, and parenchymal tissues throughout the body. The CNS modulates autonomic nervous system activity. Neurons of the CNS receive diverse afferent inputs. After central integration of afferent information, autonomic outflow is adjusted to permit the functioning of the major organ systems in concordance with the needs of the organism as a whole. Sympathetic flow is initiated from the reticular formation of the medulla oblongata, from the pons, and from centers in the hypothalamus. The adrenal medullary chromaffin tissue is also involved in the synthesis and secretion of catecholamine, 85% of which is epinephrine. Catecholamine secretion is stimulated by acetylcholine from the preganglionic sympathetic nerves. All three of the naturally occurring catecholamines, norepinephrine, epinephrine, and dopamine, also function as neurotransmitters within the CNS. Catecholamines influence effector cells by interacting with specific surface receptors coupled to G protein  $\alpha$ - and  $\beta$ -adrenergic receptors.

## AGING

### HPA Axis

Studies on the effects of normal aging on the HPA axis have revealed conflicting results. For the most part, these conflicting results could be explained; small sample size, lack of attention to diurnal variability of cortisol and ACTH concentrations or to sex differences in cortisol concentrations, and the variable of comorbid conditions are among the most important reasons for the differences in the results obtained.

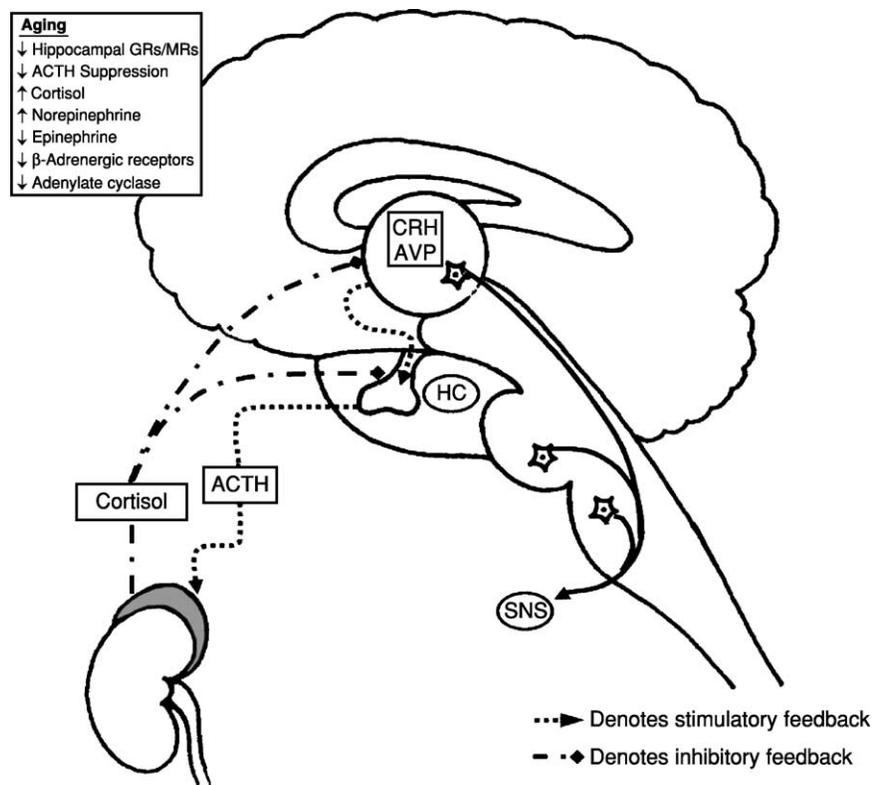
Studies that have given careful attention to diurnal variation and sex differences have, however, consistently shown a clinically significant increase in mean 24h cortisol concentration with aging. However, random serum cortisol determinations and the 24h urinary cortisol concentration remain well within the reference range. The increase in mean cortisol concentration seems to be due to the net effect of two main changes associated with aging. First, there is an increased nocturnal nadir concentration. Second, the nocturnal quiescent period is shortened; that is, the quiescent period starts later and ends sooner. Although the circadian rhythmicity seems to be maintained with aging, the relative amplitude diminishes.

The effectiveness of the negative feedback is also diminished in aging as indicated by studies of the

ACTH response to glucocorticoid suppression and CRH stimulation testing. When exogenous cortisol is used for suppression of ACTH, there is a consistent delayed ACTH suppression in older subjects. However, studies using the synthetic glucocorticoid dexamethasone for HPA axis suppression have revealed conflicting results. CRH stimulation studies have also consistently shown greater pituitary–adrenal secretory responses in older subjects. In addition, there are also sex-related differences in aging; for instance, in one study cortisol response to ACTH was higher in postmenopausal women than in premenopausal women. As mentioned, in humans, spironolactone treatment increases HPA axis activity. This effect is more pronounced in the elderly. Overall, studies consistently show an increased HPA axis secretory activity with aging (see Fig. 1).

Studies in rats have been pivotal in shedding light on the underlying mechanism of aging-associated

changes in HPA axis activity. Under basal conditions, there is a progressive age-related decline in CRH mRNA levels in the paraventricular nucleus of the hypothalamus, hypothalamic CRH content, and hypothalamic *in vitro* CRH secretion. Prolonged treatment with stress doses of corticosterone in middle-aged rats is associated with learning impairment. Also, prolonged exposure to high social stress has been shown to be associated with learning impairment, whereas adrenalectomized rats that receive low-dose corticosterone replacement do not develop learning impairments. Chronic administration of corticosterone has been associated with down-regulation of both GRs and MRs in the hippocampus. An age-related decline in both the synthesis of GRs and MRs and the uptake of their ligands has been reported. Young rats exposed to corticosterone for a prolonged period of time have shown hippocampal neuronal loss similar to that in aged hippocampus. Middle-aged rats



**Figure 1** The hypothalamus secretes corticotropin-releasing hormone (CRH) and vasopressin (AVP), which stimulate adrenocorticotrophic hormone (ACTH) synthesis and secretion from the pituitary gland. ACTH stimulates cortisol synthesis and secretion from the adrenal gland. Cortisol, in turn, inhibits hypothalamic CRH and AVP secretion and pituitary ACTH secretion. The hippocampus (HC) mediates part of the negative feedback effect of glucocorticoids on the hypothalamic–pituitary–adrenal (HPA) axis. Sympathetic flow is initiated from the reticular formation of the medulla oblongata, from the pons, and from centers in the hypothalamus. Aging is associated with hippocampal hypotrophy, a decreased number of hippocampal glucocorticoid receptors (GRs)/mineralocorticoid receptors (MRs), and increased axis activity. Overall, sympathetic nervous system (SNS) activity is decreased with aging, due to decreased  $\beta$ -adrenergic receptor density.

that were adrenalectomized have shown an improvement in hippocampal morphologic correlates of brain aging and reduced neuronal loss. Overall, these studies have led to the notion of central glucocorticoid toxicity; that is, prolonged exposure to glucocorticoid is associated with down-regulation of hippocampal GRs and MRs, leading to increased HPA axis activity and further elevation of glucocorticoid levels. Increased glucocorticoid concentration is also associated with hippocampal neuronal loss, leading to cognitive and memory impairment similar to that observed with aging.

## SNS

Plasma norepinephrine concentration is increased in older adults compared with young adults. This is due to increased secretion and decreased clearance of norepinephrine with aging. On the other hand, plasma epinephrine concentration is slightly decreased with aging. This is due to a significant decline in epinephrine secretion from the adrenal medulla, whereas plasma epinephrine clearance decreases slightly (see Fig. 1). In addition, the adrenergic receptor density, especially the  $\beta$ -adrenergic receptor density, and adenylyl cyclase activity are reduced in aging. These changes have been observed in myocytes, detrusor muscles, and adipocytes.

## MAJOR DEPRESSION

Major depression is associated with disturbances of the HPA axis and the SNS. Specifically, melancholic depression, a subtype of major clinical depression characterized by insomnia, loss of appetite, and weight loss, is associated with significantly higher cerebrospinal fluid norepinephrine levels around the clock.

Similar to the changes that are observed with aging, a higher mean 24 h plasma concentration cortisol has been observed in patients suffering from major depression, mainly due to a shortened quiescent period and an increase in the magnitude of the secretory episodes. Thus, patients suffering from major depression are exposed to higher glucocorticoid and catecholamine concentrations over time, which could theoretically accelerate the aging process.

## ALZHEIMER'S DISEASE

There is some evidence that prolonged exposure to higher glucocorticoid concentrations might be associ-

ated with cognitive impairment and Alzheimer's disease. A group of healthy elderly subjects with different cortisol concentrations were followed over a period of 4 years. The subjects with a significant increase in cortisol concentrations over the years and with high current basal cortisol concentrations were impaired on tasks measuring explicit memory and selective attention compared to subjects who did not show increased cortisol concentrations. In another study, the mean plasma cortisol concentrations after dexamethasone administration were significantly higher in aged subjects suffering from dementia than in clinically healthy aged subjects.

## SUMMARY

Aging is associated with changes in the CNS stress system, a reduced coping response to stressors, and, in some cases, memory and cognitive impairment. Chronic stress by itself induces changes in the HPA axis and SNS, leading to an accelerated aging process. Glucocorticoids are considered to be the main culprit; high concentrations of glucocorticoids are associated with synaptic loss in the hippocampus and hippocampal atrophy and thus cognitive dysfunction. Consequently, conditions that are associated with elevated glucocorticoid concentration, such as major depression, chronic stress, and Cushing's syndrome, could theoretically accelerate the aging process. Furthermore, a subset of the population with an exaggerated HPA axis response might be more prone to these changes. Likewise, interventions, whether pharmacologic or nonpharmacologic, that might decrease glucocorticoid concentrations or its actions might be proven beneficial in delaying the aging process.

## See Also the Following Articles

Aging and Longevity of Human Populations • Aging, Animal Models for • Aging, Immunology and • Alzheimer's Disease and Hormones • Autonomic Nervous System, Aging and • Functional Genomics of Aging • Neuroendocrine System and Aging • Oxidative Stress and Aging

## Further Reading

- Bradbury, M. J., Akana, S. F., and Dallman, M. F. (1994). Roles of type I and II corticosteroid receptors in regulation of basal activity in the hypothalamo-pituitary-adrenal axis during the diurnal trough and the peak: Evidence for a nonadditive effect of combined receptor occupation. *Endocrinology* **134**, 1286-1296.
- Cizza, G., Calogero, A. E., Brady, L. S., *et al.* (1994). Male Fischer 344/N rats show a progressive central impairment of

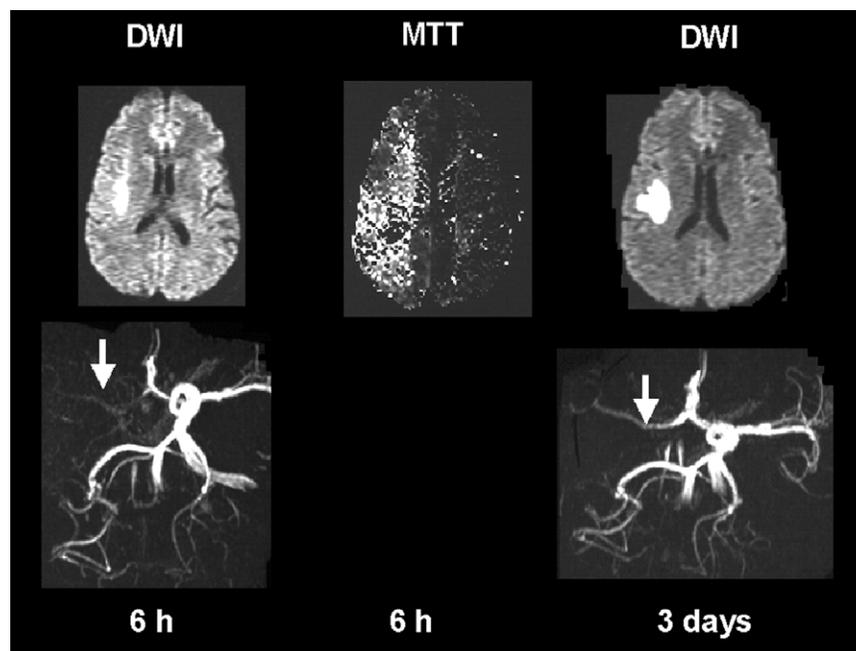
- the hypothalamic–pituitary–adrenal axis with advancing age. *Endocrinology* **134**, 1611–1620.
- De Benedictis, G., Tan, Q., Jeune, B., *et al.* (2001). Recent advances in human gene–longevity association studies. *Mech. Aging Dev.* **122**, 909–920.
- Esler, M., Kaye, D., Thompson, J., *et al.* (1995). Effects of aging on epinephrine secretion and regional release of epinephrine from the human heart. *J. Clin. Endocrinol. Metab.* **80**, 435–442.
- Ferrari, E., Arcaini, A., Gornati, R., *et al.* (2000). Pineal and pituitary–adrenocortical function in physiological aging and in senile dementia. *Exp. Gerontol.* **35**, 1239–1250.
- Heuser, I., Deuschle, M., Weber, A., *et al.* (2000). The role of mineralocorticoid receptors in the circadian activity of the human hypothalamus–pituitary–adrenal system: Effect of age. *Neurobiol. Aging* **21**, 585–589.
- Landsberg, L., and Young, J. B. (2001). Physiology and pharmacology of the autonomic nervous system. In “Principles of Internal Medicine” (E. Braunwald, A. S. Fauci, D. L. Kasper, S. L. Hauser, D. L. Longo, and J. J. Jameson, eds.), pp. 438–448. McGraw-Hill, New York.
- Luisi, S., Tonetti, A., Bernardi, F., *et al.* (1998). Effect of acute corticotropin releasing factor on pituitary–adrenocortical responsiveness in elderly women and men. *J. Endocrinol. Invest.* **21**, 449–453.
- Parker, C. R., Jr., Slayden, S. M., Azziz, R., *et al.* (2000). Effects of aging on adrenal function in the human: Responsiveness and sensitivity of adrenal androgens and cortisol to adrenocorticotropin in premenopausal and postmenopausal women. *J. Clin. Endocrinol. Metab.* **85**, 48–54.
- Van Cauter, E., Leproult, R., and Kupfer, D. J. (1996). Effects of gender and age on the levels and circadian rhythmicity of plasma cortisol. *J. Clin. Endocrinol. Metab.* **81**, 2468–2473.
- Waltman, C., Blackman, M. R., Chrousos, G. P., *et al.* (1991). Spontaneous and glucocorticoid-inhibited adrenocorticotrophic hormone and cortisol secretion are similar in healthy young and old men. *J. Clin. Endocrinol. Metab.* **73**, 495–502.
- Wilkinson, C. W., Peskind, E. R., and Raskind, M. A. (1997). Decreased hypothalamic–pituitary–adrenal axis sensitivity to cortisol feedback inhibition in human aging. *Neuroendocrinology* **65**, 79–90.

fibrinolytic activity, and increased release of B thromboglobulin. Unfortunately, tight glycemic control in diabetic patients has not been shown to reduce the risk of stroke (as either a primary or a secondary preventive measure), although it does reduce the risk of many other mostly microvascular complications. In patients with myocardial infarction, treatment of hyperglycemia (blood sugar levels  $>11$  mmol/L) was shown to decrease mortality. Recently, in a study by Gaede and colleagues published in the *New England Journal of Medicine*, the intensive concurrent treatment of various risk factors for cardiovascular disease, including diabetes, was shown to decrease the risk of stroke. Although nonfatal stroke was not the primary end point, and intensive treatment was simultaneously targeted at a number of risk factors such as diabetes, there were only 3 nonfatal strokes in the intensive treatment group and 11 in the control group. However, the specific contribution of glycemic control to this overall risk reduction is not clear.

Diabetes mellitus and acute hyperglycemia are associated with a poor outcome of stroke. Although it appears increasingly likely that hyperglycemia causes this poorer outcome, this has not been definitely established

in humans. Animal studies have shown convincingly that hyperglycemia, as compared with euglycemia, increases the extent of ischemic damage in rats and monkeys. However, in focal ischemia models, this effect was shown only for reperfused brain tissue, albeit less consistently. The negative effects of hyperglycemia on outcome of brain ischemia are probably mediated through increased lactic acidosis and increased release of excitatory amino acids (glutamate in particular) that may contribute to neuronal cell death (excitotoxicity), exaggeration of edema formation, blood-brain barrier disruption, and hemorrhagic transformations. In humans, the data underlining a causal relation between hyperglycemia and stroke are still only circumstantial. This is explained partly by the intricate relations between the pathogenesis of hyperglycemia and the evolution and type of the stroke.

First, reperfused brain tissue may be especially vulnerable to hyperglycemia, whereas anoxic tissue may even benefit from hyperglycemia. Recently, it was shown that the ischemic penumbra in stroke patients, represented as a mismatch between perfusion and diffusion on magnetic resonance imaging (MRI) (Fig. 1), is more likely to progress to infarction when



**Figure 1** Diffusion weighted imaging (DWI) sequence 6 h after the onset of a stroke showing a circumscribed area of high signal indicating disturbed diffusion. This represents infarction of the brain. The magnetic resonance (MR) angiography after 6 h shows absent flow in the right middle cerebral and carotid artery. The mean transit time (MTT) MR shows hypoperfusion of the total right middle cerebral artery territory. The area of hypoperfusion on the MMT MR minus the area of disturbed diffusion on the DWI MR is generally considered to be the penumbra. The DWI MR after 3 days shows that the hypoperfused area has not progressed to infarction. The MR angiography shows reperfusion of the right middle cerebral artery through the circle of Willis. Courtesy of Geoffrey Donnan, National Stroke Research Institute.

the patient had hyperglycemia at presentation. Also, hyperglycemia levels were related to lactate levels, which in turn were independently related to salvage of mismatch tissue. In a previous study, a poorer outcome of strokes had been reported when hyperglycemia occurred in nonlacunar strokes that may show reperfusion but not in lacunar strokes with little or no reperfusion. A special vulnerability of reperfused brain tissue to hyperglycemia is also suggested by the experiences with thrombolysis in stroke patients. In the first trial showing the benefit of thrombolysis in acute stroke patients, the negative effects of hyperglycemia on outcome of thrombolysis were also reported. Further studies have shown that it was specifically in patients with reperfusion after tissue plasminogen activator (TPA) that hyperglycemia was associated with poor outcome. In these patients, there is a relation with hemorrhagic transformation of the infarct. Interestingly, diabetes mellitus has actually been reported to carry a decreased risk of primary intracerebral haemorrhage, so it seems less likely that diabetic vasculopathy explains the increased incidence of hemorrhagic transformation in thrombolized strokes. The fact that stroke type determines the association between hyperglycemia and stroke outcome is also underlined by the fact that, in hemorrhagic strokes, hyperglycemia is not related to a poorer outcome in both nondiabetic and diabetic patients.

Second, it seems that acute hyperglycemia may be more strongly related to poor outcome than is hyperglycemia due to diabetes mellitus. In a large meta-analysis, hyperglycemia was found to have a relative risk of mortality after ischemic stroke of 3.1 in nondiabetic patients and only 1.3 in diabetic patients. In conclusion, the literature suggests a causal relation between hyperglycemia and stroke outcome rather than hyperglycemia as a parphenomenon of strokes with more severe outcome. It seems sensible to withhold intravenous fluids containing glucose during the acute phase of a stroke, and hyperglycemia should be controlled with the usual measures, especially when blood sugar levels higher than 16 mmol/L are measured. However, whether tight control of hyperglycemia should be the goal in patients with stroke cannot be determined with certainty until currently conducted trials of treatment of hyperglycemia in acute stroke patients are concluded.

Hyperglycemia occurs in 20 to 50% of stroke patients. One-third of these hyperglycemic patients were known to have diabetes mellitus. In another third, the hyperglycemia is the initial presentation of *de novo* diabetes, demonstrated by an elevation of glycosylated hemoglobin (HbA1c) levels. In the remainder (with

a normal HbA1c at the time of hyperglycemia), the hyperglycemia is considered to be a result of the stroke, but the mechanism for this is unclear. Although a general stress response may lead to hyperglycemia, this is probably not the predominant cause in patients with a stroke. Other parameters of the stress response, such as levels of catecholamines, have been shown not to be related to blood sugar levels after stroke. Because the effects of focal ischemic events on neurotransmitter release at a distance from the focal ischemia can operate in the entire ipsilateral hemisphere of the stroke, the neuroendocrine axis may well be influenced by these alterations in neuronal excitation. Alternatively, the focal ischemic brain may mediate hyperglycemia directly through as yet unclarified mechanisms. If this were to occur preferentially in inadequately reperfused brain, the association of hyperglycemia and poor outcome of stroke would represent a parphenomenon and not a causal relationship.

Although hyperglycemia will rarely be confused with stroke, the differential diagnosis between stroke and hypoglycemia can be problematic. The blood sugar level in the patient presenting with coma or focal neurological deficit is obviously of crucial importance. If it is normal, hypoglycemia is unlikely. However, in the case of focal neurological deficit, the hypoglycemia may have been corrected by the patient's oral glucose intake before the neurological signs and symptoms resolve. When the patient is comatose, established hypoglycemia will be the likely cause. More difficulties arise when the patient presents with hypoglycemia and focal neurological deficits. Hypoglycemia can occasionally present with focal neurological deficit without any accompanying symptoms of hypoglycemia. Correction of blood glucose levels will generally lead to a rapid recovery of the deficit, but excluding a transient ischemic event might be impossible in these cases. In patients with previous strokes, hypoglycemia reproduces or worsens the previous deficit. When there are accompanying signs and symptoms of hypoglycemia, the balance will definitely shift toward hypoglycemia as a cause of the neurological deficit.

## THYROID DISEASE

Both hyperthyroidism and hypothyroidism can contribute to a cardioembolic source for stroke. The most frequent event is hyperthyroidism leading to atrial fibrillation, a factor that generally leads to a fivefold increase in risk of subsequent stroke and

also a factor that, if it cannot be cured, leads to anticoagulation. Hypothyroidism can affect cardiac function and, thereby, the risk of intracardiac thrombus formation. Graves–Basedow disease is rarely associated with cerebral vasculitis, an entity also referred to as Hashimoto's encephalopathy.

Thyroid disease has not been firmly established as affecting the outcome of strokes. However, in general, a range of systemic diseases affect the outcome of stroke, and it seems appropriate to diagnose and treat thyroid disease promptly in patients in these circumstances.

Strokes have not been shown to lead to thyroid disease.

Thyrotoxicosis may feature in the differential diagnosis of stroke. Delirium and coma may be a feature of both, but the fever, tachycardia, hypotension, vomiting, and diarrhea should point toward thyrotoxicosis. Hyperthyroidism has been reported to lead to isolated corticospinal tract dysfunction, the mechanism of which is not known. Hypothyroidism may present with limb and gait ataxia as signs of cerebellar dysfunction. The onset is often more gradual than in strokes.

## PARATHYROID DISEASE

Hyperparathyroidism is known to contribute to hypertension. Patients with osteoporosis and compensatory hyperparathyroidism may have an increased risk of stroke due to this contribution. Hypoparathyroidism has not been identified as a risk factor for stroke, but it is related to MELAS syndrome. MELAS (mitochondrial myopathy, encephalopathy, lactic acidosis, stroke) is a mitochondrial disorder related to a point mutation in the transfer RNA<sup>Leu</sup>(UUR) gene. It is typified by mitochondrial (ragged red fiber) myopathy, epilepsy, lactate acidosis, and stroke-like episodes. Frequently, this syndrome is associated with hypothalamo–pituitary axis dysfunction, diabetes mellitus, and/or hyperthyroidism.

Parathyroid dysfunction has not been shown to affect the outcome of stroke or to be the result of stroke.

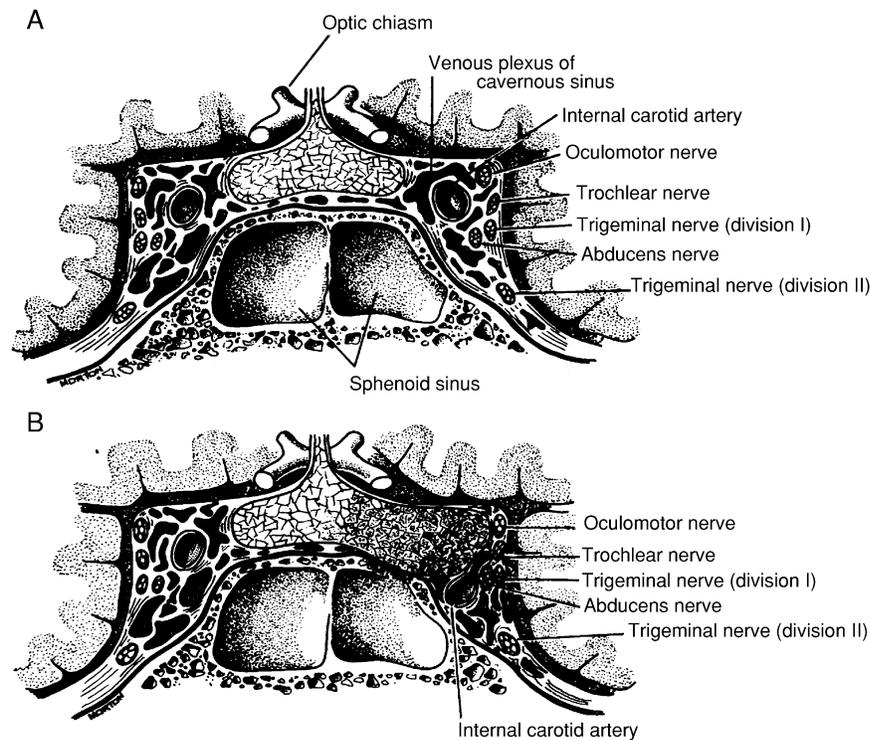
Hypercalcemia, whether as a result of hyperparathyroidism or secondary to malignancy, may be confused with a stroke when a patient presents with confusion, nausea, and vomiting. In hypercalcemia, there is fatigue, anorexia, constipation, increased urination, a short QT interval, and generally a serum calcium level higher than 2.9 mmol/L (11.5 mg/dl).

## HYPOTHALAMO–PITUITARY AXIS DISEASE

Hypopituitarism is associated with an increased risk of stroke, reflected in an excess mortality rate for cerebrovascular disease of 2.4. Because of the complexity of hormonal and metabolic disturbances in hypopituitarism, the causes of this excess mortality are poorly defined. A prothrombotic tendency due to growth hormone deficiency (which tends to persist even when hormones are adequately substituted), Cushing's syndrome, and secondary hyperthyroidism are among the factors that may be implied in the pathogenesis of strokes.

Pituitary disease has not been shown to affect the outcome of stroke, although it may affect the chances of recurrence of stroke.

Pituitary apoplexy is the term that refers to a stroke in the pituitary gland. Such strokes can be either ischemic or hemorrhagic and are often complications of pituitary adenomas or the surgery or radiotherapy thereof. The condition is probably underdiagnosed given that the pathological correlate of apoplexy has a prevalence of 1 to 3% in autopsy studies. The clinical presentation is highly variable but should be suspected in any patient with severe headache, visual field defects, ophthalmoplegia, and/or altered mental status. In some cases, there is only headache at first. This headache is accompanied by nausea, vomiting, nuchal rigidity, fever, stupor, and coma (when blood and necrotic tissue leak and cause the features of subarachnoid hemorrhage or aseptic meningitis). In tumorous apoplexy, the destruction of the pituitary in most cases leads to hypopituitarism because most underlying adenomas are endocrinologically silent. In cases of endocrinologically active adenomas, there can be spontaneous resolution of preexisting endocrinopathy after apoplexy. The slowly evolving expansion of adenomas rarely leads to cranial nerve deficit because the nerves slowly lengthen in response to this expansion. In apoplexy (Fig. 2), there is sudden compression of cranial nerves due to hemorrhage or infarction and swelling. Compression of the optic nerve leads to visual field defects and decreased visual acuity, whereas compression of nerves III, IV, and VI leads to oculomotor disturbances. Compression of the trigeminal nerve may result in facial paresthesias and absent corneal reflex. Compression of the cavernous sinus may result in proptosis and eyelid edema. Finally, there can be Horner's syndrome due to compression of the sympathetic chain and hyperpyrexia, and there can be diabetes insipidus or SIADH



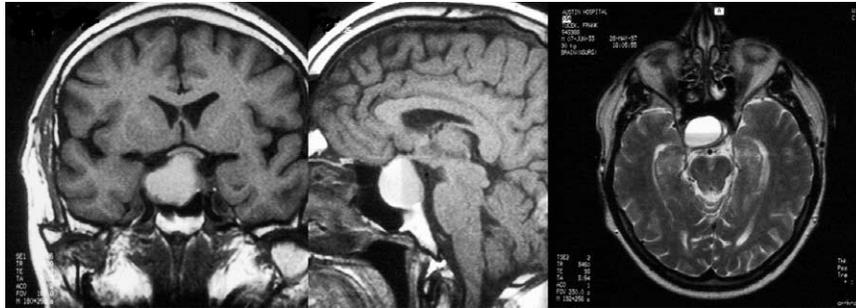
**Figure 2** (A) Posterior coronal view showing anatomical relationship of pituitary gland to optic chiasm superiorly, sphenoid sinus inferiorly, and cavernous sinus laterally. (B) Mechanism of acute compression of structures within cavernous sinus from sudden expansion of pituitary adenoma due to hemorrhage or infarction and edema. Note that the further the tumor has eroded the floor of sella turcica prior to apoplectic episode, the more likely it is that multiple structures within the cavernous sinus will be involved. Reproduced from Reid, R., Quigley, M., and Yen, S. (1985). Pituitary apoplexy. *Arch. Neurol.* **42**, 712–719.

(syndrome of inappropriate antidiuretic hormone) when the hypothalamus is compressed.

The radiological diagnosis can be made on computed tomography (CT) if the stroke is hemorrhagic or if the adenoma is sufficiently large. The investigation of choice is MRI, which shows hemorrhage and underlying adenomas in great detail (Fig. 3). Infarction can be difficult to diagnose on routine magnetic resonance sequences. Diffusion-weighted images greatly increase the accuracy of the diagnosis of pituitary infarction. Peripheral enhancement of intrasellar masses can be another less specific sign. The hypothalamic–pituitary unit can be tested as outlined elsewhere. Pregnant women have a special predisposition to apoplexy due to the prothrombotic state related to pregnancy (and the resulting 12-fold increase in risk of stroke) and the dramatic enlargement of the pituitary gland due to the proliferation of prolactin-secreting cells. The classical features of postpartum pituitary apoplexy are absence of lactation, persistent amenorrhea, and lethargy. Other predisposing factors to pituitary apoplexy include bleeding disorders, anticoagulation, upper

respiratory tract infections, trauma, carotid angiography, Cushing’s disease, diabetes mellitus, adrenalectomy, atherosclerosis, sickle cell trait, and acromegaly. Surgical intervention is advocated in the majority of patients, with radiotherapy a less likely option.

Less catastrophic is the effect of nonpituitary stroke on the hypothalamic–pituitary–adrenal (HPA) axis. Stroke is associated with increased activity of the HPA, manifested particularly by hypercortisolism. The normal regulation of cortisol secretion by adrenocorticotropic hormone (ACTH) is disturbed after stroke, a process in which cytokines seem to be implicated. Such a pathological HPA axis may exhibit an overriding function on established risk factors for cardiovascular disease, diabetes mellitus, and stroke such as abdominal obesity, hypertension, cholesterol, and triglycerides. Whether there is a causal relation between stroke and increased HPA axis activity needs to be further elucidated given that depression, anxiety, alcohol consumption, and smoking all have been shown to have similar effects. Also, changes in growth hormone, prolactin, and thyrotropin response to



**Figure 3** Axial and sagittal T1-weighted and transverse T2-weighted MRIs showing a large hemorrhage into the pituitary gland. Courtesy of G. Fitt.

thyrotropin-releasing hormone have been reported, and some of these changes may be a consequence of the hypersensitive HPA axis. They may explain some of the insulin resistance after stroke. It has been postulated that stroke in the caudate nucleus interrupts neurotransmitter pathways involved in the control of secretion of gonadotropins. Although hypercortisolism and some of the other disturbances have been related to disorientation and levels of motor impairment, the clinical relevance of many of the changes remains uncertain. In general, stroke leads to an increase in antidiuretic hormone (ADH) levels. However, this usually does not lead to hyponatremia. Patients with subarachnoid hemorrhage are especially likely to develop a syndrome of inappropriate ADH secretion (SIADH). However, it is important to distinguish SIADH from cerebral salt wasting, which is the more likely explanation for hyponatremia in patients with subarachnoid hemorrhage.

### ADRENAL GLAND DISEASE

The intermittent hypertension in pheochromocytoma increases the risk of stroke. In some cases, the pheochromocytoma is actually not identified until the patients present with a stroke. Also, Cushing's disease

has been postulated to have an increased risk of stroke due to hypercortisolism or to the effects of treatment of Cushing's disease such as external pituitary irradiation and posttreatment hypopituitarism.

### See Also the Following Articles

Cardiovascular Disease in Diabetes • Graves' Disease, Hyperthyroidism in • Hypercalcemia and Hypercalcemia Treatment • Hypercortisolism and Cushing's Syndrome • Hypertension and Diabetes • Hypopituitarism • Lactic Acidosis • Pheochromocytoma • Thyrotoxicosis: Diagnosis

### Further Reading

- Gaede, P., Vedel, P., Larsen, N., Jensen, G., Parving, H-H., and Pedersen, O. (2003). Multifactorial intervention and cardiovascular disease in patients with type 2 diabetes. *N. Engl. J. Med.* **348**, 383–393.
- Parsons, M., Barber, A., Desmond, P., Baird, T., Darby, D., Byrnes, G., Tress, B., and Davis, S. (2002). Acute hyperglycemia adversely affects stroke outcome: A magnetic resonance imaging and spectroscopy study. *Ann. Neurol.* **52**, 20–28.
- Reid, R., Quigley, M., and Yen, S. (1985). Pituitary apoplexy. *Arch. Neurol.* **42**, 712–719.
- Rosmond, R., and Bjorntop, P. (2000). The hypothalamic–pituitary–adrenal axis activity as a predictor of cardiovascular disease, type 2 diabetes, and stroke. *J. Int. Med.* **247**, 188–197.

shortly after birth with the sensory neurotoxin capsaicin lack this extrinsic innervation to the alimentary tract, with the result that there is a marked depletion of SP fibers in the gastric mucosa and around gastric submucosal blood vessels. A scattered population of SP-immunoreactive endocrine-like cells is present in the mucosa of the human small intestine and colon.

In the pancreas, a sparse distribution of SP fibers, which are probably of extrinsic origin, innervate blood vessels and acini and such fibers are also found surrounding acini and along blood vessels in the salivary glands. In the hepatobiliary system, immunoreactive fibers are found in the parenchyma of the liver and hepatic vasculature and are localized to the ganglionated and mucosal plexi of the gallbladder. In the trachea and bronchi, extrinsic SP-containing nerve fibers are found within the smooth muscle layer and around local ganglion cells and, in the nasal mucosa, within and under the epithelium and around arterioles, venules, and exocrine glands. In the human heart, SP fibers are found in close proximity to arterioles and are localized to the adventitia and to the border between the adventitia and media in a wide range of blood vessels. Within the urogenital system, immunoreactive nerve fibers of extrinsic origin are present in the urinary bladder throughout the ureter close to smooth muscle cells and around blood vessels in the kidney cortex often close to renal tubules and glomeruli. The female (uterus, oviduct, and vagina) and male (seminal vesicle, testis, epididymis, and vas deferens) genital organs are also innervated by extrinsic SP fibers.

Consistent with predominantly neuronal localization of the peptide and its rapid rate of clearance from the circulation (half-life <1 min), the plasma concentration of SP in healthy subjects is very low (<10 fmol/ml). Elevated concentrations of circulating SP (and other tachykinins) are often seen in patients with carcinoid tumors, particularly metastatic tumors of the midgut region. The synthesis and release of SP by cells of the immune system (monocytes, lymphocytes, eosinophils and macrophages) have been reported and SP fibers innervate thymus, spleen, lymph nodes, and bone marrow, supporting the concept that the tachykinins constitute a link between the central nervous system (CNS) and the immune system.

### NK<sub>1</sub> Receptor

Of the three structurally related cell surface receptors that recognize SP, the NK<sub>1</sub> receptor binds the peptide with the highest affinity, so this receptor is often termed the SP receptor. The NK<sub>1</sub> receptor is widely

but discretely distributed in different regions of the brains of mammals, although species differences are apparent. In peripheral tissues, NK<sub>1</sub> receptors are present on human pulmonary arterial blood vessels but not on bronchial smooth muscle, on circular and longitudinal smooth muscle throughout the human gastrointestinal tract, and over ganglia of the myenteric plexus. The potent sialagogic effect of SP is consistent with the high concentration of NK<sub>1</sub> receptors on rat submaxillary gland. The potent vasodilator action of SP is mediated primarily by binding to NK<sub>1</sub> receptors on the endothelium of peripheral arterial blood vessels. In the urogenital system, NK<sub>1</sub> sites are expressed on blood vessels in the urinary bladder, uterus, and ureter; in the skin, NK<sub>1</sub> sites are present at high density in postcapillary venules in the dermis. The occurrence of NK<sub>1</sub> receptors in spleen, in thymus, on arterioles and venules of the lymph nodes, and on T lymphocytes provides further evidence for an involvement of SP in immunoregulation.

### PHYSIOLOGICAL ROLES

The principal biological actions of SP are summarized in [Table I](#). It is well established that SP acts as a neurotransmitter in primary afferent sensory neurons and is involved in the perception of pain. The release of SP from small-diameter sensory "pain" fibers into the dorsal horn of the spinal cord following intense peripheral stimulation promotes central hyperexcitability and increased sensitivity to pain. Intrathecal injection of SP in mice elicits behaviors suggestive of pain sensation. However, in rats, intracerebroventricularly administered SP displays a clear-cut antinociceptive activity in the hot plate test, which is abolished by treatment with the opioid antagonist, naloxone. It has been suggested that proteolytic conversion of SP to its (1–7) fragment is a prerequisite for development of the analgesic action. Knockout mice lacking expression of the preprotachykinin A gene exhibit a surprisingly healthy phenotype and a normal response to mildly painful stimuli, e.g., the tail flick assay, but the response to moderate to intense pain was significantly reduced. Similarly, mice with a targeted deletion of the NK<sub>1</sub> receptor for SP were healthy and fertile and responded normally to acute painful stimuli but were unable to develop fully the stress-induced analgesia of control animals.

Intracerebroventricular injections of SP in rats and mice induce a wide range of behavioral responses, including increased grooming and scratching, increased hindlimb rearing, decreased aggressive behavior, and an enhancement of inhibitory avoidance

Table I A Summary of the Principal Biological Actions of Substance P with Correlations to Human Diseases

Organ system	Biological activity	Disease correlation
Central nervous system	Neurotransmitter/neuromodulator function	Migraine
	Nociception	Riley-Day syndrome Parkinson's disease Huntington's disease Alzheimer's disease Depression/anxiety
Cardiovascular system	Vasodilation	Carcinoid flush
	Plasma extravasation	
	Tachycardia	
	Cardiac output ↑	
Respiratory system	Bronchoconstriction	Asthma
	Vasodilation	
	Vascular permeability ↑	
	Mucus secretion ↑	
Salivary glands	Salivation ↑	
Gastrointestinal tract	Peristalsis ↑	Hirschprung's disease
	Water and electrolyte secretion ↑	Ulcerative colitis
	Blood flow ↑	Crohn's disease
Pancreas	Exocrine secretion ↑	
Hepatobiliary system	Bile secretion ↓	Inflammation
	Gallbladder motility ↑	
Urogenital system	Motility of ureter and bladder ↑	Urinary incontinence
	Blood flow and motility of uterus ↑	
	Renal vasodilation	
Musculoskeletal system	Neurogenic inflammation ↑	Arthritis
	Blood flow to skeletal muscle ↑	

learning. Studies in guinea pigs have suggested that blockade of central SP receptors may inhibit behavioral responses to psychological stress. A role for the SP metabolite SP-(1-7), presumably generated from SP within the CNS, in effecting these responses has been proposed.

Intra-arterial infusions of SP in humans result in a fall in diastolic blood pressure, an increase in heart rate, and an increase in cardiac output, due mostly to a greater stroke volume. The peptide is a powerful vasodilator in humans with high-dose (70 pmol/kg/min) infusions, producing a bright red flushing of the skin, particularly in the neck and face, with a subjective feeling of warmth. Intradermal injections of SP in humans produce flare, wheal, and itching that are related to the release of histamine and an increase in capillary permeability. The potent action of SP in causing plasma extravasation is mediated through the NK<sub>1</sub> receptor.

The precise actions of SP on the motility of the gastrointestinal tract are species-dependent, but

*in vitro* the peptide generally evokes a contractile response on tissues from all regions and *in vivo* is a strong stimulator of peristalsis by a mechanism that involves both a direct action on smooth muscle and a release of acetylcholine from cholinergic neurons. SP stimulates the secretion of water and electrolytes from the mucosa of the small intestine and colon of various mammals and produces an increase in short-circuit current in the guinea pig colon that is indicative of net ion transport across the tissue. The release of SP from afferent fibers innervating the stomach and gut and from intrinsic fibers of the submucous ganglia supplying submucosal arteries produces vasodilation, leading to an increase in blood flow in the mesenteric artery. A cytoprotective role for the gastric mucosa has been proposed for the SP released from primary afferent terminals.

Intra-arterial infusions of SP in dogs weakly stimulate basal output of pancreatic juice, amylase, and bicarbonate, whereas in the isolated rat pancreas, the peptide inhibits cholecystokinin-induced amylase

release and secretin-induced flow and, at high concentrations, weakly stimulates the release of insulin, glucagon, and somatostatin. However, a major physiological role for SP in the regulation of pancreatic function seems improbable. Similarly, although SP is a potent stimulator of salivary secretion in rodents, it has only weak sialagogic activity in humans.

## PATHOPHYSIOLOGY AND DISEASE

The involvement of SP in human disease is summarized in [Table I](#). A link between SP and nociception has been clearly established. Cerebral and pial arteries are richly innervated by SP fibers and it has been suggested that neurogenic inflammation within the cerebral tissues leading to the release of SP may be involved in the pathogenesis of migraine and cluster headaches. Similarly, the inflammatory response to trauma in the eye may be due to a release of SP. Patients with reduced sensitivity to pain and temperature, as in Riley-Day syndrome, are associated with a marked depletion in the SP content of the medulla oblongata and substantia gelatinosa of the spinal cord. An involvement of SP in the pathophysiology of neurodegenerative diseases is indicated by the significant decrease in the concentration of the peptide in the substantia nigra and globus pallidus of patients with Huntington's disease and Parkinson's disease. A similar decrease in the SP content in the cerebral cortex and hippocampus has been found in patients with Alzheimer's disease.

In the periphery, a role for SP in the pathogenesis of rheumatoid arthritis is suggested by the observations that joints that develop severe arthritis are densely innervated by SP-containing fibers and concentrations of the peptide in synovial fluid are increased. NK<sub>1</sub> receptors are also found on synovial endothelial cells in patients with rheumatoid arthritis. Thus, tachykinins derived from hyperactive afferent nerve endings in the arthritic joints may promote the chronic inflammatory processes. In the gastrointestinal tract, SP-containing fibers are almost absent from the aganglionic colons of patients with Hirschprung's disease. In conditions of chronic inflammatory bowel disease, such as ulcerative colitis and Crohn's disease, there is an increase in the concentration of SP in the inflamed mucosa together with a marked increase in the expression of SP-selective NK<sub>1</sub> receptor sites in

the arterioles and lymphatic tissue. SP is a potent stimulator of histamine release from mast cells in the intestines of nematode-infected rats. Conversely, chronic constipation is associated with a decrease in colonic SP levels.

The synthesis of SP in neoplastic tissue has been described for a range of neuroendocrine tumors. The peptide has been detected at relatively low concentrations in pheochromocytomas, medullary carcinomas of the thyroid, and small-cell lung carcinomas. Much higher concentrations are found in carcinoid tumors, particularly those arising in the midgut region and their metastases. It is generally accepted that the etiology of the marked cutaneous vasodilation and tachycardia seen in the carcinoid flush is multifactorial but SP (and other tachykinins) released by tumor tissue into the circulation has been implicated in some, but not all, patients.

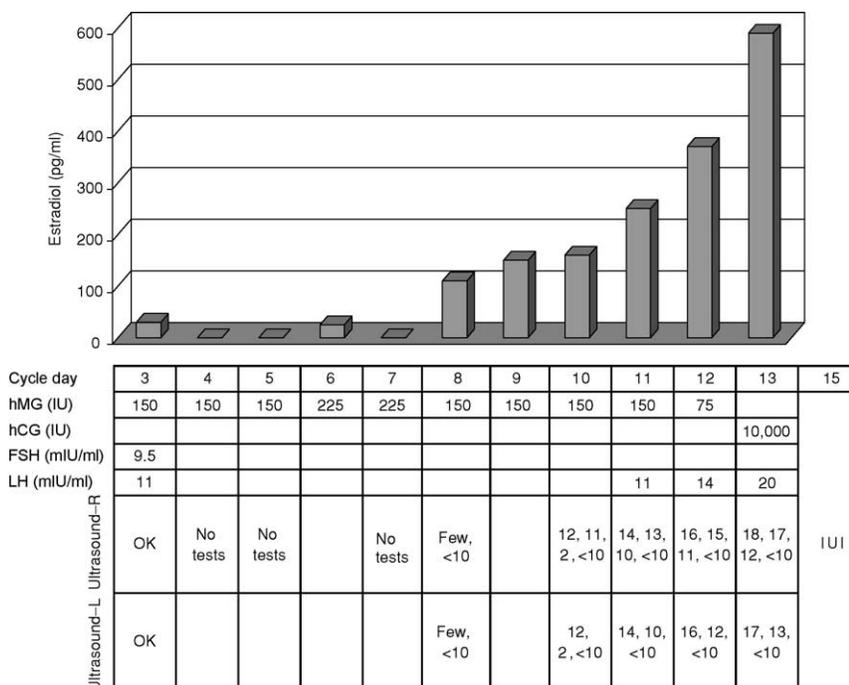
A wide range of antagonists, both peptide and non-peptide, for the NK<sub>1</sub> receptor have become available for potential use as therapeutic agents. Initial efforts at drug development were directed toward analgesics and anti-inflammatory agents but clinical trials are under way to assess the value of NK<sub>1</sub> receptor antagonists in the treatment of depression and anxiety disorders and irritable bowel syndrome and for use as anti-emetics, particularly in patients with chemotherapy-induced nausea and vomiting.

## See Also the Following Articles

Neurokinins • Neurotensin • Neurotransmitters, Overview  
• Peptide Neurotransmitters and Smooth Muscle in the Gut

## Further Reading

- Holzer, P. (1991). Capsaicin: Cellular targets, mechanisms of action, and selectivity for thin sensory neurons. *Pharmacol. Rev.* **43**, 143–201.
- Otsuka, M., and Yoshioka, K. (1993). Neurotransmitter functions of mammalian tachykinins. *Physiol. Rev.* **73**, 299–308.
- Quartara, L., and Maggi, C. A. (1997). The tachykinin NK1 receptor. I. Ligands and mechanisms of cellular activation. *Neuropeptides* **31**, 537–563.
- Quartara, L., and Maggi, C. A. (1998). The tachykinin NK1 receptor. II. Distribution and pathophysiological roles. *Neuropeptides* **32**, 1–49.
- Regoli, D., Boudon, A., and Fauchère, G. L. (1994). Receptors and antagonists for substance P and related peptides. *Pharmacol. Rev.* **46**, 551–599.



**Figure 1** Sample superovulation cycle using human menopausal gonadotropin (hMG) for ovarian stimulation, human chorionic gonadotropin (hCG) to trigger ovulation, and intrauterine insemination (IUI). Follicular size (mm), as determined by ultrasound, and serial serum estradiol, follicle-stimulating hormone (FSH), and luteinizing hormone (LH) levels are followed serially. Progesterone supplementation is started 2 days after the insemination.

are predictive of a poor ovarian response. A baseline ultrasound should also be performed to ensure that the endometrial lining is thin (menstrual) and that the ovary is free of functional cysts or any pathology that may potentially impact on follicular development. Initial doses are based on the patient's diagnosis, age, weight, and previous responses to ovarian stimulation, but typically range between 75 and 450 IU/day. Sequential serum estradiol levels and ultrasound assessment of follicle growth determine the patient's response and daily dosage requirements. Once the lead follicle reaches a preovulatory size ( $\geq 17$ –18 mm), ovulation is triggered with an intramuscular injection of hCG. Intrauterine insemination (IUI) or intercourse follows over the subsequent 24–36 h, as ovulation can be expected approximately 36 h after the administration of hCG (Fig. 1). Pregnancy rates with gonadotropins and IUI range from 10 to 20% per cycle, compared to approximately 5% for clomiphene citrate and IUI. Adjuvant medical therapies are occasionally used to augment the ovarian response during superovulation. Insulin sensitizers, such as glucophage, have been successfully employed in women with polycystic ovary disease to yield higher spontaneous and clomiphene-induced ovulation rates.

## COMPLICATIONS OF SUPEROVULATION

Use of exogenous gonadotropins presents the risk of ovarian hyperstimulation syndrome (OHSS), a multi-organ condition caused by increased capillary permeability within the peritoneal cavity with accumulation of ascites, which in severe cases can result in pulmonary edema, renal failure, thromboembolism, and, potentially, death. The incidence of OHSS can be minimized with close monitoring of the patient's serum estradiol and follicle numbers, by moderating gonadotropin dosages, and by having a low threshold for withholding the administration of hCG, which is required for the condition to manifest. Spontaneous abortion rates appear to be higher with superovulation; however, common causes of infertility, such as polycystic ovary disease, are independent risk factors for early pregnancy loss. Multiple gestation is a well-recognized complication of ovarian stimulation. The incidence of twin gestation is approximately 20% and the risk of higher order multiple pregnancies is approximately 5%. Superovulation also increases the risk of tubal pregnancy several-fold, as well as the risk of heterotopic pregnancy, where gestational sacs are

simultaneously present inside and outside the uterine cavity. Infertile patients, in general, are at increased risk for ectopic pregnancy because of their incidence of tubal disease, although increased embryo numbers and negative effects of gonadotropins on tubal motility have been implicated. The link between ovarian cancer and controlled ovarian stimulation is controversial; however, evidence seems to indicate that the increased risk of epithelial carcinomas can be attributed to independent risk factors common to this population of women, namely, that they have lower parity and reduced exposure to oral contraception.

## TECHNIQUES OF INTRAUTERINE INSEMINATION

Intrauterine insemination is the process of introducing a prepared solution of sperm into the uterine cavity, with the objective of providing a concentration of sperm in the oviduct that is adequate for fertilization. Seminal specimens are collected by masturbation into a sterile container and then allowed to liquefy at room temperature. Sperm-washing procedures and separation techniques serve to remove antigenic proteins, prostaglandins, and infectious agents, as well as nonmotile spermatozoa, leukocytes, and immature germ cells. The end result is a small volume of highly concentrated motile sperm that is safe to introduce into the sterile environment of the uterine cavity. In the past, semen was injected in or around the cervix; however, intrauterine insemination provides the highest concentration of sperm in the tube and peritoneal cavity and yields higher pregnancy rates. Fecundity rates with superovulation appear to be higher when intrauterine insemination is used than when timed intercourse is carried out. Several small studies have demonstrated an increased fecundity with the use of IUI compared to timed intercourse, in both CC and gonadotropin cycles. Controversy exists regarding whether one or two inseminations should be performed. Ideally, an insemination should be performed as close to the time of ovulation as possible. This can be accomplished by detecting the preovulatory LH surge with urine or serum testing or by triggering ovulation pharmacologically with hCG. Ultrasound assessment of follicle size is also used to determine the proximity to spontaneous ovulation and for the timing of hCG administration.

## COMPLICATIONS OF INTRAUTERINE INSEMINATION

Since the uterine cavity is normally a sterile environment, introduction of a catheter through the cervix

carries the potential for infection. Salpingitis has been described after IUI and appears to be linked to bacteria present within the sperm sample. Development of anti-sperm antibodies in the recipient is a theoretical concern, but occurs in less than 5% of women following IUI.

## INDICATIONS FOR SUPEROVULATION AND INTRAUTERINE INSEMINATION

Superovulation and IUI are used for a wide variety of both male and female reproductive disorders. They are most commonly used to treat central and peripheral ovulatory disorders, as well as unexplained infertility. When infertility is the result of a male anatomic or ejaculatory problem, IUI is an effective treatment. It also has a theoretical benefit of overcoming subtle abnormalities in sperm concentration and motility and by circumventing a potentially hostile cervical environment. In vitro fertilization (IVF) is the only option for couples with significant male factor infertility. If the sperm impairment is profound, a single healthy sperm can be selected and placed within the cytoplasm of the egg to achieve fertilization—a process called intracytoplasmic sperm injection (ICSI). Alternatively, IUI with donor sperm is a simple and inexpensive option for couples who do not wish to pursue IVF. Despite the advent of assisted reproductive techniques such as IVF and ICSI, superovulation with IUI remains the most widely used and accessible fertility treatment worldwide.

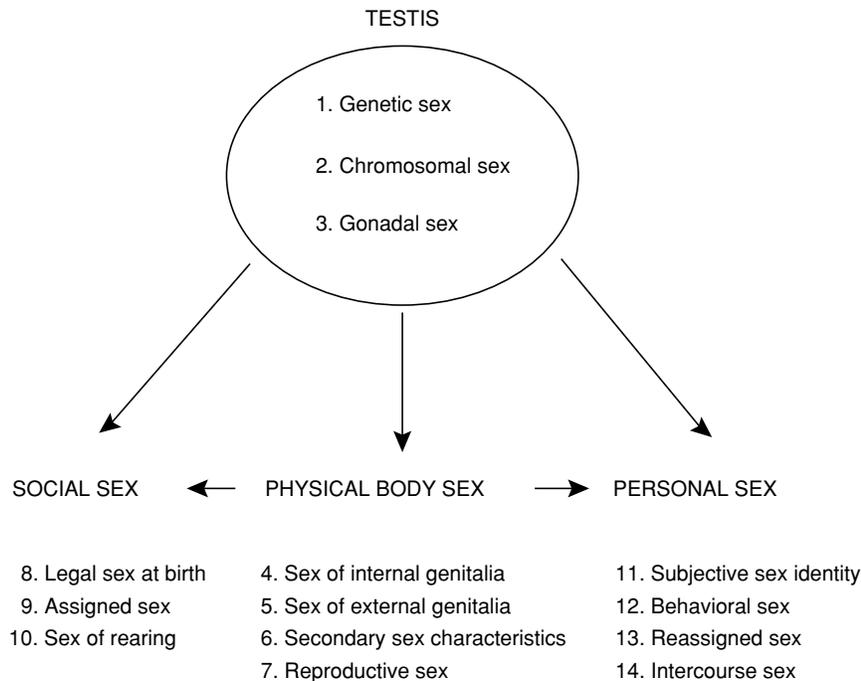
### See Also the Following Articles

Assisted Reproductive Technology (ART) • Fertilization • Gonadotropin-Induced Ovulation • Hypothalamic Anovulation, Functional • Infertility, Overview • In Vitro Fertilization (IVF) • Ovarian Failure Treatment Strategies: Egg Donation • Ovarian Hyperstimulation Syndrome • Pregnancy Endocrinology • Premature Ovarian Failure

### Further Reading

- Adashi, E. Y., Rock, J. A., and Rosenwaks, Z. (eds.). (1996). "Reproductive Endocrinology, Surgery, and Technology." Lippincott-Raven, Philadelphia, PA.
- Chaffkin, L. M., Nulsen, J. C., Luciano, A. A., and Metzger, D. A. (1991). A comparative analysis of the cycle fecundity rates associated with combined human menopausal gonadotropin (hMG) and intrauterine insemination (IUI) versus either hMG or IUI alone. *Fertil. Steril.* 55, 252–256.

- Cohlen, B. J., Vandekerckhove, P., te Velde, E. R., and Habbema, J. D. (2000). Timed intercourse versus intra-uterine insemination with or without ovarian hyperstimulation for subfertility in men. *Cochrane Database Syst. Rev.* **2**, CD000360.
- Dodson, W. C., and Haney, A. F. (1991). Controlled ovarian hyperstimulation and intrauterine insemination for treatment of infertility. *Fertil. Steril.* **55**, 457-465.
- Erickson, G. F. (1996). Physiologic basis of ovulation induction. *Semin. Reprod. Endocrinol.* **14**, 287-297.
- Guzick, D. S., Sullivan, M. W., Adamson, G. D., and Cedars, M. I. (1998). Efficacy of treatment for unexplained infertility. *Fertil. Steril.* **70**, 207-213.
- Ombelet, W., Puttemans, P., and Bosmans, E. (1995). Intrauterine insemination: A first-step procedure in the algorithm of male subfertility treatment. *Hum. Reprod.* **10**, 90-102.
- Shushan, A., and Laufer, N. (2000). Fertility drugs and ovarian cancer: What are the practical implications of the ongoing debate? *Fertil. Steril.* **74**, 8-9.
- Silverberg, K. M. (1996). Ovulation induction in the ovulatory woman. *Semin. Reprod. Endocrinol.* **14**, 339-344.
- Speroff, L., Glass, R. H., and Kase, N. G. (1999). Induction of ovulation. In "Clinical Gynecologic Endocrinology and Infertility" 6th ed., pp. 1097-1132. Lippincott Williams & Wilkins, Baltimore, MD.

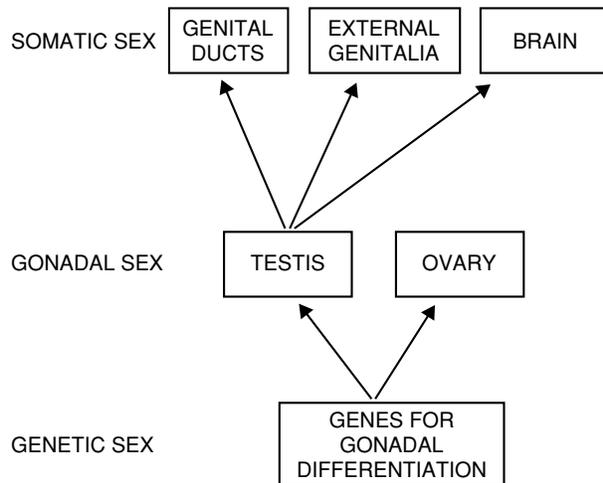


**Figure 1** Central role of the testis in the regulation of sex differentiation of organs and functions in humans and other mammals. The modalities of sex at various developmental and organizational levels are numbered from 1 to 14 roughly according to the developmental sequences. The genetic sex (1) refers to the genes that regulate the cell differentiation and development of the gonads into the female or male direction. The measurement of their activity requires a small sample of the developing tissue. The chromosomal sex (2) refers to the presence of sex chromosomes that can be studied from a blood sample. Usually, at least one Y chromosome is needed for male development. The gonadal sex (3) refers to the presence of the testis or ovary. The testes can usually be identified in a newborn baby if they have normally descended into the scrotal sacs. The sex of the internal genitalia (4) refers to the presence of epididymis and ductus deferens, whereas the sex of external genitalia (5) refers to the presence of scrotal sacs with testes. The internal and external genitalia normally develop according to regulatory factors secreted by the testis. The secondary sex characteristics (6) include shaping of the body and breasts, as well as growth pattern of the body hair, that takes place at puberty under the control of the testis and ovary together with the pituitary hormones. The reproductive sex (7) in this context refers to the ability to produce spermatozoa or egg cells for reproduction. Partially parallel with the development of body sex, the social and personal aspects of sex develop under the direct or indirect control of the testis and its hormonal products. The legal sex (8) is determined at birth in normal cases according to the anatomy of the external genitalia. In the case of ambiguity, more detailed analysis of the chromosomal and genetic status of the individual is required before the sex can be decided legally (9) in the birth certificate. In normal cases, all of the various modalities of sex develop and can be easily identified without special examinations. The sex of rearing (10) includes the behavior of parents and other related adults in relation with the growing child. The person himself or herself develops a subjective sex identity (11) according to internal physical and functional development as well as the behavior of relatives. This is soon reflected in manifestation of the behavior (12) of the child according to the established role in each society. In some cases, the sex is reassigned into the opposite sex (13) due to a misinterpretation of the sex at birth or, for example, due to a change of subjective sex identity later in life. After maturation, the sex can also be identified according to anatomical and functional roles in sexual intercourse (14). From biological, medical, and practical experience and knowledge, it is evident that any of the modalities constitutes a spectrum from normal female through different intersect states into normal male and that successive developmental phases can be independent of, and different from, the previous phases.

on the surface of the mesonephros. In terms of cellular and molecular composition, it is similar in both sexes until the seventh week of embryonic age (from fertilization) in humans and by 12.5 days in mice. The cellular differentiation proceeds by organization of irregular epithelial cords inside the growing ridge into cylindrical looped cords.

## THE TESTIS

The testicular cords consist of somatic Sertoli cells and the germ cells, which develop into spermatogonia (Fig. 4). The interstitium contains the testosterone-producing Leydig cells, undifferentiated mesenchymal cells, macrophages, developing peritubular myoid



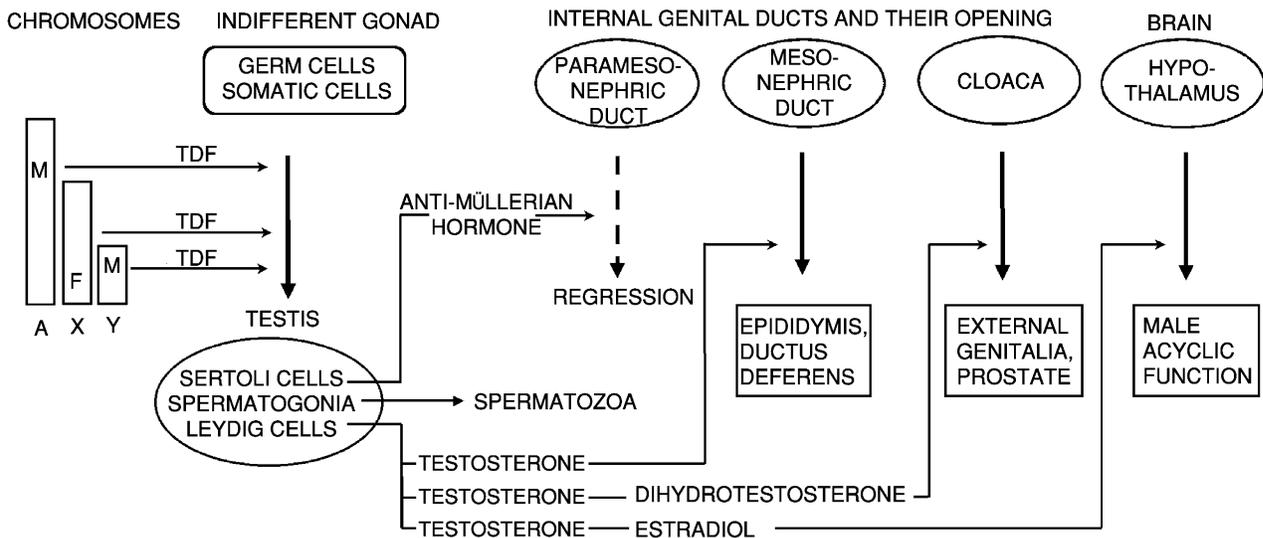
**Figure 2** Regulatory pathways of testicular differentiation in comparison with the ovary. Specific genes direct the differentiation of the gonads, but only the testis regulates the next wave of development of the internal genital ducts, external genitalia, and brain into male structures and functions. In terms of sexual differentiation, the genetic, gonadal, and somatic phases are completed by or immediately after birth.

cells, and blood vessels. The most important regulatory hormones and their functions are illustrated in Fig. 4.

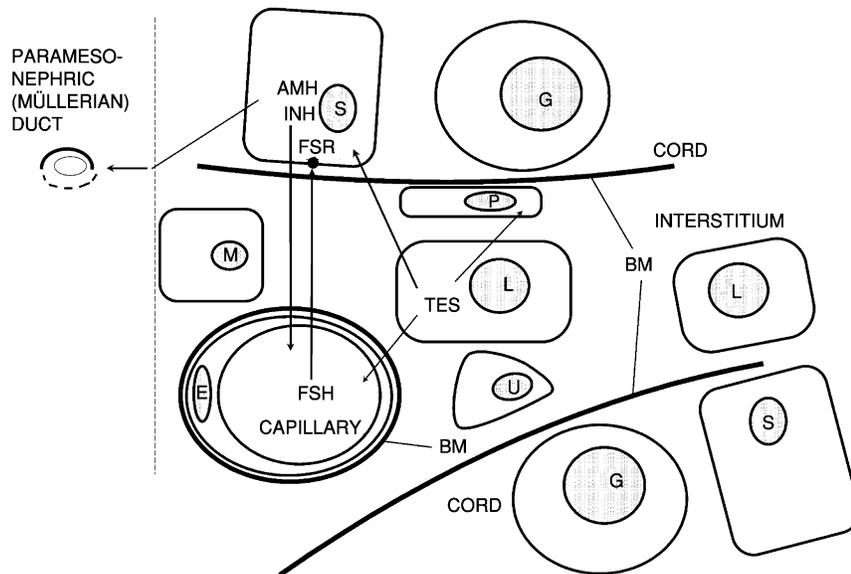
**REGULATORY MECHANISMS**

Several gene products are required for normal testicular differentiation. They are collectively referred to as testis differentiation factors (Figs. 3 and 5). Best known are Sry and Sox9, which regulate the early phase of testicular differentiation of cords, and Wt1, Gata4, and Sf1, which guide the functional differentiation of Sertoli and Leydig cells for hormone production.

After cellular differentiation, the Sertoli cells produce anti-Müllerian hormone, which causes regression of the paramesonephric (Müllerian) ducts. Otherwise, they would develop into female oviduct, uterus, and vagina (Fig. 3). Testosterone from the Leydig cells masculinizes the primordia of the male genitals and brain (Fig. 3).



**Figure 3** Schematic drawing of the differentiation of the testis and its subsequent activities in the regulation of the development of genital organs and brain into the male direction. The regulatory genes of the testis differentiation factors (TDF) are located in autosomes (A), X chromosomes (X), and Y chromosome (Y). The genes have male (M) or female (F) directing functions. Under the influence of these gene products, the indifferent gonads, consisting of germ cells and undifferentiated somatic cells, develop into a microscopically identifiable testis with cylindrical cords, consisting of Sertoli cells and spermatogonia and the interstitial tissue containing the Leydig cells. The Sertoli cells secrete anti-Müllerian hormone, which causes the female paramesonephric duct to regress and in this way prevents the development of the oviduct, uterus, and vagina. The Leydig cells secrete testosterone, which via the circulation reaches the embryonic mesonephric duct and stimulates its differentiation into epididymis and ductus deferens. In a similar fashion after conversion into dihydrotestosterone, the prostate and external genitalia develop from the cloacal region of the embryo. Another conversion of testosterone into estradiol takes place in the brain, where the regulatory neurons in the hypothalamus will develop male acyclic function and in this way prevent the later start of a menstrual cycle, which develops in the female brain without perinatal exposure to testosterone action. As indicated in Fig. 2, the internal and external genitalia, together with the brain, have an autonomous development program in the female direction and in this way do not require any known effect from the ovary.



**Figure 4** Schematic drawing of the cells and their histological organization in an embryonic testis. The testicular cords consist of Sertoli cells (S) and spermatogonia (G). The identification letter for each cell type is located in the shaded nucleus of the cell. The nucleus is surrounded by white cytoplasm of the cell outlined with a solid line drawn roughly to indicate the shape of the cell. A short arch of the circular outline of the cord is shown with a thick line labeled as basement membrane (BM). The cord itself is an elongated and curved cylinder that can be outlined by completing the arches of basement membrane into full circles with imagination. Between the cylindrical cords, the interstitial tissue consists of Leydig cells (L), peritubular developing myoid cells (P), undifferentiated mesenchymal cells (U), blood capillaries formed by endothelial cells (E) surrounded with their basement membrane (BM), and (toward the end of the embryonic period) phagocytosing macrophages (M). The embryonic Sertoli cells produce anti-Müllerian hormone (AMH), which diffuses into the adjacent paramesonephric duct and causes its regression (see Fig. 3). During the late embryonic period, the Sertoli cells have a receptor (FSR) for the follicle-stimulating hormone (FSH) produced by the pituitary gland. The Sertoli cells also secrete inhibin (INH), which via the circulation is carried into the pituitary gland, where it inhibits the secretion of FSH. The interstitial Leydig cells secrete testosterone, which is locally transported into peritubular cells and Sertoli cells and which via the blood circulation is transported through the capillaries and reaches the other more distant target organs. The mature fetal Leydig cells are regulated by luteinizing hormone (LH) via the LH receptor (LHR) in their plasma membrane.

## DISEASES

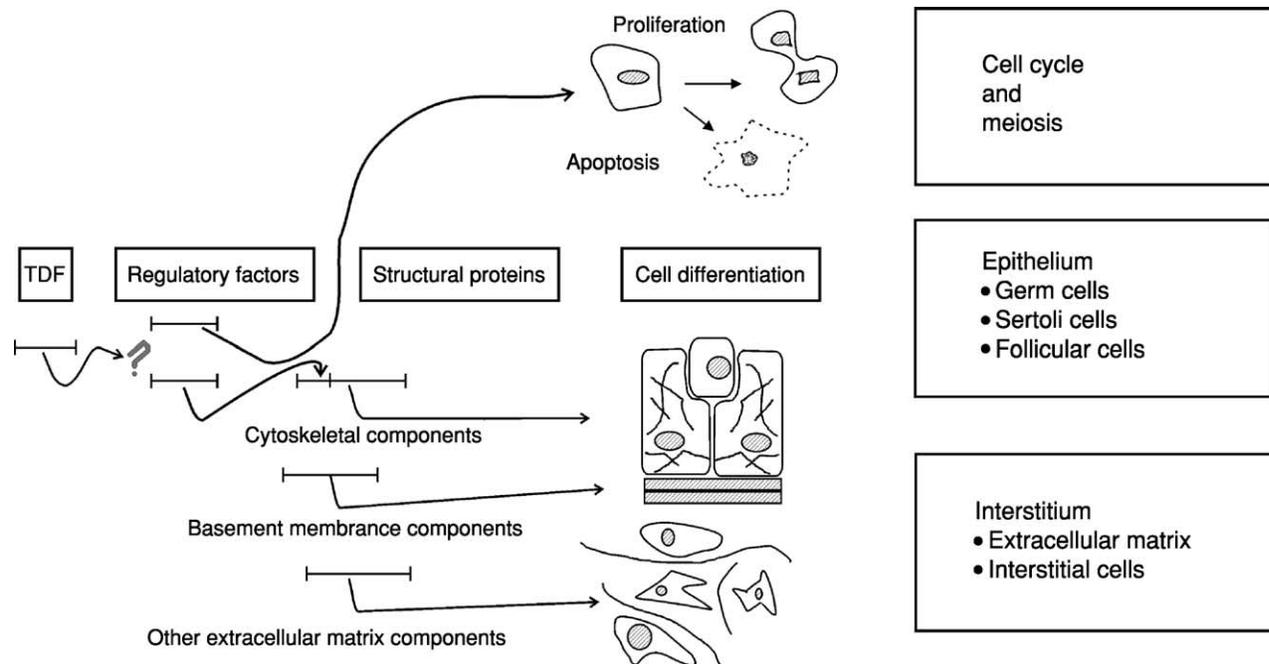
The defects in regulatory genes cause a developmental disorder that reflects the function of the gene. In fact, many genes have been identified through analysis of these defects found in human patients. The ultimate disorder, a total lack of the testis, is the most severe defect and leads to subsequent structural and functional anomalies that can be deduced from the regulatory relationships illustrated in Figs. 1, 2, and 3.

The most common (and usually mild) disorder is cryptorchidism, where the testes of the newborn are still undescended in the abdominal cavity or in the inguinal canal. For subsequent personal health and fertility, it is important to bring the testes down to the scrotal sacs because the undescended testes are prone to cancers and remain infertile at adult age.

Testicular neoplasms are rare. The most common are germ cell tumors (95%), of which the seminomas

apparently originate from the early embryonic primordial germ cells. The Leydig cell tumors can cause strange symptoms that result from the effects of their abnormal hormones on target organs such as the genital organs, hair, and pituitary gland.

In comparison with the embryonic ovary (Fig. 2), the testis has a cardinal role in the regulation of sexual differentiation of the individual both before and after birth (Fig. 1). The endocrine activity enters a quiescent phase after birth and recovers at puberty together with the other major role of the testis, that is, the production of spermatozoa by spermatogenesis. This includes removal of active inhibition of meiotic divisions during the embryonic and prepubertal phases. Meiosis is a reductive type of cell division that is required for haploid germ cell production in both sexes; thus, it allows reorganization of genes and maintenance of constant chromosome number in the offspring.



**Figure 5** Regulatory pathways through which the testis differentiation factors (TDF) produced by the testis differentiation genes guide and regulate, via a cascade of successive regulatory steps, the numerical growth of the cells by proliferation through mitosis and after postnatal maturation by meiotic divisions in the spermatogenesis. In contrast to further cell divisions, some cells are directed to death by apoptosis. The primary targets of the testis differentiation factors are not known (?). The segment lines depict genes, and the arrows show the route and fate of the gene products. Some of the regulatory factors are apparently transcription factors of the next gene in the regulatory chain, and some are signal molecules that activate regulatory pathways in their target cells. At the end of the pathway, the final proteins are structural, secretory, and other proteins for the cells and the extracellular matrix, for example, cytoskeletal components for the regulation of the cell shape and motility, basement membrane components to separate and support the epithelial tissues (e.g., testicular cords in the testis), and extracellular matrix components (e.g., collagens and fibronectin in the interstitium of the testis). The interstitial cells have protein components different from those in epithelial cells. The rate of cell cycle, apoptosis, and later meiotic divisions determine the rate of growth of the testis. The internal organization of the originally undifferentiated epithelial tissue into the testicular cords, which comprise the germ cells and Sertoli cells, is a diagnostic feature of testicular differentiation at the light microscopic level. The follicular cells in the ovary correspond to the Sertoli cells in the testis and so constitute the epithelial component of the ovarian follicles around the oocytes. The line of the cytoskeletal component gene is divided into the segments where the left represents the promoter region and the right represents the actual coding sequence of the gene. These details are present in all genes, even though they are not shown in the figure.

## See Also the Following Articles

Agonadism, Male and Female • Endocrine Disrupters and Male Sexual Differentiation • Gonadotropins and Testicular Function in Aging • Müllerian Inhibiting Substance: New Insights • Sexual Maturation, Male • Spermatogenesis, Endocrine Control of • Testicular Tumors • Undescended Testes

## Further Reading

MacLean, H. E., Warne, G. L., and Zajac, J. D. (1997). Intersex disorders: Shedding light on male sexual differentiation beyond SRY. *Clin. Endocrinol.* **46**, 101–108.

Payne, A. H., Hardy, M. P., and Russell, L. D. (eds.) (1996). “The Leydig Cell.” Cache River Press, Vienna, IL.

Pelliniemi, L. J., and Fröjdman, K. (2001). Structural and regulatory macromolecules in sex differentiation of gonads. *J. Exp. Zool.* **290**, 523–528.

Pelliniemi, L. J., Fröjdman, K., Sundström, J., Pöllänen, P., and Kuopio, T. (1998). Cellular and molecular changes during sex differentiation of embryonic mammalian gonads. *J. Exp. Zool.* **281**, 482–493.

Russell, L. D., and Griswold, M. D. (eds.) (1993). “The Sertoli Cell.” Cache River Press, Clearwater, FL.

Sharpe, R. M. (2001). Hormones and testis development and the possible adverse effects of environmental chemicals. *Toxicol. Lett.* **120**, 221–232.

Söder, O. (ed.) (2003) “The Developing Testis.” Karger, Basel, Switzerland.

Toppari, J., and Kaleva, M. (1999). Maldescendus testis. *Horm. Res.* **51**, 261–269.

Vilain, E. (2002). Anomalies of human sexual development: Clinical aspects and genetic analysis. *Novartis Found. Symp.* **244**, 43–53.

**Table 1 Histologic Classification of Testicular Germ Cell Tumors**

Seminoma
Nonseminoma
Embryonal carcinoma
Teratoma
Mature
Immature
Mature or immature with malignant transformation
Choriocarcinoma
Yolk sac tumor (endodermal sinus tumor)
Mixed germ cell tumor (may include seminomatous component)

carcinoma, teratoma, choriocarcinoma, or yolk sac tumor (endodermal sinus tumor). Seminoma may also be a component of a mixed GCT. If a tumor has both seminomatous and nonseminomatous components, it is classified as a nonseminomatous GCT and managed accordingly.

Nonseminomatous GCTs containing embryonal carcinoma with or without elements of choriocarcinoma may produce human chorionic gonadotropin (hCG), an important serum tumor marker. Elevations in alpha-fetoprotein (AFP), another important serum marker, are seen only in nonseminomatous GCTs and may be observed in tumors with embryonal and/or yolk sac components. Thus, a tumor is diagnosed as a seminoma only if there is no nonseminomatous component present (i.e., only pure seminoma) and if the AFP level is normal.

Pure seminomatous GCTs occur most frequently during the fourth decade of life. They are composed of uniform, large polyhedral cells surrounded by connective tissue stroma. Seminomas may contain syncytiotrophoblasts that are the source of increased serum concentrations of hCG; however, seminomas with syncytiotrophoblasts are not classified as nonseminomatous.

## GENETICS

Cytogenetic and molecular genetic analyses have shown that an isochromosome of the short arm of chromosome 12, *i*(12p), is a specific marker of GCTs. It is detectable in approximately 80% of cases. *i*(12p)-negative tumors show excess 12p genetic material by fluorescent *in situ* hybridization (FISH) or other appropriate genetic techniques. Thus, excess 12p genetic material is present in all GCTs, including

ITGCN, and represents one of the earliest genetic changes associated with malignant transformation of germ cells. In some circumstances where tumor histogenesis is uncertain, identification of excess 12p copy number is an important diagnostic marker indicating germ cell origin.

## DIAGNOSIS

The majority of patients with testicular GCTs present with scrotal discomfort, painless swelling, or testicular hardness. The pathognomonic finding of a painless testicular mass is present in only a minority of patients. Less commonly, patients present with back pain and/or nipple tenderness. In addition to a thorough bilateral testicular examination, the physical exam should focus on palpation for abdominal or cervical lymphadenopathy and gynecomastia (due to production of hCG).

Scrotal ultrasonography is necessary in the evaluation of a testicular mass with a high sensitivity for detection of a testicular tumor. Once a testicular tumor is suspected, all patients must undergo a radical inguinal orchiectomy with ligation of the spermatic cord at the internal ring. Embryologically, the testes originate in the genital ridge and descend during fetal life through the abdomen and inguinal canal into the scrotum; thus, the vascular and lymphatic drainage of the testes is to the renal or great vessels and retroperitoneal lymph nodes, respectively.

Because regional metastases usually first appear in the retroperitoneal lymph nodes below the renal vessels, computed tomographic (CT) imaging of the abdomen and pelvis is required. Right testicular tumors metastasize initially to interaortocaval lymph nodes, whereas left testicular tumors metastasize to para-aortic nodes. CT chest radiography is also required to evaluate for pulmonary nodules or supra-diaphragmatic adenopathy that may occur in the presence or absence of retroperitoneal disease.

Serum tumor markers, including AFP, hCG, and lactate dehydrogenase (LDH), have independent prognostic value and must be determined before, during, and after treatment. AFP concentrations may be elevated in other conditions, including hepatocellular carcinoma, liver disease, and gastrointestinal malignancies. hCG consists of  $\alpha$ - and  $\beta$ -subunits, with the  $\alpha$ -subunit sharing homology with several pituitary hormones, including luteinizing hormone (LH). False-positive elevations in hCG may result from cross-reactivity of the radioimmunoassay antibody with LH. An increasing AFP and/or hCG level in

a patient with a diagnosis of a GCT nearly always indicates active disease.

## STAGING

Because of the prognostic significance of AFP, hCG, and LDH, the American Joint Committee on Cancer incorporated serum tumor markers into the standard "TNM" staging system (now TNMS). Stage I includes patients with tumors with or without vascular or lymphatic invasion as well as patients with persistently elevated tumor markers in the absence of clinical or radiographic evidence of metastases. Stage II includes patients with retroperitoneal nodal involvement with or without increased tumor marker levels. Stage III includes patients with distant metastases or high serum tumor marker levels.

## MANAGEMENT

After radical inguinal orchiectomy, patients with early-stage seminoma (retroperitoneal disease less than 5 cm in diameter) are treated with radiation therapy to the retroperitoneal and ipsilateral pelvic lymph nodes. Approximately 4 to 10% of patients with stage I and II seminoma will relapse after radiation therapy. Chemotherapy will cure more than 90% of patients who relapse. Overall, approximately 99% of patients with early-stage seminomas are cured. Patients with bulky retroperitoneal lymph node involvement (>5 cm in diameter) initially receive chemotherapy.

Clinical stage I nonseminomatous GCTs can be cured in more than 95% of cases. Standard treatment options for clinical stage I disease without lymphatic or vascular invasion (stage 1A) include surveillance or retroperitoneal lymph node dissection (RPLND). The rationale for a rigorous surveillance program is based on a high cure rate with chemotherapy in the event of relapse. The likelihood of relapse is significantly higher ( $\geq 50\%$ ) in patients with lymphatic, vascular, scrotal, or spermatic cord invasion (stage IB), and RPLND is preferable to surveillance. The presence of pure embryonal carcinoma may also be an adverse prognostic factor. Depending on the extent of disease, tumor marker levels, and tumor-related symptoms, patients with stage II nonseminomatous GCTs are treated initially with RPLND, chemotherapy, or both.

Patients with advanced nonseminomatous GCTs are initially treated with chemotherapy. Patients are staged according to the likelihood of cure with chemotherapy. Good, intermediate, and poor risk groups have been identified based on primary site, presence or absence of nonpulmonary visceral metastases, histology, and serum tumor marker concentration (Table II). Prior to the mid-1970s, chemotherapy for GCTs was only minimally effective. In 1977, Einhorn and Donohue reported impressive results with a chemotherapy regimen composed of cisplatin, vinblastine, and bleomycin (PVB) in patients with disseminated disease. In a series of subsequent clinical trials that investigated cisplatin-containing chemotherapy regimens over the ensuing 20 years, etoposide and cisplatin (EP) and EP with bleomycin (BEP) have been established as the standard treatment regimens

Table II International Germ Cell Cancer Collaborative Group Risk Classification

	Nonseminoma	Seminoma
<b>Good risk</b>	Testicular or retroperitoneal primary AFP < 1000 ng/ml HCG < 5000 U/L LDH < 1.5 × upper limit of normal No nonpulmonary visceral metastases	Any primary site Nonpulmonary visceral metastases absent Normal AFP Any hCG Any LDH
<b>Intermediate risk</b>	Testicular or retroperitoneal primary AFP 1000–10,000 ng/ml HCG 5000–50,000 U/L LDH 1.5–10 × upper limit of normal No nonpulmonary visceral metastases	Any primary site Nonpulmonary visceral metastases present Normal AFP Any hCG Any LDH
<b>Poor risk</b>	Mediastinal primary tumor Any nonpulmonary visceral metastases AFP > 10,000 ng/ml HCG > 50,000 U/L LDH > 10 × upper limit of normal	None

for previously untreated patients with advanced disease. Good risk patients may be treated with either four cycles of EP or three cycles of BEP. This strategy avoids the potential pulmonary, neurological, vascular, and myelosuppressive toxicity associated with bleomycin in patients who are most likely to be cured. Intermediate and poor risk patients should receive four cycles of BEP.

In advanced nonseminomatous GCTs treated with cisplatin-containing chemotherapy, surgical resection of all sites of residual disease, including RPLND and a resection of any additional disease sites, is necessary. If viable residual GCT is present in any resected specimen, two additional cycles of chemotherapy are administered to maximize the cure rate. A persistently elevated tumor marker (AFP or hCG) after chemotherapy is associated with viable GCT. These patients are usually treated with additional chemotherapy rather than immediate surgical resection.

## FUTURE DIRECTIONS

In spite of the remarkable improvement in disease outcome over the past 25 years or so, 20 to 30% of patients with advanced GCTs do not achieve a durable remission with firstline therapy. Patients who do not have a complete remission (CR) or who relapse after a CR are treated with salvage chemotherapy or high-dose chemotherapy with stem cell rescue. In the United States, a prospective, randomized, cooperative group clinical trial is evaluating the role of high-dose chemotherapy with peripheral stem cell rescue as firstline therapy in patients with poor or intermediate risk disease. In addition, ongoing clinical trials of chemotherapy combinations, such as paclitaxel, ifosfamide, and cisplatin (TIP) and paclitaxel plus gemcitabine, have shown promise in the treatment of patients with refractory GCT.

## See Also the Following Articles

Gonadotropins and Testicular Function in Aging • Gonadotropin-Secreting Tumors • Prostate Cancer • Pseudoprecocious Puberty, Male • Testes, Embryology of

## Further Reading

- Bhatia, S., Abonour, R., Porcu, P., Seshadri, R., Nichols, C. R., Cornetta, K., and Einhorn, L. H. (2000). High-dose chemotherapy as initial salvage chemotherapy in patients with relapsed testicular cancer. *J. Clin. Oncol.* **18**, 3346–3351.
- Bosl, G. J. (1999). Germ cell tumor clinical trials in North America. *Semin. Surg. Oncol.* **17**, 257–262.
- Bosl, G. J., and Motzer, R. J. (1997). Testicular germ-cell cancer. *N. Engl. J. Med.* **337**, 242–253.
- Chaganti, R. S. K., Houldsworth, J., and Bosl, G. J. (2000). Molecular biology of adult male germ cell tumors. In "Comprehensive Textbook of Genitourinary Oncology" (N. J. Vogelzang, W. U. Shipley, P. T. Scardino, and D. S. Coffey, eds.), 2nd ed., pp. 891–896. J. B. Lippincott, Philadelphia.
- Hinton, S., Catalano, P., Einhorn, L. H., Loehrer, P. J., Sr., Kuzel, T., Vaughn, D., and Wilding, G. (2002). Phase II study of paclitaxel plus gemcitabine in refractory germ cell tumors (E9897): A trial of the Eastern Cooperative Oncology Group. *J. Clin. Oncol.* **20**, 1859–1863.
- International Germ Cell Cancer Collaborative Group (1997). International Germ Cell Consensus Classification: A prognostic factor-based staging system for metastatic germ cell cancers. *J. Clin. Oncol.* **15**, 594–603.
- Mazumdar, M., Bajorin, D. F., Bacik, J., Higgins, G., Motzer, R. J., and Bosl, G. J. (2001). Predicting outcome to chemotherapy in patients with germ cell tumors: The value of the rate of decline of human chorionic gonadotropin and alpha-fetoprotein during therapy. *J. Clin. Oncol.* **19**, 2534–2541.
- Motzer, R. J., Mazumdar, M., Sheinfeld, J., Bajorin, D. F., Macapinlac, H. A., Bains, M., Reich, L., Flombaum, C., Mariani, T., Tong, W. P., and Bosl, G. J. (2000). Sequential-dose intensive paclitaxel, ifosfamide, carboplatin, and etoposide salvage therapy for germ cell tumor patients. *J. Clin. Oncol.* **18**, 1173–1180.
- Motzer, R. J., Sheinfeld, J., Mazumdar, M., Bains, M., Mariani, T., Bacik, J., Bajorin, D., and Bosl, G. J. (2000). Paclitaxel, ifosfamide, and cisplatin second-line therapy for patients with relapsed testicular germ cell cancer. *J. Clin. Oncol.* **18**, 2413–2418.

60–80 mg/kg/day) of deferoxamine are used. These bone lesions include metaphyseal lesions, usually in younger patients, and spinal lesions (lordosis or kyphosis), generally in adolescents. However, after 10 years of age, despite the fact that adequate levels of hemoglobin are maintained, many of the  $\beta$ -thalassemia major children start having decelerated growth. In the pubertal children, there may be a reduced growth spurt with marked deceleration for which iron overload may be responsible. In this age group, truncal shortening, most likely due to hypogonadism secondary to iron deposition, may also be found.

The hormonal causes of growth retardation in  $\beta$ -thalassemia major children is complex, with the main causes being hypogonadism and hypothyroidism. However, it has become apparent that growth hormone (GH) also plays a role in their abnormal growth. It still remains unclear how the GH and insulin-like growth factor-1 (IGF-1) axis may play a role. Several authors reported reduced serum concentrations of IGF-1 in the presence of a normal GH response to provocative studies, and this led to the speculation that GH insensitivity is most likely the cause of the abnormal growth.

Our group reported that the IGF-1 and IGF-binding protein-3 (IGFBP-3) generation tests in children with  $\beta$ -thalassemia major help to distinguish the causes of their growth retardation. By using these generation tests, we noted that most of the  $\beta$ -thalassemia major patients had growth retardation secondary to disturbances in GH secretion rather than GH insensitivity. Classic GH deficiency and GH neurosecretory dysfunction are the causes of low IGF-1 and short stature in many  $\beta$ -thalassemia major children. Growth hormone treatment (hGH) in short  $\beta$ -thalassemia major has shown that these children can benefit from this therapy. This seems to be true in the  $\beta$ -thalassemia major children who have GH deficiency, both classic GH deficiency (Fig. 1) and GH neurosecretory dysfunction (Fig. 2). However, the growth of these children after hGH therapy is not as good as that of their nonthalassemic peers, possibly because a higher dose of hGH is required to achieve an optimal growth result. But because growth hormone is a natural antagonist of insulin, glucose metabolism parameters must be monitored closely in all  $\beta$ -thalassemia major children receiving hGH therapy.

Histological examination at autopsy of the endocrine glands of  $\beta$ -thalassemia patients has shown mild to moderate siderosis and a reduced number of cells in the pituitary gland as well as fibrosis and siderosis of the thyroid gland and gonads. This is consistent with the view that many  $\beta$ -thalassemia major patients

may have primary or secondary endocrine gland dysfunction secondary to iron overload. Undernutrition is also an important cause of low IGF-1 and associated growth disturbances in this group. This may be compounded by zinc deficiency in certain cases of undernutrition. Good nutrition with adequate vitamins and trace elements, along with calcium and vitamin D supplementation, can increase bone density and also enhance the growth of  $\beta$ -thalassemia major patients.

## HYPOTHYROIDISM

The reported incidence of thyroid disorders in  $\beta$ -thalassemia major varies greatly in different countries, probably due mostly to the difference in iron chelation therapy and to the different genetic varieties of the defect of the  $\beta$ -chain. The frequency of primary hypothyroidism, which is usually compensated, can vary from 10 to 50%, also depending on the age of the patients selected, the criteria used to diagnose hypothyroidism, and the availability of deferoxamine (which is used as an iron chelating agent). The incidence of incipient hypothyroidism, as judged by an exaggerated thyrotropin (TSH) in response to thyrotropin-releasing hormone (TRH) in  $\beta$ -thalassemia major patients of all ages in Greece, was found to be 35%. These patients can also have an underlying defect present in the thyroid without an apparent thyroid abnormality. This can be detected by a substantial increase in serum TSH when these patients are exposed to pharmacological doses of iodide. The vast majority of  $\beta$ -thalassemia patients (64%) who developed subclinical hypothyroidism during iodide administration developed permanent hypothyroidism during a 5-year follow-up. This can be of major importance to the thalassemic patients who commonly present with heart failure and cardiac arrhythmias. The antiarrhythmic drug amiodarone, which contains iron, can cause subclinical hypothyroidism. This can be extremely detrimental to  $\beta$ -thalassemia major patients with cardiac problems, and these patients need to be treated with thyroxine while taking amiodarone.

The cause of thyroid dysfunction in  $\beta$ -thalassemia major seems to be iron overload, chronic anemia, or tissue damage by hypoxia that is detrimental to the thyroid tissue. Therefore, frequent screening for thyroid dysfunction in  $\beta$ -thalassemic major patients is mandatory to help these individuals maintain optimal health.

## HYPOGONADISM

Hypogonadism is common in  $\beta$ -thalassemia major and seems to be the most frequent endocrine

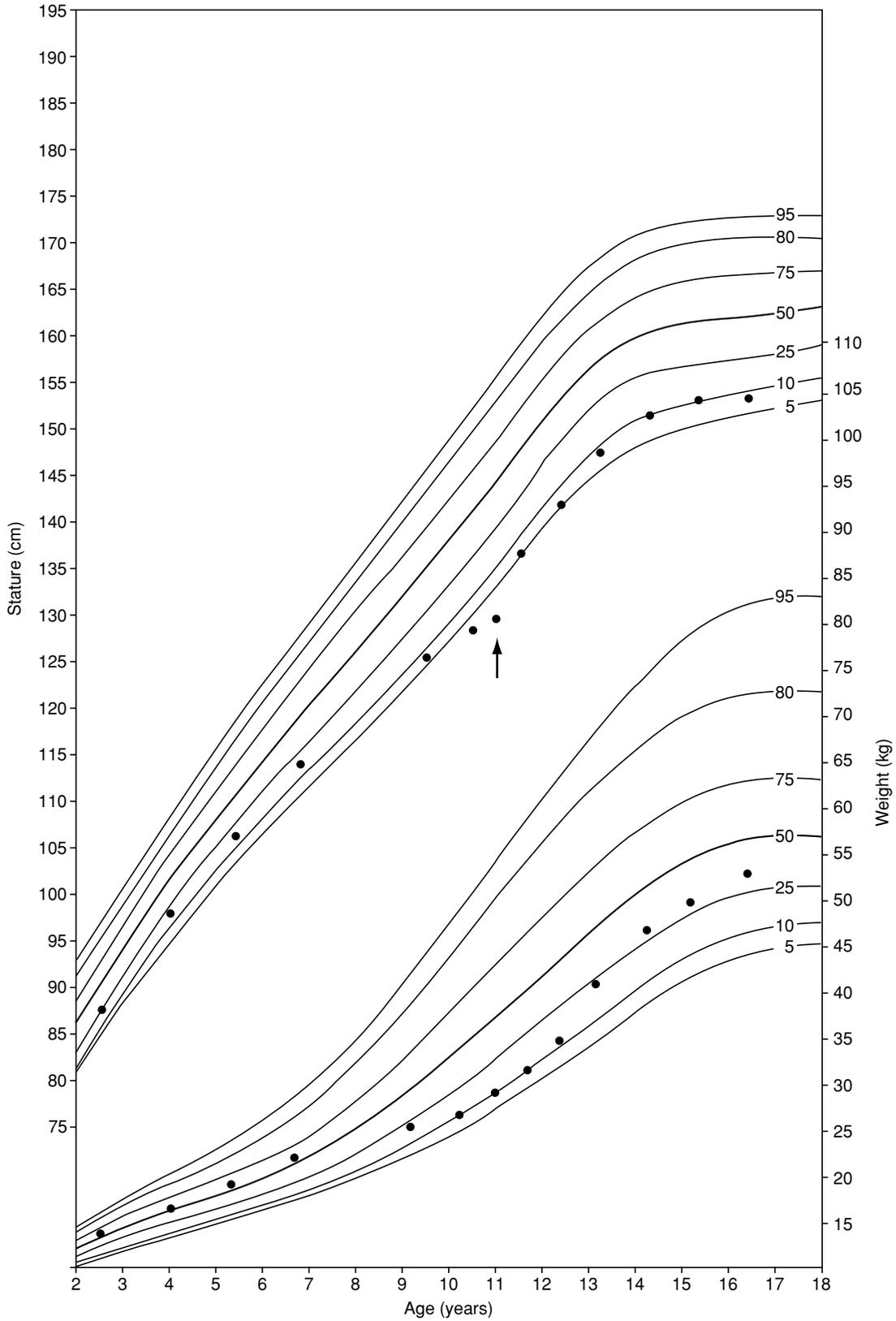


Figure 1 Growth curve of a  $\beta$ -thalassemia major female with classic growth hormone deficiency before and after therapy with hGH (arrow).

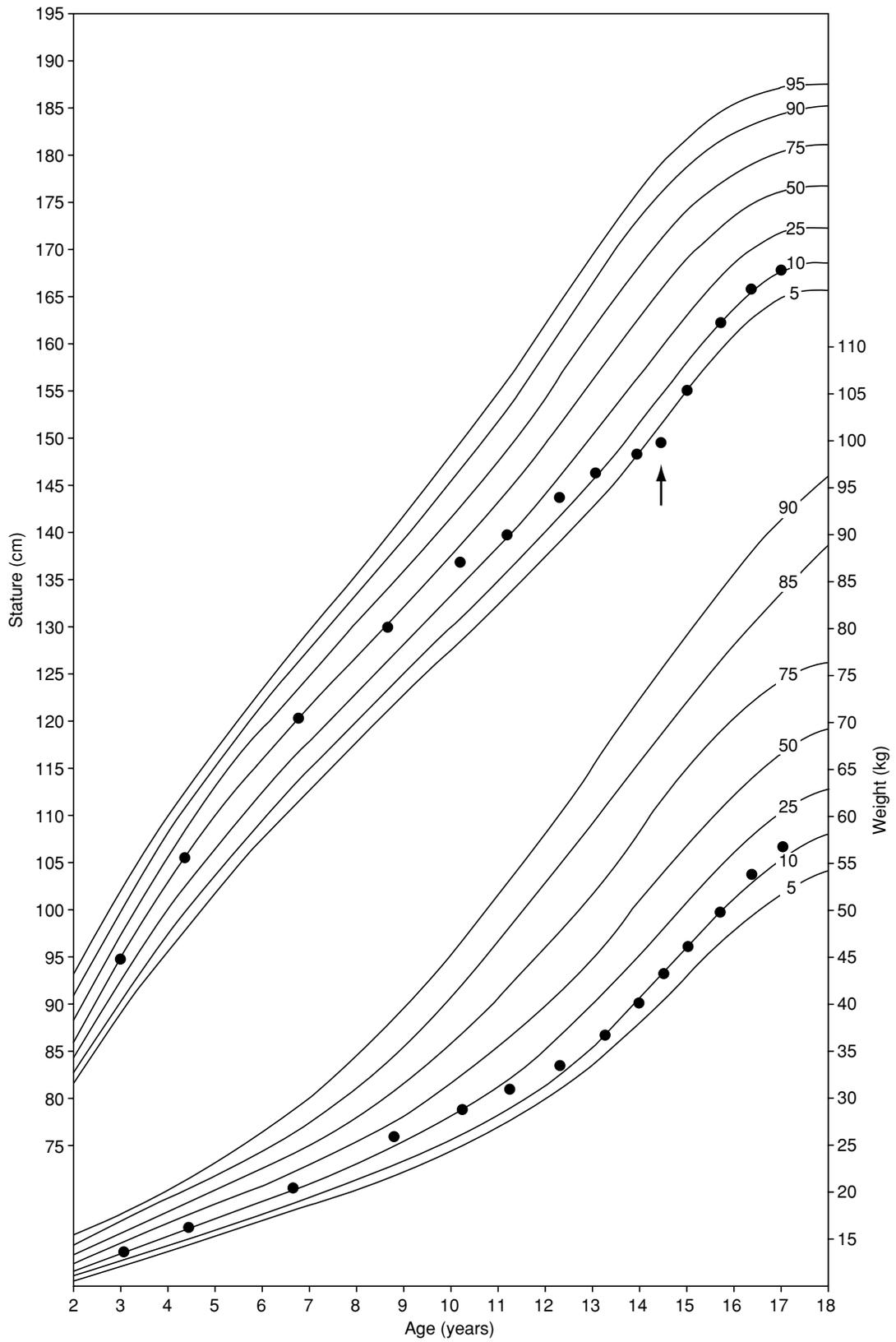


Figure 2 Growth curve of a  $\beta$ -thalassemia male with GH neurosecretory dysfunction before and after therapy with hGH (arrow).

abnormality found in this group. Most  $\beta$ -thalassemia major patients have hypogonadotropic hypogonadism due to hypothalamic and pituitary dysfunction resulting from iron deposition in these areas. This can be compounded by gonadal insufficiency due to iron deposition in the ovaries and testes. These abnormalities can cause severe psychological problems in young developing  $\beta$ -thalassemia major patients.

The pituitary–gonadal axis dysfunction can present in different ways with delayed puberty, arrested puberty at various Tanner stages, or complete lack of sexual development. The incidence of these abnormalities can be quite high. Up to 47% of girls and 51% of boys present with hypogonadism, 23% of girls present with secondary amenorrhea, and 13% of patients of both sexes present with sexual maturation arrest. Because sex steroids play a major role in truncal growth, any impairment in sexual maturation can cause severe truncal growth problems and a loss of the pubertal growth spurt. Studies have shown that up to 80% of short  $\beta$ -thalassemia major children and 40% of normal-statured  $\beta$ -thalassemia patients present with a short trunk. The etiology of the short trunk is not only hypogonadism but also iron overload and deferoxamine-induced bone dysplasia. Consensus guidelines for the treatment of hypogonadism have been compiled to promote near-normal sexual development and optimal growth. They are as follows:

A. Puberty is pharmacologically induced with ethinylestradiol (2.5–5.0  $\mu$ g orally for 4–6 months) in girls over 13 years of age who have a bone age of 10 to 11 years. If there is no evidence of further spontaneous pubertal progression, the sex steroid dose is progressively increased every 6 months, followed by hormonal replacement therapy with estrogens and progesterone.

B. Pubertal induction in boys is induced pharmacologically (25–50 mg/month of depot testosterone intramuscularly for 3–6 months) when they are over 14 years of age and have a bone age of 11 to 12 years. If puberty does not progress, the sex steroid doses are increased. In young adults, 100 mg depot testosterone twice monthly achieves the desired virilizing effect.

C. If GH insufficiency is also present, hGH treatment should be started at a dose of 0.6 to 0.8 IU/kg/week (0.20–0.26 mg/kg/week). With sex steroid treatment, sexual maturation and growth can be achieved and optimal height can be attained.

Alternative treatment to induce puberty in males has also been successful with human chorionic gonadotropin (hCG). Good results in testicular development and

fertility have been reported with doses of hCG (500–2000 IU twice weekly) plus human urinary follicle-stimulating hormone (75 IU two or three times/week).

## CALCIUM AND BONE METABOLISM ABNORMALITIES

In  $\beta$ -thalassemia, major hypoparathyroidism often can be completely asymptomatic, with hypocalcemia detected only in routine laboratory examinations. However, in certain cases, the symptoms of hypocalcemia in  $\beta$ -thalassemia major can be extremely severe, with signs of tetany, seizures, and even cardiac failure.

Magnetic resonance imaging (MRI) of the brain can sometimes show calcium deposits at the thalamus, basal ganglia, and cerebral hemispheres. The treatment of choice is oral administration of oral calcium and calcitriol.

## DIABETES MELLITUS

Diabetes mellitus (DM) is an endocrine complication in  $\beta$ -thalassemia major that is not infrequent. Its frequency progresses with age and can vary from 2 to 24% in various studies.

Both insulin-dependent diabetes mellitus (IDDM) due to insulin deficiency and non-insulin-dependent diabetes mellitus (NIDDM) due to liver function impairment or insulin resistance are found in  $\beta$ -thalassemia major patients.

Our group showed that before the onset of DM in  $\beta$ -thalassemia major, there is an impairment of glucose sensing of the islet beta cells of the pancreas, with the loss of the first-phase response of insulin to intravenous glucose similar to what occurs during the prediabetic state in IDDM and NIDDM in patients without  $\beta$ -thalassemia major. Oral glucose tolerance tests (OGTT) in  $\beta$ -thalassemia major patients also seem to exhibit significantly higher plasma glucose levels and more sustained glycemic responses. Evaluation by the homeostasis model assessment (HOMA) also showed a clear reduction in both beta cell function and insulin resistance indexes. What is of interest is that the impaired glucose tolerance (IGT) found in  $\beta$ -thalassemia major seems to be reversible since a significant decrease in IGT has been reported to follow bone marrow transplantation.

Risk factors that are associated with DM are lack of compliance to chelation therapy, iron overload, and the presence of cirrhosis and severe fibrosis. A decreased incidence of DM has been observed with ferritin levels of less than 2500 ng/ml.

Guidelines for the treatment of DM in  $\beta$ -thalassemia major are as follows. First, for patients with normal or impaired OGTT but increased or delayed insulin secretion, diet and exercise are recommended. Second, for patients with impaired or diabetic OGTT with an adequate but delayed insulin response, an oral hypoglycemic agent is indicated. Third, for patients with insulin deficiency and a diabetic OGTT, insulin is recommended.

## CONCLUSION

The improved life expectancy of  $\beta$ -thalassemia major patients has brought to light a variety of endocrine problems for which early detection and treatment, together with an aggressive iron chelation program, are imperative so that  $\beta$ -thalassemia major patients can live longer and higher quality lives.

## See Also the Following Articles

Delayed Puberty and Hypogonadism, Female • Delayed Puberty, Male • Diabetes, Type 1 • Diabetes, Type 2 • Growth Hormone (GH) • Hypocalcemia, Therapy • Hypothyroidism, Systemic Manifestations of • Insulin-like Growth Factors

## Further Reading

Alexandrides, T., Georgopoulos, N., Yarmenitis, S., and Vagenakis, A. G. (2000). Increased sensitivity to the inhibitory effect of excess iodide on thyroid function in patients with  $\beta$ -thalassemia major and iron overload and the subsequent development of hypothyroidism. *Eur. J. Endocrinol.* **143**, 319–325.

- Caruso-Nicoletti, M., De Sanctis, V., Cavallo, L., Raiola, G., Ruggiero, L., Skordis, N., and Wonke, B. (2001). Management of puberty for optimal auxological results in  $\beta$ -thalassemia major. *J. Pediatr. Endocrinol. Metab.* **14**, 939–944.
- Chrysis, D. C., Alexandrides, T. K., Koromantzou, E., Georgopoulos, N., Vassilakos, P., Kiess, E., Kratsch, I., Beratis, N. G., and Spiliotis, B. E. (2001). Novel application of IGF-I and IGFBP-3 generation tests in the diagnosis of growth hormone axis disturbances in children with beta-thalassemia. *Clin. Endocrinol. (Oxf.)* **54**(2), 253–259.
- Gamberini, M. R., Fortini, M., Gilli, G., Testa, M. R., and De Sanctis, V. (1998). Epidemiology and chelation therapy effects on glucose homeostasis in thalassaemic patients. *J. Pediatr. Endocrinol. Metab.* **11**, 867–869.
- Kattamis, C., Liakopoulou, T., and Kattamis, A. (1990). Growth and development in children with thalassemia major. *Acta Pediatr. Scand.* **366**(Suppl.), 111–117.
- Ladis, V., Theodorides, C., Palamidou, F., Frissiras, S., Berdousi, N., and Kattamis, C. (1998). Glucose disturbances and regulation with glibenclamide. *J. Pediatr. Endocrinol. Metab.* **11**, 871–878.
- Mariotti, S., Loveselli, A., Murenu, S., Sau, F., Valentino, L., Mandas, A., Vacquer, S., Martino, E., Balestrieri, A., and Lai, M. E. (1999). High prevalence of thyroid dysfunction in adult patients with beta-thalassemia major submitted to amiodarone treatment. *J. Endocrinol. Invest.* **22**, 55–63.
- Petrou, G., Makri, M., Armoni, A., Spiliotis, B. E., and Alexandrides, T. K. (1995).  $\beta$ -cell glucose non-sensing in patients with thalassemia major and glucose intolerance. In "Endocrine Disorders in Thalassemia" (S. Ando and C. Brancati, eds.), pp. 157–160. Springer-Verlag, Berlin.
- Soliman, A. T., El Banna, N., and Ansari, B. M. (1998). GH response to provocation and circulating IGF-I and IGF-binding protein-3 concentrations: The IGF-I generation test and clinical response to GH therapy in children with beta-thalassemia. *Eur. J. Endocrinol.* **138**, 394–400.
- Spiliotis, B. E., August, G. P., Hung, W., Sonis, W., Mendelson, W., and Bercu, B. B. (1984). Growth hormone neurosecretory dysfunction: A treatable cause of short stature. *J. Am. Med. Assoc.* **251**, 2223–2230.

**Table 1 Thirst Regulation in Older Adults: Homeostatic and Nonhomeostatic Factors**

<b>Homeostatic</b>
Plasma hyperosmolarity
Hypovolemia
Plasma renin activity
Peripheral and central vasopressin levels
Oropharyngeal metering
<b>Nonhomeostatic</b>
Hedonic qualities of fluid
Palatability of drink
Fluid preference
Availability
Sociocultural influences
Lifestyle habits
Social isolation
Functional status
Associated comorbidity

axis that may exert a central modulatory effect on thirst and vasopressin release.

Of interest is an additional thirst regulatory mechanism described as oropharyngeal metering. Evidence indicates that the act of swallowing may be the afferent limb of a conceptual pathway that ultimately determines the volume of fluid ingested in response to dipsogenic stimuli. Although the exact pathophysiology is unclear, these reflexes are thought to account for the rapid assuaging of thirst following drinking, even prior to the restoration of euolemia and normal tonicity. Additionally, these reflexes may also be involved in the dampening of vasopressin secretion that occurs during drinking. It is likely that the teleological role of this reflex is to bar the excessive ingestion of water.

Ultimately, drinking is a voluntary and highly coordinated function indicating involvement of multiple complex neural pathways. However, the cortical pathways responsible for the cortical control of drinking and dipsogenesis remain poorly defined. This is likely due to the confounding effect of multiple psychosocial and behavioral influences on drinking habits.

## THIRST AND AGING: HOMEOSTATIC REGULATION

Older adults are more likely to develop hyperosmolar dehydration following water deprivation. Reasons for the increased risk of dehydration in older adults include a reduction in total body water with aging and

age-related impairment in renal tubular concentrating ability. Dipsogenic influences therefore assume increased significance in fluid homeostasis with aging.

Age-related hypodipsia is well recognized. Additionally, compensatory dipsogenic fluid repletion, following a period of water deprivation, proceeds at a much slower rate in older adults. However, available evidence indicates that fluid deficits are eventually fully replaced. Experimental data obtained from human studies of hypertonic saline infusion reveal comparatively lower levels of thirst and less water intake in older adults than in younger adults. These findings support the theory of age-related blunting of osmoreceptor sensitivity. Hypertonic saline infusion in older adults results in increased vasopressin release, suggesting an age-related increase in the sensitivity of central vasopressin neuroreceptors. Thus, it is plausible that age-related attenuation of the thirst response is more likely due to altered circulating vasopressin levels. Indeed, there is a loss of circadian rhythm resulting in nocturnal peaking of vasopressin secretion in older adults. However, studies have yielded conflicting results regarding age-related changes in circulating baseline and diurnal vasopressin levels.

Several other hormones have been implicated in the etiology of age-related hypodipsia. Intravenous infusion of pharmacological levels of atrial natriuretic peptide has been shown to inhibit osmotically induced dipsogenesis and vasopressin secretion. Similarly, data relating the relatively higher levels of circulating norepinephrine (NE) levels to the known anti-dipsogenic effect of NE suggest a possible role for NE in the genesis of age-related hypodipsia.

Altered opioid sensitivity may also play a causal role in age-related physiological hypodipsia. Administration of naloxone has been shown to suppress fluid intake after overnight water deprivation in younger subjects. However, the anti-dipsogenic effect of naloxone is far less pronounced in older adults, raising the possibility of reduced opioid receptor sensitivity with aging.

The effect of aging on oropharyngeal metering is unknown. Pharyngeal dysphagia, resulting from esophageal dysmotility, is a recognized complication of presbyesophagus. However, the effect of this condition on oropharyngeal metering as a fluid homeostatic mechanism requires exploration. Similarly, other age-related oropharyngeal responses are ill-defined. Traditionally, taste and tactile sensations are considered the primary sensory modalities perceived by the oral and lingual surfaces. However, degrees of moistness are perceived within the oral cavity. It is plausible that persons with age-related dysgeusia

might also exhibit an attenuated response to dryness of the oral mucosa. Further research is needed in this area.

### THIRST AND AGING: NONHOMEOSTATIC REGULATION

Much of the research on thirst centers on homeostatic responses to fluid deprivation and altered plasma tonicity, possibly because fluid balance in the younger, healthy adult is notably threatened only under conditions that overtly disrupt homeostatic regulation. In contrast, older adults not only are subject to the effects of physiological attenuation of fluid homeostasis, but also face the threat of compromised nonhomeostatic dipsogenic influences.

The significance of nonhomeostatic factors is highlighted by studies of *ad libitum* fluid intake, indicating that physiological regulation of fluid intake rarely comes into play in free-living healthy individuals, regardless of age. Data indicate that, when fluid is available, in both young and old adults, thirst results in fluid intake well before hypovolemic or hypertonic deficits occur. Dipsogenic triggering events are undoubtedly complex and intricate. Multiple nonhomeostatic dipsogenic factors include hedonic influences such as appearance, ambience, fluid preferences, and palatability. Thus, age-related sensory impairments, such as dysgeusia, anosmia, and impaired vision, might detract notably from the hedonic qualities of fluids. Psychological influences, sociocultural expectations, health awareness, and lifestyle habits also affect the type and quantity of fluid consumed. Most importantly, data indicate a direct correlation between food intake and fluid intake, introducing age-related and pathological anorexia as additional factors in threatening fluid balance in older adults. Increased isolation and impaired functional status may further restrict access to fluid. Acute illness and chronic disease have a notable, albeit variable, effect on fluid balance in older adults. An added threatening variable to fluid balance is the effect of medication and polypharmacy on fluid balance.

### DEHYDRATION IN THE ELDERLY: CLINICAL OVERVIEW

Dehydration is highly prevalent among older adults, occurring in approximately 1 million elders each year admitted to acute care facilities in the United States. Annually, health care expenditure on dehydration among persons older than 65 years exceeds 1.2 billion

**Table II Complications of Dehydration in the Elderly**

Fatigue
Reduced exercise tolerance
Delirium
Hypotension
Constipation
Pressure ulcers
Falls
Dental caries
Hypernatremia
Nephrolithiasis
Urinary incontinence
Dysuria
Renal failure
Increased risk of uro-epithelial malignancies

dollars. Complications of dehydration are myriad, including delirium, falls, and incontinence (Table II). Associated mortality rates approach 50%. Multiple clinical factors predispose the older adult to dehydration. Although an exhaustive review of the clinical causes, diagnosis, and treatment of dehydration is beyond the scope of this article, it is pertinent to note that most of the causes of dehydration are avoidable and easily reversible. Prompt clinical detection is, however, dependent on increased awareness of the risk of dehydration in geriatric patients. Reliance on clinical diagnosis of dehydration is confounded by the poor predictive value of traditional clinical indices of dehydration. Poor skin turgor and lack of axillary sweating correlate poorly with dehydration in older persons. Similarly, although dry oral mucous membranes are the most sensitive signs of dehydration in older patients, adverse effects of medication and mouth breathing may produce identical changes. Furthermore, available data show that orthostatic hypotension in older patients does not reflect dehydration severity as medications, age-related changes, and prolonged bed rest may also result in orthostasis via autonomic dysfunction. Thus far, selected laboratory indices of dehydration, namely, serum osmolality >300, blood urea nitrogen (BUN) >20, or BUN:creatinine ratio >20, are the most reliable indices of dehydration.

Ultimately, prevention of dehydration is the most cost-effective method to thwart this potentially fatal, yet easily treatable, condition. New dietary guidelines for older Americans recommend consumption of at least 2 liters of fluid daily. However, there is limited evidence to support this arbitrary figure. Available evidence-based data support the benefits of ingestion

of at least 1.5 liters of fluid daily. It is unclear whether regular consumption of fluids in excess of this amount may result in dilutional hyponatremia and other consequences of water overload. Treatment of dehydration mandates due attention to restoration of normal plasma tonicity and correction of the fluid deficit. Older persons who are unable to ingest fluids orally may require parenteral fluid administration. Intravenous administration is the traditional method. However, health professionals should be aware of the utility of alternative routes, such as percutaneous gastrostomy, that may be preferred in older adults with coexisting anorexia or dysphagia. Subcutaneous fluid infusion (hypodermoclysis) is a convenient, simple, relatively safe, and cost-effective method that may be particularly effective in frail, institutionalized persons at increased risk of dehydration.

Effective management of dehydration in the older adult mandates meticulous attention to pathophysiological stressors in the context of preexisting hypodipsia. Ultimately, the ideal therapeutic goal is the development of an individualized fluid maintenance and hydration intervention strategy for all older adults, healthy or otherwise.

### See Also the Following Articles

Aging and Longevity of Human Populations • Atrial Natriuretic Factor and Family of Natriuretic Peptides •

Neuroendocrine System and Aging • Norepinephrine Transporter • Renin

### Further Reading

- Dasgupta, M., Binns, M., and Rochon, P. (2000). Subcutaneous fluid infusion in a long term care setting. *J. Am. Geriatr. Soc.* **48**, 795–799.
- Horani, M. H., and Morley, J. E. (1998). Pathophysiology of hypodipsia. In “Hydration and Aging” (B. Vellas, J. L. Albarade, and P. J. Garry, eds.), pp. 119–132. Springer, New York.
- Kenney, W. L., and Chiu, P. (2001). Influence of age on thirst and fluid intake. *Med. Sci. Sports Exerc.* **33**, 1524–1532.
- Lindemann, R. D., Romero, L., Liang, H. C., Baumbartner, R. N., Koehler, K. M., and Garry, P. J. (2000). Do elderly persons need to be encouraged to drink more fluids? *J. Gerontol. Med. Sci.* **55A**, M349–M352.
- Morley, J. E. (2000). Water, water everywhere and not a drop to drink. *J. Gerontol. Med. Sci.* **55A**, M359–M360.
- Ritz, P., and Investigators of the Source Study (2001). Chronic cellular dehydration in the aged patient. *J. Gerontol. Med. Sci.* **56A**, M361–M365.
- Stachenfield, N. S., DiPietro, L., Nadel, E., and Mack, G. W. (1997). Mechanism of attenuated thirst in aging: Role of central volume receptors. *Am. J. Physiol.* **272**, R148–R157.
- Wakefield, B., Mentes, J., Diggelmann, L. K., and Culp, K. (2002). Monitoring hydration status in elderly veterans. *West J. Nurs. Res.* **24**, 132–142.
- Wilson, M. G. (1998). The management of dehydration in the nursing home. In “Hydration and Aging” (B. Vellas, J. L. Albarade, and P. J. Garry, eds.), pp. 181–200. Springer, New York.

Table I Summary of Thyroglobulin Characteristic Sequences

Thyroglobulin (amino acid residue nr)	Characteristics
-19	5' amino acid
-19 ↔ -1	Signal peptide
1	5' amino acid after signal peptide cleavage
59, 93, 181, 467, 479, 512, 731, 799, 930, 1203, 1332, 1348, 1699, 1757, 1852, 1896, 2105, 2233, 2274, 2565	Glycosylation sites
5, 1291, 2554, 2747	Iodothyronine acceptor sites
130, 847, 1448	Iodothyronine donor sites
	Domains with cysteine-rich repeated sequences
12 ↔ 1191, 1492 ↔ 1546	Type 1
1437 ↔ 1484	Type 2
1584 ↔ 2168	Type 3
2192 ↔ 2716	Acetylcholine esterase homologous domain
1440 ↔ 1443, 1454 ↔ 1457, 1471 ↔ 1474	Thioredoxin boxes
789 ↔ 1173	Heparin-binding domain
2749	3' amino acid

Glycosylation is a key event in Tg maturation. Approximately three-quarters of the potential glycosylation sites in Tg are occupied. Human Tg contains different types of carbohydrate units. Common are the "polymannose" units consisting of mannose and N-acetylglucosamine and the "complex" unit that has a core of three mannose residues with several chains of N-acetylglucosamine, galactose, and fucose or sialic acid extending from it. Both types of carbohydrate complexes are linked to peptide through an asparagine-N-acetylglucosamine bond. Moreover, human thyroglobulin (hTg) contains two additional carbohydrate units linked to the peptide chain through the hydroxyl group of serine or threonine.

Tg contains several tyrosine residues that are preferentially iodinated. By coupling of two iodotyrosine residues, one functioning as a donor residue and the other as acceptor residue, thyroxine is generated at the acceptor site. The selectivity of tyrosine residues that become iodinated appears to be strongly dependent on the stereo-specific structure of the molecule. The main product is thyroxine and in much lesser quantities, 3,5,3'-triiodothyronine (I3) and 3,3',5' triiodothyronine (reverse I3) are formed.

The function of the 11 type 1 cysteine-rich repeats in the molecule, which is described in bovine and human Tg, is not clear. It has been shown that these repeats are protein modules that can act as pH-dependent binders and reversible inhibitors of the proteases implicated in Tg degradation and thyroxine release. Acetylcholine esterase is known to interact with the cell membrane in the nervous system, and a similar role for the homologous domain at the

carboxy-terminal end of the Tg molecule has been proposed for apical membrane interaction. Unfortunately, the mutation in the cog/cog mouse model, which is located in this domain, leads to ER storage of an apparently misfolded Tg molecule and, therefore, does not directly contribute to the elucidation of the function of this domain.

Recently, three conserved thioredoxin boxes have been identified in mammalian Tg, and studies on a bovine Tg fragment revealed a role for these boxes in self-assisted disulfide-bond formation leading to the intramolecular cross-linking of luminal Tg.

In rat Tg, the heparin-binding domain is proposed to bind to megalin that participates in the endocytosis of Tg-T4 complexes from the follicular lumen.

## THYROGLOBULIN SYNTHESIS DEFECTS

The Tg synthesis defect has been studied extensively in three animal species. In Afrikaner cattle, a homozygous nonsense mutation in exon 9 results in truncated Tg protein of 75 kDa due to the conversion of arginine 697 to a premature stop codon. In this case, an alternatively spliced mRNA lacking exon 9 sequence also is present, encoding a Tg protein of 250 kDa. In Dutch goats, a homozygous nonsense mutation in exon 8 changes tyrosine 296 to a stop and results in a truncated Tg protein of 40 kDa, causing hypothyroidism with goiter. Administration of extra iodine to homozygous mutant Dutch goats restores euthyroidism, but goiter persists. Furthermore,

in a mouse model (*cog/cog* mouse), congenital goiter is linked to the Tg locus and is caused by the transition of leucine 2366 in the Tg protein to proline. This defect results in an ER storage disease.

Patients who are suspected of having disorders in the synthesis of Tg are moderately to severely hypothyroid. The plasma Tg concentration is usually low, especially in relation to the thyroid-stimulating hormone (TSH) concentration, and intravenous injection of TSH does not increase the plasma Tg concentration. Patients classified in the category “Tg synthesis defects” often have abnormal circulating iodoproteins, mainly iodinated albumin, and excrete iodopeptides of low molecular weight in the urine.

Mutations in the Tg gene have been elucidated only in four human families. A homozygous mutation at the acceptor splice site of intron 3 gives an “in-frame” deletion of exon 4 from the Tg mRNA, resulting in an aberrant Tg protein lacking the hormonogenic site tyrosine 130 and a characteristic Cys–Trp–Cys repeat. Another patient with a homozygous in-frame mRNA deletion of 138 nucleotides has been described. The preferential accumulation of a Tg mRNA alternative splice product with an in-frame deletion of exon 22, linked to a homozygous nonsense mutation at position 1510 and resulting in an ER storage disease similar to the *cog/cog* mice, has also been reported. Furthermore, a homozygous nonsense mutation in the Tg mRNA of moderately hypothyroid patients with goiter has been described. A homozygous nonsense mutation in exon 7 results in the conversion of arginine 277 to a premature stop codon and subsequently results in the synthesis of a truncated Tg protein of 30 kDa. As has been seen in Dutch goats, the truncated Tg glycoprotein can still be iodinated and is able to synthesize thyroid hormone.

In a study with six patients with the clinical classification “Tg synthesis defects”, no mutations are

present in the Tg mRNA. Therefore, the clinical entity “Tg synthesis defects” cannot be restricted to the Tg molecule itself, indicating that under this clinical entity other proteins involved in Tg synthesis and processing can be affected.

### See Also the Following Articles

Iodine • Iodine Deficiency • Thyroid Autoimmunity • Thyroid Carcinoma • Thyroid Disease, Epidemiology of • Thyroid Disease, Genetic Factors in • Thyroid Hormone Action

### Further Reading

- Dunn, J. T. (1999). Thyroglobulin: Chemistry and biosynthesis. In “Werner’s and Ingbar’s The Thyroid” (L. E. Braverman and J. D. Utiger, eds.), pp. 91–104. J. B. Lippincott, Philadelphia.
- Kim, P. S., and Arvan, P. (1998). Endocrinopathies in the family of endoplasmic reticulum (ER) storage diseases: Disorders of protein trafficking and the role of ER molecular chaperones. *Endocrine Rev.* **19**, 173–202.
- Klein, M., Gestmann, I., Berndorfer, U., Schmitz, A., and Herzog, V. (2000). The thioredoxin boxes of thyroglobulin: Possible implications for intermolecular disulfide bond formation in the follicle lumen. *Biol. Chem.* **381**, 593–601.
- Mendive, F. M., Rivolta, C. M., Moya, C. M., Vassart, G., and Targovnik, H. M. (2001). Genomic organization of the human thyroglobulin gene: The complete intron–exon structure. *Eur. J. Endocrinol.* **145**, 485–496.
- Mercken, L., Simons, M. J., DeMartynoff, G., Swillens, S., and Vassart, G. (1985). Presence of hormonogenic and repetitive domains in the first 930 amino acids of bovine thyroglobulin as deduced from the cDNA sequence. *Eur. J. Biochem.* **147**, 59–64.
- Van de Graaf, S. A. R., Cammenga, M., Ponne, N. J., Veenboer, G. J. M., Gons, M. H., Orgiazzi, J., De Vijlder, J. J. M., and Ris-Stalpers, C. (1999). The screening for mutations in the thyroglobulin cDNA from six patients with congenital hypothyroidism. *Biochimie* **81**, 425–432.
- Van de Graaf, S. A. R., Ris-Stalpers, C., Pauws, E., Mendive, F. M., Targovnik, H. M., and De Vijlder, J. J. M. (2001). Up to date with human thyroglobulin. *J. Endocrinol.* **170**, 307–321.

Table 1 Types of Thyroid Autoimmunity<sup>a</sup>

Type	Thyroid antibodies	Thyroid function	Other features
Focal thyroiditis	TG or TPO Abs in most cases at low titer	Normal or subclinical hypothyroidism	May progress to overt hypothyroidism
Silent thyroiditis	TPO Ab in most cases, TGAb variable	Transient thyrotoxicosis and/or hypothyroidism	Most commonly occurs 1 year after delivery as postpartum thyroiditis; may progress to overt hypothyroidism
Hashimoto's (goitrous) thyroiditis	TG and TPO Ab, usually high titer; occasional TBAb	Euthyroid or hypothyroidism	Associated rarely with thyroid lymphoma
Atrophic thyroiditis (primary myxedema)	TG and TPO Ab at variable titer; TBAb in up to 20% of cases	Hypothyroidism	May evolve from Hashimoto's thyroiditis
Graves' disease	TSAb in all cases; TPO Ab in 80% and lower frequency of TGAb	Hyperthyroidism	Strongly associated with TAO

<sup>a</sup>Abbreviations used: Ab, antibody; TG, thyroglobulin; TPO, thyroid peroxidase; TBAb, TSH receptor-blocking antibodies; TSAb, TSH receptor-stimulating antibodies; TAO, thyroid-associated ophthalmopathy.

variety of soft tissue changes, including periorbital edema, proptosis, diplopia, and, in severe cases, optic nerve compression. Approximately 90% of patients with TAO have Graves' disease, 5% have autoimmune hypothyroidism, and almost all of the remainder have subtle evidence of thyroid autoimmunity.

## EPIDEMIOLOGY

Approximately 5% of pregnancies are followed by postpartum thyroiditis, although clinically significant features occur in only a small percentage of these cases. The prevalence of autoimmune hypothyroidism in women is 1 or 2% and in men it is 5–10 times less frequent. Approximately 2% of individuals with thyroid autoantibodies and presumed focal thyroiditis progress to overt autoimmune hypothyroidism annually; the rate is more than double if there are thyroid autoantibodies and subclinical hypothyroidism. The peak incidence of autoimmune hypothyroidism occurs in individuals between 50 and 60 years of age.

Graves' disease accounts for 60–80% of all cases of thyrotoxicosis, with a prevalence of approximately 1% in women, peaking in those aged 30–60 years. In men, the frequency is 5- to 10-fold lower than that in women. Approximately 50% of patients with Graves' disease have clinically obvious TAO, but subclinical evidence of this complication can be found in more than 90% of cases by using imaging techniques.

## PATHOLOGICAL FEATURES

In Hashimoto's thyroiditis, the thyroid is enlarged, usually symmetrically, due to a diffuse infiltration by lymphocytes, plasma cells, and macrophages, together

with the formation of lymphoid germinal centers. Thyroid cells undergo hyperplasia and oxyphil metaplasia, with the latter leading to the formation of Hürthle or Askanazy cells. Progressively, however, thyroid cell destruction occurs, together with variable amounts of fibrosis. In primary myxedema (atrophic thyroiditis), there is more extensive fibrosis and follicular destruction and minimal or moderate lymphoid infiltration.

In Graves' disease, there is hypertrophy and hyperplasia of the thyroid follicles, which have columnar, folded epithelium and less colloid than usual. There is a variable degree of lymphoid infiltration, sometimes with germinal center formation. All these changes are largely reversed by antithyroid drug treatment. The changes seen in TAO consist of variable lymphocytic infiltration of the extraocular muscles and orbital connective tissue, edema, and, in the later stages, fibrosis and possibly muscle atrophy. Fat content may increase, especially late in the disease process, and mast cells may be prominent. The levator palpebrae muscle fibers are enlarged. Lymphocytic infiltration of the eyelids is uncommon, whereas this can often be seen in the lacrimal glands.

## DETERMINANTS OF THYROID AUTOIMMUNITY

Like almost all autoimmune disorders, those associated with thyroid autoimmunity are the result of a complex interaction between a large number of genetic, environmental, and endogenous factors, with the same disease resulting from different combinations of factors in different patients. The contribution of genetic factors to disease susceptibility is unknown, but

**Table II Diseases Associated with Thyroid Autoimmunity**

Endocrine disorders
Type 1 diabetes mellitus
Addison's disease
Premature ovarian failure
Lymphocytic hypophysitis
Autoimmune polyglandular syndrome types 1 and 2
Organ-specific autoimmunity
Vitiligo
Alopecia areata
Pernicious anemia
Celiac disease and dermatitis herpetiformis
Myasthenia gravis
Autoimmune serositis
Chronic active hepatitis
Primary biliary disorders
Non-organ-specific autoimmunity
Rheumatoid arthritis
Systemic lupus erythematosus
Sjögren's syndrome
Relapsing polychondritis
Systemic sclerosis
Other conditions
Breast cancer
Depression or bipolar affective disorder
Recurrent miscarriage

concordance rates in monozygotic twins are approximately 20–30%, considerably higher than in dizygotic twins. Different autoimmune thyroid diseases may occur in the relatives of a proband in a single family, and there is a well-known set of associations with other autoimmune diseases (Table II), suggesting that these share common susceptibility factors, most likely genetic.

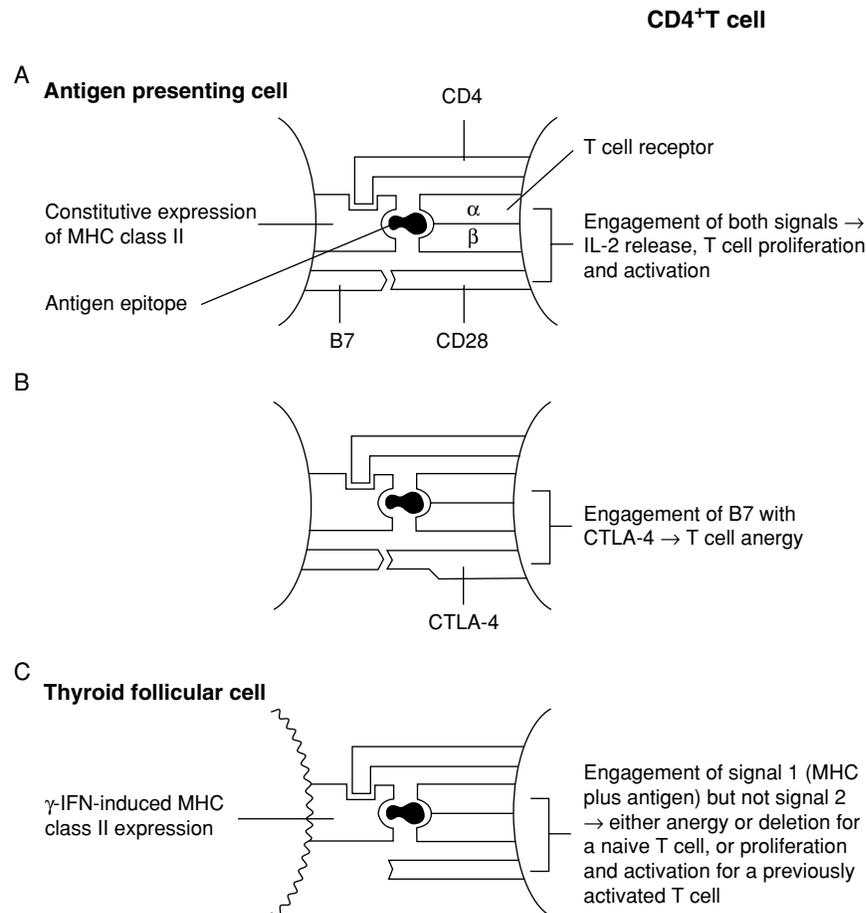
## Genetic Factors

In common with the majority of autoimmune diseases, the highly polymorphic alleles of the major histocompatibility complex (MHC), called HLA (human leukocyte antigen) in man, have been the most intensively studied and are the most frequently associated genetic factors in thyroid autoimmunity (Table III). There are several reasons why HLA alleles might confer susceptibility to autoimmune thyroiditis. The MHC class II genes are crucial for antigen presentation (Fig. 1), and it is possible that only certain class II molecules can bind particular autoantigenic epitopes and thus initiate an autoimmune response (determinant selection). Particular class II alleles may also be critical for the thymic selection of T cells during early life, either permitting the evolution of autoreactive T cells (positive selection) or preventing their development (negative selection); in the latter case, the allele would have a protective effect. Alternatively, some MHC class II alleles may determine selection of an increasingly better defined subpopulation of regulatory T cells, previously called suppressor cells. Finally, the effects of the HLA-DR3 specificity are nonspecific and seem to be related to a general enhancement of immune responsiveness, which in turn may be related to the effects of polymorphisms in other genes in linkage disequilibrium, such as complement components or the cytokine tumor necrosis factor.

However, it is clear that many individuals possess HLA-DR alleles that confer susceptibility but never develop autoimmune thyroid disease. Conversely, approximately half of individuals with autoimmune thyroid disease do not have any known HLA allele associated with susceptibility. The inevitable conclusion is that other genes must be involved, and these may act independently or interact with MHC-encoded susceptibility, making the determination of their effect more difficult. This genetic diversity is

**Table III Major HLA Associations in Thyroid Autoimmunity**

Condition	Ethnic group	Association
Autoimmune hypothyroidism	Caucasian	HLA-DR3, DR4
	Chinese	HLA-B46, DR9
	Japanese	HLA-DR53, DR9
Postpartum thyroiditis	Caucasian	HLA-DR5, possibly DR3
Graves' disease	Caucasian	HLA-DR3 (DR17 and DR18 subtypes)
	Japanese, Korean	HLA-DR5, DR8, DPB1*0501 (DQB1*0501 protective)
	U.S. Black	None
	South African Black	HLA-DR1, DR3
Thyroid-associated ophthalmopathy	Caucasian and Japanese	As for Graves' disease; HLA-DPB1*0201 may be protective



**Figure 1** Pathways of antigen presentation leading to CD4<sup>+</sup> T-cell stimulation (A, acute antigen presentation) or anergy (B, chronic antigen presentation; C, antigen presentation by endocrine cell). Reproduced from Weetman, A. P. (2003). Autoimmune thyroid disease: propagation and progression. *Eur. J. Endocrinol.* **148**, 1–9.

also demonstrated by the three to four times higher concordance rate for Graves' disease in monozygotic twins than in HLA-identical siblings.

The only other unequivocal genetic factor known to be associated with susceptibility is polymorphism in the *CTLA-4* gene, although it is unclear whether this is related to effects from the gene or closely related genes in linkage disequilibrium. Certainly, *CTLA-4* is an attractive candidate autoimmunity gene because the CTLA-4 molecule is critical to the termination of immune responses (Fig. 1) and CTLA-4 knockout mice develop a lymphoproliferative syndrome associated with widespread lymphocytic infiltration of tissues. The susceptibility conferred by *CTLA-4* polymorphisms is modest and similar in both Graves' disease and autoimmune hypothyroidism; despite initial observations, these polymorphisms do not confer any heightened risk of complications such as TAO. Moreover, the same polymorphisms are shared with

other disorders, such as type 1 diabetes mellitus, vitiligo, and Addison's disease, and therefore seem to confer a general propensity to develop organ-specific autoimmunity, which in turn may partially explain the frequent occurrence of the disorders.

Further progress in understanding the genetic basis of thyroid autoimmunity will only be achieved by performing large-scale family studies, either by testing candidate genes as these are discovered in other immunological disorders or by performing genomewide scanning in a search for previously unsuspected loci. Some progress has been made in this area, with certain candidate loci such as TSH-R being excluded from having a major role, whereas several potential loci have been identified, particularly on chromosomes 14q31, 18q21, 20q13, and Xp11. However, the power of these studies is insufficient to provide conclusive evidence for linkage, and several hundred families will need to be tested to obtain adequate power for

confidence. Even then, localizing the relevant susceptibility gene within a large locus will take much time and effort.

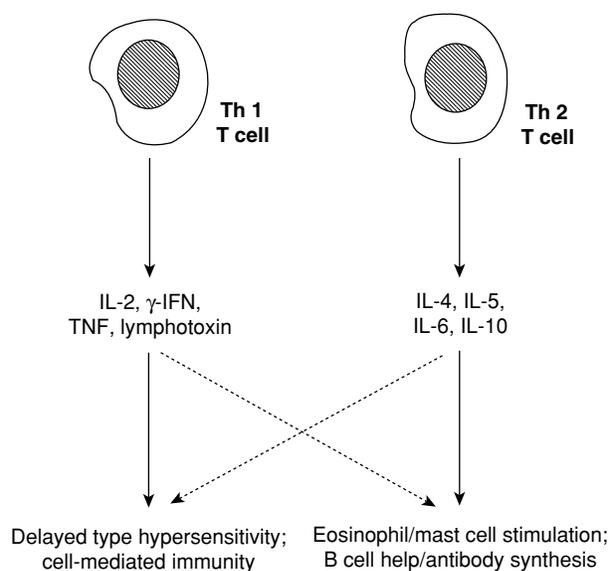
### Environmental Factors

The importance of environmental factors in causing autoimmune thyroiditis is illustrated by the incomplete concordance in monozygotic twins and the lack of any family history of autoimmunity in many patients. Genetic factors are more likely to exert their effects in early onset cases, and environmental factors, which require exposure of the individual, likely operate at later stages in life, but matters are even more complex because there are certainly interactions between genetic and environmental factors. Our inadequate knowledge stems largely from the difficulty of analyzing such heterogeneous interactions.

Infections are frequently suggested to play a role in producing autoimmunity by causing release of autoantigens from damaged tissue, by molecular mimicry in which epitope sequences of the organisms provoke an immune response against matching host protein sequences, or by altering target cell surface molecules that render the cell susceptible to autoimmune attack. However, there is no good evidence that thyroid autoimmunity is precipitated by infection and, indeed, subacute thyroiditis caused by viral infection of the thyroid is characterized by recovery of normal thyroid function. *Yersinia* infection has been examined as a potential trigger because there is evidence that the organism contains TSH-R-like sequences, but any role is likely to be small at best because few patients with Graves' disease have evidence of infection. There is even less reliable support for the involvement of retroviruses.

Iodide intake may determine the frequency of thyroid autoimmunity because there is epidemiological evidence that increases in dietary iodide are associated with an increase in the incidence of thyroid autoantibodies, autoimmune hypothyroidism, and Graves' disease. Animal models of autoimmune thyroiditis have provided support for such an enhancing role; conversely, iodide depletion appears to be protective. Excess iodide may be toxic to thyroid cells or may have immunological effects, including enhanced immunogenicity of thyroid autoantigens.

Major stress may precipitate Graves' disease, at least based on evidence from retrospective surveys, and such an effect could be produced by the neuroendocrine consequences of stress, especially on glucocorticoids. Smoking is weakly associated with Graves' disease and is the strongest known determinant of TAO. The



**Figure 2** Reciprocal regulation of T helper (Th) cell subsets. The dotted lines represent an inhibitory pathway.

reasons for these associations are not known, but it is possible that smoking may either modulate the immune response in some way (e.g., by affecting cytokine production) or enhance orbital fibroblast responses secondary to partial hypoxia.

At low doses, radiation may initiate thyroid autoimmunity because both Graves' disease and autoimmune hypothyroidism have been reported after mantle irradiation for lymphoma, and there is an increase in thyroid autoantibodies in the survivors of irradiation from nuclear weapon or reactor fallout. Some drugs may also precipitate thyroid autoimmunity, particularly in those who are genetically predisposed; examples include lithium,  $\alpha$ -interferon, and other cytokines. For example, Graves' disease occurs in one-third of multiple sclerosis patients treated with monoclonal antibodies to deplete T cells; it is likely that during T-cell recovery from this treatment, the cytokine profile of the emerging T cells shifts from a T helper cell 1 (Th1) to a Th2 pattern, with the latter provoking autoantibody production (Fig. 2). However, it is unclear why the TSH-R specifically is targeted.

### Endogenous Factors

Pregnancy is associated with amelioration of autoimmune thyroiditis, and patients with Graves' disease can usually stop drug treatment during the last trimester. However, a rebound in the autoimmune process occurs during the year after delivery. This is

when postpartum thyroiditis emerges and Graves' disease appears with an increased incidence. These changes are largely unexplained but presumably relate in part to massive fluctuations in various hormones during and after pregnancy. Sex hormones are likely the major determinants of the increased risk of thyroid autoimmunity in women, but the causes of the increased risk in Turner's and Down's syndromes are unknown.

## THYROID AUTOANTIGENS

Understanding any autoimmune disease is contingent on a thorough knowledge of the nature of the autoantigens involved and their B- and T-cell epitopes. In addition to aiding in the determination of the etiology and pathogenesis of an autoimmune disease, autoantigen identification is central to assays for autoantibodies that may be markers or predictors of disease or for the delineation of particular subsets of patients, and it may lead to novel treatments, such as that based on modified T-cell peptide epitopes. Indeed, our poor understanding of the pathogenesis of TAO is due directly to an inadequate understanding of the autoantigens involved. A summary of the main thyroid autoantigens is shown in [Table IV](#).

### Thyroglobulin

This was the first tissue autoantigen to be identified, based on the pioneering work by Witebsky and Rose, who immunized rabbits with thyroid extract in adjuvant, eliciting thyroglobulin antibodies and thyroiditis, and work by Doniach and Roitt, who first showed the presence of thyroglobulin antibodies in patients with Hashimoto's thyroiditis. Thyroglobulin is the thyroid hormone storage system; tyrosine molecules at hormonogenic sites are iodized to form T3 and T4. The protein is secreted into the follicular lumen, in which it is endocytosed and hydrolyzed to release

T3 and T4. Iodination of the molecule is important for its immunogenicity in animal models of autoimmune thyroiditis: there is no clear evidence that this is important in humans.

Two major epitopes and one minor conformational epitope exist on each thyroglobulin subunit for human autoantibodies, but the B-cell immune response becomes even more diverse with chronicity and increasing antibody titer, and some Hashimoto autoantibodies bind to linear determinants of thyroglobulin. The majority of thyroglobulin antibodies belong to the IgG-1 and IgG-4 subclasses but do not fix complement because the spacing of the autoantigenic epitopes prevents cross-linking. In the past, thyroglobulin autoantibodies were detected by immunofluorescence or hemagglutination assays, but these have generally been superseded by enzyme-linked immunosorbent assays (ELISAs) and radioimmunoassays, which express levels in units related to standard antibody preparation. The most sensitive assays indicate that low levels of thyroglobulin antibodies can be detected in up to 20% of the healthy female population.

### Thyroid Peroxidase

There are two forms of thyroid peroxidase, TPO-1 and -2, but only the first has enzymatic activity and is expressed on the cell surface, where it catalyzes the iodination and coupling of tyrosine molecules to form T3 and T4. Originally, autoantibodies against TPO were called microsomal antibodies because they bound to an antigen in this subcellular fraction, although they were also recognized from immunofluorescence to bind to the apical microvillar surface of the thyroid cells. It was subsequently shown that binding of these complement-fixing microsomal antibodies could be inhibited by TPO-specific monoclonal antibodies, and studies of recombinant TPO confirmed that essentially all the binding to thyroid microsomes is directed against TPO.

**Table IV** Key Properties of Thyroid Autoantigens

	Chromosomal location	Size (kDa)	Protein type	Function	Homology
Thyroglobulin	8	660	Iodinated glycoprotein	T3 and T4 prohormone and storage	Acetylcholinesterase
Thyroid peroxidase	2	105–110	Hemoprotein enzyme	Iodination and coupling of tyrosine	Myeloperoxidase
TSH receptor	14	85	G protein-coupled receptor	Thyroid stimulation by TSH	Related receptors (e.g., luteinizing hormone)
Na <sup>+</sup> /I <sup>-</sup> symporter	19	65–77	Transmembrane transporter	Iodide uptake	Na <sup>+</sup> /glucose transporter

Further work localized two linear B-cell epitopes, C-2 (amino acids 590–622) and C-21 (amino acids 710–722), but these are only recognized by autoantibodies arising late in the disease process. In the early phase of any autoimmune response involving TPO, autoantibodies are directed against confirmational epitopes in two overlapping domains, A and B, and relative antibody binding to each of these domains is fairly constant over time. This narrow response seems to be related to the marked restriction of the immunoglobulin variable (V) region gene usage in TPO autoantibody formation, with each domain being recognized by antibodies encoded by different V gene sequences, indicating a genetic component to the control of autoantibody production. In contrast, T-cell epitopes are heterogeneous, with several different epitopes typically recognized by T cells from patients and only partially shared between patients. Such diversity is typical of a chronic immune response in which only one dominant T-cell epitope may initiate the process, but this is followed by “spreading” of the response to involve other epitopes over time, including so-called cryptic epitopes that only become exposed during the phase of tissue destruction.

TPO antibodies are measured using the same techniques as those used to measure thyroglobulin antibodies and are also found in healthy subjects in the absence of overt thyroid disease. They are predominantly of the IgG-1 and IgG-4 subclasses. Approximately one-fourth of patients with thyroid autoimmunity have TPO antibodies that can also bind thyroglobulin, so-called TGPO antibodies. Their pathological significance is unknown.

## TSH-R

This receptor is similar to other G-coupled receptors and comprises two subunits, corresponding to the extracellular and transmembrane domains, joined by disulfide bonds. The fundamental importance of the TSH-R to Graves' disease was first established by the identification of a long-acting thyroid stimulator (LATS) in the serum of Graves' patients that matched the thyroid stimulatory action of TSH but had a much longer time course and, hence, could not be TSH. LATS was soon identified as an IgG antibody and assays were developed for these thyroid-stimulating antibodies that relied on the ability of IgG to simulate cyclic AMP production—that is, the second messenger produced from TSH-R interaction with its ligand. Other intracellular signaling pathways may be activated by certain TSH-R antibodies, which may lead to goiter formation.

Considerable effort has been directed at identifying the B-cell epitopes on TSH-R recognized by autoantibodies, but the results are not conclusive. However, multiple species of TSH-R autoantibodies exist and some actually prevent TSH from stimulating the receptor without causing any stimulation themselves. The effect of these blocking antibodies is hypothyroidism. Other “neutral” TSH-R antibodies bind but have no functional effects. Although the detection of TSH-R-stimulating and-blocking autoantibodies depends on bioassays, most commonly measuring the cyclic AMP responses of the FRTL-5 rat thyroid cell line or Chinese hamster ovary cells transfected with TSH-R, these are cumbersome and not suited to rapid diagnostic testing. Antibodies that bind to TSH-R can be measured simply in solid-phase assays that estimate competition between antibody and radiolabeled TSH for binding to the receptor. Results from assays of these so-called TSH-binding inhibiting immunoglobulins (TBIs) correlate well with the results of thyroid-stimulating antibodies in thyrotoxic patients, but TBI is also positive when blocking antibodies are present. Therefore, a positive TBI cannot be equated with the presence of the TSH-R-stimulating antibodies that cause Graves' disease.

There are multiple B-cell epitopes in the TSH-R that are generally conformational, comprising discontinuous segments that overlap the TSH binding site. Antibodies that bind to the carboxyl terminus of the extracellular domain have blocking activity, and those against the amino terminus have stimulating activity. Further dissection of these epitopes may allow the development of solid-phase assays for TSH-R-stimulating and -blocking antibodies that would render bioassays obsolete. In some Graves' patients, the TSH-R-stimulating antibody response is restricted in terms of light and heavy-chain use, implying genetic control of these strongly pathogenic antibodies. Another noteworthy feature is the very low concentration of these antibodies in sera, which makes their study difficult.

T-cell epitopes are heterogeneous and no clear dominant epitope has been identified. Nonetheless, the production of TSH-R antibodies is T-cell dependent, and presumably there are such dominant epitopes in the earliest phase of disease.

## Other Autoantigens

The Na<sup>+</sup>/I<sup>-</sup> symporter is recognized by antibodies in 5–30% of patients with Graves' disease or autoimmune hypothyroidism. Some of these antibodies have weak blocking effects on symporter activity *in vitro*, but *in vivo* iodide uptake is not a rate-limiting step in

thyroid hormone synthesis and it is unknown whether there is any pathogenic role for these antibodies. A poorly characterized second colloid antigen (i.e., in addition to thyroglobulin) has been recognized for decades based on immunofluorescence studies with patient sera, but its nature is obscure. It has been suggested that a set of thyroid growth-stimulating and growth-inhibiting antibodies, separate from TSH-R antibodies, cause goitrous and atrophic thyroiditis, but this hypothesis is controversial and these antibodies are poorly characterized.

The biggest enigma is the nature of the orbital autoantigen(s) in TAO. To explain the close association between this condition and thyroid autoimmunity, especially Graves' disease, the existence of an orbital autoantigen that cross-reacts with one in the thyroid has been proposed. The most attractive candidate is the TSH-R, which is expressed by the preadipocyte subset of fibroblasts; however, whether this is sufficient to explain pathogenesis is debated. A wide variety of extraocular muscle and fibroblast antigens have been suggested during the past 20 years based on ELISA and immunoblotting studies, and it seems likely that several autoantigens are targets of the autoimmune response, at least in the later stages of disease, akin to the situation in autoimmune thyroiditis and type 1 diabetes mellitus.

## PATHOGENESIS

### Animal Models

A wide variety of animal models have yielded invaluable insights into the pathogenesis of autoimmune

thyroid disease (Table V). Immunization of mice with thyroglobulin in Freund's adjuvant allowed the demonstration of the crucial role of MHC class II in controlling responsiveness as measured by the production of lymphocytic thyroiditis. It also showed that the production of thyroglobulin antibodies is under additional genetic control. Certain strains of mice (H-2<sup>k,s</sup>) are so susceptible to autoimmune thyroiditis that immunization with thyroglobulin alone will induce disease, indicating that normal, healthy animals have thyroid autoreactive T cells that have failed to undergo thymic deletion or tolerance. From this it can be inferred that such T cells are kept under control, either through the actions of regulatory T cells or through the mechanism of clonal ignorance, so that there is no response because autoantigen is normally sequestered or because appropriate costimulatory signals, including cytokines, are not normally present. T cells, but not antibodies, from animals with immunization-induced thyroiditis can transfer the disease to recipients, and the critical effector cells appear to be CD4-dependent CD8<sup>+</sup> cells.

One mechanism whereby T-cell-mediated regulation might occur is through the reciprocal inhibition of T helper cell subsets (Fig. 2). This has been shown in experiments in which Th1 responses, which are normally associated with autoimmune thyroiditis, were suppressed and Th2 responses enhanced by  $\gamma$ -interferon deficiency. Immunization with thyroglobulin under these conditions led to an eosinophilic granulomatous thyroiditis rather than lymphocytic thyroiditis, and high levels of thyroglobulin antibodies were induced.

Table V Animal Models of Thyroiditis<sup>a</sup>

Model	Species	Features
Immunization with TG or TSH-R and adjuvant	Mouse, rat, rabbit, guinea pig, monkey	TGAb or TBII, thyroiditis, euthyroid
T-cell depletion (thymectomy, with or without irradiation, cyclosporin A, depletion and reconstitution)	Mouse, rat	TGAb, thyroiditis, euthyroid; reversed by T <sub>R</sub> cells from healthy animals
TSH-R cDNA immunization	Mouse	TBII, rare TSAb, thyroiditis, euthyroid in BALBc; TSAb and hyperthyroidism plus thyroiditis in outbred mice
Immunization with fibroblasts transfected with MHC class II and TSH-R or TPO	Mouse	TBII including TSAb or TPO Ab, no thyroiditis; animals with TSAb thyrotoxic
Spontaneous thyroiditis	Obese strain (OS) chickens, BB and Buffalo strain rats, NOD and MRL <i>Ipr/Ipr</i> mice	TGAb, thyroiditis, and in OS chickens hypothyroidism; incidence of thyroiditis in Buffalo strain rats increased by 3-methylcholanthrene

<sup>a</sup>Abbreviations used: TG, thyroglobulin; TPO, thyroid peroxidase; TSH-R, TSH receptor; TSAb, TSH-R-stimulating antibodies; TBII, TSH-binding inhibiting immunoglobulins; Ab, antibodies; T<sub>R</sub>, regulatory T cells.

Another type of T-cell-mediated regulation has been shown in mice or rats subjected to T-cell depletion by thymectomy and other methods. The best characterized example is genetically susceptible rats, which develop autoimmune thyroiditis and thyroglobulin antibodies when subjected to neonatal thymectomy and sublethal irradiation. Disease can be prevented by the transfer of T cells from healthy donors, clearly indicating the presence of suppressor T cells in these animals. The concept of suppressor cells suffered from criticism and a decade of neglect after it proved impossible to characterize the mechanisms responsible for their antigen recognition and function. However, rebranded as regulatory T cells ( $T_R$ ), it has become clear that such cells do exist in the  $CD4^+$  population, are generated in the thymus, and express high levels of the interleukin-2 (IL-2) receptor (CD25).

The antigen specificity of these  $T_R$  cells is unclear but unlikely to be identical to the epitopes recognized by the effector T cells that they regulate. It is possible that  $T_R$  cells respond to epitopes on a separate autoantigen from the one generating the effector response, but this nonetheless allows their localization within a target organ, and their possible activation by the same antigen presenting cell that activates the effector T cell could allow "linked suppression." The exact mechanism whereby suppression is mediated is also unknown, but it probably involves cell contact as well as the release of inhibitory cytokines, such as IL-10, IL-13, and transforming growth factor- $\beta$ . Other types of regulatory T cells may exist, especially in the  $CD8^+$  population, and the relative importance of these various pathways, especially in the intact animal, will be the subject of intense research in the future since the potential for enhancing  $T_R$  function has been recognized to provide a novel way of controlling autoimmune responses.

Another valuable finding from the thymectomy/irradiation model of thyroiditis is that although disease is induced only in genetically susceptible animals, there is an important interaction with environmental factors. In particular, animals raised under germ-free conditions do not develop thyroiditis, whereas those that were germ free but were then given normal gut microflora developed thyroiditis with the expected incidence. In addition, disease in these animals, and in those with immunization-induced thyroiditis, can be modulated by sex steroids: The frequency and severity of thyroiditis are increased in castrated males or those given estrogens and decreased in female animals given testosterone.

The closest models to that of Hashimoto's thyroiditis are those that develop spontaneously in

genetically susceptible animals (Table V). Dietary iodide and other environmental factors have important permissive roles in these highly inbred strains. However, progress has been much slower in producing animal versions of Graves' disease, and only recently have closely corresponding models been developed. Perhaps the most interesting is the creation of pathological features suggestive of TAO in a specific strain of mouse (BALB/c) given TSH-R-activated T cells from donor animals immunized with TSH-R. Another mouse strain (NOD) developed a destructive thyroiditis under the same conditions, and this difference may be explained by the relative bias toward a Th2 response in BALB/c mice and a Th1 response in NOD mice.

## Autoimmune Hypothyroidism

### Cell-Mediated Effects

As previously discussed, T cells play a critical role in the pathogenesis of autoimmune thyroiditis in animals, and considerable evidence supports a similar role in humans.  $CD4^+$  and  $CD8^+$  T cells accumulate in the thyroid and the majority of these cells express markers of activation. Although some studies have suggested that these T cells are clonally restricted, others have found diverse uses of T-cell receptor  $V\alpha$  and  $V\beta$  genes, and this type of polyclonal response is in agreement with the known multiplicity of autoantigens and their epitopes and the likely spreading of antigenic determinants over time. Circulating T cells have an increased percentage of HLA-DR $^+$  T cells, marking a state of activation, and the proportion of  $CD8^+$  T cells is probably decreased, but these are features of many autoimmune disorders and may have no autoantigen-specific cause or effect.

Circulating and intrathyroidal T cells from patients with autoimmune hypothyroidism can be activated by and proliferate in response to thyroglobulin or thyroid peroxidase, and many ingenious attempts have been made to use such assays to demonstrate a defect in suppressor cell function as an underlying feature in pathogenesis. However, firm evidence of any such defect remains elusive, and no studies have examined the potential role of  $CD4^+/CD25^+$   $T_R$  cells in human thyroid autoimmunity. Although a predominantly Th1-type response would be predicted in autoimmune hypothyroidism, a complex and heterogeneous cytokine profile has been found when the intrathyroidal lymphocyte population has been examined. Of course, thyroid autoantibody synthesis, which is likely Th2 dependent, is a prominent feature of this condition.

The mechanism by which thyroid-specific T cells are activated has been a major focus of attention for the past two decades. Conventional antigen presentation in the initiation of an immune response is primarily a function of dendritic cells, which can process antigen and present it with appropriate costimulatory molecules (Fig. 1). B cells and macrophages can also present antigen under appropriate conditions, and for B cells there is the additional feature that specific antigens can be processed because these are taken up by binding to antigen-specific antibody on the B-cell surface. The nature of the cell presenting antigen, and particularly the cytokines it secretes, plays a key role in determining whether a Th1 or Th2 response is mounted.

When it was discovered that thyroid follicular cells express MHC class II molecules in autoimmune hypothyroidism and Graves' disease, there was considerable interest in the idea that this could be the precipitant of the autoimmune process because the thyroid cell would be capable of presenting thyroid autoantigens to any specific T cells in the vicinity. It is now clear that such class II expression is a secondary rather than primary event, depending on the production of  $\gamma$ -interferon by locally infiltrating T cells, which clearly must have been activated by other antigen presenting cells. Moreover, the initiation of an immune response depends on the delivery of a second costimulatory signal to T cells at the same time as the first signal, namely MHC class II molecule complexed with antigenic epitope (Fig. 1). Failure to provide this second signal leads to anergy or even deletion—that is, the opposite outcome to activation. In general, thyroid cells do not appear to express B7, the most

important costimulatory molecule, although there is a single report suggesting its expression in Hashimoto's thyroiditis. However, it seems that MHC class II expression by thyroid cells, from a teleological standpoint, is a protective mechanism to provide peripheral tolerance, for instance, in the event of viral thyroiditis.

Once activated, T cells can be restimulated by cells presenting antigen in the absence of B7 and thus, in the setting of autoimmune thyroiditis, MHC class II expression of thyroid cells may have an important role in expanding the intrathyroidal pool of T cells and thus exacerbating the autoimmune process (Fig. 1). It is also possible that differences in the regulation of T cells at this critical stage of disease (e.g., through genetically determined variability in class II expression or cytokine availability) may determine whether a focal thyroiditis progresses to full-blown disease.

Following these observations, it has become clear that thyroid follicular cells express a vast array of immunologically active molecules, including adhesion molecules and cytokines (Table VI). Many of these are induced by the cytokines known to be present in the intrathyroidal milieu in autoimmune thyroid disease, especially IL-1, tumor necrosis factor, and  $\gamma$ -interferon. Their expression tends to exacerbate the autoimmune response, although in some cases, such as the up-regulation of complement regulatory proteins, the cell is protected from immune attack.

How thyroid cells are destroyed is unresolved. Data from animal models indicate that CD8<sup>+</sup> cytotoxic T cells kill thyroid follicular cells, and perforin-containing cells exist in the thyroid infiltrate in human autoimmune thyroid disease. Recently, attention has focused on apoptosis as a major mechanism for

**Table VI Immunological Active Molecules Expressed by Thyroid Cells in Autoimmune Thyroiditis<sup>a</sup>**

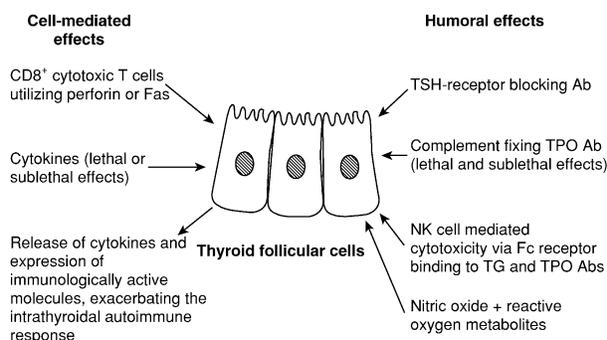
Molecule	Function	Upregulatory factors
MHC class I	Recognition by CD8 <sup>+</sup> cytotoxic T cells	$\gamma$ -IFN, TNF increase basal expression
MHC class II	Recognition by CD4 <sup>+</sup> helper T cells	$\gamma$ -IFN induces expression; enhanced by TNF
Adhesion and homing molecules (ICAM-1, LFA-3, Hermes-1, CD44, NCAM)	Binding of T and NK cells	$\gamma$ -IFN, IL-1, TNF
Cytokines (IL-1, IL-6, IL-8, IL-13, IL-15)	Multiple effects on bystander lymphocytes	$\gamma$ -IFN, IL-1, TNF
Complement regulatory proteins (CD46, CD55, CD59)	Prevention of lethal complement attack	$\gamma$ -IFN, IL-1, TNF
CD40	Binds CD40 ligand on T cells, acting as a costimulator	$\gamma$ -IFN, IL-1
Nitric oxide	Tissue injury	$\gamma$ -IFN, IL-1, TNF (decreased by IL-4, $\alpha$ -IFN, TGF- $\beta$ )
Fas/Fas ligand	Apoptosis	IL-1

<sup>a</sup>Abbreviations used: IFN, interferon; TNF, tumor necrosis factor; IL, interleukin; TGF, transforming growth factor; ICAM-1, intercellular adhesion molecule-1; NCAM, neural cell adhesion molecule.

thyroid cell loss. Thyroid follicular cells in Hashimoto's thyroiditis overexpress Fas (CD95), likely the result of locally released cytokines, especially IL-1. This could lead to an increased chance of thyroid cell destruction through interaction with CD8<sup>+</sup> T cells that bind to Fas and thereby trigger apoptosis. Several groups have described expression of Fas ligand (FasL) on thyroid follicular cells in autoimmune thyroid disease, although normally this molecule is restricted to CD8<sup>+</sup> T cells and sites of immunological privilege (e.g., trophoblast and testis), where it prevents T lymphocyte recognition of antigen by deleting these T cells through Fas engagement. In some studies, IL-1 up-regulated FasL expression by Hashimoto's thyroid cells *in vitro*, leading to the suggestion that exposure to this cytokine *in vivo* may cause suicide or fratricide within the Fas-positive thyroid cell population. However, questions have been raised over technical aspects of these experiments and, even if correct, the proportion of FasL<sup>+</sup> thyroid cells is often low. Up-regulation of Bcl-2 and other anti-apoptotic intracellular proteins, especially in Graves' disease, may limit thyroid cell death. Finally, the opposite view of FasL expression could be taken; namely, that such expression may protect thyroid cells from immune attack by creating the conditions of immunological privilege. Much work remains to be done before the roles of Fas/FasL and other cytotoxic pathways in the thyroid can be determined.

### Humoral Effects

Overwhelming evidence indicates that thyroid antibodies alone cannot cause thyroid cell destruction. Thyroglobulin antibodies do not fix complement. These antibodies and those against TPO and the Na<sup>+</sup>/I<sup>-</sup> symporter can cross the placenta and yet neonates born to mothers with high levels of such antibodies have normal thyroid function, as do many adults who have detectable thyroid antibodies. Experimental data indicate that these antibodies cannot access their antigens within an intact thyroid follicle. Nonetheless, there may be a secondary role for TPO antibodies: they cause tissue injury once the follicular architecture is disrupted by cytokines or T-cell-mediated cytotoxicity. Also, there is *in vitro* evidence that thyroglobulin and TPO antibodies can effect antibody-dependent, cell-mediated cytotoxicity by engaging Fc receptors on natural killer cells. There is certainly widespread complement activation within the thyroid in both Hashimoto's thyroiditis and Graves' disease. Although heightened expression of complement regulatory proteins (Table VI) may protect the thyroid cells



**Figure 3** Pathogenetic mechanisms in autoimmune hypothyroidism. Ab, antibody; TPO, thyroid peroxidase; TG, thyroglobulin.

from overt lysis, sublethal complement attack impairs the metabolic response of thyroid cells to TSH and induces cytokine production, which may exacerbate the autoimmune process.

TSH-R-blocking antibodies, in contrast, have pathogenic effects, causing hypothyroidism. This is most clearly demonstrated in the transient neonatal hypothyroidism that follows placental transfer. This effect lasts only as long as the antibodies persist and is therefore a functional rather than cytotoxic effect. As many as 20% of patients with hypothyroidism may have TSH-R-blocking antibodies, and these appear particularly common in patients with atrophic rather than goitrous thyroiditis in Asia, whereas in other regions TSH-R-blocking antibodies show no such dichotomy. Unexplained fluctuations in these antibodies (and sometimes in coincident TSH-R-stimulating antibodies) account for most, if not all, cases of hypothyroidism that show spontaneous recovery. The role of other potentially cytotoxic molecules, such as lymphotoxin- $\alpha$ , nitrous oxide, and reactive oxygen metabolites, in causing thyroid injury is unclear (Fig. 3).

## Graves' Disease

### Cell-Mediated Effects

Similar findings as those for autoimmune thyroiditis are found for Graves' disease. Circulating T cells show an increase in HLA-DR<sup>+</sup> and a decrease in CD8<sup>+</sup> subsets, and the thyroid infiltrate comprises CD4<sup>+</sup> and CD8<sup>+</sup> cells with little evidence of clonal restriction. Thyroglobulin-, TPO-, and TSH-R-specific T cells can be demonstrated in the circulation and thyroid infiltrate, but there is no evidence of a specific defect in their regulation. The same issues have been raised regarding antigen presentation as those for autoimmune hypothyroidism, and there is a mixed

in thyrocyte cytokine profile with evidence of both a Th1 and a Th2 response. These observations help explain the transition of Graves' disease to autoimmune hypothyroidism seen in some patients spontaneously and in others years after treatment with antithyroid drugs. There is also indirect evidence that hypothyroidism is more likely to occur after radioiodine or surgery in Graves' patients with the most marked autoimmune responses against thyroglobulin and TPO.

### **Humoral Effects**

Graves' disease is the prime exemplar of type V hypersensitivity, a disease caused by antibodies with a stimulatory action. The role of TSH-R-stimulating antibodies was discussed previously. Virtually all patients have detectable TSH-R-stimulating antibodies, providing the most sensitive assays are used. The few who do not most likely have antibodies but in an amount that is below the current limit of detection. The level of TSH-R-stimulating antibodies correlates loosely with the severity of hyperthyroidism due to the ongoing intrathyroidal T-cell-mediated events, the secondary effects of TPO antibody causing thyroid injury, and the simultaneous presence of TSH-R-blocking antibodies in some patients. There is a much clearer relation between the maternal level of TSH-R antibodies in women with Graves' disease in the last trimester of pregnancy and the occurrence of neonatal thyrotoxicosis, suggesting that the first two processes, which do not affect fetal or neonatal thyroid, are most important.

### **Effects of Treatment**

Spontaneous remission of Graves' disease appears to be unusual, unless autoimmune hypothyroidism intervenes. Accurate data are not available, but clinical remission was seen in only 10–20% of patients treated decades ago with beta-blockers, and these patients had the mildest degree of thyrotoxicosis. Remission of other receptor antibody-mediated diseases, such as myasthenia gravis, is also rare. Therefore, the 40–50% remission rate of Graves' disease after treatment with the antithyroid drugs carbimazole, methimazole, or propylthiouracil indicates an immunomodulatory effect of these agents. TSH-R antibody levels decline during treatment and remain low in those who do not relapse. However, levels of nonthyroid antibodies do not decline. This apparent paradox is most likely explained by the actions of antithyroid drugs on the thyroid cells, including suppression of cytokine, reactive oxygen metabolite, and prostaglandin production. As a result, the thyroid lymphocytic infiltrate declines

and the production of thyroid autoantibodies is curtailed. Remission is not expected in patients in whom the autoimmune process has spread beyond the thyroid gland because antithyroid drugs do not easily influence any extrathyroidal autoreactivity; this is in accord with the observation that patients with the most severe disease do poorly on antithyroid drugs.

Subtotal thyroidectomy usually leads to cessation of thyroid antibody production, although this may continue in some patients due to the persistence of specific B cells in the bone marrow and lymph nodes. Thyroid antibodies show a characteristic rise and then fall during the year after  $^{131}\text{I}$  therapy: this may be due to release of autoantigen from the damaged thyroid, followed by death of the intrathyroidal lymphocyte population. Activated T cells increase in the circulation in the first few months after  $^{131}\text{I}$  and these cells, by homing to the orbit, may be responsible for the exacerbation of TAO seen in up to 15% of patients after such treatment. This worsening can be prevented by glucocorticoids.

### **Thyroid-Associated Ophthalmopathy**

The pathogenesis of ophthalmopathy is complex and unclear. The extraocular muscles are the main focus of the autoimmune process, but occasional patients appear to have a primary expansion of the retro-ocular connective tissue, including fat. A separate process is likely involved in the clinical sign of lid retraction (which is more extreme than that in any type of thyrotoxicosis), at least based on the histological findings. There is no evidence that tissue destruction plays a role in initiating these changes, although extraocular muscle cell damage may occur late in severe disease; thus, attempts to find cytotoxic autoantibodies or lymphocytes seem ill founded. Apart from lid retraction, which could be mediated by the sympathetic overactivity secondary to thyrotoxicosis, no convincing signs of TAO have been reported in neonates born to women with TAO, arguing against a primary role for orbital autoantibodies. Attempts to find such autoantibodies are justifiable, however, because they may provide information on the nature of orbital autoantigens, and these autoantibodies may contribute to secondary events in TAO.

The most likely explanation for TAO is that fibroblasts in the extraocular muscles and orbital connective tissue are activated by cytokines from the infiltrating T cells and macrophages. IL-1, tumor necrosis factor, and  $\gamma$ -interferon all affect fibroblast function and may cause the increase in expression of HLA-DR and other immunologically active molecules

found in TAO. More important, they enhance glycosaminoglycan production, leading to water trapping and edema. An identical or related pathway of cytokine-induced activation may lead to fibrosis later in disease, the so-called burnt-out stage of TAO. Smoking may exacerbate this pathogenetic process by affecting cytokine secretion or action or by inducing partial hypoxia, which increases glycosaminoglycan production. Indeed, as the orbital contents increase, the pressure effects within the confines of the bony orbit may lead to partially hypoxic and congested conditions with impaired venous drainage, which in turn may cause further edema in addition to being a proinflammatory environment. In such a scenario, relatively modest initial increases in orbital tissue volume may quickly increase as in congestive ophthalmopathy. Anatomical differences between the capacity of the two orbits in an individual may accommodate different degrees of swelling, leading to the asymmetric form of TAO seen in 5–10% of cases. Another factor accounting for TAO is the likely heightened sensitivity of the orbital fibroblast population to cytokine activation compared to that of fibroblasts from other sites. Together with the presumed intraorbital location of a thyroid cross-reactive autoantigen, this susceptibility may explain the localization of disease.

Approximately 1% of patients with Graves' disease develop thyroid dermopathy, often situated over the shins, where it is called pretibial myxedema. These patients almost always have marked TAO, and dermopathy seems to be a more flagrant form of the same process, in which immunologically mediated fibroblast activation occurs within the dermis, presumably via the action of cytokines. Dermopathy is usually localized to sites where fluid retention and injury occur,

suggesting that a cascade of events similar to those of TAO may cause worsening of disease.

### See Also the Following Articles

Autoimmune Polyglandular Syndrome • Depression, Thyroid Function and • Goitrogens, Environmental • Graves' Disease • Immune System, Hormonal Effects on • Smoking and the Thyroid • Sodium Iodide Symporter • Thyroglobulin • Thyroid, Aging and • Thyroid Disease, Epidemiology of • TSH Receptor (Thyrotropin Receptor)

### Further Reading

- Allahabadia, A., and Gough, S. C. L. (1999). The different approaches to the genetic analysis of autoimmune thyroid disease. *J. Endocrinol.* **163**, 7–13.
- Bretz, J. D., and Baker, J. R., Jr. (2001). Apoptosis and autoimmune thyroid disease: Following a TRAIL to thyroid destruction? *Clin. Endocrinol.* **55**, 1–11.
- Davidson, A., and Diamond, B. (2001). Autoimmune diseases. *N. Engl. J. Med.* **345**, 340–350.
- Goodnow, C. C. (2001). Pathways for self-tolerance and the treatment of autoimmune diseases. *Lancet* **357**, 2115–2121.
- Kamradt, T., and Mitchison, N. A. (2001). Tolerance and autoimmunity. *N. Engl. J. Med.* **344**, 655–664.
- Kourilsky, P., and Truffa-Bachi, P. (2001). Cytokine fields and the polarization of the immune response. *Trends Immunol.* **22**, 502–509.
- Rapoport, B., and McLachlan, S. M. (2001). Thyroid autoimmunity. *J. Clin. Invest.* **108**, 1253–1259.
- Weetman, A. P. (2000). Graves' disease. *N. Engl. J. Med.* **343**, 1236–1248.
- Wiersinga, W. M., and Prummel, M. F. (2000). Editorial: Pathogenesis of Graves' ophthalmopathy—Current understanding. *J. Clin. Endocrinol. Metab.* **86**, 501–503.
- Wucherpfennig, K. W. (2001). Mechanisms for the induction of autoimmunity by infectious agents. *J. Clin. Invest.* **108**, 1097–1104.

carcinomas in patients with no history of radiation and in 60 to 85% of papillary carcinomas in patients following irradiation, diagnosed either in children in Belarus exposed to radiation after the Chernobyl accident or in patients who were exposed to external radiation in childhood. The frequency of TRK rearrangements is at least twofold lower. In papillary carcinomas, RAS mutations are found in 15% of cases and B-RAF mutations in 40–60% of cases; thus, abnormalities in the RET/RAS/B-RAF/mitogen-activated protein kinase pathway are found in 80% of cases, suggesting an important role.

Activating point mutations of the RAS genes are found with a similarly high frequency in both thyroid adenomas and follicular carcinomas, suggesting that RAS mutations represent an early event in thyroid tumorigenesis. PPAR $\gamma$ -PAX8 rearrangements were found in 26–63% of follicular carcinomas, mostly of the widely invasive subtype and more rarely in follicular adenomas.

Activating mutations of the genes encoding the thyrotropin receptor and the  $\alpha$ -subunit of the G<sub>s</sub> protein have been reported in some follicular carcinomas. Inactivating point mutations of the p53 tumor suppressor gene are rare in differentiated thyroid carcinomas, but are common in undifferentiated (anaplastic) thyroid carcinomas.

### Thyroid Irradiation

External irradiation to the neck during childhood increases the risk of papillary carcinoma. The latency period between exposure and diagnosis of thyroid carcinoma is at least 5 years. The risk is maximal at approximately 20 years, remains high for approximately 20 years, and then decreases gradually. The risk is increased after a mean dose to the thyroid of as little as 10 cGy. Above this dose, there is a linear relationship between the dose (up to 1500 cGy) and the risk of carcinoma. Beyond this point, the risk per Gray decreases, probably due to cell killing. A major risk factor is a young age at irradiation; over the age of 15 or 20 years, the risk is not increased. In children exposed to 1 Gy to the thyroid, the excess risk of thyroid carcinoma is 7.7.

The risk of thyroid carcinoma is not increased in patients given <sup>131</sup>I for diagnosis or therapy. However, the number of patients exposed to <sup>131</sup>I for medical reasons during childhood is too small to exclude a carcinogenic effect of <sup>131</sup>I at a young age. Conversely, a direct tumorigenic effect on the thyroid of radioactive isotopes of iodine, both <sup>131</sup>I and short-lived isotopes, has been strongly suggested by the

increased incidence of papillary carcinomas in children in the Marshall Islands after atomic bomb testing and in Belarus and the Ukraine after the Chernobyl accident. In Belarus and the Ukraine, approximately 1500 cases have been reported in children and adolescents who were younger than the age of 10 years at the time of the accident, which corresponds in some regions to a 100-fold increase in incidence compared with nonirradiated children.

### Other Factors

In countries where iodine intake is adequate, papillary and follicular cancers represent more than 80% of all thyroid carcinomas, with the papillary histologic type being the more frequent (60 to 80% of cases). There is no increase in the incidence of thyroid carcinomas in countries where iodine intake is low, but there is a relative increase in follicular and anaplastic carcinomas.

A high incidence of papillary carcinomas has been reported in patients with adenomatous polyposis coli and Cowden's disease (multiple hamartoma syndrome). Approximately 5% of cases of papillary carcinoma are familial and several loci for predisposing genes have been identified in these families.

## PATHOLOGY

### Papillary Carcinoma

Papillary carcinoma is an unencapsulated tumor with papillary and follicular structures that is characterized by changes in cell nuclei, such as overlap of the nuclei, large nuclei, a ground glass appearance, longitudinal nuclear grooves, and invaginations of the cytoplasm into the nuclei. Recognized histologic variants are the encapsulated cell, the follicular cell, the tall cell, the columnar cell, the clear cell, and the diffuse sclerosing variants; they are classified as papillary carcinomas due to their typical nuclear features. The tumor is multicentric in 20 to 80% of cases, depending on the care used to examine the thyroid, and is bilateral in approximately one-third of cases. It spreads through the lymphatics within the thyroid, to the regional lymph nodes, and less frequently to the lungs.

### Follicular Carcinoma

Follicular carcinoma is characterized by follicular differentiation but is devoid of the nuclear changes characteristic of papillary carcinoma. These carcinomas are encapsulated and invasion of the capsule and

vessels are the key features distinguishing follicular carcinomas from follicular adenomas. Two forms are recognized according to the pattern of invasion: a minimally invasive form and a widely invasive form. The growth pattern may also vary, from a well-differentiated pattern with macrofollicular structures to a poorly differentiated pattern with areas of solid growth and a high degree of atypia (trabecular or insular carcinoma). Hürthle cell (oxyphilic or oncocytic) carcinoma is a cytologic variant of follicular carcinoma. Multicentricity and lymph node involvement are less frequent than in papillary carcinoma and distant metastases to the lungs and bones stem from hematologic spread.

## DIAGNOSIS

Most differentiated thyroid carcinomas present as asymptomatic thyroid nodules, but occasionally the first signs of the disease are lymph node metastases and rarely lung or bone metastases. Hoarseness, dysphagia, cough, and dyspnea are suggestive of advanced stages of the disease.

At physical examination, the carcinoma, usually single, is firm, freely moveable during swallowing, and not distinguishable from a benign nodule. A thyroid nodule should be suspected of being a carcinoma when it is found in children or adolescents or in patients over the age of 60 years; in men, when it is hard and irregular; when ipsilateral lymph nodes are enlarged or compressive symptoms are present; and when there is a history of a progressive increase in size. Virtually all patients are clinically euthyroid and have normal serum thyrotropin concentrations.

Whatever the presentation, fine-needle aspiration cytology is the best test for distinguishing between benign and malignant thyroid nodules (Table I). Provided an adequate specimen is obtained, three cytologic results are possible: benign, malignant, and indeterminate (or suspicious). False-negative results, usually from sampling errors or interpretive errors, and false-positive results are rare. Among indeterminate results,

**Table I Probability of Malignancy Based on FNA Cytology**

Cytology	Percentage of results (%) [mean (range)]	Probability of malignancy (range)
Inadequate	17 (15–20)	10–20
Benign	70 (53–90)	1–2
Suspicious	10 (5–23)	10–20
Malignant	4 (1–10)	>95

only 10–20% are from malignant nodules, reflecting the difficulty of differentiating benign follicular adenomas from their malignant counterparts. Thyroid ultrasonography is useful for assessing the size of the nodule and detecting other nodules and to guide the fine-needle biopsy of small or poorly palpable nodules.

## PROGNOSTIC FACTORS

The overall 10-year survival rates for middle-aged adults with thyroid carcinomas are approximately 80 to 95%. Five to 20% of patients have local or regional recurrences and 10 to 15% have distant metastases. Prognostic indicators of recurrence and of death are age at diagnosis, histologic type, and extent of the tumor.

There are many scoring systems for thyroid carcinoma, among which the pathological tumor node metastases (pTNM) staging system is the most widely accepted (Table II). Based on this system, 80 to 85% of patients are classified as being at low risk for cancer-specific mortality. Some patients have a higher risk of recurrence, even if their risk of cancer-specific mortality is low. They include young (<16 years) and older (>45 years) patients, those with certain histologic subtypes (among papillary carcinomas, the tall cell, the columnar cell, and diffuse sclerosing variants, and among follicular carcinomas, the widely invasive and poorly differentiated histologic types and the Hürthle cell carcinomas), and those with large tumors, extension of the tumor beyond the thyroid capsule, or lymph node metastases. Therefore, the extent of initial treatment and follow-up should be adapted according to these prognostic indicators.

## INITIAL TREATMENT

### Surgery

The goal of surgery is to remove all neoplastic tissue in the neck. Therefore, the thyroid gland and affected cervical lymph nodes should be resected.

Although some controversy still exists regarding the extent of thyroid surgery, there are strong arguments in favor of a total or near-total (leaving no more than 2 to 3 g of thyroid tissue) thyroidectomy for all patients. Total or near-total thyroidectomy reduces the recurrence rate compared with more limited surgery because many papillary carcinomas are multifocal and bilateral. Removal of most if not all of the thyroid gland facilitates total ablation with <sup>131</sup>I. The argument against total thyroidectomy is that it increases the risk of surgical complications, such as recurrent

Table II TNM Staging System for Papillary and Follicular Carcinoma of the Thyroid

	Definition of TNM	
	1992	2002
Primary tumor (T)		
<b>T0</b>	No evidence of primary tumor	No evidence of primary tumor
<b>T1</b>	Tumor $\leq 1$ cm limited to the thyroid	Tumor $\leq 2$ cm limited to the thyroid
<b>T2</b>	Tumor $>1$ to $\leq 4$ cm limited to the thyroid	Tumor $>2$ to $\leq 4$ cm limited to the thyroid
<b>T3</b>	Tumor $>4$ cm limited to the thyroid	Tumor $>4$ cm limited to the thyroid or any tumor with minimal extrathyroid extension (e.g., extension to sternothyroid muscle or perithyroid soft tissues)
<b>T4</b>	Any size extending beyond the thyroid capsule	
<b>T4a</b>		Tumor of any size with extension beyond the thyroid capsule and invades any of the following: subcutaneous soft tissues, larynx, trachea, esophagus, recurrent laryngeal nerve
<b>T4b</b>		Tumor invades prevertebral fascia or mediastinal vessels or encases carotid artery
Regional lymph node (N) (In the 2002 system, to classify a tumor as N0 or N1, at least six lymph nodes should be examined at histology; otherwise, the tumor is classified as Nx)		
<b>N0</b>	No regional lymph node metastasis	No regional lymph node metastasis
<b>N1</b>	Regional lymph node metastasis	Regional lymph node metastasis
<b>N1a</b>		Metastases in pretracheal and paratracheal region, including prelaryngeal and delphian lymph nodes
<b>N1b</b>		Metastases in other unilateral, bilateral, or contralateral cervical or upper mediastinal lymph nodes
Distant metastases (M)		
<b>M0</b>	No distant metastasis	No distant metastasis
<b>M1</b>	Distant metastasis	Distant metastasis
TNM Staging		
		<b>Age <math>&lt;45</math> years</b>
<b>Stage I</b>	Any T, any N, M0	Any T, any N, M0
<b>Stage II</b>	Any T, any N, M1	Any T, any N, M1
<b>Stage III</b>	None	None
<b>Stage IV</b>	None	None
		<b>Age <math>\geq 45</math> years</b>
<b>Stage I</b>	T1, N0, M0	T1, N0, M0
<b>Stage II</b>	T2-T3, N0, M0	T2, N0, M0
<b>Stage III</b>	T4, N0, M0, or any T, N1, M0	T3, N0, M0 or any T1-3, N1a, M0
<b>Stage IV</b>	Any T, any N, M1	
<b>Stage IVA</b>		T1-3, N1b, M0 or T4a, any N, M0
<b>Stage IVB</b>		T4b, any N, M0
<b>Stage IVC</b>		Any T, any N, M1

laryngeal nerve injuries and hypoparathyroidism, and yet often some thyroid tissue remains, as detected by postoperative scanning with  $^{131}\text{I}$ .

In low-risk patients (those with papillary carcinomas less than 1 cm in diameter, if unifocal and intralobar, and those with a small minimally invasive follicular carcinoma), a lobectomy may be appropriate.

Surgery of the lymph nodes is routinely performed in patients with papillary carcinomas. It includes dissection of the central compartment (paratracheal and tracheoesophageal areas) and may also include dissection of the supraclavicular area and the lower third of the jugulocarotid chain. A modified neck dissection is performed if there are palpable lymph node metastases in the jugulocarotid chain. Dissection

is preferable to lymph node picking. This type of lymph node dissection was proven to improve the recurrence and survival rates in several series and several arguments support its routine use in patients with papillary carcinomas: approximately two-thirds have lymph node metastases, more than 80% of whom have involvement of the central compartment; metastases are difficult to detect in lymph nodes located behind the vessels or in the paratracheal groove. In the case of follicular carcinomas, lymph node metastases are less frequent, being found in approximately 35% of cases. A lymph node dissection is performed only if follicular carcinoma is already diagnosed or if palpable lymph nodes are present.

### <sup>131</sup>I Therapy

<sup>131</sup>I therapy is given postoperatively for three reasons: it destroys normal thyroid remnants, thereby increasing the sensitivity of subsequent <sup>131</sup>I total body scans and the specificity of measurements of serum thyroglobulin for the detection of persistent or recurrent disease; it may destroy occult microscopic carcinoma, thereby decreasing the long-term recurrence rate, and it permits a postablative <sup>131</sup>I total body scan, a sensitive tool for persistent carcinoma.

Postoperative <sup>131</sup>I therapy should be used selectively (Table III). In low-risk patients, the long-term prognosis after surgery alone is so favorable that <sup>131</sup>I ablation is not usually recommended. However, patients who are at high risk of recurrence are routinely

**Table III Indications for <sup>131</sup>I Ablative Treatment in Patients with Thyroid Carcinoma after Initial Surgery**

No indication:
Low risk of cancer-specific mortality and low risk of relapse: pTNM stage I
Indication:
Definite
Distant metastases
Incomplete tumor resection
Complete excision of tumor but high risk of mortality/recurrence: pTNM stages II and III
Probable
Age <18 years
Histologic subtype: papillary tall cell, columnar cell, or diffuse sclerosing; follicular widely invasive or poorly differentiated follicular
Elevated serum Tg concentrations >3 months after surgery on L-thyroxine treatment

treated with <sup>131</sup>I, because it decreases both the rate of recurrence and the rate of death.

Postoperatively, <sup>131</sup>I therapy is administered 4 to 6 weeks after surgery, during which no thyroid hormone treatment is given. A diagnostic total body scan with 2 mCi (74 MBq) <sup>131</sup>I is performed only when thyroidectomy has been partial. Another total body scan is performed 4 to 7 days later and thyroxine therapy is initiated. Total ablation may be verified by a <sup>131</sup>I total body scan 6 to 12 months later with 2 mCi (74 MBq).

Total ablation (no visible uptake) is achieved after administration of both 100 mCi (3700 MBq) and 30 mCi (1000 MBq) in more than 80% of patients who had a near-total thyroidectomy. After less extensive surgery, ablation is achieved in only two-thirds of patients with 30 mCi (1000 MBq). Therefore, near-total thyroidectomy should be performed in all patients who are to be treated with <sup>131</sup>I. Total ablation requires that a dose of at least 300 Gy be delivered to thyroid remnants and a dosimetric study allows the <sup>131</sup>I dose to be administered to be estimated more precisely.

### External Radiotherapy

External radiotherapy to the neck and mediastinum is indicated only in patients older than 45 years in whom surgical excision is incomplete or impossible and in whom the tumor tissue does not take up <sup>131</sup>I.

### FOLLOW-UP

The goals of follow-up after initial therapy are to maintain adequate thyroxine therapy and to detect persistent or recurrent thyroid carcinoma. Most recurrences occur during the first years of follow-up, but some occur late. Therefore, follow-up is necessary throughout the patient's life.

### Thyroxine Treatment

The growth of thyroid tumor cells is controlled by thyrotropin [or thyroid-stimulating hormone (TSH)] and inhibition of thyrotropin secretion with thyroxine improves the recurrence and survival rates. Therefore, thyroxine should be given to all patients with thyroid carcinoma, whatever the extent of thyroid surgery and other treatment. The initial effective dose is between 2.0 and 2.4 μg/kg body weight in young adults; children require higher doses. The adequacy of therapy is monitored by measuring serum TSH 3 months after

therapy is begun, with the initial goal being a serum TSH concentration of  $\leq 0.1 \mu\text{U/ml}$  and a serum free-triiodothyronine concentration within the normal range, to avoid overdosing.

### Early Detection of Recurrent Disease

#### **Clinical and Ultrasonographic Examinations**

Palpation of the thyroid bed and lymph node areas should be routinely performed but is poorly sensitive. Ultrasonography is performed in patients at high risk of recurrence and in any patient with suspicious findings. Palpable lymph nodes that are small, thin, or oval, that are found in the posterior neck chains, and that decrease in size after an interval of 3 months are considered benign. Serum thyroglobulin is undetectable in 20% of patients receiving thyroxine treatment who have isolated lymph node metastases and undetectable values do not exclude metastatic lymph node disease. These false-negative results can be found with neck ultrasonography, with an ultrasound-guided node biopsy performed in suspicious cases for cytology and thyroglobulin measurement in the fluid aspirate.

#### **X Rays**

Chest X rays are no longer routinely performed in patients with an undetectable serum thyroglobulin concentration. The reason is that virtually all patients with abnormal X rays have detectable serum thyroglobulin concentrations.

#### **Serum Thyroglobulin Determinations**

Thyroglobulin (Tg) is a glycoprotein that is produced only by normal or neoplastic thyroid follicular cells. It should not be detectable in patients who have had total thyroid ablation and its detection in such patients signifies the presence of persistent or recurrent disease.

The sensitivity of good Tg assays is 1 ng/ml or even less. The results can be artifactually altered by serum anti-Tg antibodies, which are found in approximately 15% of patients with thyroid carcinoma, and these antibodies should always be sought by radioimmunoassay or by a recovery test. In immunoradiometric assay (IRMA) methods, interference induces falsely reduced or false-negative serum Tg measurements.

The production of Tg by both normal and neoplastic thyroid tissue is in part TSH-dependent. Thus, when interpreting the serum Tg value, the serum TSH value should be taken into account, as well as the presence or absence of thyroid remnants (Table IV). When serum Tg is detectable during thyroxine

**Table IV Percentages of Patients with Detectable ( $>1 \text{ ng/ml}$ ) Serum Thyroglobulin Concentrations during Thyroxine Treatment and after Discontinuation of Thyroxine According to the Presence or Absence of Normal Thyroid Tissue<sup>a</sup>**

	Total ablation		Total thyroidectomy	
	On	Off	On	Off
L-T4 treatment	On	Off	On	Off
Complete remission <sup>b</sup>	<2	10	7	20
Lymph node metastases	80	~90	—	—
Distant metastases with normal X rays	95	~100	—	—
Large distant metastases <sup>c</sup>	~100	~100	—	—

<sup>a</sup>Detectable serum thyroglobulin concentrations are  $>1 \text{ ng/ml}$ . Serum thyroglobulin values are highly dependent on the assay. In this study, an immunoradiometric method with a sensitivity of 1 ng/ml was used.

<sup>b</sup>Most detectable serum thyroglobulin concentrations were  $<5 \text{ ng/ml}$ .

<sup>c</sup>Most serum thyroglobulin concentrations were  $>10 \text{ ng/ml}$ .

treatment, it will increase after the treatment is discontinued.

The serum Tg concentration is an excellent prognostic indicator. Less than 1% patients not receiving thyroxine with undetectable serum Tg and normal neck ultrasonography relapsed after more than 10 years of follow-up. Conversely, 20–80% of patients with serum Tg concentrations above 40 ng/ml after thyroxine withdrawal and with no other evidence of disease have detectable foci of  $^{131}\text{I}$  uptake in the neck or at distant sites after administration of therapeutic doses of  $^{131}\text{I}$ .

#### **$^{131}\text{I}$ Total Body Scan**

The results of  $^{131}\text{I}$  total body scan (TBS) depend on the ability of neoplastic thyroid tissue to take up  $^{131}\text{I}$  in the presence of high serum TSH concentrations, which are achieved by withdrawing thyroxine for 4 to 6 weeks. However, the resulting hypothyroidism is poorly tolerated by some patients. This can be attenuated by substituting the more rapidly metabolized triiodothyronine for thyroxine for 3 weeks and withdrawing it for 2 weeks. The serum TSH concentration should be above some arbitrary value ( $30 \mu\text{U/ml}$ ) in patients managed in this way; if it is not,  $^{131}\text{I}$  administration should be delayed until it is. Intramuscular injection of recombinant human thyrotropin (rhTSH) is an alternative [0.9 mg administered intramuscularly for 2 consecutive days with  $^{131}\text{I}$  administration (4 mCi) on the day following the second injection and TBS and serum Tg determination 2 days later], because thyroxine treatment needs not be discontinued and side

effects are minimal. Its efficiency for the detection of persistent and recurrent disease is comparable to that of thyroxine withdrawal in most patients.

When  $^{131}\text{I}$  scanning is planned, patients should be instructed to avoid iodine-containing medications and iodine-rich foods and urinary iodine should be measured in doubtful cases. Pregnancy must be excluded in women of childbearing age. For routine diagnostic scans, from 2 to 5 mCi (74 to 185 MBq)  $^{131}\text{I}$  is given; higher doses may reduce the uptake of a subsequent therapeutic dose of  $^{131}\text{I}$ . The scan is performed and uptake, if any, is measured 72 h after the dose using a double-head gamma camera equipped with high-energy collimators. False-positive results are rare.

### Post $^{131}\text{I}$ Therapy Total Body Scans

Assuming equivalent fractional uptake after administration of a diagnostic and a therapeutic dose of  $^{131}\text{I}$ , uptake too low to be detected with 2 to 5 mCi (74–185 MBq) may be detectable after the administration of 100 mCi (3700 MBq). Thus, a total body scan should be routinely performed 4 to 7 days after a high dose. This is also the rationale for administering a therapeutic dose of  $^{131}\text{I}$  in patients with elevated Tg levels ( $>10$  ng/ml off thyroxine treatment), even if the diagnostic scan is negative.

### Follow-up Strategy

If the total body scan performed after the administration of  $^{131}\text{I}$  to destroy the thyroid remnant does not show any uptake outside the thyroid bed, physical examination is performed and serum TSH and Tg are measured during thyroxine treatment 3 months later. A diagnostic  $^{131}\text{I}$  total body scan was performed after thyroxine withdrawal, 6 to 12 months later. Reports have clearly demonstrated that  $^{131}\text{I}$  TBS does not provide any useful information in these patients. Thus, the current protocol includes a determination of serum Tg following rhTSH stimulation and neck ultrasonography (Fig. 1).

If serum thyroglobulin following TSH stimulation is undetectable and neck ultrasonography is normal, low-risk patients are considered cured, and the dose of thyroxine is decreased to maintain a low but detectable serum TSH concentration (0.1 to 0.5  $\mu\text{U}/\text{ml}$ ). In higher-risk patients, higher doses of thyroxine are given, the goal being a serum TSH concentration of  $\leq 0.1$   $\mu\text{U}/\text{ml}$ . Clinical, ultrasonographic, and biochemical evaluation is performed annually; any other testing is unnecessary as long as the patient's serum Tg concentration is undetectable.

If serum Tg is detectable following TSH stimulation, two attitudes can be considered: in those patients with a relatively low Tg level, another determination is obtained 6 to 36 months later, because with longer follow-up serum Tg became undetectable in two-thirds of these patients; in those with high serum Tg levels and in those with increasing serum Tg at consecutive determinations, a high dose, 100 mCi (3700 MBq), of  $^{131}\text{I}$  is administered with a TBS 3–5 days later. If no uptake is detected, an  $^{18}\text{F}$ -fluorodesoxyglucose (FDG) positron emission tomography scan may be performed.

In patients receiving thyroxine in whom serum Tg becomes detectable, serum Tg should be measured after TSH stimulation. If the serum Tg concentration increases above 10 ng/ml, 100 mCi  $^{131}\text{I}$  should be given.

In low-risk patients who have had a near-total or a total thyroidectomy but who were not given  $^{131}\text{I}$  postoperatively, the follow-up protocol is based on Tg determinations and neck ultrasonography.

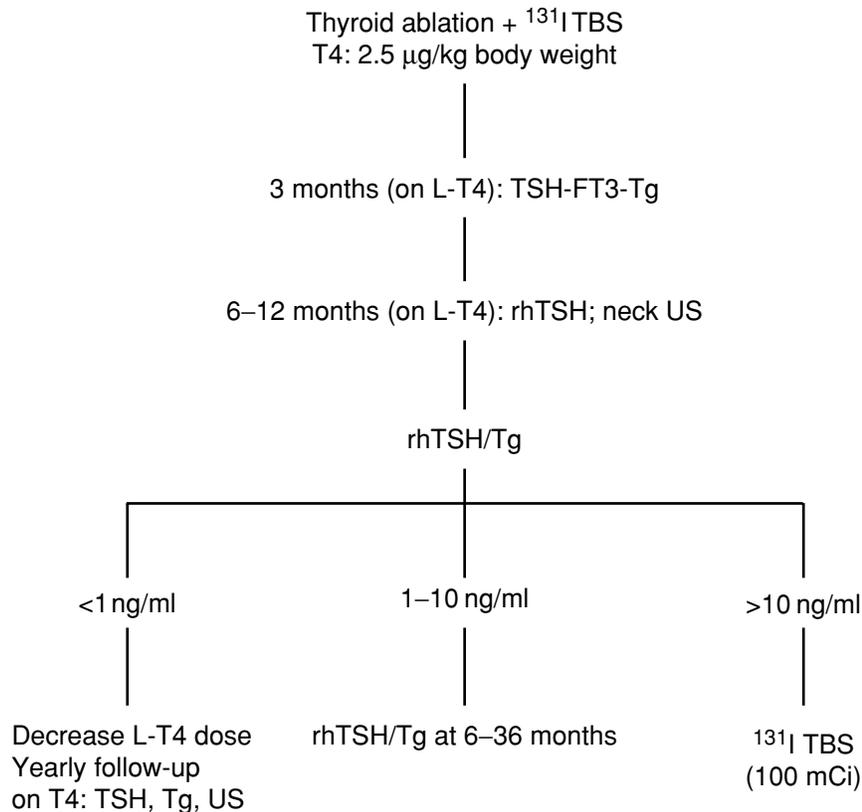
In low-risk patients who underwent a lobectomy, yearly follow-up consists of a neck examination and of serum Tg determination during thyroxine treatment. Ultrasonography will show abnormalities in the remaining lobe in most patients with detectable Tg concentrations. If the abnormality is small, fine-needle biopsy may be impossible and surgery is frequently the only option.

## LOCAL AND REGIONAL RECURRENCES

Local or regional recurrences occur in 5 to 20% of patients with differentiated thyroid carcinomas. Some are related to incomplete initial treatment (in a thyroid remnant or in lymph nodes) and others indicate tumor aggressiveness (in the thyroid bed after total thyroidectomy or in soft tissues).

A local or regional recurrence that is palpable or easily visualized with ultrasonography or computerized tomography (CT) scan should be excised. Total excision may be facilitated by total body scanning 4 days after administration of 100 mCi (3700 MBq)  $^{131}\text{I}$ , because additional tissue that should be excised may be identified. Surgery is performed 1 day later, preferably using an intraoperative probe. The completeness of resection is verified 1 to 2 days after surgery by another total body scan.

External radiotherapy is indicated in patients with soft tissue recurrences that cannot be completely excised and that do not take up  $^{131}\text{I}$ .



**Figure 1** Follow-up of patients after total thyroid ablation, based on serum thyroglobulin determinations and neck ultrasonography. The point at which a decision is made depends on the assay used to measure serum Tg. TBS, total body scan; TSH, thyrotropin; T4, thyroxine; Tg, thyroglobulin; US, neck ultrasonography; L-T4, L-thyroxine treatment.

## DISTANT METASTASES

Distant metastases, mostly in the lungs and bones, occur in 10 to 15% of patients with differentiated thyroid carcinomas. Lung metastases are most frequent in young patients with papillary carcinoma and are almost the only site of distant spread in children. Bone metastases are more common in older patients and in those with follicular carcinoma. Other less common sites are the brain, liver, and skin.

## Diagnosis

Clinical symptoms of lung involvement are uncommon. In contrast, pain, swelling, or fractures occur in more than 80% of patients with bone metastases. The pattern of lung involvement may vary from macronodular to diffuse infiltrates. The latter, when not detected by chest X ray, are usually diagnosed with <sup>131</sup>I total body scan and may be confirmed by CT; enlarged mediastinal lymph nodes are often present in

patients with papillary carcinomas, especially children. Bone metastases are osteolytic and are often difficult to visualize on X rays; bone scintigraphy may show decreased or moderately increased uptake, and bone involvement is better visualized by CT or magnetic resonance imaging. Nearly all patients with distant metastases have a high serum Tg concentration, unless the metastases are not visible on X rays, and two-thirds of patients have <sup>131</sup>I uptake in the metastases.

## Treatment

Palliative surgery is required for bone metastases when there are neurological or orthopedic complications or a high risk of such complications. Surgery may also be useful to debulk large tumor masses.

Patients with metastases that take up <sup>131</sup>I should be treated with 100 to 150 mCi (3700 to 5550 MBq) every 4 to 6 months. The effective radiation dose, which depends on the ratio between total uptake and

the mass of thyroid tissue, and outcome of  $^{131}\text{I}$  therapy are correlated. For this reason, higher doses [200 mCi (7400 MBq) or more] have been advocated in patients with bone metastases, but their effectiveness remains to be demonstrated. Lower doses [1 mCi/kg (37 MBq) body weight] are given to children. There is no limit to the cumulative dose of  $^{131}\text{I}$  that can be given to patients with distant metastases, although above a cumulative dose of 500 mCi (18,500 MBq), further  $^{131}\text{I}$  therapy has little benefit but the risk of leukemia increases significantly. External radiotherapy is given to patients with bone metastases visible on X rays. Chemotherapy is poorly effective and should be reserved for patients with progressive and nonfunctioning metastases.

### Treatment Results

Complete responses have been obtained in 45% of patients with distant metastases with  $^{131}\text{I}$  uptake, more frequently in younger patients and in those with small pulmonary metastases.

Overall survival after the discovery of distant metastases is approximately 40% at 10 years. Young patients with well-differentiated tumors that take up  $^{131}\text{I}$  and have metastases that are small when discovered have a more favorable outcome. When the tumor mass is considered, the location of the metastases, be it the lungs or bone, has no independent prognostic influence. The poor prognosis of patients with bone metastases is linked to the bulkiness of their lesions. The prognostic importance of the small metastases at their discovery has led to the administration of 100 mCi (3700 MBq) doses of  $^{131}\text{I}$  to patients with elevated serum Tg concentrations and no other evidence of disease.

### Complications of Treatment with $^{131}\text{I}$

#### **Acute Side Effects**

Acute side effects (nausea, sialadenitis) after treatment with  $^{131}\text{I}$  are common but usually mild and they resolve rapidly. Radiation thyroiditis is usually trivial, but if the thyroid remnant is large, the patient may have enough pain to warrant corticosteroid therapy for a few days. Tumors in certain locations—brain, spinal cord, and paratracheal region—may swell in response to TSH stimulation or after  $^{131}\text{I}$  therapy, causing compressive symptoms. Radiation fibrosis may develop in patients with diffuse lung metastases and eventually prove fatal if high doses [ $>150$  mCi (5550 MBq)] are administered at short intervals ( $<3$  months).

#### **Genetic Defects and Infertility**

Particular attention must be paid to avoid administration of  $^{131}\text{I}$  to pregnant women.

After  $^{131}\text{I}$  treatment, spermatogenesis may be transiently depressed and women may have transient ovarian failure. Genetic damage induced by exposure to  $^{131}\text{I}$  before conception has been a major subject of concern. However, the only anomaly that has been reported is an increased frequency of miscarriages in women treated with  $^{131}\text{I}$  during the year preceding the conception. Therefore, it is recommended that conception be postponed for 1 year after treatment with  $^{131}\text{I}$ . There is no evidence that pregnancy affects tumor growth in women receiving adequate thyroxine therapy, who should be monitored carefully (at least every 2 months) during pregnancy.

#### **Carcinogenesis and Leukemogenesis**

Mild pancytopenia may occur after repeated  $^{131}\text{I}$  therapy, especially in patients with bone metastases also treated with external radiotherapy. The overall relative risk of secondary carcinoma and of leukemia was found to be increased in patients treated with a high cumulative dose of  $^{131}\text{I}$  [ $>500$  mCi (18,500 MBq)] or in association with external radiotherapy. Because the dose–effect relationship is linear, any therapeutic dose of (30–100 mCi)  $^{131}\text{I}$  may increase this risk.

### CONCLUSION

Most patients with papillary or follicular carcinomas can be cured. However, both initial treatment and follow-up should be individualized according to prognostic indicators and thereafter to any evidence of disease in order to improve the patient's quality of life.

### ANAPLASTIC CARCINOMA

Anaplastic carcinomas represent less than 5% of thyroid carcinomas. They occur in elderly patients with a long-standing goiter. They are one of the most aggressive carcinomas in humans and prognosis can be improved by the use of combined treatment modalities.

Usually, anaplastic carcinoma manifests by a rapid increase in the size of a thyroid nodule, frequently with enlarged neck lymph nodes. Compressive symptoms are frequent. The tumor does not concentrate radioiodine and does not produce thyroglobulin. Treatment should be initiated as soon as possible and includes surgery with gross resection of all tumor masses present in the neck and followed by a

combination of chemotherapy and radiation therapy. The authors' protocol combines chemotherapy with adriamycin ( $60 \text{ mg/m}^2$ ) and cisplatin ( $120 \text{ mg/m}^2$ ) every 4 weeks for six cycles with radiation therapy ( $1.25 \text{ Gy}$  twice a day for 5 days per week for a total dose of  $40 \text{ Gy}$  to the neck and upper mediastinum) between the second and third cycles of chemotherapy.

When surgery is feasible, local control is obtained in the majority of patients and long-term survival of up to 40% of patients with no evidence of distant metastases has been reported. This protocol prevents death from local invasion and suffocation and most deaths are due to distant metastases.

### See Also the Following Articles

Goitrogens, Environmental • Iodine, Radioactive • Irradiation, Thyroid and • Medullary Thyroid Carcinoma •

Parathyroid Cancer • Parathyroid Glands, Pathology • Thyroglobulin • Thyroid, Aging and • Thyroid Disease, Epidemiology of

### Further Reading

- Mazzaferrri, E. L., and Kloos, R. T. (2001). Current approaches to primary therapy for papillary and follicular thyroid cancer. *J. Clin. Endocrinol. Metab.* **86**, 1447–1463.
- Schlumberger, M. (1998). Papillary and follicular thyroid carcinoma. *N. Engl. J. Med.* **338**, 297–306.
- Schlumberger, M. J., Filetti, S., and Hay, I. D. (2002). Nontoxic goiter and thyroid neoplasia. In "Williams Textbook of Endocrinology" (R. P. Larsen, H. M. Kronenberg, S. Melmed, and K. S. Polonsky, eds.), 10th ed., pp. 457–490. W. B. Saunders, Philadelphia, PA.
- Sherman, S. I. (2003). Thyroid carcinoma. *Lancet* **361**, 501–511.
- Thyroid Cancer Task Force. (2001). AACE/AAES medical/surgical guidelines for clinical practice: Management of thyroid carcinoma. *Endocr. Pract.* **7**, 203–220.

Table 1 Physiological Changes in Pregnancy and the Influence on Thyroid Function Tests<sup>a</sup>

Physiologic change	Thyroid function test change
Increased thyroid gland size (women in areas of insufficient iodine intake)	Increased serum thyroglobulin
Estrogen-mediated increase in serum levels of TBG	Elevated serum total T <sub>4</sub> and T <sub>3</sub> concentrations
Increased plasma volume	Increased T <sub>4</sub> and T <sub>3</sub> pool
First-trimester elevation in serum hCG	Transient reduction in serum TSH concentration and possible increase in free T <sub>4</sub>
Increased expression of type 3 5-deiodinase (D <sub>3</sub> ) in placenta and uterus	Accelerated degradation of T <sub>4</sub> and T <sub>3</sub>

<sup>a</sup>Abbreviations used: T<sub>4</sub>, thyroxine; T<sub>3</sub>, triiodothyronine; hCG, human chorionic gonadotropin; TBG, thyroxine-binding globulin.

TSH that corresponds to the peak of serum hCG in the first trimester. Generally, serum TSH remains in the normal range, although subnormal values were reported in approximately 15% of women from an area of relative iodine insufficiency. The reduction of serum TSH is likely the result of an increase in circulating free thyroid hormones at this time. The increase in free T<sub>4</sub> level is most likely mediated by direct stimulation of the thyroidal TSH receptor by hCG.

## HYPOTHYROIDISM DURING PREGNANCY

The previously described physiological factors that influence thyroid function during pregnancy—increased iodine clearance, increased serum TBG concentration, and increased D<sub>3</sub> activity in the placenta and uterus—are a “stress” on the thyroid gland. Women with normal thyroid glands in iodine-sufficient areas can compensate for these changes, but women in iodine-insufficient areas or those with underlying hypothyroidism may not be able to maintain euthyroxinemia. Although the fetal thyroid gland begins to function at 10–12 weeks of development, maternal thyroid status may influence fetal development throughout pregnancy.

### Iodine Deficiency

Iodine deficiency remains an important cause of maternal hypothyroidism in a number of areas throughout the world, including parts of central Africa and Asia. It is estimated that as many as 500 million people live in areas of iodine deficiency, despite a worldwide program to eradicate iodine deficiency. The World Health Organization recommends that pregnant women receive 200 µg of iodine per day. Since iodine clearance increases during pregnancy, pregnant women in areas of relative iodine insufficiency are at

an increased risk of iodine deficiency. In the United States, a survey conducted from 1988 to 1994 indicated that, compared to a similar study conducted from 1971 to 1974, the percentage of women of child-bearing age with low urinary iodine (<5 µg/dl) increased from 4 to 15% and that of pregnant women increased from 1 to 7%.

### Hashimoto's Disease

The most common cause of hypothyroidism during pregnancy is chronic lymphocytic thyroiditis, also referred to as Hashimoto's thyroiditis. There is a marked female predominance of Hashimoto's thyroiditis, with a female to male ratio of approximately 7:1. Hypothyroidism can also be seen as a consequence of radioiodine or surgical treatment of Graves' disease, thyroid cancer, or goiter. Transient hypothyroidism in the postpartum period is most often due to postpartum lymphocytic thyroiditis. Hypothyroidism can also occur after external radiation for cervical neoplasms and may also be caused by the medications lithium, amiodarone, or interferon.

Women may first be diagnosed with hypothyroidism during pregnancy. A survey of thyroid function studies of a sample of healthy individuals without known thyroid disease showed a high incidence of undiagnosed hypothyroidism. Thyroid tests were performed from 1988 to 1994 and showed that 4.3% of the group had undiagnosed subclinical hypothyroidism and 0.3% had undiagnosed overt hypothyroidism. Positive anti-thyroid peroxidase antibodies were found in 11.3% of the population without known thyroid disease. The incidence of abnormalities was significantly higher in women than in men. Approximately one-fourth of hypothyroid women have menstrual irregularity, with oligomenorrhea more common than hypomenorrhea and menorrhagia.

## Complications

A variety of maternal and fetal complications of pregnancy have been associated with hypothyroidism. Although severe hypothyroidism can interfere with conception, women have become pregnant even when profoundly hypothyroid. Correction of subclinical hypothyroidism with thyroxine replacement has been associated with conception in women with ovulatory defects. Although complications of pregnancy are not uniformly seen in hypothyroidism, a number of complications are observed with a higher frequency in women who are hypothyroid during pregnancy. Among women who experience a miscarriage, a high percentage have positive anti-thyroid antibodies, and some have mild or overt hypothyroidism. Although it is possible that reduced thyroid hormone levels and thyroid autoantibodies play a pathogenic role in miscarriage, it is more likely that thyroid autoantibodies are a marker for antibodies to other fetal or placental components that result in miscarriage. Other complications of pregnancy associated with hypothyroidism include pregnancy-induced hypertension and preterm delivery. In a study of 150 pregnancies complicated by overt or subclinical hypothyroidism, the incidence of miscarriage and preterm delivery was much higher in patients with incompletely treated hypothyroidism than in those receiving adequate thyroxine replacement. Term deliveries were seen in approximately 20% of women with inadequately treated hypothyroidism and approximately 95% of women receiving adequate thyroxine replacement at the time of delivery. [Table II](#) summarizes the effects of hypothyroidism on pregnancy and fertility.

## Increased Thyroxine Dose

In women known to be hypothyroid, the thyroxine dose should be titrated to a low-normal TSH prior to

conception, if possible. A number of studies suggest that maternal euthyroxinemia is important in early gestation, prior to fetal thyroid gland functioning at 10–12 weeks. Approximately 50–70% of hypothyroid women require an increase in thyroxine dose during pregnancy. Women who are completely athyreotic (as a consequence of radioiodine therapy or surgical thyroidectomy) have the least thyroid reserve and are likely to need the greatest increase in thyroxine dose. Women with some thyroid reserve will need only a small increment or no change in thyroxine dose. The average required increase in thyroxine dose is approximately 50 µg. It is also important to caution women about taking calcium and iron supplements simultaneously with thyroxine since this may interfere with thyroxine absorption. Many women may take such supplements for the first time during pregnancy. Thyroid function should be evaluated in the first trimester with subsequent monitoring every 4–6 weeks, with dose adjustments as necessary to keep the serum TSH normal. The dose can be returned to the prepregnancy dose immediately after delivery.

## Fetal Development

The influence of hypothyroidism on fetal development is controversial. When identified and started on thyroxine replacement soon after birth, infants with congenital hypothyroidism have generally been found to have normal intellectual and neurological development. Some investigators have identified subtle defects of cognitive function in treated hypothyroid children. The placenta does not usually allow thyroxine to cross between mother and fetus. In congenital hypothyroidism, however, there is evidence that in the presence of a large maternal-to-fetal gradient, thyroxine will cross. This permits sufficient thyroxine for the fetus for development. The intellectual development of the offspring of hypothyroid mothers

**Table II Hypothyroidism and Fertility/Pregnancy**

Maternal complications	Fetal complications
Ovulatory dysfunction	Low birth weight
Miscarriage (highly associated with positive anti-TPO <sup>a</sup> antibodies regardless of thyroid status)	Perinatal mortality (some studies show a small increase; others show no difference)
Pregnancy-induced hypertension	Increase in congenital anomalies (studies vary, with some showing a small increase and others no difference)
Placental abruption	Intellectual development (some studies show a small reduction in IQ at 8 years of age)
Preterm delivery	

<sup>a</sup>TPO, thyroid peroxidase.

is thought by some to be modestly reduced. One study showed a small but significant reduction in the IQ of 8-year-old children born to mothers with untreated hypothyroidism.

## HYPERTHYROIDISM DURING PREGNANCY

### Graves' Disease

Graves' disease is the most common clinically recognized cause of hyperthyroidism in pregnant women. The majority of patients have previously identified Graves' disease, although it can also develop for the first time during pregnancy. Most women with hyperthyroidism have oligomenorrhea or amenorrhea.

### Complications

Graves' disease is associated with a variety of complications during pregnancy (Table III). There can also be significant fetal complications as a consequence of antithyroid drug treatment. Preterm labor and infants small for gestational age are associated with Graves' disease. Approximately 1 in 100 infants from mothers with Graves' disease will have neonatal Graves' disease. Since this results from maternal thyroid-stimulating immunoglobulin (TSI) crossing the placenta, the fetus is at risk even if the mother has been definitively treated with radioiodine or surgery. Some advocate measuring maternal TSI levels in the serum to determine the risk of neonatal Graves' disease, although the predictive value of this measurement is not known.

### Treatment

There is controversy regarding the appropriate treatment for Graves' disease in anticipation of pregnancy,

and the recommendations vary with the severity of disease and the region of the world. Some advocate definitive treatment of Graves' disease with radioiodine prior to pregnancy. This eliminates the possibility of Graves' disease complicating pregnancy, although thyroxine dose during pregnancy must be carefully monitored and adjusted. Some endocrinologists recommend against radioiodine treatment in women of reproductive age because of fear of radiation dose to the ovaries. Although limited data are available, in follow-up of a small group of women treated as children with high-dose radioiodine for thyroid cancer, there was no increase in the incidence of congenital anomalies. In most cases of mild to moderate Graves' disease in women of reproductive age, treatment with antithyroid drugs is used.

The primary treatment for Graves' disease during pregnancy is antithyroid drugs. Most authorities recommend a treatment goal of keeping the free T<sub>4</sub> level at the upper normal to slightly elevated level and not attempting to normalize the serum TSH level. Studies that have correlated fetal thyroid hormone levels with maternal levels indicate that when the mother is treated to euthyroidism, there is a much greater risk of fetal hypothyroxinemia. The total dose of antithyroid drug should be limited to restrict transplacental passage. The choice of drug is controversial. A limited number of studies suggest that propylthiouracil (PTU) has less transplacental passage and is found in the breast milk to a lesser extent than methimazole. Methimazole therapy has also been associated with aplasia cutis, a rare congenital defect manifest as loss of cutaneous structures in the crown of the head. In Europe, the most popular drug is carbimazole, which is metabolized to methimazole and has not been associated with aplasia cutis. There is no compelling evidence to designate any single drug as superior in pregnancy, but most endocrinologists in the United States use PTU.

In general, Graves' disease becomes less active during pregnancy as general immunosuppression occurs in the later stages. The dose of medication required for treatment is often considerably less in the second and third trimesters. Postpartum, however, there can be a flare of Graves' disease, as seen with many other autoimmune diseases. Some women may have their first manifestation of Graves' disease in the postpartum period. This must be differentiated from postpartum thyroiditis. Although radioiodine uptake distinguishes between these conditions, it cannot be easily measured in women while they are nursing. TSH receptor antibodies can be tested to support a diagnosis of Graves' disease. Sequential thyroid studies show spontaneous improvement or transition to a hypothyroid state

Table III Hyperthyroidism and Pregnancy

Maternal complications	Fetal complications
Preeclampsia	Low birth weight
Preterm labor	Increased perinatal mortality
	Neonatal hypothyroidism (due to excessive treatment of mother with antithyroid drugs)
	Neonatal hyperthyroidism (due to transplacental passage of maternal thyroid-stimulating immunoglobulin)

in those with thyroiditis, whereas those with Graves' disease have persistent hyperthyroidism.

### HUMAN CHORIONIC GONADOTROPIN-ASSOCIATED HYPERTHYROIDISM

There is an exponential increase in serum hCG during the first trimester of pregnancy, with peak levels at 10–12 weeks of pregnancy. *In vitro* studies using cells transfected with the human TSH receptor demonstrate that hCG binds the TSH receptor and activates adenylate cyclase. This is likely responsible for the reduction in serum TSH seen in most women in the first trimester of pregnancy. There are also a number of clinical manifestations of excess hCG (Table IV).

### Hyperemesis Gravidarum

Hyperemesis gravidarum is characterized by prolonged and severe nausea and vomiting in early pregnancy and is associated with a loss of 5% body weight, dehydration, and ketosis. It occurs in approximately 1–1.5% of pregnancies, is more prevalent in Asian women than in Caucasians, and is more common with multiple gestations. Abnormalities of serum chemistries include hyponatremia, hypokalemia, hypochloremic alkalosis, and abnormalities of liver function. There is a positive correlation between the severity of vomiting and serum hCG level.

The majority of hyperemesis patients have a suppressed serum TSH level and increased free thyroxine concentration. Free thyroxine and TSH levels normalize when the hyperemesis resolves in the second trimester. In general, these patients do not have clinical features of hyperthyroidism or goiter. The hyperthyroidism is most likely due to a less sialylated hCG, which stimulates the TSH receptor to a greater extent, although it has a reduced serum half-life.

In most patients, the increased thyroid function of hyperemesis gravidarum is self-limited and subsides with the disappearance of vomiting. In a small percentage of patients, however, there is clear clinical evidence of hyperthyroidism; this has been termed transient hyperthyroidism of hyperemesis gravidarum or gestational thyrotoxicosis. In these situations, the diagnosis of Graves' disease should be excluded. Treatment with antithyroid drugs does not influence the course of hyperemesis gravidarum, even when associated with elevated thyroid function tests. If they are used, these drugs should be discontinued as soon as thyroid function tests return to normal and the vomiting subsides.

### Trophoblastic Tumors

Clinical thyrotoxicosis has been associated with excessive hCG produced by trophoblastic tumors, hydatidiform mole, and choriocarcinoma. The hCG produced by these tumors is used as a marker for the diagnosis and management of these patients. The serum concentration of hCG in affected patients may be severalfold higher than the peak levels of normal pregnancy. Molar pregnancies are estimated to occur in approximately 1 in 1500 pregnancies in the United States and 1 in 1000 pregnancies in the United Kingdom, and they are significantly more common in Asian and Latin America populations.

Molar pregnancies are usually diagnosed before 20 weeks of gestation because of abnormal vaginal bleeding. Toxemia and hyperemesis gravidarum may also occur in molar pregnancy with greater frequency than in normal pregnancy. Thyroid function abnormality ranges from subclinical hyperthyroidism to increased free thyroxine levels with minimal clinical features of hyperthyroidism and even to severe thyrotoxicosis causing atrial fibrillation and congestive heart failure. In general, the manifestations of hyperthyroidism probably depend on the severity and duration of

Table IV Spectrum of hCG Action in Pregnancy<sup>a</sup>

Nature of hCG elevation	Clinical consequences and changes in thyroid function
Physiologic elevation in first trimester	Mild elevation of serum free T <sub>4</sub> , mild reduction in serum TSH concentration
Exaggerated first-trimester hCG elevation (commonly seen with multiple gestations)	Greater increase in serum free T <sub>4</sub> and greater reduction in serum TSH concentration; associated with hyperemesis gravidarum
Pathologic elevation from hydatidiform mole or choriocarcinoma	Marked elevation in serum free T <sub>4</sub> and free T <sub>3</sub> concentrations; suppression in serum TSH; clinical thyrotoxicosis observed in some patients
Increased "sensitivity" to hCG due to a TSH receptor mutation	Gestational thyrotoxicosis

<sup>a</sup>Abbreviations used: hCG, human chorionic gonadotropin; TSH, thyroid-stimulating hormone.

the increased thyroid hormone levels, which are in turn proportional to the hCG level, and the hCG level is proportional to the mass of the tumor.

### TSH Receptor Mutation Sensitive to hCG

A very rare but informative cause of gestational thyrotoxicosis was recently reported. A woman and her mother both had a similar history of hyperemesis and hyperthyroidism during pregnancy. The TSH receptor had an A-to-G heterozygous mutation at codon 183 in exon 7, resulting in substitution of arginine for lysine in the middle portion of the extracellular domain of the TSH receptor. *In vitro* studies indicated that cells transfected with the mutant TSH receptor were much more sensitive to hCG than those expressing the wild-type TSH receptor. The hyperthyroidism during pregnancy in this family is most likely due to activation of this mutant TSH receptor by hCG.

### See Also the Following Articles

Graves' Disease • Hashimoto's Disease • Iodine Deficiency • Irradiation, Thyroid and • Pregnancy Endocrinology • Thyroid Disease, Epidemiology of • Thyroid Disease, Genetic Factors in • Thyroid Disorders in the Elderly

### Further Reading

- Abalovich, S., Gutierrez, S., Alcaraz, G., Maccallini, G., and Garcia, A. (2002). Overt and subclinical hypothyroidism complicating pregnancy. *Thyroid* **12**, 63–68.
- Brent, G. A. (1999). Maternal hypothyroidism: Recognition and management. *Thyroid* **9**, 661–665.
- Glinoe, D. (1997). The regulation of thyroid function in pregnancy: Pathways of endocrine adaptation from physiology to pathology. *Endocr. Rev.* **18**, 404–433.
- Goodwin, T. M., Montoro, M., Mestman, J., Pekary, A. E., and Hershman, J. M. (1992). The role of chorionic gonadotropin in transient hyperthyroidism of hyperemesis gravidarum. *J. Clin. Endocrinol. Metab.* **75**, 1333–1337.
- Haddow, J. E., Palomaki, G. E., Allan, W. C., *et al.* (1999). Maternal thyroid deficiency during pregnancy and subsequent neuropsychological development of the child. *N. Engl. J. Med.* **341**, 549–555.
- Hallowell, J. G., Staehling, N. W., Flanders, W. D., *et al.* (2002). Serum TSH, T4, and thyroid antibodies in the United States population (1988 to 1994): National Health and Nutrition Examination Survey (NHANES III). *J. Clin. Endocrinol. Metab.* **87**, 489–499.
- Mandel, S. J., Larsen, P. R., Seely, E. W., and Brent, G. A. (1990). Increased need for thyroxine during pregnancy in women with primary hypothyroidism. *N. Engl. J. Med.* **323**, 91–95.
- Mandel, S. J., Brent, G. A., and Larsen, P. R. (1994). Review of antithyroid drug use during pregnancy and report of a case of aplasia cutis. *Thyroid* **4**, 129–133.
- Rodien, P., Bremont, C., Raffin Sanson, M. D., *et al.* (1998). A mutant thyrotropin receptor, hypersensitive to human chorionic gonadotropin, responsible for familial gestational thyrotoxicosis. *N. Engl. J. Med.* **339**, 1823–1826.
- Yoshimura, M., and Hershman, J. M. (1995). Thyrotropic action of human chorionic gonadotropin. *Thyroid* **5**, 425–434.

treatable cause of mental retardation. There is an inverse relationship between age at diagnosis and IQ in later life. In iodine-replete areas, 85% of the cases are due to sporadic developmental defects of the thyroid gland (thyroid dysgenesis), such as the arrested migration of the embryonic thyroid (ectopic thyroid) or a complete absence of thyroid tissue (athyreosis). The remaining 15% have thyroid dysmorphogenesis defects transmitted by an autosomal recessive mode of inheritance. Iodine deficiency ( $<25\mu\text{g}/\text{day}$ ), particularly in preterm infants, accounts for many cases in Europe, Asia, and Africa. Clinical diagnosis occurs in less than 5% of newborns with hypothyroidism because symptoms and signs are often minimal. As a result, it is not possible to predict which infants are likely to be affected. Without prompt diagnosis and treatment, most affected children gradually develop growth failure, irreversible mental retardation, and a variety of neuropsychological deficits.

### Spontaneous Hypothyroidism

A cross-sectional study of 2779 subjects of the community of Whickham, a mixed urban and rural area in northeast England, in the 1970s documented the prevalence of thyroid disorders. Other cross-sectional studies from Europe, the United States, and Japan have corroborated the findings of this survey. Therefore, its findings have a broad application to iodine-replete communities. The prevalence of spontaneous hypothyroidism is between 1 and 2%, it is more common in older women, and it is 10 times more common in women than in men. In an iodine-deficient community of Pescopagno, Italy, the prevalence of newly diagnosed overt hypothyroidism was 0.3% of 573 women (autoimmune thyroiditis confirmed as etiology) and 0% of 419 men. The testing of hospital inpatients, predominantly elderly women, might reveal a higher proportion of unsuspected hypothyroidism, but this is not supported by the available studies, which confirm a prevalence of 2%.

### Subclinical Hypothyroidism

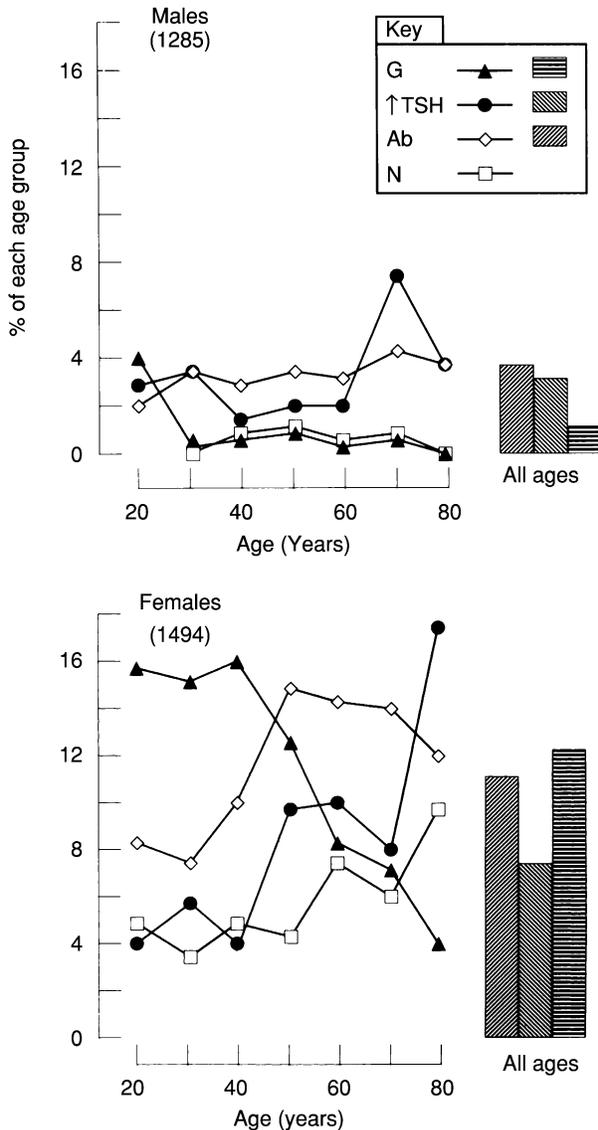
Subclinical hypothyroidism represents a compensated state in which increased thyrotropin (TSH) output is required to maintain normal circulating thyroid hormone levels. An elevated serum TSH is a sensitive indicator of some degree of thyroid failure and there is a clear inverse relationship with free thyroxine ( $\text{FT}_4$ ) levels.

A substantial proportion of the population, particularly elderly women who live in iodine-replete areas, have circulating thyroid antibodies [antithyroid peroxidase (microsomal) and anti-thyroglobulin antibodies] and normal thyroid function. The presence of these antibodies correlates with the presence of focal thyroiditis in biopsy and in postmortem material of patients with no evidence of hypothyroidism during life. Patients with hypothyroidism caused by either atrophic or goitrous autoimmune thyroiditis usually have high serum titers of the same antibodies. Serum TSH concentrations do not change as a function of age among adult men, but in women older than age 40 the concentrations increase markedly (Fig. 1). If women with anti-thyroid antibodies are excluded, there is no age-related increase. Nearly all older women with elevated serum TSH values have subclinical hypothyroidism. With respect to epidemiological studies, the definition of subclinical hypothyroidism varies from any increase in serum TSH to values  $>10\text{mU}/\text{liter}$  or, more stringently, a serum TSH value  $>10\text{mU}/\text{liter}$  and a positive test for circulating thyroid antibodies in serum.

In the Whickham survey, 8% of women (10% of women older than age 55) and 3% of men had subclinical hypothyroidism. A cross-sectional screening survey of 25,682 subjects older than age 18 attending a health fair in Colorado found that 9% of the population, excluding 1525 on thyroxine therapy, had an elevated serum TSH level; of these,  $<1\%$  had overt hypothyroidism. Community studies of elderly persons have confirmed the high prevalence in this age group, with approximately 10% of subjects older than age 60 having serum TSH values higher than the normal range. In iodine-deficient Pescopagno, the prevalence of subclinical hypothyroidism was slightly lower, 4% of women and 3% of men, but anti-thyroid antibodies were as prevalent, although at lower titers, as in iodine-replete communities. Subclinical hypothyroidism is found at higher frequency (18% in Iceland and 24% in Hungary) in areas where iodine intake is high, but most cases are not of autoimmune origin.

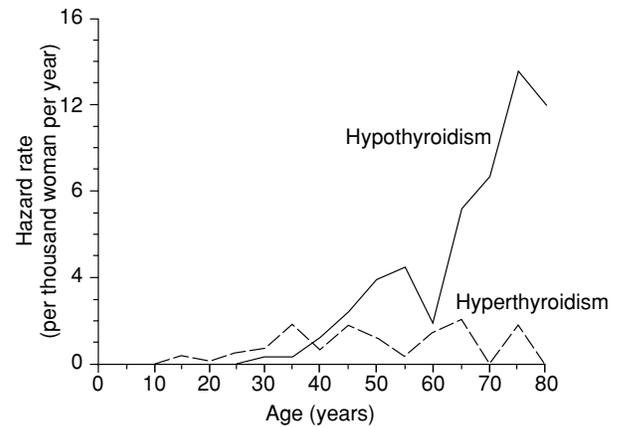
### Incidence of Hypothyroidism

In the 20-year follow-up of the Whickham cohort, the mean annual incidence of spontaneous hypothyroidism in surviving women was 3.5 per 1000, increasing to 4.1 per 1000 if all cases are included, including those who had received destructive treatment for hyperthyroidism. The probability of a



**Figure 1** Age and sex distribution of thyroid cytoplasmic antibodies (Ab), raised serum TSH greater than 6 mU/liter ( $\uparrow$ TSH), visible goiter (G), and nodules (N) in the Whickham survey. From Tunbridge *et al.* (1977). The spectrum of thyroid disease in the community: The Whickham Survey. *Clin. Endocrinol.* 7, 485. Reproduced with permission from Blackwell Science Ltd.

woman developing hypothyroidism at a particular time (i.e., the hazard rate) increased with age to 14 per 1000 in women aged 75–80 (Fig. 2). The mean annual incidence in men (all spontaneous except for one case of lithium-induced hypothyroidism) was 0.6 per 1000. Either raised serum TSH or positive thyroid antibodies alone or in combination were associated with a significantly increased risk of hypothyroidism in women and men. The odds were greatly increased

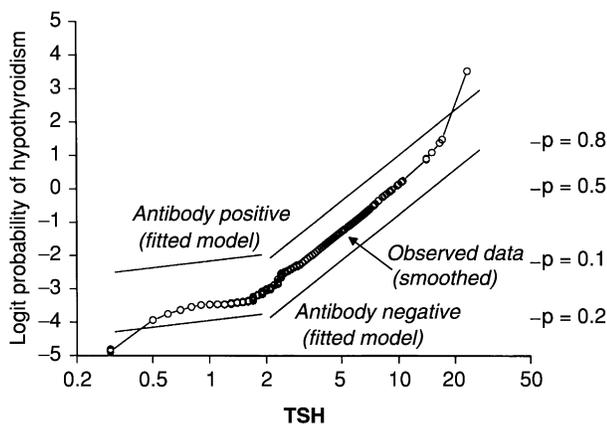


**Figure 2** Age-specific hazard rates for the development of overt hyperthyroidism and hypothyroidism in women at 20-year follow-up of the Whickham Survey. From Vanderpump *et al.* (1995). The incidence of thyroid disorders in the community: A twenty-year follow-up of the Whickham Survey. *Clin. Endocrinol.* 43, 60. Reproduced with permission from Blackwell Science Ltd.

when both risk factors were present and each had a similar effect. In the surviving women, the annual risk of developing spontaneous hypothyroidism was 4% in those who had both raised serum TSH levels and were thyroid-antibody positive, 3% if only serum TSH was raised, and 2% if only thyroid antibodies were positive. The cumulative incidence of hypothyroidism over 20 years was 55, 33, and 27%, respectively. The probability of developing hypothyroidism in women increases linearly when serum TSH is higher than 2 mU/liter; this is further increased if women are anti-thyroid antibody positive or decreased if anti-thyroid antibody negative (Fig. 3). The development of hypothyroidism also correlated with the strength of titer of anti-thyroid microsomal antibodies at first survey. A positive family history of any form of thyroid disease, the presence of a goiter at either the first or the follow-up survey, or parity at first survey were not associated with an increased risk of hypothyroidism. The prognostic importance of positive thyroid antibody tests and increasing serum TSH levels has been confirmed in other cohort studies.

### Iatrogenic Hypothyroidism

After destructive treatment for hyperthyroidism (either radioiodine therapy or surgery), the incidence of overt hypothyroidism is greatest in the first year. Hypothyroidism occurs 2–4 weeks after total thyroidectomy, but the time course after subtotal thyroidectomy in Graves' hyperthyroidism is more



**Figure 3** Probability for development within 20 years of hypothyroidism with increasing values of serum TSH at first Whickham Survey in 912 surviving women. From Vanderpump *et al.* (1995). The incidence of thyroid disorders in the community: A twenty-year follow-up of the Whickham Survey. *Clin. Endocrinol.* **43**, 63. Reproduced with permission from Blackwell Science Ltd.

variable. Factors that influence outcome are age, gland size (small glands are more likely to lead to hypothyroidism postoperatively), remnant size, and iodine intake (high intake is associated with recurrence). Radioiodine therapy for Graves' hyperthyroidism, toxic or nontoxic nodular goiter, or autonomously functioning thyroid adenomas can cause hypothyroidism months or years later. A fixed high dose of radioiodine (550 MBq), administered with the intent of ablating the thyroid and inducing early hypothyroidism, results in permanent hypothyroidism in at least 50% of patients by 1 year. Subclinical hypothyroidism is commonly found either postradioiodine therapy or postsurgery in up to 50% of apparently euthyroid patients. It may be evident for only a few months, but more often it represents a stage in the progression toward overt thyroid failure. If serum TSH remains raised, then the rate of progression toward overt hypothyroidism is 2–6% per year after either treatment. Treatment of Graves' disease with antithyroid drugs alone is also associated with the eventual development of hypothyroidism from either autoimmune thyroiditis or the presence of TSH-blocking antibodies in 5–20% of cases.

Other iatrogenic causes of hypothyroidism are surgery and external irradiation for head and neck cancer and drugs used to treat nonthyroid conditions, including amiodarone, lithium carbonate, interferon- $\alpha$ , and interleukin-2. In addition, patients with hypothyroidism who are taking thyroxine may become hypothyroid if given drugs that decrease thyroxine absorption (e.g., cholestyramine and iron salts) or drugs that

increase its clearance (e.g., phenytoin and carbamazepine or an estrogen-induced increase in the serum concentration of thyroxine-binding globulin). Poor compliance with thyroxine therapy or suboptimal treatment may also result in hypothyroidism.

## HYPERTHYROIDISM

### Prevalence

The prevalence of hyperthyroidism in women is between 0.5 and 2% and is 10 times more common in women than in men in iodine-replete communities. The prevalence data on elderly persons indicate a range from 0.4 to 0.9%. The prevalence of past and present hyperthyroidism in iodine-deficient Pescopagno was high at 3% (approximately one-third of these patients had a diffuse goiter), and the frequency in men and women was similar.

Hospital inpatients and even outpatients are selected populations. Isolated alterations in serum TSH concentrations [either slightly low (0.1–0.5 mU/liter) or high (5–20 mU/liter)] occur in approximately 15% of such patients due to the lability of TSH secretion in response to nonthyroidal illness or drugs. Approximately 2 or 3% of hospitalized patients have serum TSH concentrations that are suppressed (<0.1 mU/liter) or elevated (>20 mU/liter), but less than half have an underlying thyroid disorder. The reported prevalence rates for previously undiagnosed hyperthyroidism, between 0.3 and 1%, are consistent with community surveys. An accurate diagnosis can be achieved if clinical indications for measuring thyroid function exist, an accurate drug history is taken, the abnormal serum TSH is subsequently confirmed, and there is a reciprocal change in serum FT<sub>4</sub>.

### Subclinical Hyperthyroidism

More than 10 years ago, the introduction of assays for serum TSH that were sensitive enough to distinguish between normal and low concentrations allowed subjects with subclinical hyperthyroidism to be identified. This is defined as a low serum TSH concentration in the presence of normal serum T<sub>4</sub> concentrations and the absence of hypothalamic or pituitary disease, nonthyroidal illness, or ingestion of drugs that inhibit TSH secretion, such as glucocorticoids or dopamine.

In cross-sectional studies, 2–4% of subjects have below normal serum TSH (detection limit, 0.05 mU/liter), of whom half are on thyroxine therapy. Those with low but detectable levels may recover spontaneously when retested. In studies using more sensitive

serum TSH assays (detection limit, 0.01 mU/liter), approximately 2% of subjects have subnormal serum TSH and 1% have an undetectable serum TSH. The Colorado study, which screened more than 25,000 healthy volunteers with a similarly sensitive serum TSH assay, also found 2% of the population to have a subnormal serum TSH, with more than half on thyroxine therapy. In short follow-up studies, the incidence of hyperthyroidism has been calculated at 5% per year.

The prevalence of subnormal serum TSH (detection limit, 0.01 mU/liter and excluding those on thyroxine therapy) in Pescopagano was significantly higher (6%) due to functional autonomy from nodular goiters. In iodine-deficient Jutland, 10% of a random sample of 423 subjects aged 68 had a below normal serum TSH compared to 1% of a similar sample of 100 subjects in iodine-rich Iceland. Subclinical hyperthyroidism was not detected in a sample of elderly nursing home residents in an iodine-rich region of Hungary.

### Incidence of Thyrotoxicosis

In the 20-year follow-up of the Whickham cohort, the mean annual incidence of hyperthyroidism in women was 0.8 per 1000. No new cases were detected in men. The estimated probability of developing hyperthyroidism in women at a particular time (the hazard rate) averaged 1.4 per 1000 between the ages of 35 and 60 (Fig. 2). Other cohort studies provide comparable incidence data. The incidence data for overt hyperthyroidism in men and women from large population studies are comparable at 0.4 per 1000 women and 0.1 per 1000 men per year, although the age-specific incidence varies considerably. The peak age-specific incidence of Graves' disease was between 20 and 49 years in two studies but increased with age in Iceland and peaked at 60–69 years in Malmö, Sweden. The only available data for a black population from Johannesburg suggest a 10-fold lower annual incidence of hyperthyroidism in black Africans than in European whites: 0.09 per 1000 women and 0.007 per 1000 men.

### POSTPARTUM THYROIDITIS

Approximately 10% of women entering antenatal clinics at 16 weeks of gestation are anti-thyroid peroxidase antibody (anti-thyroid microsomal antibody) positive, which is comparable to the prevalence of thyroid antibodies in community surveys. Postpartum thyroiditis (PPT) is a transient, destructive autoimmune thyroiditis that occurs between weeks 12 and 16

postpartum in 40–50% of women. The presence of anti-thyroid microsomal antibodies early in pregnancy increases the risk of developing PPT. It presents as a temporary, usually painless, episode of hypothyroidism, occasionally preceded by a short episode of hyperthyroidism.

A proportion of women progress to permanent hypothyroidism following an episode of PPT, particularly those with high antibody titers. In one study, 17% of Japanese patients developed hypothyroidism during a 5- to 16-year follow-up period (mean, 9 years), whereas in Wales, approximately one-fourth of women with PPT became hypothyroid within 5 years after delivery. It is not clear whether pregnancy alters the final incidence of autoimmune thyroid disease or merely quickens the development of thyroid disease.

### DIFFUSE AND NODULAR GOITER

The most common thyroid disease is simple (diffuse) goiter. The clinical grading of thyroid size is subjective and imprecise. The World Health Organization (WHO) grading system recognizes that an enlarged thyroid gland may be palpably but not visibly enlarged. Examiner variation is greatest in deciding whether a thyroid that is palpable but not visible is normal (WHO stage O-A) or enlarged (WHO stage O-B). Interexaminer variation may also lead to differences in classifying goiter as diffuse or multinodular. There is also considerable overlap between the five WHO grades based on clinical criteria and thyroid volume estimated by ultrasonography. Ultrasonography has been used in epidemiological studies to assess thyroid size, resulting in much higher estimates of goiter prevalence than in studies in which goiter size was assessed by physical examination.

Most studies define a thyroid that is visible and palpable as a goiter (WHO grade 1 or higher). Considerable regional variations exist, even in nonendemic goiter areas. In cross-sectional surveys, the prevalence of diffuse goiter declines with age, the greatest prevalence is in premenopausal women, and the ratio of women to men is at least 4:1. In the Whickham survey, 16% of the cohort had small but easily palpable diffuse or multinodular goiters. In men, the prevalence of goiter declined with age from 7% in those younger than age 25 to 4% in those age 65–74. No goiters were detected in men older than age 75. Among the women, 26% had a goiter; the frequency ranged from 31% in those younger than age 45 (mostly diffuse) to 12% in those older than age 75 (who had a higher proportion of nodular

goiter) (Fig. 1). Longitudinal studies confirm the decreasing frequency of diffuse goiter with age.

This decline in frequency of diffuse goiters with age is in contrast to the increase in frequency of thyroid nodules and anti-thyroid antibodies with age (Fig. 1). A nodular goiter may be solitary or part of a multinodular goiter. In the Whickham survey, fewer than 1% of the men but 5% of the women had thyroid nodules detected clinically, and the frequency increased to 9% in women older than age 75. In 5234 subjects older than age 60 who were studied for a 5-year period in Framingham, Massachusetts, the prevalence of single thyroid nodules was 3% and that of multinodular goiter nodules was 1%. A higher prevalence of nodular goiter is found in areas of iodine deficiency in Europe, such as Italy, Germany, and Denmark. The only available longitudinal data suggest an annual incidence for nodules of 1 per 1000 and that, once formed, they tend to remain present and benign for a long period of time. Ultrasonography can detect thyroid nodules in up to one-third of subjects screened and is thus too sensitive as a screening tool.

## THYROID CANCER

The clinical presentation of thyroid cancer is usually as a solitary thyroid nodule or increasing goiter size. Although thyroid nodules are common, thyroid cancers are rare. The four major histological types are papillary, follicular, medullary, and anaplastic, and each displays a different epidemiology. The annual incidence of all thyroid cancers ranges between 1 and 10 per 100,000 population in most countries and is two to four times more frequent in women than men.

Papillary and follicular tumors, which comprise 60–90% of the total, are rare in children and adolescents, but their incidence increases with age in adults. Papillary thyroid carcinoma (PTC) is the most common thyroid malignancy and worldwide constitutes 50–90% of differentiated follicular cell-derived thyroid cancers. Papillary thyroid microcarcinomas (diameter,  $\leq 1$  cm) are found in 4–36% of adults post-mortem in population-based studies. The reported increase in the incidence of these carcinomas in recent years may be attributed to an improvement in pathological techniques. Most diagnoses of PTC occur in patients 30–50 years old (median age, 44 years), and the majority (60–80%) occur in women. Follicular thyroid cancer occurs relatively infrequently compared to papillary cancer and accounts for approximately 15% of all thyroid cancers. In contrast, there is an increased frequency of follicular to

papillary carcinoma (5:1) in iodine-deficient endemic goiter areas. It tends to be a malignancy of older persons, with a mean age of 50 years in most studies. External radiation exposure, particularly in childhood, is a major risk factor for papillary cancer. Four years after the nuclear accident at Chernobyl, a dramatic increase in the incidence of childhood thyroid cancer (almost exclusively papillary tumors) was recorded in the regions most exposed. There is no documented association between radioiodine therapy for thyrotoxicosis and subsequent development of thyroid cancer in adults.

Medullary thyroid cancer (MTC) occurs in both sporadic and hereditary forms. The highest incidence of sporadic disease occurs in the fifth decade. Hereditary MTC can be inherited as an autosomal dominant trait with a high degree of penetrance associated with multiple endocrine neoplasia type 2 syndrome or as familial MTC without any other endocrinopathies. It can be diagnosed before clinical presentation by genetic and biochemical screening. Anaplastic thyroid cancer is very rare and is more frequent in populations with endemic goiter. Thyroid lymphoma is also uncommon, constituting approximately 2% of extranodal lymphomas and occurring predominantly in older women. Up to one-third of patients have a history of goiter, whereas some have established autoimmune thyroiditis and may be taking thyroxine therapy.

## SCREENING FOR THYROID DISEASE

In the 1970s, screening programs for congenital hypothyroidism were developed in which TSH was measured in heel-prick blood specimens to detect this condition as early as possible. Individuals with atrial fibrillation or hyperlipidemia should have an assessment of thyroid function at least once to detect hypothyroidism. There is a high frequency of asymptomatic thyroid dysfunction in unselected patients with diabetes, and it has been calculated that performing a test of thyroid function in the annual review of diabetic complications is cost-effective. The threefold increase in the prevalence of thyroid antibodies in patients with breast cancer suggests that it may be worth screening this group for thyroid dysfunction. There is no consensus on whether healthy women should be screened for PPT. However, because women with type 1 diabetes are three times more likely to develop postpartum thyroid dysfunction, it is recommended that all such diabetic women should be tested in the first trimester for thyroid antibodies.

Any woman with a history of PPT should be offered annual surveillance of thyroid function due to the possible long-term risk of permanent hypothyroidism. Because of the high prevalence of hypothyroidism in people with Down's syndrome and Turner's syndrome, an annual check of thyroid function is recommended. Thyroid function tests are indicated every 6 months for those receiving amiodarone and lithium and every 12 months following head and neck irradiation. Following destructive treatment for thyrotoxicosis by either radioiodine or surgery, indefinite surveillance for the development of hypothyroidism is required. Thyroid function should be assessed approximately 4–8 weeks post-treatment and then every 3 months up to 1 year and annually thereafter.

After hypothyroidism has been diagnosed and the appropriate dose of thyroxine has been established, the dose remains constant for most patients. During pregnancy, there may be a need to increase the dose by at least 50  $\mu\text{g}$  daily to maintain a normal serum TSH, which should be measured each trimester. In order to improve compliance and account for possible variations in dosage caused by concomitant drug treatment, serum TSH should be measured annually. Access to a computerized thyroid disease register improves the surveillance of patients treated for thyroid dysfunction.

Controversy exists regarding whether healthy adults would benefit from screening for autoimmune thyroid disease. The prevalence of unsuspected overt thyroid disease is low, but a significant number of subjects tested have evidence of mild thyroid failure or excess. Is subclinical thyroid disease of sufficient clinical importance to warrant screening, and, once detected and confirmed, can therapy be justified?

In the absence of the confounding effects of non-thyroidal illness or drugs, a normal serum TSH concentration has a high predictive value in ruling out thyroid disease in the healthy ambulant subject. In unselected populations, serum TSH has a sensitivity of 89–95% and specificity of 90–96% for overt thyroid dysfunction compared to cases confirmed by history, examination, and additional testing. Normal concentrations of serum TSH may be recorded in hypothyroidism secondary to pituitary or hypothalamic disease, but both these situations are rare. In nearly all populations screened, a serum TSH higher than 5 mU/liter is accepted as being unequivocally increased. The standard follow-up investigation for subjects with a serum TSH higher than 5 mU/liter is the measurement of serum FT<sub>4</sub>. However, the measurement of thyroid peroxidase antibodies in subjects with a borderline raised serum TSH may be

justified, and they should be offered long-term follow-up evaluation as part of the screening program with repeat serum TSH measurements. The ideal follow-up interval has not been established, but epidemiological studies suggest that retesting every 3 years is probably adequate, providing a mechanism for recall is established. The potential risks of lifelong thyroxine therapy are precipitating symptoms of ischemic heart disease or the development of osteoporosis, but the actual risks of properly monitored thyroxine therapy to maintain a normal serum TSH are almost certainly negligible.

Treatment of overt hypothyroidism with thyroxine is cheap and effective. There is evidence that in subjects with subclinical hypothyroidism, nonspecific symptoms and psychometric scores can be improved by thyroxine with a retrospective awareness of disability and that thyroxine may have beneficial effects on cardiovascular function and lipoprotein metabolism. No high-quality randomized, controlled trials of a complete screening program for subclinical hypothyroidism exist. A cost-utility analysis using a computer decision model assessed the consequences and costs of including serum TSH screening with cholesterol screening. It concluded that testing women age 35 or older with repeat serum TSH every 5 years for 50 years would be beneficial. More than half of the presumed benefit in quality-adjusted life-years was accounted for by preventing progression to overt hypothyroidism, 30% by improving associated mild symptoms, and 2% by preventing cardiovascular disease. This may be an overestimate because the 20-year follow-up of the Wickham cohort found no association with cardiovascular outcomes. Another assumption of the model was that only subjects with increased serum TSH levels and positive thyroid antibodies are at risk of progression to overt hypothyroidism. In the computer model, the cost of detecting subclinical hypothyroidism was \$9223 for women and \$22,595 for men per quality-adjusted life-year, but this cost was heavily dependent on the cost of the TSH assay kit (\$10–50). The cost-effectiveness of screening for mild thyroid failure was comparable with that of other commonly performed preventive and therapeutic health practices, such as hypertension, exercise, breast cancer screening, and estrogen replacement therapy, in women of the same age group, providing a similar increase in quality-adjusted years. The cost-benefit analysis did not justify the additional cost of measuring FT<sub>4</sub> with a serum TSH in the screening process or allow for the extra costs of detecting, investigating, and potentially treating subclinical hyperthyroidism. Few subjects screened will have overt

hyperthyroidism, but the consequences of finding a suppressed serum TSH have to be addressed even if it is not worth screening for hyperthyroidism. In addition to the possible risk of subsequent development of overt hyperthyroidism, other possible long-term complications include atrial fibrillation and osteoporosis. There is no consensus regarding the treatment of subclinical hyperthyroidism, even though it has been strongly argued that therapy with antithyroid drugs or radioiodine may be indicated. Any potential benefits of therapy in subclinical hyperthyroidism must be weighed against the significant morbidity associated with the treatment of hyperthyroidism.

From the available evidence, the following recommendations may be justified for an iodine-replete community:

Screening for thyroid dysfunction in women younger than age 50 and in men is not warranted in view of the relatively low point prevalence of unsuspected overt thyroid dysfunction.

Case-finding in women during menopause or during visits to a primary care physician with nonspecific symptoms is justified due to the high prevalence of subclinical hypothyroidism.

If increased serum TSH is found at screening, then measurement should be repeated 2 months later together with FT<sub>4</sub> measurement after excluding non-thyroidal illness, drugs, etc.

Treatment with thyroxine is recommended if the serum TSH is  $\geq 10$  mU/liter, irrespective of whether FT<sub>4</sub> is low.

Subjects with a serum TSH between 5 and 10 mU/liter and normal FT<sub>4</sub> are at increased risk of developing hypothyroidism, and repeat measurement of serum TSH is warranted at least every 3 years if not annually.

If a suppressed serum TSH is found at screening, it should be remeasured 2 months later, and if it is still suppressed, FT<sub>3</sub> should be measured.

After thyroxine replacement is initiated, for whatever indication, long-term follow-up with at least an annual measurement of serum TSH is required.

## See Also the Following Articles

Antithyroid Drugs • Hyperthyroidism, Subclinical • Hypothyroidism, Congenital • Hypothyroidism, Subclinical • Iodine Deficiency • Thyroid Autoimmunity • Thyroid Carcinoma • Thyroid Disease and Pregnancy • Thyroid Disease, Genetic Factors in • Thyroid Disorders in the Elderly • Thyrotoxicosis, Overview of Causes

## Further Reading

- Dunn, J. T. (1997). Adverse effects of iodine deficiency and its eradication by iodine supplementation. In "Contemporary Endocrinology: Diseases of the Thyroid" (L. E. Braverman, ed.), pp. 349–360. Humana Press, Totowa, NJ.
- Grebe, S. K. G., and Hay, I. D. (1995). Follicular thyroid cancer. *Endocrinol. Metab. Clin. North Am.* **24**(4), 761–801.
- Laurberg, P., Pedersen, I. B., Knudsen, N., Ovesen, L., and Andersen, S. (2001). Environmental iodine intake affects the type of nonmalignant thyroid disease. *Thyroid* **11**, 457–469.
- Lazarus, J. H., Hall, R., Othman, S., Parkes, A. B., Richards, C. J., McCulloch, B., and Harris, B. (1996). The clinical spectrum of postpartum thyroid disease. *Q. J. Med.* **89**, 429–435.
- Powe, N. R., Danese, M. D., and Ladenson, P. W. (1997). Decision analysis in endocrinology and metabolism. *Endocrinol. Metab. Clin. North Am.* **26**, 89–111.
- Spencer, C. A. (1988). Clinical utility and cost-effectiveness of sensitive thyrotropin assays in ambulatory and hospitalized patients. *Mayo Clinic Proc.* **63**, 1214–1222.
- Tunbridge, W. M. G., and Vanderpump, M. P. J. (2000). Population screening for autoimmune thyroid disease. *Endocrinol. Metab. Clin. North Am.* **29**, 239–252.
- Tunbridge, W. M. G., Evered, D. C., Hall, R., Appleton, D., Brewis, M., Clark, F., Grimley Evans, J., Young, E., Bird, T., and Smith, P. A. (1977). The spectrum of thyroid disease in the community: The Wickham Survey. *Clin. Endocrinol.* **7**, 481–493.
- Vanderpump, M. P. J., and Tunbridge, W. M. G. (2000). The epidemiology of thyroid diseases. In "Werner and Ingbar's The Thyroid: A Fundamental and Clinical Text" (L. E. Braverman and R. D. Utiger, eds.), 8th ed., pp. 467–473. Lippincott-Raven, Philadelphia.
- Vanderpump, M. P. J., Tunbridge, W. M. G., French, J. M., Appleton, D., Bates, D., Clark, F., Grimley Evans, J., Hasan, D., Rodgers, H., Tunbridge, F., and Young, E. T. (1995). The incidence of thyroid disorders in the community: A twenty-year follow-up of the Wickham Survey. *Clin. Endocrinol.* **43**, 55–69.

## INTRODUCTION

When ancient vertebrates migrated from the iodine-rich oceanic environment to the iodine-poor terrestrial environment, evolution provided the new organisms' thyroid with an extremely complex and specialized machinery, completely devoted to the avid collection and storage of iodine, for the sole purpose of continuously providing enough substrate for the synthesis of thyroid hormone. As a very specialized organ, the thyroid therefore expresses, along with common housekeeping genes, a unique subset of genes, whose products are mostly enzymes involved in iodine metabolism and/or thyroid hormone synthesis. On one hand, some of these enzymes serve as unique and possibly confined targets (antigens) for the immune aggression characteristic of autoimmune thyroid diseases (AITD), which are very common conditions in humans, whose etiology is still incompletely understood. The susceptibility to these diseases has long been recognized to be largely hereditary in nature. Because of the high prevalence in the general population and because of the large amount of available data on genetic factors, AITD are given special emphasis in this article.

On the other hand, a wide range of thyroid disorders have been described in which a single inherited point mutation in one thyroid-specific gene causes clinically significant derangements in thyroid physiology and function, with or without hereditary goiter. Given their strong negative effect on survival, the latter conditions are distinctly rare but are usually inherited in a Mendelian fashion and therefore represent an important part of the genetics of thyroid disease. Thyroid cancer and goiter are also quite prevalent. Although epidemiological studies indicate that environmental factors are by far predominant in causing these disorders, emerging evidence shows significant inherited susceptibility to these disorders as well.

## THE MONOGENIC DISEASES OF THE THYROID GLAND

### Inherited Diseases of the Thyrotropin Receptor

The human thyrotropin receptor (TSHR) was cloned in 1989. The TSHR is a protein receptor, one of the G protein-coupled, seven-transmembrane domain receptors. The TSHR gene is located on chromosome 14q31. The primary structure of the TSHR is closely related to that of the follicle-stimulating hormone and luteinizing hormone receptors. Activation of the

TSHR is the key event leading to both increased thyroid hormone secretion and growth. Because of its central role in the regulation of thyroid physiology, the TSHR has been the focus of constant research in the history of endocrinology. In newer studies, however, inherited defects of the TSHR have been described as a cause of clinically relevant thyroid disease.

### Activating Mutations of the TSHR

Site-directed mutagenesis initially showed that single point mutations in critical regions of the TSHR were capable of constitutively activating the TSHR, in the absence of its natural ligand, thyrotropin (or thyroid-stimulating hormone (TSH)). Such mutations, if found *in vivo*, would be expected to cause thyroid growth and hyperthyroidism. Indeed, naturally occurring, activating mutations of the TSHR have been described in a significant fraction of human toxic adenomas and toxic multinodular goiters. Mutations of the TSHR causing distinct hyperfunctioning nodules are always somatic, i.e., are found only within the nodular tissue. The nodules containing the mutation therefore represent monoclonal expansions of an initially mutated single thyroid cell and the condition is not hereditary. Since activation of the TSHR induces thyroid growth but also differentiation, these neoplasms are only very rarely malignant. More rarely, however, germ-line activating mutations of the TSHR have been described. The phenotype caused by these mutations is termed autosomal-dominant nonautoimmune hyperthyroidism [Online Mendelian Inheritance in Man (OMIM) 603372]. Although in many cases, this disease is inherited in a dominant fashion, at least half of the cases described thus far (>20) in the literature are sporadic; i.e., the mutation is found only in the index case and not in the parents (*de novo* mutation). The classical phenotype is characterized by severe congenital or neonatal hyperthyroidism and goiter. TSHR antibodies are usually negative and this helps differentiate the disorder from the other relevant cause of neonatal hyperthyroidism, the transplacental passage of maternal TSHR-stimulating antibodies. Treatment with anti-thyroid drugs can usually control the hyperthyroidism, but relapses are the rule. Thyroid growth is not contrasted by anti-thyroid drugs and the goiter evolves from initially diffuse to multinodular early during childhood, requiring surgery. Milder cases have been described, with later onset of hyperthyroidism and goiter. Several mutations of the TSHR have been identified as being responsible for the syndrome. The relationship of genotype to phenotype has been found to be somewhat inconsistent in large families with multiple

members carrying the same mutation. This indicates that other factors (genetic or environmental) play a role in determining the phenotype, at least with some of the known mutations.

More rarely, inherited mutations of the TSHR cause an inordinately responsiveness of the receptor to the placental hormone  $\beta$  human chorionic gonadotropin ( $\beta$ -HCG), which is closely related to TSH. In these cases, the effect of the mutation becomes evident only when high levels of  $\beta$ -HCG are present, i.e., during pregnancy, hence the appellation of familial gestational hyperthyroidism for this peculiar disease (OMIM 603373). Other than in pregnancy, the carriers of this rare mutation have a completely normal thyroid function.

#### **Inactivating Mutations of the TSHR**

Inactivating mutations of the TSHR manifest themselves with the clinical phenotype of resistance to TSH (OMIM 275200). Patients have often been detected at birth with severe primary congenital hypothyroidism (elevated TSH and low thyroxine) and a normally located but hypoplastic thyroid. In other cases, only mild, subclinical hypothyroidism has been found. Some of these patients have been noted to have an increased thyroglobulin level, an unexplained finding. Mutations leading to this disease have a variable effect on the receptor function, when studied *in vitro*. As expected with such a variable phenotype, several different mutations have been found; some completely abolish the expression of the receptor, whereas others only incompletely impair its responsiveness to TSH. The disease was initially described as autosomal-recessive, as predicted in many "loss-of-function" mutations. Initially reported cases have most often been compound heterozygotes, i.e., carrying two different mutations on each copy of the gene, but true homozygotes have also been found, usually in offspring of consanguineous marriages. However, since the first recognition of the syndrome, more and more heterozygous cases are being reported, with mild subclinical hypothyroidism, no goiter, and no evidence of thyroid autoimmunity. In one report of 10 such cases, 4 of the patients did indeed harbor mutations of the TSHR, 3 of which were heterozygous.

#### **Inherited Defects of Thyroid Hormone Biosynthesis and Processing**

Defective synthesis of thyroid hormone from the thyroid results in overt or compensated primary hypothyroidism. Since thyroid growth is largely independent

of the enzymatic pathways involved in the biosynthesis of thyroid hormone, ongoing TSH elevation as a consequence of low circulating thyroid hormone results in goiter. Therefore, the clinical phenotype of this group of diseases is characterized by hereditary goiter, usually early in onset, and various degrees of impairment of thyroid function, ranging from euthyroidism to overt primary hypothyroidism. Additional clinical features may be observed when the defect has consequences on other organs, as in Pendred's syndrome. The phenotype may be further characterized by the application of sophisticated thyroid function tests, such as the perchlorate challenge test or the measurement of thyroid hormone by-products. Because of their great impact on the overall health and survival of affected individuals, some of these diseases are quite rare and mostly recessive. However, their recognition has greatly enhanced the understanding of thyroid gland metabolism.

#### **Pendred's Syndrome**

Pendred's syndrome (OMIM 274600) is an autosomal-recessive disease characterized by congenital deafness, early-onset goiter, and euthyroidism (occasionally subclinical hypothyroidism). The prevalence of Pendred's syndrome is estimated at 1/1000. The sensory deafness is often due to a malformation of the inner ear in which the cochlea is replaced by a single cavity (Mondini's defect). However, an enlarged vestibular aqueduct, endolymphatic sac, and endolymphatic duct on magnetic resonance imaging (MRI) of the ear have been shown to be more specific signs. Except for its unusually early onset, the goiter is clinically not distinguishable from the more common endemic multinodular goiter. The pathology of the thyroid, however, displays marked hyperplasia and nodularity, sometimes suggestive of cancer. The incidence of true thyroid cancer, however, does not seem to be increased in these patients. The diagnosis of the disease relies on the classical perchlorate discharge test, which is highly sensitive but relatively nonspecific as other defects of thyroidal iodine organification result in a positive test. However, the combination of a positive perchlorate discharge test with typical malformations on MRI imaging of the inner ear establishes the diagnosis. The disease has been mapped by linkage analysis to chromosome 7q31. The cloned gene (termed PDS) encodes a protein (pendrin) that has features consistent with a transmembrane protein and is largely expressed in the adult human thyroid and (to a lesser extent) in the human cochlea and kidney. Although its function has not been fully clarified, pendrin seems to be an iodine-chloride transporter.

### **Sodium–Iodide Symporter Defect (OMIM 601843)**

This extremely rare disease is characterized by an autosomal-dominant mode of inheritance, severe congenital hypothyroidism, and multinodular goiter. The diagnosis is clinically established on the basis of a very low thyroidal uptake of radioiodine in the presence of goiter. An additional feature is a partial response in terms of thyroid hormone production to high oral doses of potassium iodide. The disease is caused by homozygous inactivating mutations of the sodium–iodide symporter gene, located on chromosome 19p33.2–p12.

### **Thyroglobulin Defects (OMIM 274900)**

Patients with this autosomal-recessive condition display a wide range of thyroid dysfunction, ranging from severe hypothyroidism to euthyroidism. Goiter is invariably present and the thyroidal radioiodine uptake is elevated. Thyroglobulin is a very large protein whose gene is located on chromosome 8q24.2–q24.3. Several structural (qualitative) and quantitative defects of thyroglobulin have been detected in patients with the disease. Usually, patients with qualitative defects have elevated circulating thyroglobulin levels, whereas patients with quantitative defects have low thyroglobulin levels.

### **Thyroperoxidase Defects (OMIM 274500)**

Patients with thyroperoxidase defects have severe congenital hypothyroidism and various degrees of goiter. The disease is autosomal-recessive and is diagnosed by a positive perchlorate discharge test, in the absence of clinical features of Pendred's syndrome. The gene for thyroperoxidase has been mapped to chromosome 2p25 and has been found to be mutated in many but not all patients with this clinical presentation, suggesting some degree of heterogeneity in the syndrome.

### **Thyroid Hormone Coupling Defect (OMIM 247700)**

This is a poorly defined group of disorders, also characterized by goiter and various degrees of hypothyroidism in which metabolic studies seem to suggest an altered coupling of iodotyrosil residues on thyroglobulin. Because part of this reaction is probably mediated by thyroperoxidase, there is a partial overlap with deficits of that enzyme and the phenotype could also be explained by alterations in thyroglobulin structure.

### **Dehalogenase Defects (OMIM 274800)**

Deiodinases (or dehalogenases) are a group of enzymes capable of deiodinating metabolites of thyroid hormone degradation, such as diiodotyrosine (DIT)

and monoiodotyrosine (MIT). As such, they greatly contribute to the intrathyroidal pool of iodine by deorganifying “used” organic iodine and making it available for new hormone synthesis. Patients with a dehalogenase defect are incapable of reusing MIT or DIT and develop hypothyroidism and goiter secondary to urinary loss of iodine. The defect is inherited in an autosomal-dominant fashion and is diagnosed by an elevated urinary excretion of administered labeled diiodotyrosine and monoiodotyrosine. Also characteristic is a prompt resolution of hypothyroidism when dietary supplementation with high doses of iodine is given.

## **Developmental Defects of the Thyroid and Congenital Hypothyroidism**

Congenital hypothyroidism (CH) has an incidence of approximately 1 in 3000–4000 newborns. Universal screening for the disease is available in most countries and allows early detection and treatment. Although CH can be observed in any severe form of the defects of thyroid synthesis described above, it is most often due to an abnormal *in utero* development of the thyroid gland, varying from its complete absence (thyroid agenesis) to various degrees of ectopy and hypoplasia (thyroid dysgenesis). Hereditary forms of the latter are very rare, most likely due to the severity of the consequences of CH, which include infertility. A number of genes involved in thyroid organogenesis have been identified and their role in the pathogenesis of CH has been partially clarified. The thyroid transcription factor 1 (TTF-1; OMIM 600635), mapped to chromosome 14q21, has been shown in animal models to be necessary for normal early thyroid organogenesis. Despite its central role, mutations in TTF-1 have been identified in only a few patients with congenital hypothyroidism due to thyroid dysgenesis. TTF-2 (on chromosome 9q22; OMIM 241850) is also highly expressed during thyroid ontogenesis. A familial case of thyroid agenesis associated with cleft palate and spiky hair has been shown to be due to heterozygous missense mutations of TTF-2 and was named Banforth-Lazarus syndrome. Paired-box gene 8 (PAX-8; OMIM 167415) is another transcription factor involved in thyroid organogenesis and regulation of the transcription of thyroid-specific genes, such as thyroglobulin and thyroid peroxidase. Nonsense and missense mutations of PAX-8 have been observed in two sporadic cases and one familial case of thyroid dysgenesis. Despite these encouraging findings, the etiology of the large majority of cases

of CH and thyroid dysgenesis remains unknown. It is likely that, in addition to other genetic factors, environmental factors play an important role.

## GENETIC FACTORS IN THYROID AUTOIMMUNE DISEASES

### The Clinical Phenotype

It is important to recognize that these diseases represent a group of diseases rather than a homogenous entity and that several clinical features may occur independently of one another, defining a variety of possible phenotypes. In general, the AITD are defined by the presence of a thyroid lymphocytic infiltrate, associated with serological evidence of thyroid autoimmunity in the form of circulating antibodies reactive to thyroid antigens and various degrees of thyroid dysfunction, ranging from profound hypothyroidism, as in the case of atrophic Hashimoto's thyroiditis (HT; OMIM 140300), to severe hyperthyroidism, as in the typical Graves' disease (GD; OMIM 275000). A number of other features may or may not be associated, such as the presence of Graves' ophthalmopathy, an autoimmune disease of the orbital tissues, typically observed in a relevant number of patients with hyperthyroid GD but occasionally also found in patients with HT. Pretibial myxedema is a puzzling inflammatory process of the dermis, localized to the pretibial regions, that is also associated (more rarely than ophthalmopathy) with GD. Finally, depending on various factors that are not completely understood, the thyroid gland may be enlarged as a consequence of either massive lymphocytic infiltration or of ongoing stimulation, it may be normal in size, or it may be strikingly reduced in size, as in the case of primary myxedema.

### Evidence for a Role of Genetic Factors in the Pathogenesis of Autoimmune Thyroid Disease

It is a common experience in the practice of endocrinology to observe familial clustering of HT and GD. These diseases are quite prevalent in the general population (prevalence of approximately 1%). Therefore, given this high prevalence, random occurrence of the disease in more than one family member could account for the clinical observation, without necessarily indicating the presence of a genetic predisposition. However, a number of lines of evidence strongly

indicate an important influence of genetic factors in the etiology of these diseases. One indication of the presence of a genetic influence in the predisposition to AITD comes from family studies. In general, when there is such an influence, the prevalence of the disease in first-degree relatives of patients with the disease (probands) is significantly higher than in the general population.

In early studies conducted in the late 1950s and 1960s, a high incidence of thyroid autoantibodies was observed in first-degree relatives of patients with HT or GD. The prevalence of positive anti-thyroglobulin or anti-microsomal antibody tests in relatives ranged from 45 to 55%, compared to a general population prevalence of approximately 15%. Although those studies were exposed to a number of potential selection biases, they provided the first indication that at least one component of the disease phenotype (e.g., the formation of thyroid autoantibodies) had a possible inherited cause. Indeed, more stringent classical segregation analysis in families with autoimmune thyroid disease confirmed these earlier results, indicating a Mendelian dominant mode of inheritance for thyroid autoantibodies, at least in some families. It is interesting that such a model of inheritance would in fact result in a prevalence of the phenotype in first-degree relatives of patients of approximately 50%, as in the initial studies. In a further refinement of these findings, some investigators have indicated that in some families not only the predisposition to form thyroid autoantibodies is strongly hereditary, but also that even the fine molecular specificity of such antibodies can be inherited.

In summary, there seems to be a clear role for genetic factors in the formation of thyroid antibodies, at least in families in which the clinical disease exists in one member. In these families, the thyroid autoantibody trait seems to be inherited in a classical Mendelian fashion as an autosomal-dominant trait, with high penetrance. The situation was different when normal subjects with thyroid autoantibodies but without clinical HT were selected as probands. In one such study, only 30% of relatives had a positive result, compared to approximately 16% expected from population data, indicating at best a polygenic inheritance with low penetrance. Thus, the relatively common anti-thyroid autoantibody phenotype may represent the consequence of a strong genetic influence only in families with overt AITD.

The presence of circulating thyroid autoantibodies represents only one aspect of the AITD phenotype. Less clear data are available when one looks at the inheritance of the full AITD phenotype. Only

approximately 18% of relatives of patients with HT were found to be affected in one study, although the percentage also includes second-degree relatives. Another study observed that 33% of siblings of patients with HT were affected. Even lower percentages have been reported in GD, where a prevalence of 5 to 10% has been reported in siblings of patients. By dividing the prevalence of a disease in first-degree relatives of probands with the disease by the prevalence of the same disease in the population of origin, one can obtain the relative risk for the disease in sibs, also termed  $\lambda(s)$ . This measure provides an estimate of the importance of genetic factors in a disease. In nongenetic diseases,  $\lambda(s)$  equals 1, whereas in highly inheritable and penetrant, monogenic disorders, it can rank as high as several hundred. The  $\lambda(s)$  for HT can be estimated from available data to range from 20 to 45, quite a high number, indicating that indeed a significant genetic component in the etiology of HT exists. A somewhat lower value of 7.5 to 10 has been estimated for GD. It should be noted, however, that these numbers are somewhat artificial in that they are extrapolated from different studies, whereas no large population-based studies directly aimed at obtaining unbiased data are available.

Another way of estimating the role of the genetic contribution to the etiology of a disease is studying twins. In theory, although both dizygotic and identical twins share as much of the environmental influences as possible, including the intrauterine milieu, only dizygotic twins share approximately 100% of the genome, whereas dizygotic twins share on the average only 50% of the genome. For example, an autosomal-dominant, fully penetrant disease should be 100% concordant (i.e., present in both twins) in identical twins and only 50% concordant in dizygotic twins. By measuring the concordance rate of a disease in identical twins and comparing it to the rate observed in dizygotic twins, investigators can obtain very precise data on the relevance of genetic factors, the penetrance, and the mode of inheritance of disease. In both GD and HT, studies in twins have shown concordance rates well below 100% in identical twins (approximately 30% in GD and 55% in HT). However, much lower concordance rates (close to 0) have been observed in dizygotic twins. Thus, whereas a lower concordance rate in dizygotic twins indicates the presence of genetic factors, the less-than-100% concordance rate observed in identical twins indicates a role for environmental factors as well (incomplete penetrance). Interestingly, 80% of identical twins and 40% of dizygotic twins of HT patients had circulating thyroid autoantibodies. Again, these data suggest a dominant mode of

inheritance with high (80%) penetrance for the thyroid autoantibody trait, when part of a general predisposition to thyroid autoimmune disease.

Family and twin studies confirm the clinical impression of a genetic predisposition to AITD. However, the mode of inheritance of the full clinical phenotype seems to be complex rather than simple Mendelian inheritance. Admittedly, knowledge of the mode of inheritance of AITD is still very limited. In the search for the susceptibility genes, it is reasonable to consider two hypotheses. According to one hypothesis, the overall genetic susceptibility is provided by a relatively (>10) large number of disease-associated gene variations at many genetic loci (risk factors). When a sufficient number of these variants, in any combination, are inherited by an individual and appropriate environmental events take place, a threshold is reached and the disease develops. Given the small contribution to the overall susceptibility provided by these putative genes, they are more easily detected by sensitive association studies. If, in contrast, only a few (<10) major "necessary" genes exist, the disease will develop only in those subjects who will inherit the complete set of susceptibility genes. These genes are best detected by linkage analysis. Indeed, if many risk factors exist, it will be very difficult to develop efficient prediction tools to be used in clinical practice.

### Predisposing Genes in AITD

The evidence summarized above has stimulated in the past few decades quite a large number of studies aimed at the identification of the genes involved in the genetic predisposition to AITD. Earlier studies mostly employed the method of genetic association analysis, using known genetic loci. Several candidate loci have been investigated, mostly genes of thyroid autoantigens and genes involved in the immune response. Since the development of large-scale genome sequencing data obtained through the Human Genome Project, powerful linkage analyses have been applied to the search for AITD genes. Data obtained with this tool are therefore as yet largely unconfirmed but quite promising. Given the relative simplicity of association studies, these have been most often used in order to detect putative inherited risk factors for AITD. Several genetic variations in different loci have been found to be positively associated with AITD. However, only two of those loci have been thus far convincingly confirmed in several studies, the human leukocyte antigen (HLA) gene complex and cytotoxic

T lymphocyte antigen 4 (CTLA-4). The HLA has been extensively studied given its complexity and its relevance to the regulation of the immune system.

### **The HLA Genes and AITD**

The HLA gene complex is located on the short arm of the sixth chromosome (6p21) and encodes a large number of genes mostly (but not only) involved in the regulation of the immune response. A detailed description of the HLA complex is beyond the scope of this article and has been previously published elsewhere. In general, genes within the HLA complex have been subdivided into three classes. Class I includes the HLA class A, B, and C antigens, which are widely expressed on several tissues. Class I genes are mostly involved in direct cytotoxic reactions leading to the killing by the immune system of epithelial cells carrying exogenous antigens, mostly viral in nature. Class II includes the HLA DR, DP, and DQ antigens, mostly expressed on immune cells. These antigens are mostly involved in the presentation of antigen within the immune system. This process is central to the development of the normal immune response and, depending on several factors, it will lead to either the selection and amplification of antigen-specific T and B cells or to the silencing of these cells (tolerance). Class III includes genes coding for complement factors, heat shock proteins, tumor necrosis factor  $\alpha$ , and several other proteins not directly involved in the immune response. As such, they truly represent genes that happen to be located within the HLA system. The HLA genes (especially class I and class II) are highly polymorphic and some alleles show striking linkage disequilibrium. Because of its function, the HLA complex has been widely studied in relation to autoimmune diseases. The AITD are no exception. GD was initially shown to be associated with the HLA-B8 allele in several studies involving Caucasian patients with relative risks from 1.5 to 3. However, the HLA-B8 antigen has been subsequently recognized to be in linkage disequilibrium with the class II allele HLA-DR3. Indeed, additional association studies have shown a slightly higher relative risk (2.5–4) when this allele was studied in Caucasian GD patients. It was concluded that the HLA-DR3 allele was more relevant than HLA-B8 in determining the inherited susceptibility to GD. GD has also been found to be associated with (relative risk 3.8) the allele HLA-DQA1\*0501, also in linkage disequilibrium with DR3, as well as with the allele DRB1\*0301, in linkage disequilibrium with DR3. HT has been occasionally found to be associated with HLA-DR3, HLA-DR4, and HLA-DR5, with similarly low relative risks,

ranging from 2 to 7. It is interesting that different haplotypes of the HLA have been involved, as GD and HT are often found in different members of the same family, portending a common genetic background. In summary, association studies have demonstrated (because of their consistency) a significant role for the HLA system in the genetic susceptibility to the AITD. The HLA gene complex has also been studied in linkage studies, although much less often. In keeping with the low relative risks obtained in association studies, most studies, including a whole genome screen analysis, have shown no evidence in favor of linkage with the HLA. Since one possible explanation for low relative risks in association studies with negative multi-family linkage studies is the presence of genetic heterogeneity, linkage to the HLA was sought in families in which the proband was DR3 positive. Even in this extremely selected subgroup of families, there was no evidence of cosegregation of the disease phenotype with the HLA gene complex. The only positive linkage analysis involved patients with GD and IDDM, a patient population that does not really reflect the common phenotype of AITD. In summary, genes within the HLA complex have been consistently shown to be involved in the genetic predisposition to AITD. However, as exemplified by low relative risks and negative linkage, the HLA complex failed to explain the large hereditary predisposition to the diseases shown by family studies. Indeed, the haplotypes found to be associated with AITD are quite prevalent in the studied control populations, indicating that most of the people who carry those haplotypes do not develop the disease. Conversely, only a fraction of the affected subjects carry the associated alleles. This shows that mutations within the HLA system are not absolutely required or sufficient to cause AITD, representing risk factors with a small overall effect.

### **CTLA-4 and AITD**

The human CTLA-4 gene is located on chromosome 2q33 and its product is a protein belonging to the immunoglobulin gene superfamily. The CTLA-4 protein is expressed on T cells only after antigen or mitogen stimulation. It subsequently interacts with nearby activated T cells carrying a CTLA-4 receptor (B-7) and causes their silencing or death. Therefore, CTLA-4 acts as a key down-regulator of the immune response and represents an excellent candidate for an autoimmunity gene. As a proof of its importance, mice defective in CTLA-4 function (CTLA-4 $^{-/-}$ ) develop rapidly fatal multiorgan lymphocytic infiltrates, a condition that resembles human autoimmune diseases. CTLA-4 therefore represents an excellent

candidate for a gene predisposing to autoimmunity. Although there are no known variations or polymorphisms within the coding sequence of human CTLA-4, three polymorphic markers within or close to the CTLA-4 gene have been well characterized and used in association studies to test the CTLA-4 gene as a candidate predisposing gene in AITD. The first described allelic system is a typical microsatellite marker, located in the 3'-untranslated region of exon 3 of the gene. One of the alleles of the system (106 bp long) was found more often in GD chromosomes (27%) than in control chromosomes (14%), yielding a relative risk of 2.8. Subsequent studies using this marker have confirmed an association of this particular allele with AITD. Subsequent studies have employed single nucleotide polymorphisms (SNPs), which are less polymorphic and therefore more useful in association studies. Two SNPs within the CTLA-4 gene are known. The first polymorphism is an A-to-G transition at position 49 of the coding sequence, causing a threonine/alanine substitution at codon 17 of the protein. The G allele shows strong linkage disequilibrium with the 106 bp allele of the CTLA-4 microsatellite marker. This means that the two alleles are found associated together on chromosomes at a higher frequency than expected. Indeed, the G allele has been shown to be associated with GD, HT, and even postpartum thyroiditis in a number of studies, with only a few exceptions. Like the microsatellite marker, the G allele yields small relative risks in AITD, in the range of 2 to 4. It is thus far unclear whether the homozygous state confers additional risk, but the available data seem to indicate that the effect is dominant. The G allele is quite prevalent in the general population (in fact, it is the most prevalent allele, with a prevalence of up to 85% among the Japanese), much more frequent than the AITD themselves. It is therefore intuitive to conclude that this polymorphism plays a small, although probably significant role in the genetic predisposition to AITD. Two studies have reported a significant increase in the prevalence of the G allele in GD patients with ophthalmopathy as compared to GD patients without clinically relevant eye disease. However, the only available family-based study indicates that there is no genetic predisposition to Graves' ophthalmopathy existing independently of the general predisposition to GD.

Interestingly, the G allele has been associated with a variety of other autoimmune disorders, raising some questions about the specificity of these findings. The third known polymorphism of the CTLA-4 gene is a C-to-T (C/T) (SNP) transition at position -318 with respect to the start codon of the gene. This

polymorphism does not appear to confer additional risk over that derived from the 49 A/G SNP. The CTLA-4 locus has also been extensively analyzed by linkage analysis. Initial studies employing the CTLA-4 3' microsatellite marker in multiplex families with AITD excluded linkage of the CTLA-4 gene to the full AITD phenotype, i.e., with the clinical diseases. However, when the same data set was analyzed for linkage to the CTLA-4 locus but extending the definition of affected subjects to family members with positive thyroid antibody tests, even in the absence of relevant goiter or thyroid dysfunction, a high likelihood of linkage was observed and the model was consistent with an autosomal-dominant, highly penetrant trait. These results are in keeping with previous results obtained with a different type of linkage analysis, which showed positive linkage of CTLA-4 to GD but did not allow for the distinction between subjects with positive antibody tests and patients with the full-blown disease.

In summary, polymorphisms near the CTLA-4 gene represent a weak but significant risk factor for the development of AITD as shown by association studies. Moreover, evidence from linkage studies suggests that the CTLA-4 region (2q33) contains a major determinant of the genetic susceptibility to the development of thyroid autoantibodies, as a distinct part of the AITD phenotype, and that this major determinant is CTLA-4. However, CTLA-4 alone is not sufficient to explain the whole genetic predisposition to the development of overt AITD. Moreover, it is still unclear how these polymorphisms affect the function of CTLA-4, especially since they are located outside the coding region of the gene and are therefore not expected to cause major changes in its function. Promising but preliminary data from functional studies have shown that lymphocytes from CTLA-4 49 G/G carriers show a lower degree of CTLA-4-dependent T cell inhibition, regardless of the presence of AITD, a finding that provides some indication that this polymorphism may indeed underlie an altered immune system. It is also interesting to note that both of the genes (HLA and CTLA-4) that have been identified as risk factors for AITD participate in the same process, i.e., antigen presentation.

#### **Gene Loci Identified by Whole Genome Screens**

The availability of extensive mapping data from the Human Genome Project and of automated sequencing systems has allowed the performance of relatively fast analyses of the whole genome to search for susceptibility genes for common diseases. Some laboratories

Table I Putative Susceptibility Loci and Maximum LOD Scores for the AITD Detected in Linkage Studies and Their Chromosomal Location

Linked phenotype	Chromosomal location	Name of locus	LOD score	Comments
AITD	6p	AITD-1	2.9	
AITD	5q31-q33	—	3.1	
AITD	18q21	IDDM6	3.46	
AITD-HT	8q24	—	3.7	Confirmed in independent studies, contains thyroglobulin
GD	14q31	GD-1	2.5	
GD	20q11.2	GD-2	3.5	Confirmed in independent studies
GD	Xq21	GD-3	2.5	
GD	Xp11	—	2.21	
HT	13q32	HT-1	2.1	
HT	12q22	HT-3	1.7	Statistical evidence of heterogeneity
Thyroid antibody GD	2q33	TAb-1	3.6	Confirmed in independent studies, contains CTLA-4

have addressed the identification of susceptibility genes for the AITD with this powerful tool. Because of the inherent complexity of these studies, results obtained thus far should be considered largely preliminary, at least until mutated genes are isolated from the linked regions and firmly demonstrated to cause the diseases. However, a critical mass of data is rapidly accumulating from different groups of patients and populations and has begun to indicate regions in the genome highly likely to contain susceptibility genes. Table I summarizes putative loci identified thus far.

An initial genome scan performed in a large set of Caucasian families with AITD has shown three putative gene loci for GD, two putative loci for HT, and one common locus for GD and HT. Interestingly, the same genome screen identified the region containing CTLA-4 as being significantly linked to the presence of circulating antibodies, as summarized above. In general, LOD (logarithm of differences) scores of these initially reported loci were relatively low, although statistically significant. Moreover, given the relatively small (56 families) size of the sample, large areas of the genome gave indeterminate results; i.e., linkage could not be proved or disproved, leaving room for more loci to be identified. A subsequent whole genome screen performed in Japanese patients with AITD using sib-pair analysis (which involves some relevant statistical differences from classical linkage analysis as exemplified above by the case of CTLA-4) has shown linkage of AITD to 5q31-q33 and 8q23-q24. The locus on 8q23-q24 has subsequently been confirmed by an extension to 104 families of the first whole genome screen. This

chromosomal location is particularly interesting, since it contains the thyroglobulin gene. Indeed, association studies have shown that at least one allele of a polymorphic marker very close to the thyroglobulin gene is not only linked but also associated with AITD, strongly suggesting that thyroglobulin may represent a susceptibility gene for the AITD.

The susceptibility to AITD is largely genetically determined, although the inheritance pattern is complex, as is the case with most genetically determined, frequent disorders. Researchers have started to unravel the gene variations responsible for the AITD, and in the future, accurate prediction in at-risk individuals will likely be available.

## GENETIC FACTORS IN SIMPLE AND MULTINODULAR GOITER

The main etiologic factor in goiter is indisputably iodine deficiency, a worldwide problem that affects almost 1 billion people around the world. In areas with iodine deficiency, the prevalence of goiter in the general population is >10%. Epidemiologists have traditionally distinguished “endemic” goiter, observed in geographic areas of iodine deficiency, from “sporadic” goiter, observed in areas of iodine sufficiency. The accuracy of the distinction is, however, questionable, as the clinical phenotype of the two forms of the disease is indistinguishable, except for areas of extreme iodine deficiency, where large goiters are associated with cretinism and hypothyroidism. In addition to iodine deficiency, other factors are likely involved in the pathogenesis of this common disease. This hypothesis stems from several circumstantial observations. In

areas with iodine deficiency, not all exposed subjects develop goiter. Universal iodine supplementation strikingly reduces but does not abolish the disease and goiter is observed in the absence of iodine deficiency, as mentioned above. Additional environmental factors, such as tobacco smoking, naturally occurring goitrogens, and pharmacologic goitrogens, have been demonstrated. Sex hormones are also likely to be involved as goiter is more prevalent in women living in areas with mild iodine deficiency. The observation of familial clustering of the disease is a daily experience of all endocrinologists, especially in areas with mild or no iodine deficiency, and this has led to the hypothesis that inherited factors coincide in the predisposition to the development of goiter. The familial clustering of the disease has also been reported in several epidemiological surveys, mostly from areas of endemic iodine deficiency. As in the case of the AITD, a number of twin studies have been performed. These have shown that the environmental factors are largely predominant when cases are drawn from endemic regions, as shown by similar concordance rates between dizygotic and monozygotic twins. However, a large, population-based study in twins has shown significantly different and overall higher concordance rates between monozygotic and dizygotic twins (Table II). Interestingly, the intrinsic population prevalence of goiter in this latest study is much lower than in the two older studies, performed in the 1960s in areas of iodine deficiency. One possible explanation for these apparently discrepant data is that once the effect of the major environmental factor (iodine deficiency) is removed from the population, cases with a stronger genetic effect emerge. In this view, the distinction between endemic and sporadic goiter seems to be justified. Despite evidence in favor of relevant genetic influences, these have not been clarified yet. As in many common diseases, there is no identifiable mode of inheritance and multiple, relatively frequent gene variants are likely to play a role. Thus far, only a few studies have addressed this problem on a large scale. Studies in single large kindreds

with highly prevalent and apparently dominant transmission of multinodular goiter have indicated linkage of the phenotype to a locus (termed MNG-1) located on chromosome 14q31. The same studies have ruled out the TSHR (which resides in the same chromosomal region) from the linked region. MNG-1 has been confirmed as a susceptibility locus in at least another large family. Interestingly, MNG-1 overlaps with GD-2, a susceptibility locus for GD. It is therefore possible that the region contains a novel thyroid-specific growth factor and/or antigen. Similarly, a point mutation in exon 10 of the thyroglobulin gene causing a substitution of histidine for glutamine at codon 870 has been described and linked to goiter in three families, in the absence of hypothyroidism. Finally, an Italian family with suggestions of X-linked transmission has been described. Indeed, linkage analysis has shown evidence of linkage of goiter to chromosome Xp22. These findings are encouraging but were obtained in a small subset of families with a seemingly large genetic effect and their epidemiological relevance to goiter found in the general population is unknown. Monogenic disorders of thyroid hormone metabolism represent a rare cause of hereditary goiter, with or without hypothyroidism. These conditions, described in some detail above, are distinct from the commonly encountered sporadic or endemic goiter in that there is usually an earlier appearance of goiter, a more distinct familial pattern (usually autosomal-recessive), and several degrees of thyroid dysfunction. Although it is conceivable that more subtle defects in any of these genes might be involved in the pathogenesis of goiter, extensive screening of the general population for association and/or linkage to known genes is lacking.

In summary, genetic predisposition is probably important in the etiology of simple goiter, both sporadic and endemic. In endemic areas, though, the effect of environmental factors (iodine deficiency) is predominant and widespread, making goiter very prevalent. In contrast, in areas with normal iodine intake, genetic factors are more relevant, but with a smaller

Table II Available Studies in Twins in Populations with Different Prevalences of Goiter

Year of the study	Geographic location	Concordance rate in monozygotic twins	Concordance rate in dizygotic twins	Population prevalence of goiter
1967	United Kingdom	24%	12%	25.3%
1967	Greece	89%	73%	53.0%
1999	Denmark	42%	13%	1.5%

Note. In populations with a lower prevalence, a higher concordance rate is observed in monozygotic twins when compared with dizygotic twins, suggesting a stronger genetic predisposition.

prevalence. As a consequence, the prevalence of goiter is strikingly reduced and the phenotype more clearly clustered in families. It is likely that multiple different genes are involved in the predisposition to simple goiter, but thus far only a few of these have been identified.

## GENETIC FACTORS IN NONMEDULLARY THYROID CANCER

The etiology of thyroid cancer of follicular origin (OMIM 188550) is poorly understood, although several somatic mutations consistently associated with the neoplasm have been described. Exposure to ionizing radiation, especially during childhood, is a well-recognized environmental risk factor, accounting for only a minority of cases. Most other cases appear to be sporadic. In some instances, familial clustering has been observed. Although in many such cases, random occurrence of this relatively common disease in members of the same family can explain the observation, in some extended families, a clear pattern of inheritance has been observed, suggesting the existence of genetic factors. Clinically, familial nonmedullary thyroid cancer (NMTC) has been reported to be more aggressive and more often multifocal than its sporadic counterpart. A few population-based studies have shown an increased risk of NMTC in relatives of probands. For example, the Connecticut Tumor Registry study has demonstrated a fivefold increased risk in relatives of index cases with thyroid cancer. Such studies, however, are subject to several selection biases and may also reflect environmental influences shared by families. It is the opinion of many experts that a genetic predisposition to isolated nonmedullary thyroid cancer exists in only a few families. In two members of the family in which MNG-1 was identified, thyroid cancer was found. It is tempting to speculate that the same gene driving follicular cell hyperplasia in that family may also increase the risk for thyroid malignancy. Moreover, distinct loci have been identified in a few families. In a large multigeneration family with familial NMTC, linkage to a 20 cM segment on chromosome 1q21 was established. In two affected members, papillary renal neoplasms were also observed, suggesting a syndromic effect of the identified locus. In another large kindred with several cases of papillary thyroid cancer only, linkage was suggested with chromosome 2q21. The authors were able to confirm their results in a data set of 80 smaller families with more than one case of NMTC, obtaining significant evidence for linkage at the same locus, although the presence

of benign thyroid disease was included in the analysis. Even stronger evidence was found when the analysis was limited to families with the follicular variant of papillary thyroid cancer. A rare form of thyroid neoplasm, thyroid tumors with oxyphilia, has also been linked to chromosome 19p31. The phenotype in this case is represented by benign or malignant thyroid nodules, with prominent oxyphilia. The same locus (termed TCO1) was found to be linked to papillary thyroid carcinoma in a subsequent study in several families. Several genes are known to reside within that chromosomal location, but none has been found to be mutated in patients with NMTC thus far.

An increased risk of thyroid cancer has also been observed as a component of known hereditary syndromes. Cowden's disease (OMIM 158350) is an autosomal-dominant disease, characterized by multiple hamartomas of the skin, breast, thyroid, brain, and endometrium. Facial trichilemmomas are a distinctive feature of the syndrome. Other findings include craniomegaly, scrotal tongue, and cerebellar neoplasias. There is some overlap with related syndromes such as Lermite-Duclos syndrome and Ruvalcaba's syndrome, so that some investigators believe they are in fact synonyms. A significant increase in the incidence of breast cancer is seen in Cowden's syndrome families. Thyroid cancer, of both the papillary and follicular types, is also highly prevalent, as is benign thyroid nodular disease, although the magnitude of the relative risk is not known. Cowden's syndrome has been mapped to chromosome 10q23.31 and the mutated gene, termed PTEN for phosphatase and TENsin homologue (OMIM 601728), has been isolated. Several mutations of PTEN have been found in families with Cowden's syndrome, without a clear-cut genotype-phenotype relationship. Moreover, in several families with many features of the syndrome, no mutation of the gene has been found, indicating that some degree of heterogeneity exists. Several *in vitro* studies suggest that PTEN acts as a tumor suppressor gene. This observation is in keeping with an autosomal-dominant mode of inheritance and with the observation that most of the mutations identified thus far cause a loss of function. The PTEN protein is expressed in several tissues and is believed to function as a tyrosine phosphatase. As such, the protein is believed to be able to dephosphorylate activated regulator proteins, thus down-regulating proliferation pathways.

An increased risk of NMTC has also been observed in familial adenomatous polyposis (FAP; OMIM 175100), also termed Gardner's syndrome. FAP is inherited in an autosomal-dominant fashion. The most

prominent feature of the syndrome is early-onset colonic cancer and multiple widespread, preneoplastic polyps of the colon. Extracolonic manifestations include congenital hypertrophy of the retinal pigment epithelium in up to 90% of affected persons and has been used as a marker of the disease before genetic testing became available. Other features of the syndrome include benign bone tumors and hepatoblastomas. In women with the syndrome, the risk of papillary thyroid cancer has been estimated to be 160-fold higher than in the general population. Papillary thyroid cancer in the setting of FAP is often detected before age 30, is multifocal, and displays unusual pathological features. The gene responsible for FAP has been mapped to chromosome 5q21 and was designated adenomatous polyposis of the colon (APC). As expected by the autosomal mode of inheritance, APC functions as a tumor suppressor gene and detected mutations associated with the phenotype cause a loss of function of the gene product. The adenomatous polyposis of the colon (APC) gene product ( $\beta$ -catenin) is expressed on the membrane of several epithelial cell types. It has been hypothesized that loss of  $\beta$ -catenin function causes a loss of contact inhibition or a loss of cell adhesion mechanisms, leading to uncontrolled cell growth.

In summary, although the vast majority of cases of NMTC are sporadic, a few cases present in the setting of recognized hereditary syndromes. Knowledge of these clinical conditions is important to the clinician, in order to perform appropriate screening for associated conditions and to provide sound genetic counseling to individuals at risk.

## FAMILIAL MEDULLARY THYROID CANCER

Approximately 7% of thyroid cancers originate from the parafollicular cell lineage. These cancers are termed medullary thyroid cancer (MTC) and account for approximately 15% of all thyroid cancer-related deaths. In contrast to NMTC, MTC is quite often familial (in approximately one-fourth of the cases). Familial medullary thyroid cancer (FMTC) arises in the setting of multiple endocrine neoplasia (MEN) type II syndromes or as an isolated disease. In MEN IIA (OMIM 171400), FMTC is associated with pheochromocytomas and parathyroid adenomas, and in MEN IIB (OMIM 162300), patients present a marfanoid habitus, mucosal neuromas, and ganglioneuromatosis, in addition to the above-described features (see Fig. 1). In isolated FMTC (OMIM 155240), there are no

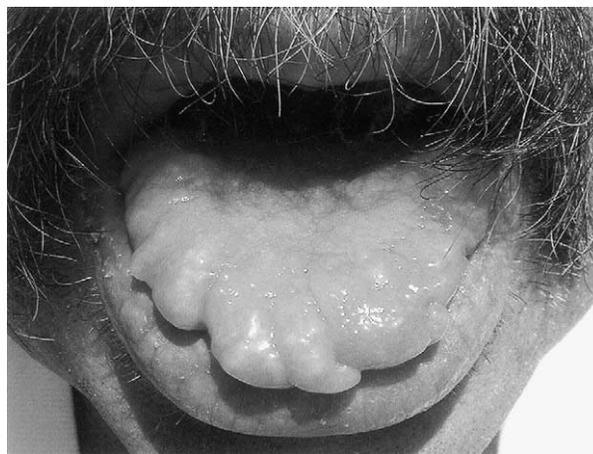


Figure 1 Typical lingual neuromas in a patient with MEN IIB.

extrathyroidal manifestations. The inheritance of these closely related conditions is autosomal-dominant, with high penetrance. In MEN IIA and MEN IIB, MTC is found in almost 100% of carriers by the end of third decade of life if sought by biochemical screening, whereas pheochromocytomas occur in 50% of carriers and parathyroid tumors in 25% of carriers. Despite these clearly distinct phenotypes, all three conditions are due to germ-line mutations in the RET proto-oncogene sequence. The RET proto-oncogene maps to chromosome 10q11.2 and encodes a membrane-bound tyrosine kinase receptor. Causative mutations detected thus far are almost entirely limited to a few cysteine residues located in the putative extracellular domain. It is hypothesized that these mutations induce critical conformational changes in the gene product, causing constitutive activation of the receptor with ongoing production of intracellular proliferative signals. Interestingly, there is quite a strict genotype-phenotype relationship. Mutations involving codons 609, 611, 618, 620, and 634 are found exclusively in families with MEN IIA or with FMTC and mutations involving codons 768, 804, and 891 have been found only in families with FMTC, although even in these families occasional cases of pheochromocytoma and hyperparathyroidism are found. Since the molecular genetics of the two conditions partially overlap, other genes must be involved in determining the phenotype. In contrast, mutations at codon 918 are found in almost all families with MEN IIB. Somatic missense mutations at codon 918 are also found in one-half of cases of sporadic MTC. The RET gene is also mutated in many cases of Hirschsprung's disease, although most mutations detected in this disorder induce the loss of function of the mutant allele. Surprisingly,

Hirschsprung's disease has been found in some families with MEN IIA.

Detailed knowledge of the molecular genetics of these aggressive diseases has greatly enhanced the ability of clinicians to detect persons at high risk for multiple endocrine neoplasias. Genetic testing for RET mutations has become widely available for family members of patients and allows early, preventive treatment, which is expected to greatly increase the life expectancy of patients.

### See Also the Following Articles

Iodine • Sodium Iodide Symporter • Thyroid Autoimmunity • Thyroid Carcinoma • Thyroid Disease and Pregnancy • Thyroid Disease, Epidemiology of • Thyroid Disorders in the Elderly • Toxic Multinodular Goiter

### Further Reading

- Alsanea, O., and Clark, O. H. (2001). Familial thyroid cancer. *Curr. Opin. Oncol.* **1**, 44–51.
- Barbesino, G., and Chiovato, L. (2000). The genetics of Hashimoto's disease. *Endocrinol. Metab. Clin. N. Am.* **2**, 357–374.
- Braverman, L. E., and Utiger, R. D. (2000). "The Thyroid," 8th ed. Lippincott Williams & Wilkins, Philadelphia, PA.
- Brix, T. H., and Hegedus, L. (2000). Genetic and environmental factors in the aetiology of simple goiter. *Ann. Med.* **3**, 153–156.
- Kopp, P. (2001). The TSHR and its role in thyroid disease. *Cell Mol. Life Sci.* **9**, 1301–1322.
- Macchia, P. E. (2000). Recent advances in understanding the molecular basis of primary congenital hypothyroidism. *Mol. Med. Today* **1**, 36–42.
- Malchoff, C. D., and Malchoff, D. M. (2002). The genetics of hereditary nonmedullary thyroid carcinoma. *J. Clin. Endocrinol. Metab.* **6**, 2455–2459.
- Tomer, Y., and Davies, T. F. (1997). The genetic susceptibility to Graves' disease. *Bailliere Clin. Endocrinol. Metab.* **3**, 431–450.

the elderly. Hypercholesterolemia is present in the majority of overt hypothyroid patients. Cardiovascular system abnormalities include cardiomegaly secondary to pericardial effusion, bradycardia, diastolic hypertension, and atherosclerosis. The myopathy of hypothyroidism causes proximal muscle pain and stiffness and weakness. Hypothyroidism also causes sleep apnea. Anemia—microcytic, normocytic, or macrocytic—is a well-characterized hematologic feature of hypothyroidism.

Thyroxine requirements are decreased in the elderly because of a decline in the metabolism of thyroid hormone. In elderly patients with coexisting cardiovascular disease, starting treatment with full replacement doses can result in exacerbation of angina and worsening of underlying heart disease. The starting dose of thyroxine should be small, 12.5–25 µg/day. The dose should be adjusted at 6- to 8-week intervals by an increment of 12.5–25 µg until the patient is euthyroid and the serum TSH is in the mid-normal range. Patients with central hypothyroidism should be monitored by FT<sub>4</sub> measurement and not serum TSH.

Drugs that block thyroxine absorption and increase thyroxine requirements are calcium carbonate, ferrous sulfate, cholestyramine, colestipol, sucralfate, and aluminum hydroxide. Rifampin, carbamazepine, phenytoin, and sertaline (Zoloft) accelerate T<sub>4</sub> clearance and increase the serum TSH level in patients on previously ideal replacement doses. Estrogen may also increase thyroxine requirements.

### Subclinical Hypothyroidism

In an English survey, the prevalence of subclinical hypothyroidism was 17.4% in women older than age 75. In a survey in the United States, the prevalence of serum TSH elevation was 8.5% in women and 4.4% in men older than age 55. The causes of subclinical hypothyroidism are the same as those of overt hypothyroidism. Subclinical hypothyroidism increases the risk for myocardial infarction two- or threefold and increases low-density lipoprotein cholesterol significantly when serum TSH is increased threefold above normal. The potential benefits of treating mild thyroid failure include prevention of progression to overt hypothyroidism, reduction of elevated serum cholesterol, and improvement in the quality of life. Replacement therapy is recommended for all patients with serum TSH concentrations greater than twice the upper limit of normal. It is also recommended if there are any clinical features of depression, fatigue, hyperlipidemia, or goiter. The goal of therapy is to normalize the serum TSH concentration. In older

people, 12.5–25 µg of levothyroxine is recommended as the initial dose. With minimal TSH elevation and the absence of clinical features, patients should be followed at intervals of 6 months for worsening of the condition, indicating a need for treatment.

### HYPERTHYROIDISM

The prevalence of hyperthyroidism varies from 0.5 to 2.3% in the elderly. Approximately one-sixth of all hyperthyroid patients are older than age 60. Graves' disease is the most common cause of hyperthyroidism in the elderly. TSH receptor antibodies are detectable in approximately 80% of untreated Graves' patients. Toxic multinodular goiter is more common in the elderly and has been reported in almost half of older patients with hyperthyroidism, especially in regions of relative iodine deficiency. Thyrotoxicosis can be precipitated in patients with nontoxic multinodular goiter by administration of a large iodine load such as radiocontrast agent. Some autonomously functioning thyroid adenomas have a mutation in the TSH receptor that results in chronic activation of the follicular cell. Amiodarone may cause destructive thyroiditis and thyrotoxicosis. Hyperthyroidism resulting from a TSH-secreting pituitary adenoma or pituitary resistance to thyroid hormone is very rare.

Elderly patients often lack typical features of hyperthyroidism, including goiter. The absence of the typical hypermetabolic manifestations of hyperthyroidism is termed apathetic hyperthyroidism. The dominant clinical findings may be weight loss, atrial fibrillation, or heart failure.

Serum TSH level is suppressed and FT<sub>4</sub> and FT<sub>3</sub> are elevated in hyperthyroidism. However, the serum T<sub>3</sub> level was found to be increased in only half of hyperthyroid patients between 75 and 95 years of age because of reduced conversion of T<sub>4</sub> to T<sub>3</sub> in peripheral tissues. A thyroid radioiodine uptake is useful to detect the conditions causing thyrotoxicosis with low thyroid uptake: thyroiditis, exogenous intake of thyroid hormone, or iodine-containing drugs.

Radioactive iodine-131 is the most common therapy of Graves' hyperthyroidism and toxic multinodular goiter. The usual doses are 5–15 mCi of <sup>131</sup>I for Graves' disease and 15–50 mCi for large multinodular glands. Beta-adrenergic blockers are used to control symptomatic tachycardia, tremor, anxiety, and muscle weakness, and they are discontinued when the patient is euthyroid. Hypothyroidism is a common consequence after radioiodine treatment, with an eventual incidence of more than 50%.

Definitive therapy with an antithyroid drug, propylthiouracil or methimazole, is appropriate for otherwise healthy elderly patients. The recurrence rate is significantly less with advanced age than in younger patients. Surgical thyroidectomy is only advised if there are obstructive symptoms from a large goiter or the presence of a nodule that is suspicious for malignancy. Although the mortality from subtotal thyroidectomy is very low, the complications of recurrent laryngeal nerve damage and hypoparathyroidism can result in lifelong disability.

### Subclinical Hyperthyroidism

Subclinical hyperthyroidism is defined as a state of suppression of serum TSH with normal free thyroxine and triiodothyronine levels in a patient who lacks clinical features of thyrotoxicosis. The causes of subclinical thyrotoxicosis are the same as those of overt thyrotoxicosis. In a study of patients older than 55 years of age, 0.7% had endogenous subclinical hyperthyroidism. Two meta-analyses found a significant loss of bone density in postmenopausal women with suppressed serum TSH. Subclinical hyperthyroidism has been associated with an increased frequency of nervous symptoms and an increased risk of dementia and Alzheimer's disease.

Therapy should be considered for any patient with subclinical hyperthyroidism who has mental symptoms, osteoporosis, atrial fibrillation, or cardiac disease. A trial of antithyroid drugs to normalize the serum TSH level is warranted. In patients with more severe features, such as atrial fibrillation, ablation of the hyperfunctioning thyroid with radioactive iodine is preferable.

### Thyroid Nodules

Thyroid nodules, either solitary or multiple, increase in frequency with advancing age. Ninety percent of women older than age 60 and 60% of men older than age 80 have a nodular thyroid gland. Thyroid nodules in asymptomatic individuals (incidentalomas) are identified more frequently by ultrasonography rather than by examination of the gland by palpation.

Most thyroid nodules do not cause symptoms. Pain may occur with a hemorrhage into a preexisting colloid nodule or a benign adenoma. Rapid growth over a period of weeks is suspicious of malignancy, and persistent hoarseness may indicate recurrent laryngeal nerve invasion by tumor. A hard and fixed nodule is more likely to be malignant, but many papillary carcinomas or

follicular tumors are soft or cystic. Lymphadenopathy is strongly suggestive of malignancy.

Low serum TSH concentration in the setting of a nodular goiter suggests the presence of either an autonomously functioning adenoma or a toxic multinodular goiter. Positive anti-peroxidase antibody indicates lymphocytic thyroiditis that may present as a nodule. Thyroid ultrasound is capable of identifying palpable nodules as small as 2 mm. The clinical significance of small nodules detected by ultrasonography is uncertain. Solitary incidentalomas larger than 1.5 cm should probably be biopsied under ultrasound guidance. Fine-needle aspiration (FNA) biopsy is the most important diagnostic test, with accuracy, sensitivity, and specificity of 98 or 99%. In a large series of FNA biopsy of thyroid nodules, benign cytology was found in 69% (mainly colloid goiter), malignant cytology in 3.5%, and suspicious cytology in 10%. The suspicious category consists of variants of follicular neoplasm, but follicular adenomas are approximately 10-fold more common than follicular carcinomas. In a patient with a follicular nodule, a radioiodine scan may be helpful. "Hot" or functional nodules are rarely malignant. The presence of nuclear atypia in a follicular lesion has a 44% prevalence of malignancy, and the absence of nuclear atypia denotes a benign lesion. Positive immunostaining for galectin-3 correlates with malignancy; immunostaining for galectin-3 and other proteins may improve the differential diagnosis of suspicious lesions.

Treatment of the thyroid nodule depends on the functional state of the nodule and cytologic diagnosis. If the cytology indicates malignancy or is strongly suspicious for malignancy, the nodule should be removed surgically. In the 10% of suspicious cytologic findings, approximately one-fourth of patients who go to surgery are found to have a malignant lesion. Altogether, only 6% of thyroid nodules are malignant. The hyperfunctioning hot nodule is treated with radioiodine ablation or surgery. The vast majority of thyroid nodules are benign and should be managed medically. Medical management with thyroxine suppression therapy is based on the assumption that growth of the nodule depends on TSH. Spontaneous regression of thyroid nodules may occur. Use of suppressive therapy of benign thyroid nodules has been challenged in the past few years due to the failure of some studies to show a significant decrease in nodule size and concern about reducing mineral bone density. However, several studies have shown >50% reduction in nodule size in 40% of patients with a single nodule. Generally, patients are followed by palpation at intervals of 3 months. Ultrasound examination

provides more objective assessment of growth or shrinkage of a nodule.

## THYROID CANCER

Thyroid cancer accounts for 1.6% of all new cancers in the United States and causes 0.4% of all cancer deaths. It is classified into five major types: papillary, follicular, medullary, anaplastic, and thyroid lymphoma. Most thyroid cancers are indolent and grow slowly over years, whereas a few grow aggressively and cause death within 1 year. Thyroid carcinomas tend to be more aggressive and poorly differentiated in the elderly compared to younger patients.

Papillary carcinoma accounts for 80% of all thyroid cancers, and follicular carcinoma accounts for 10%. These differentiated cancers are more aggressive in older patients. Hurthle cell carcinoma is considered a variant of follicular thyroid carcinoma and carries an even worse prognosis. Extension of the tumor through the thyroid capsule and into the surrounding structures is associated with poorer prognosis. Cervical lymph node metastases occur in approximately 50% of patients with papillary carcinoma and are associated with only a slightly higher rate of recurrence and mortality.

Surgery, either near total or total thyroidectomy, is the initial treatment of choice for patients with differentiated carcinoma. Near total thyroidectomy is performed for extensive unilateral tumors with local metastases. Total thyroidectomy is performed for patients with extensive multifocal disease with metastases to the cervical lymph nodes, contiguous neck structures, or distant sites. The main disadvantage of total thyroidectomy is the higher incidence of hypoparathyroidism.

Radioiodine therapy is used as an adjunct to surgery to treat patients with residual or recurrent papillary cancer in the neck. Thyroid hormone in a suppressive dose is given after thyroidectomy to reduce the recurrence rate. TSH stimulates growth of thyroid tumors that contain TSH receptors. The dose of thyroxine should be adjusted to keep TSH suppressed without causing clinical thyrotoxicosis. The degree of suppression is based on the staging of the patient. In patients with a good prognosis, TSH should be suppressed to the slightly subnormal range. In patients with worse prognosis, which includes many of the elderly, TSH should be suppressed to  $<0.05$  mU/liter without causing clinical thyrotoxicosis.

Medullary carcinoma accounts for 2–4% of thyroid cancers and is derived from the calcitonin-secreting

parafollicular cells. Elevated serum calcitonin levels establish the diagnosis and correlate with tumor mass. Approximately 80% of medullary carcinomas are sporadic and usually occur after age 50.

Anaplastic carcinoma of the thyroid, the most aggressive and lethal cancer, comprises 2% of all thyroid carcinomas and may be derived from a well-differentiated thyroid carcinoma. Its peak occurrence is in the seventh decade. Treatment includes surgery followed by external radiation and chemotherapy; the survival rate is only 20% at 12 months.

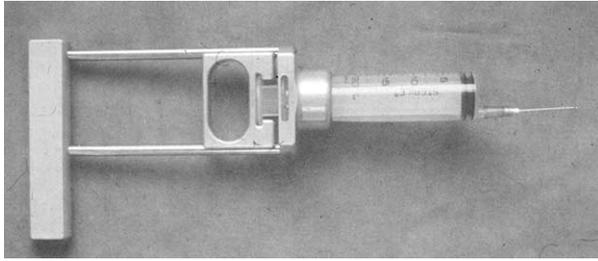
Thyroid lymphoma accounts for approximately 1% of thyroid malignancies and is always accompanied by chronic lymphocytic thyroiditis. The usual clinical presentation is a rapidly enlarging thyroid mass in an older woman with a long history of Hashimoto's thyroiditis. Definitive diagnosis usually requires an open biopsy. Treatment with external radiation and four to six courses of chemotherapy almost always produces a permanent remission.

## See Also the Following Articles

Graves' Disease • Hyperthyroidism, Subclinical • Hypothyroidism, Subclinical • Thyroid Carcinoma • Thyroid Disease and Pregnancy • Thyroid Disease, Epidemiology of • Thyroid Disease, Genetic Factors in • Thyroid Nodule

## Further Reading

- Bagchi, N., Brown, T. R., and Parish, R. F. (1990). Thyroid dysfunction in adults over age 55 years. *Arch. Intern. Med.* **150**, 785–787.
- Cooper, D. S. (2001). Subclinical hypothyroidism. *N. Engl. J. Med.* **345**, 260–266.
- Doucet, J., et al. (1994). Does age play a role in clinical presentation of hypothyroidism? *J. Am. Geriatr. Soc.* **42**, 984–986.
- Gharib, H., and Goellner, J. R. (1993). Fine needle aspiration biopsy of the thyroid: An appraisal. *Ann. Intern. Med.* **118**, 282–289.
- Hershman, J. M., et al. (1993). Serum thyrotropin and thyroid hormone levels in elderly and middle-aged euthyroid persons. *J. Am. Geriatr. Soc.* **41**, 823–828.
- Ladenson, P. W., et al. (2000). American Thyroid Association guidelines for detection of thyroid dysfunction. *Arch. Intern. Med.* **160**, 1573–1575.
- Lin, J. D., et al. (2000). Characteristics of thyroid carcinomas in aging patients. *Eur. J. Clin. Invest.* **30**, 147–153.
- Nordyke, R. A., Gilbert, F. I., Jr., and Harada, A. S. (1988). Graves' disease. Influence of age on clinical findings. *Arch. Intern. Med.* **148**, 626–631.
- Toft, A. D. (2001). Subclinical hyperthyroidism. *N. Engl. J. Med.* **345**, 512–516.
- Yamada, T., et al. (1994). Age-related therapeutic response to antithyroid drugs in patients with hyperthyroid Graves' disease. *J. Am. Geriatr. Soc.* **42**, 513–516.



**Figure 1** The syringe pistol used for thyroid FNA. Different disposable syringes and needles can be used.

strongly to the better selection of suspicious thyroid nodules that really need to be assessed by FNA cytology. When FNA is performed, aspirated material is placed onto a glass slide, fixed using standard techniques (95% alcohol, spray fixed or air-dried), and stained using Papanicolaou, hematoxylin and eosin, or May–Grünwald–Giemsa (MGG) staining methods. Cytological preparations fixed in cold acetone for 10 min can be used for a wide variety of immunocytochemical studies.

## FNA IN NORMAL THYROID IN METABOLIC DISORDERS AND THYROIDITIS

### Normal Thyroid

The thyroid gland is a bilobate structure located in the midportion of the neck, in front to the larynx and trachea. The lobes are connected medially by an isthmus, which lies across the trachea anteriorly below the level of the cricoid cartilage. In approximately 40% of cases, a narrow projection of thyroid tissue, the pyramidal lobe, may be present. This lobe, extending upward from the isthmus and lying on the surface of the thyroid cartilage, represents the vestige of the thyroglossal duct. The normal weight of the thyroid gland in a middle-aged adult is 15–25 g.

Microscopically, the thyroid gland is characterized by a follicular structure. The functional unit of the thyroid gland is the follicle, which is a ring-like structure lined by cubical follicular cells. The lumen of the follicles contains an eosinophilic viscous material known as colloid, in which concentrated thyroglobulin is present. Each follicle is surrounded by a richly vascularized stroma. The diameter of thyroid follicles is approximately 200  $\mu\text{m}$ , but there is considerable variation in size, depending on the functional activity of the gland and age of the patient. Consequently, the height of the follicular epithelium varies from 3 to

20  $\mu\text{m}$ . Flattened follicular cells and larger follicles are visible when the gland is relatively inactive, whereas cubical follicular cells and smaller follicles are generally observed in the functionally active gland. Follicular cubical cells are involved in secretion of colloid into the follicular lumen and may contain apical secretory vacuoles. Follicular columnar cells resorb thyroglobulin-rich colloid and excrete active thyroid hormones into blood vessels. In this case, basal vacuoles can be observed. The parafollicular cells (also called C cells) differ from thyrocytes in their morphology, function, and location. They are larger, triangle-shaped cells with clear cytoplasm and secrete calcitonin.

Parafollicular C cells are located along the central axes of the upper two-thirds of each lateral lobe and comprise less than 0.1% of the total thyroid cellular mass. The cytologic appearance of normal thyroid is variable, depending on the functional state of the gland and the age of the patient. Aspirated material typically consists of follicular cells generally associated with scant colloid. Sometimes, thyrocytes are dispersed or arranged in small clusters. Well-constituted follicular structures are rarely observed. Normal thyrocytes show a pale cytoplasm, a centrally located oval or round nucleus containing fine granular chromatin, and one or two small nucleoli. Colloid appears as proteinaceous material, which can be stained from gray-green to pink using the Papanicolaou method, red-violet using the MGG method, and pink using conventional hematoxylin and eosin staining.

### Metabolic Disorders

#### **Amyloidosis**

Extracellular amyloid deposit in the thyroid gland can be detected in the majority of cases of secondary amyloidosis as a localized form of disease or as part of a systemic disease. Cytological material obtained by FNA may contain scattered normal thyrocytes and fragments of cyanophilic amyloid, which can be stained with Congo red and visualized by using a polarized-light microscope.

#### **Hemochromatosis**

In this pathological condition, thyroid can be involved by massive iron overload, which confers a brown color to the gland. The cytoplasm of follicular cells, as well as macrophages and fibroblasts, show intense deposit of hemosiderin. This yellowish-brown pigment can be stained blue by the Perls staining method. Focal iron deposits, which should not be confused with hemochromatosis, can be detected in areas of

hemorrhage that are commonly present in large adenomas, carcinomas, and nodular goiters, particularly after a previous thyroid FNA evaluation.

## Thyroiditis

This inflammatory condition of the thyroid is common in clinical practice. Three different types of thyroiditis may be distinguished: acute, subacute, and chronic. However, the nosological spectrum of thyroiditis is much more complex if autoimmune thyroid disorders are included. For the latter, there is no internationally accepted classification. The cytological features of the most common thyroiditis are described here.

### Acute Thyroiditis

Acute (suppurative) thyroiditis is rare. The pathogenesis is infectious; bacteria (gram-positive cocci) are the most common agents involved, but fungus- (*Candida*, *Aspergillus*, and *Cryptococcus*), virus- (*Cytomegalovirus*), and protozoa (*Pneumocystis carinii*)-induced acute thyroiditis have been also observed in immunosuppressed hosts.

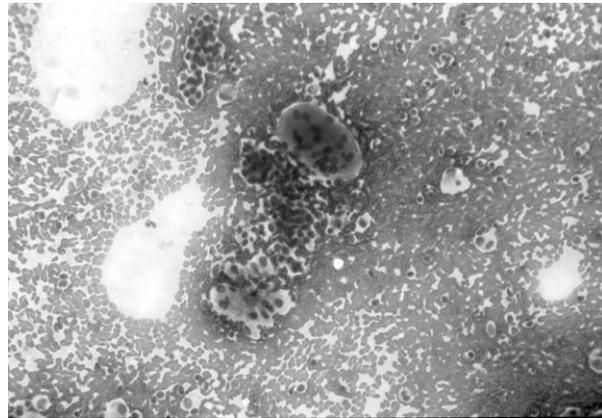
In acute suppurative thyroiditis, cytological material obtained by FNA contains normal follicular cells, sometimes with degenerative changes, abundant neutrophils, and macrophages. In the background, cellular debris and fibrin are observed.

### Subacute Thyroiditis

Subacute granulomatous or de Quervain's thyroiditis is a nonspecific, self-limited inflammatory condition of the thyroid that is thought to be of viral origin. It may account for 5% of thyroid abnormalities. This form of thyroiditis occurs mostly in middle-aged woman, who present with painful swelling of the gland and fever, generally associated with hyperthyroidism-related symptoms. Subacute thyroiditis commonly resolves spontaneously in 2 or 3 months. A typical cytological feature at the onset of de Quervain's thyroiditis is the presence of multinucleated giant cells, grouped in granuloma-like arrangement. Degenerated thyrocytes, epithelioid cells, lymphocytes, and granulocytes on a background containing cellular debris and colloid complete the cytological picture.

### Chronic Thyroiditis

This term includes a heterogeneous group of lesions showing histological as well as pathophysiological differences: tuberculosis (TBC)-related granulomatous thyroiditis and other chronic granulomatous thyroiditis; Riedel's thyroiditis; focal lymphocytic thyroiditis; and chronic autoimmune thyroiditis, including Hashimoto's thyroiditis and its related entities,



**Figure 2** Chronic granulomatous thyroiditis after thyroid surgery. A multinucleated giant cell (foreign body giant cell) in granuloma-like arrangement, scattered normal follicular cells, mature lymphocytes, epithelioid histiocytes, and blood cells are visible (Papanicolaou staining;  $\times 250$ ). Kindly provided by M. R. Giovagnoli, Sant' Andrea Hospital, Rome, Italy.

Graves' disease, and other autoimmune thyroid disorders.

TBC-related chronic granulomatous thyroiditis is caused by mycobacterium tuberculosis and can be diagnosed during the course of secondary tuberculosis. On FNA, specific epithelioid granulomas with caseification are present. Different types of granulomas can be observed in other chronic thyroid diseases as well as in de Quervain's thyroiditis, sarcoidosis, mycosis, vasculitis, syphilis, and after thyroid surgery. In the latter, foreign body giant cells are commonly observed (Fig. 2).

Riedel's thyroiditis is a chronic fibrosing multifocal inflammation involving thyroid as well as extrathyroid structures. The thyroid shows a diffuse fibrosis with petrous consistence. Follicular structures are replaced by sclerosis; consequently, FNA material is very poor in follicular cells, whereas inflammatory cells and fibroblasts are prevalent. This condition needs to be distinguished from other fibrosing thyroid lesions (i.e., the late stage of Hashimoto's disease, malignant fibrous tumors, scleroderma, postradiation thyroid fibrosis, and posttraumatic fibrosis).

Focal lymphocytic thyroiditis may be associated with several thyroid (i.e., nodular goiter and papillary carcinomas) and not thyroid diseases. Histologically, thyroid gland shows focal areas of inflammatory cells, represented mostly by mature lymphocytes. FNA is usually not diagnostic; scattered lymphocytes and typical follicular cells may be observed.

Manifestations of autoimmune thyroiditis encompass a wide spectrum of functional abnormalities and morphological changes of the gland. Although some

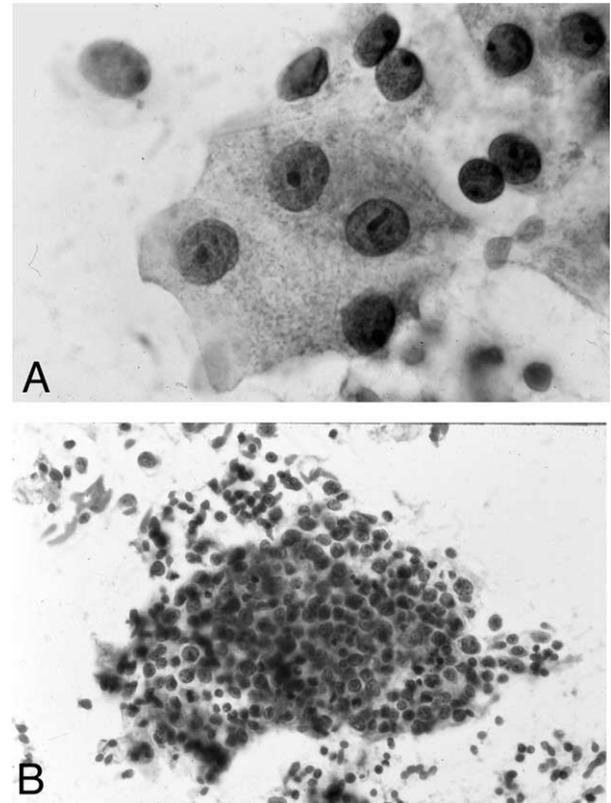
clinical and histological features may be considered specific for certain autoimmune thyroid diseases, at least three anatomoclinical entities probably represent different stages of the same disease. Patients with chronic autoimmune thyroiditis present with hypothyroidism, goiter, or both. The hallmark of autoimmune thyroid diseases is the presence of anti-thyroid peroxidase antibodies and anti-thyroglobulin antibodies. In most cases, the presence of such autoantibodies precedes the clinical onset of the disease. General diagnostic criteria for autoimmune thyroiditis include a positive test for circulating thyroid autoantibodies, elevated serum concentration of thyrotropin (TSH), and the presence of lymphocytic infiltration of the thyroid gland.

In patients affected by autoimmune thyroiditis, FNA should be carefully evaluated and mostly directed at analyzing clinically suspicious nodules or rapidly enlarging goiters. In fact, during autoimmune thyroiditis, follicular cells may show oxyphilic changes and cytological atypia, which may be easily misinterpreted as malignancy, leading to unnecessary surgery. Cytological features of the various manifestations of autoimmune thyroid diseases are described in detail next.

Hashimoto's thyroiditis was first described by Hashimoto in 1912, who reported four patients with goiter histologically characterized by diffuse lymphocytic infiltration, atrophy of follicles, fibrosis, and eosinophilic changes (oxyphilic-metaplasia) in groups of follicular cells, also called Hürthle cells or oncocytes. This disease affects women more frequently than man, and it may be associated with hypothyroidism, euthyroidism, or, occasionally, hyperthyroidism. Hashimoto's thyroiditis is considered the most common form of noniatrogenic hypothyroidism.

In its classical presentation, hypertrophic lymphocytic thyroiditis, the thyroid gland may show a hypofunctional goiter. Histologically, thyroid follicles are small and atrophic and contain scant colloid. Frequently, oxyphilic cells are present. These cells have abundant eosinophilic and granular cytoplasm, due to the presence of a large number of mitochondria, and pleomorphic nuclei (Fig. 3A). Hyperplastic solid areas of oncocyte cells may be present. An intense infiltration of lymphoplasmocytic cells, sometimes organized in germinal centers, regenerative follicles, and interstitial fibrosis, complete the histological picture.

Cytological smears obtained by FNA commonly consist of a large number of mature lymphocytes and plasma cells but also centrofollicular lymphoid cells. Of course, this lymphoid infiltrate is not specific for Hashimoto's thyroiditis, but when it surrounds small



**Figure 3** Hashimoto's thyroiditis. (A) Groups of oncocytes showing fine granular cytoplasm and mild pleomorphic nuclei. (B) Incipient Hashimoto's thyroiditis, in which mature lymphocytes and plasmacytes surround a small cluster of normal follicular cells, showing slight pleomorphic nuclei. (Papanicolaou staining; A,  $\times 1000$ ; B,  $\times 250$ ). Kindly provided by M. R. Giovagnoli, Sant' Andrea Hospital, Rome, Italy.

clusters of follicular and oncocyte cells, showing pleomorphic nuclei, it is quite suggestive for this disease. Fibrosis and the destruction of follicular cells are of variable intensity and evolution. Some Hashimoto's goiters become atrophic over years.

The so-called juvenile lymphocytic thyroiditis (adolescents' goiter) is viewed as the incipient stage of Hashimoto's thyroiditis. This diffuse goiter is difficult to distinguish from euthyroid sporadic goiter. Lymphoid cell infiltration is mild, oncocyte changes are focal or absent, and thyroid autoantibodies are not constant. Consequently, FNA will show normal lymphoid cells and a variable amount of thyrocytes. Scattered Hürthle cells may be observed sporadically (Fig. 3B).

Fibrous lymphocytic thyroiditis generally represents the final stage of Hashimoto's thyroiditis and is observed at an older age. The goiter is mild and associated with severe hypothyroidism. Histologically, the thyroid shows severe fibrosis, marked follicular

atrophy, oncocytic and squamous metaplasia, and lymphocytic infiltration. FNA yields little material consisting of fibroblasts, follicular cells, oncocytic cells, and lymphocytes.

Although the aforementioned cytological features may be variably observed in all different forms of Hashimoto's thyroiditis, clinical and laboratory evaluations will support the cytological diagnosis.

Sometimes, Hashimoto's thyroiditis may present as a palpable thyroid nodule, which can result from hyperplasia of follicular and/or oncocytic cells, heavy focal lymphoid infiltrate, the presence of fibrosis, or the development of cancer. It is noteworthy that association with a true thyroid carcinoma or lymphoma has been reported in approximately 20% of cases of Hashimoto's thyroiditis. This possibility should be always considered during the evaluation of FNA. An accurate evaluation of thyroid FNA cytology is very important to avoid unnecessary surgical procedures. In fact, because nuclear pleomorphism is easily observed in thyrocytes and/or oncocytic cells derived from this inflammatory condition, a misdiagnosis of thyroid cancer is not rare. Moreover, in the course of hypertrophic lymphocytic thyroiditis, in which lymphoid infiltrate is the main feature and follicular cells are scantily represented, FNA cytology may simulate a malignant lymphoma.

Graves' disease should not be considered an autoimmune thyroiditis. Diffuse symmetrical goiter, hyperthyroidism, and exophthalmos characterize this disorder. Graves's disease is not an all-thyroid disease since it also involves the musculoconjunctival retrobulbar orbital content and the subcutaneous tissue in the anteroexternal part of the legs, foot dorsum, and first metacarpal area (pretibial myxedema and hand acropachy). Thyroid hyperfunction is mediated by stimulating anti-TSH receptor antibodies. Histologically, follicles are lined by columnar or cubical thyrocytes showing numerous resorption vacuoles. Lymphoid infiltrate is occasionally seen.

For this pathological condition, FNA does not represent a useful diagnostic procedure because diagnosis is generally obtained on clinical bases. Cytological smears by FNA show typical follicular cells and scant colloid. The cytoplasm is pale and finely granular, and marginal vacuoles are frequently observed. When MGG staining is used, these vacuoles show pink, granular material (flame cells). These cytological features should be not considered specific for Graves' disease; they are common findings in all hyperfunctional thyroid conditions.

The spectrum of autoimmune thyroid diseases also includes radiation-induced thyroiditis, postpartum

thyroiditis, autoimmune polyendocrine syndromes involving the thyroid gland, and iatrogenic autoimmune thyroiditis. For all these conditions, thyroid FNA cytology is not useful for the clinical diagnosis.

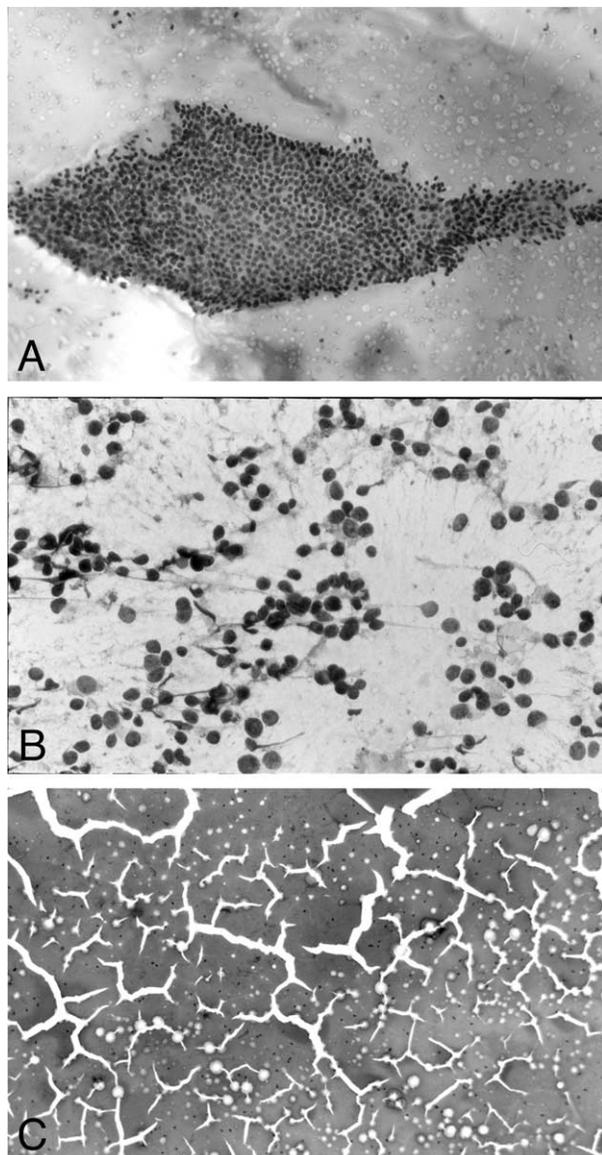
## FNA IN BENIGN AND MALIGNANT FOLLICULAR LESIONS

Hyperplasia is the most common thyroid disease, and it can be considered a compensatory response of the thyroid gland to thyroxin deficit. Different types of hyperplastic lesions can be distinguished. One type is diffuse colloid hyperplasia, in which proliferating follicular cells surround small or medium-sized follicles filled with little colloid. When thyrocytes cease to proliferate, an increased amount of colloid may be detected within follicles, which become larger. The thyroid gland is uniformly enlarged, and cytological smears obtained by FNA show typical follicular cells and abundant colloid.

Nodular hyperplasia is a common condition in which poorly delimited nodules of different sizes enlarge the thyroid gland asymmetrically. Hyperplastic nodule can be solitary, but the most frequent clinical picture is the so-called multinodular goiter. This condition is very common in iodine-deficient areas. Histologically, different types of hyperplastic nodules can be considered: colloid or hypocellular nodules, adenomatoid or hypercellular nodules, and papillary nodules. In hypocellular nodules, follicles are large and thyrocytes cubical or flat. FNA will provide abundant colloid, sparse follicular cells sometime arranged into follicles or monolayers (Fig. 4), scattered lymphocytes, and blood cells. When colloid is dense, typical cracked aspects can be appreciated after conventional staining. (Fig. 4C). Follicular cells show pale cytoplasm, homogeneous nuclei, uniformly distributed fine chromatin, and inconspicuous nucleoli. Isolated cells may look like lymphocytes due to their cytoplasm fragility.

In adenomatoid nodules, a large number of small follicles lined with cylindrical cells are present. In this case, FNA-cytology shows scanty colloid and a large number of follicular cells often arranged in monolayers and small follicles. Thyrocytes show abundant cytoplasm and small variation in nuclear size. Nuclear pleomorphism and prominent nucleoli may be observed in older patients and in patients treated with antithyroid drugs. In adenomatoid nodules associated with hyperthyroidism, cytoplasmic vacuoles may be observed.

Adenomatoid nodular hyperplasia, follicular adenoma, and well-differentiated follicular carcinoma represent the "gray lesions" of thyroid FNA cytology.



**Figure 4** Nodular hyperplasia. (A) Colloid nodular hyperplasia showing a monolayer of normal and uniform follicular cells in a background of thin colloid. Scattered lymphocyte-like nuclei of isolate thyrocytes are visible. (B) Nodular hyperplasia showing numerous typical follicular cells, both isolate or forming small follicles. Nuclei are small and chromatin is finely distributed. Colloid is not present. (C) In nodular colloid goiter, abundant and dense colloid shows a cracked effect in conventional cytological preparations (Papanicolaou staining; A and C,  $\times 100$ ; B,  $\times 250$ ). Kindly provided by A. Vecchione, University La Sapienza, Rome, Italy.

In fact, no definitive and reliable cytomorphological criteria exist for discriminating among these entities.

In nodules with papillary hyperplasia, papillary structures (tufts) are observed histologically and normal thyrocytes arranged in papillary pattern may

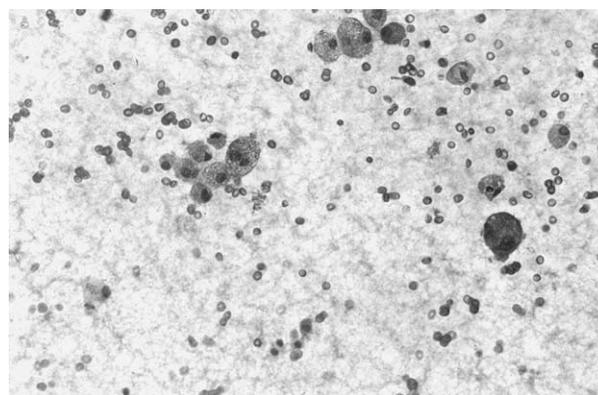
be visible in cytological smears. This condition must be differentiated from papillary carcinomas. Occasionally, nuclear grooving and intranuclear cytoplasmic inclusions, which are considered cytological features of papillary carcinomas, may be observed in cellular smears from nodular goiters.

Dysormonogenic nodular hyperplasia is a congenital condition caused by a defect in the biosynthesis of thyroid hormones. A thyroid goiter develops in the first months of life. FNA shows scant colloid, a large number of follicular cells arranged in groups, and marked anisokaryosis. Preoperative cytological diagnosis of this condition is impossible without clinical information. This cytological picture can be easily misinterpreted as follicular neoplasm or as a poorly differentiated carcinoma.

In general, all the aforementioned hyperplastic conditions may be associated with regressive phenomena, ischemic necrosis, giant multinucleated cells, foamy macrophages, inflammatory cells, fibrosis, calcification, and even ossification. Hyperplastic nodules associated with single or multiple cystic spaces (colloid cysts) are a common finding.

Aspirates from colloid cysts yield a variable amount of liquid, which may be dark brown and dense (chocolate cyst), hemorrhagic, or yellow. Cytological smears may contain sparse follicular cells showing degenerative changes and a large number of foamy macrophages (Fig. 5).

The preoperative diagnosis of follicular neoplasms represents the “gray zone” in thyroid FNA cytology. In fact, distinction between hyperplastic/adenomatous nodules, well-differentiated follicular carcinomas, and follicular variants of papillary carcinoma is difficult



**Figure 5** Cystic nodule. FNA material shows scattered follicular cells with degenerative changes (vacuolated cytoplasm) and foamy macrophages. Numerous red blood cells are present in the background (Papanicolaou staining;  $\times 400$ ).

using this method, even in the hands of cytologists with extensive experience. Because of the specific criteria required for the diagnosis of follicular carcinoma, particularly the unequivocal demonstration of capsular penetration and vascular invasion, well-differentiated follicular malignancies are also sometimes difficult to distinguish from adenomas histologically. For this reason, when thyroid FNA is performed on follicular lesions, the diagnostic report is never definitive and is commonly worded "follicular nodule not otherwise specified." As a consequence, the majority of these lesions are referred to surgery more for diagnosis rather than for therapeutic necessity. The social cost of this clinical approach to the preoperative characterization of thyroid nodules is quite high for patients as well as for the health care system, considering that less than 10% of the resected lesions will be definitively classified as carcinomas. Attempts to improve the diagnostic accuracy of FNA cytology for follicular thyroid nodules, following strict instructions for obtaining adequate specimens or including clinical parameters (i.e., gender, dimension of the nodule, and character of the gland by palpation), have been made, but no clinical, radiological, or laboratory tests have been demonstrated to be sensitive and specific enough to reliably distinguish whether a follicular lesion identified by FNA is benign or malignant. The possibility of improving the diagnostic accuracy of thyroid FNA cytology using immunocytochemical detection of tumor-associated antigens is discussed later.

Among follicular adenomas, different histological types are recognized: normofollicular, macrofollicular (or colloid), microfollicular (or fetal), solid-trabecular (or embryonic), hyalinizing trabecular, atypical, oncocytic (or Hürthle cell adenoma), and other rare variants. Histologically, these lesions are clearly encapsulated, and the adjacent thyroid parenchyma is compressed.

Follicular carcinoma represents 12–15% of thyroid carcinomas in the general population; however, in areas of endemic goiter and iodine deficiency, this type of thyroid cancer is much more frequent. Two histological types of follicular carcinoma are recognized: a microinvasive form, in which a minimal capsular invasion can be observed on histological bases, and a widely invasive form, showing prominent capsular and/or vascular invasion. These tumors may have normo-, micro-, or macrofollicular or trabecular morphology.

As discussed previously, there are no distinctive cytological features that can be used for a reliable distinction between benign and malignant follicular lesions. FNA cytology from normo- and macrofollicular

adenomas is also undistinguishable from that of the aforementioned hyperplastic nodule. When normo- and microfollicular adenomas are evaluated by FNA, overlapping cytological characteristics as those for well-differentiated follicular carcinomas are the rule.

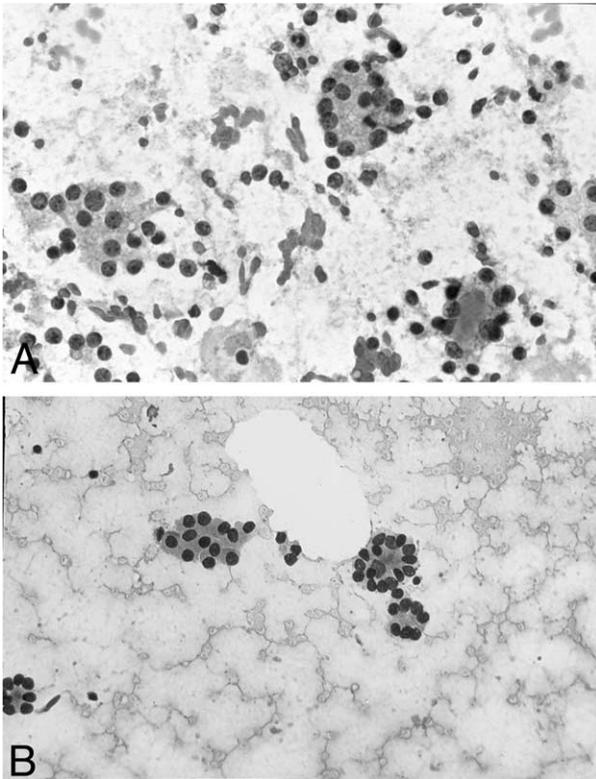
In general, the presence of large pools of colloid in cytological smears may be suggestive of benign follicular proliferation; in fact, follicular carcinomas are generally associated with the presence of scanty colloid. However, normo- and macrofollicular carcinomas in which follicular cells are arranged in large follicles filled with abundant colloid have been reported. In thyroid follicular malignancies, both cellular characteristics and nuclear features are not diagnostic.

In fact, bizarre follicular cells with marked cytological atypia and anisokaryotic nuclei may be observed in FNA from the so-called atypical adenomas. Although the presence of prominent nucleoli is strongly suggestive of carcinoma, their absence does not rule out a thyroid malignancy. Therefore, distinction between benign follicular lesions and well-differentiated follicular carcinoma cannot be supported by specific and exclusive cytomorphological criteria (Fig. 6).

Thus, for the reasons mentioned previously, FNA cytology of follicular thyroid nodules cannot be considered a reliable diagnostic procedure. The use of monoclonal antibodies specific for tumor-associated molecules has consistently improved FNA performance.

Oncocytic cells (also called Hürthle cells, oxyphil cells, or Askanazy cells) derive from metaplastic changes of follicular cells. These cells can be observed in association with several types of thyroid disease, including Hashimoto's thyroiditis, hyperplastic goiters, adenomas, and carcinomas. However, adenomas and carcinomas may arise from oncocytic cells, showing histological features of follicular (Fig. 6B) or papillary neoplasms (Fig. 7C). Cytological smears obtained by FNA usually show a rich population of oncocytic cells with little or no colloid. These cells show little cohesiveness; they can be dispersed or form small clusters. Oncocytic cells show a fine granular eosinophilic cytoplasm with well-defined contour and large, eccentric, and slightly pleomorphic nuclei. As mentioned for the other follicular neoplasms, also for the oncocytic follicular lesions FNA cytology has a limited ability to distinguish between adenoma and follicular carcinoma.

A high percentage of oncocytic cells, a monomorphic aspect of the smear with small nuclear



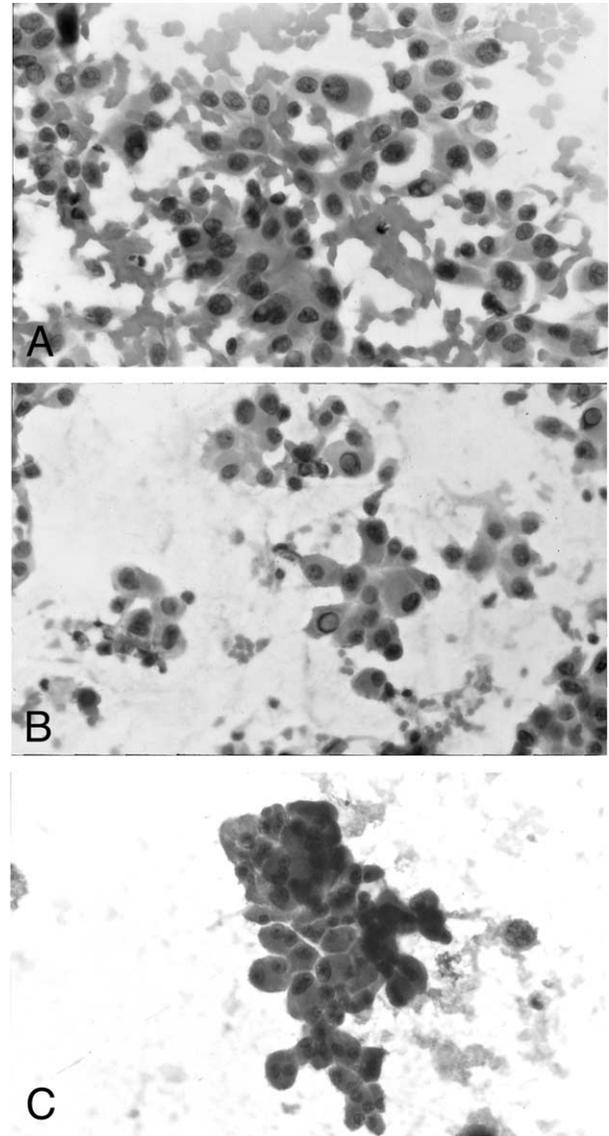
**Figure 6** (A) In this not otherwise specified follicular lesion, thyrocytes are both dispersed and organized in small follicular structures. Their cytoplasm is pale with poorly defined limits. Nuclei are round and uniform, and chromatin appears fine granular and uniformly dispersed. Some inconspicuous nucleoli are visible. (B) In this oncocyctic follicular lesion, oxyphilic cells with granular eosinophilic cytoplasm, hyperchromatic, and eccentric nuclei are arrayed in microfollicular structures. No reliable cytological criteria of malignancy can be defined (Papanicolau staining;  $\times 250$ ). Kindly provided by M. R. Giovagnoli, Sant' Andrea Hospital, Rome, Italy.

pleomorphism, and the absence of lymphocytic infiltrate with stromal and cellular debris help to differentiate oncocyctic neoplasms from thyroid lesions containing nonneoplastic oncocyctic cells (i.e., in Hashimoto's thyroiditis). Ischemic necrosis is a common finding in oncocyctic tumors after FNA biopsy.

## FNA IN OTHER THYROID MALIGNANCIES

### Papillary Carcinoma

Papillary carcinoma is the most common type of thyroid cancer, accounting for 65–80% of thyroid cancers in the United States. This malignancy is generally diagnosed in young females, and in most cases a clear association with radiation exposure can be



**Figure 7** Papillary carcinoma. (A) Numerous neoplastic cells arranged in micropapillary structures. Cells show irregular nuclei and enlarged cytoplasm. Cytoplasmic nuclear inclusions and grooves are visible in some cells. (B) In this papillary carcinoma, nuclear inclusions are easily identified. (C) Papillary Hürthle cell carcinoma composed by large polygonal cells with granular cytoplasm and well-defined contours. Nuclei are slightly eccentric and polymorphic and show fine granular chromatin. Nucleoli are visible in some cells (Papanicolau staining;  $\times 400$ ). Kindly provided by M. R. Giovagnoli, Sant' Andrea Hospital, Rome, Italy.

demonstrated. Histologically, the two morphological hallmarks of papillary carcinomas are the presence of papillae and nuclear changes. Papillae consist of a central fibrovascular stalk lined by neoplastic epithelium formed by cubical, columnar, or tall follicular cells. These structures may show a complex branching

configuration. Psammoma bodies are frequently associated with thyroid papillary carcinomas. These round, calcific concretions exhibit concentric laminations and probably derive from necrotic tumor cells on which successive layers of calcium salt have been deposited. Abundant fibrous stroma and inflammatory infiltration are also common features of papillary carcinoma. The nuclei of the papillary carcinoma cells exhibit characteristic changes, which represent very useful diagnostic markers. These nuclear changes are ground glass appearance, presence of pseudo-inclusions, and presence of irregular outlines and grooves. Both papillae and nuclear changes are easily visible on histology in the classical forms of papillary carcinomas.

FNA from papillary carcinomas yield highly cellular smears. Colloid is generally scant or absent, but if present it is very dense and shows the so-called "bubble gum" appearance.

Papillary structures or small clusters of neoplastic cells arranged in micropapillae may be observed together with monolayers and dispersed thyrocytes. Follicular structures are occasionally seen, and they are frequently observed in the follicular variant of papillary carcinoma. Papillary carcinoma cells can be cubical, columnar, or pleomorphic. They have abundant cytoplasm and occasionally may show oncocyctic changes. The nuclei are round, oval, and polymorphic and contain finely granular chromatin and inconspicuous nucleoli. The aforementioned nuclear changes are generally visible on the majority of cytological preparations, with the exception of nuclear clearing with ground glass appearance (called orphan Annie's eyes). This effect, which represents an artifact of fixation and embedding, is only visible in fixed and paraffin-embedded tissues. Although these features can be considered diagnostic when observed in the presence of well-defined papillary structures (Fig. 7), they are neither constant nor specific. In fact, nuclear changes can be observed in other primary thyroid malignancies, in metastatic tumors, and, unfortunately, also in nonneoplastic thyroid lesions (i.e., in Hashimoto's thyroiditis).

Several histological variants of papillary carcinoma have been described. Papillary microcarcinoma is an occasional finding during postsurgical examination of the thyroid as well as during autopsy. This malignancy is less than 1 cm, but it can be discovered as a cervical lymph node metastasis. FNA evaluation of the primary tumor is generally uncommon.

The follicular variant of papillary carcinoma shows an exclusively or almost exclusively follicular pattern of growth. As discussed before, for follicular thyroid lesions, FNA interpretation is very difficult. In fact, the follicular variant may be easily confused

with nodular hyperplasia, follicular adenoma, and well-differentiated follicular carcinoma.

Cystic papillary carcinoma is a cystic neoplasm that should always be considered when FNA from large cystic lesions identifies a fluid unusually rich in follicular cells or when the lesion is still palpable after FNA drainage of the thyroid cyst.

Diffuse sclerosing papillary carcinoma represents a papillary malignancy in which a strong desmoplastic reaction of the stroma is present. In this lesion, marked inflammatory infiltration and squamous metaplasia are common findings.

The tall-cell variant of papillary carcinoma is characterized by cylindrical cells with cyanophilic and eosinophilic cytoplasm and a high nucleus:cytoplasm ratio. Papillary structures are visible and are lined by pseudostratified columnar cells. Nuclear incisions are rare and pseudo-inclusions are not present. As previously discussed, papillary carcinoma may show oncocyctic changes. When this aspect is prevalent, the occurrence of a papillary oncocyctic tumor should be considered. In this case, papillary structures are lined with neoplastic oncocyctic cells, granular eosinophilic cytoplasm, and pleomorphic and eccentric nuclei (Fig. 7C).

### Poorly Differentiated (Insular) Carcinomas

This histological type of thyroid carcinoma arises from follicular cells, but some authors consider it to be the undifferentiated solid form of both papillary and follicular carcinoma. In fact, insular carcinomas with aspects of transition toward a papillary or follicular cancer are sometimes observed on histology, suggesting the possibility that these malignancies may represent a continuous spectrum of the same disease. Interestingly, insular carcinoma, although common in central Europe, is very rare in the United States.

Histologically, poorly differentiated thyroid carcinomas are characterized by solid nests of neoplastic cells of variable size, which are surrounded by thin stromal septa. FNA from these lesions shows a large number of malignant cells organized in large, medium, and small clusters or trabeculae. Dispersed elements are also present, and loss of cellular cohesion is evident. Neoplastic cells contain enlarged round nuclei with granular chromatin and prominent nucleoli. Sometimes, microfollicular structures may be observed. The cytological background presents cellular debris, necrosis, and no colloid. Nuclear changes similar to those observed in papillary carcinomas may occasionally be seen.

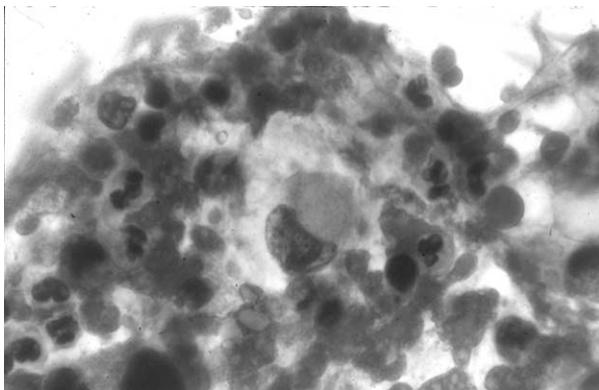
## Anaplastic Carcinoma

This tumor represents less than 10% of all thyroid cancers. It is a very aggressive malignancy composed by undifferentiated cells. Anaplastic carcinoma is generally observed in older patients and can be considered exceptional in patients younger than age 50.

This tumor grows fast and produces a rapid thyroid enlargement with compression of adjacent anatomic structures. Histologically, squamous-epithelioid, fusiform, and giant neoplastic cells can predominate or be observed in different proportions. Extensive necrotic areas are frequently seen. The hallmarks of this tumor are nuclear pleomorphism, frequent atypical mitoses, necrosis, and vascular invasion. Differential diagnosis with the rare primitive thyroid sarcoma may be necessary in some instances. FNA cytology is variable, depending in the predominant cell type. However, pleomorphic and polymetric neoplastic cells are easily recognized in a dirty background in which cellular debris, necrotic material, and blood cells are present. Irregular nuclei and prominent nucleoli are common findings (Fig. 8).

## Medullary Carcinoma

This malignancy, which constitutes approximately 12% of all thyroid cancers, does not derive from follicular epithelium but from calcitonin-producing cells (parafollicular C cells). It can occur sporadically or as a familial disease, often in association with other endocrine neoplasms (MEN syndrome). The presence of medullary carcinoma is almost invariably associated with a high serum level of calcitonin. This



**Figure 8** Anaplastic carcinoma. Neoplastic cells show enlarged polymorphic and polymetric nuclei. An inflammatory component rich in neutrophils and neutrophilic cannibalism are visible (Papanicolou staining;  $\times 1000$ ).

serological parameter is a useful indicator of disease. Histologically, medullary carcinoma shows polymorphic and polymetric cells with uniform round or oval nuclei and granular amphophilic or eosinophilic cytoplasm, which are arranged in nests and cords. Connective tissue that surrounds the neoplastic cell nests is generally rich in amyloid. These neoplasms are not capsulated and may be multifocal. Cellular smears obtained by FNA contain isolated and loosely cohesive neoplastic cells, which are highly variable in size and shape. They include round, plasmacytoid (with eccentric nucleus), fusiform, epithelioid, and sometimes multinucleated and binucleated cells. In some cases, nuclear inclusions may be observed. If present, amyloid appears as amorphous material, which specifically stains with Congo red and can be visualized under a polarized light microscope. Differential diagnosis with anaplastic carcinomas and oncocytic neoplasms can be difficult. However, the presence of isolated polymorphic cells, plasmacytic-like cells, and a high calcitonin serum level are diagnostic for medullary carcinoma.

Calcitonin-producing neoplastic cells can be visualized by using immunocytochemical techniques and specific monoclonal antibodies to calcitonin.

## Lymphoma

Primary lymphoma of the thyroid gland is rare, but approximately 15% of patients with systemic lymphoma present secondary thyroid involvement. When lymphoma is present, a monomorphic population of lymphoid cells with sparse typical follicular cells can be obtained by FNA. Differential diagnosis with thyroiditis (particularly Hashimoto's thyroiditis) should always be considered. Immunocytochemical demonstration of a monoclonal lymphocyte population using specific monoclonal antibodies to differentiate clusters of lymphocyte antigens will confirm the diagnosis.

## Metastasis to the Thyroid Gland

Several malignancies can produce thyroid metastasis by contiguous invasion of primary tumors (e.g., larynx and pharynx) or after hematogenous dissemination of other malignant tumors (e.g., melanoma, lymphoma, and carcinomas from lung, breast, kidney, colon, etc.). In all these cases, FNA cytology will show variable populations of neoplastic cells mixed with normal thyrocytes. The clinical history is helpful for formulating a correct diagnosis.

## ANCILLARY TECHNIQUES FOR IMPROVING THE DIAGNOSTIC ACCURACY OF THYROID FNA CYTOLOGY

Ultrasound evaluation of the thyroid gland is an important diagnostic tool with which FNA cytology should always be integrated. The ability to predict malignancy of nonfollicular neoplasms has been demonstrated by using a multiple logistic regression analysis based on five echographic features: margin, shape, echo structure, echogenicity, and the presence of calcification. Interestingly, in a preliminary study in which an ultrasonographic contrast agent called Levovist was used, it was possible to distinguish carcinomas from benign nodules by comparing the times of appearance of the contrast.

The use of specific monoclonal antibodies (mAbs) directed to tumor-associated antigens has contributed greatly to the improvement in diagnostic accuracy of conventional cytology for several neoplastic diseases. Interestingly, Galectin-3, a  $\beta$ -galactosyl-binding protein normally expressed in macrophages, is detectable in follicular thyroid cells only after malignant transformation. Using a specific mAb to galectin-3, it is possible to distinguish between benign (galectin-3-negative) and malignant (galectin-3-positive) thyroid nodules. This immunocytochemical approach to the preoperative evaluation of thyroid nodules has been recently validated in a large international, multicenter study. Sensitivity and specificity of immunodetection of galectin-3 in discriminating between benign and malignant thyroid lesions are 99 and 98%, respectively, and the positive predictive value and diagnostic accuracy are 92 and 99%, respectively. The integration of galectin-3 immunostaining with conventional cytomorphological and clinical diagnostic procedures represents a sensitive and reliable diagnostic approach for the preoperative identification of thyroid carcinomas. This method improves the diagnostic accuracy of conventional cytology, particularly for distinguishing between benign and malignant follicular lesions.

HBME-1 identifies a monoclonal antibody (mAb) directed to a biochemically unknown epitope expressed in mesothelioma and in several adenocarcinomas. Although this reagent seems to be focally reactive with some hyperplastic thyroid lesions also, it is commonly used in clinical practice as a diagnostic tool for thyroid cancer detection.

Molecular and immunochemical approaches directed to identify specific gene profiles and related

proteins in thyroid malignancies have been extensively reported in the recent literature. Several experimental data, mostly arising from molecular analysis of papillary thyroid carcinomas, show a clear pattern of over-expressed genes associated with this malignancy. Among these genes, GALS3L (galectin-3), FN1 (fibronectin-1), KRT19 (keratin 19), and MET gene (c-met related protein) have been found consistently over-expressed in papillary thyroid carcinoma.

Forty to 80% of papillary carcinomas are associated with rearrangements of two genes, which are not usually expressed in normal thyrocytes: *RET*, a neurotrophic factor with tyrosine kinase activity, and *NTKR1*, a nerve growth factor receptor with tyrosine kinase activity. Rearrangement of these genes with parts of other genes, or abnormally spliced promoters, leads to a constitutive activation of tyrosine kinase, which may activate uncontrolled cell growth and probably neoplastic transformation. It is noteworthy that these gene rearrangements are very common after exposure to radiation.

Specific translocation [t(2;3) (q13;p25)] involving the thyroid transcription factor PAX8 and the peroxisome proliferator-activated receptor PPAR $\gamma$ 1 results in an aberrant fusion transcript, the expression of which seems to be restricted to follicular carcinomas. If these findings are validated in large multicenter studies, the availability of specific oligonucleotide primers to be used in polymerase chain reactions, as well as of mAbs directed to these aberrant and/or overexpressed proteins, could improve the diagnostic accuracy of conventional thyroid FNA cytology.

## CONCLUSION

The application of thyroid FNA for the preoperative characterization of thyroid nodules has greatly enhanced the ability of the clinician to select appropriate patients for surgery. With this technique, most papillary carcinomas, undifferentiated carcinomas, most forms of thyroiditis, and cystic indolent lesions can be easily characterized. However, even cytologists with extensive experience with thyroid FNA find malignant follicular lesions difficult to distinguish from benign ones because capsular and/or vascular invasion, which are the distinctive hallmarks of follicular cancer, cannot be appreciated on cytological bases. In these cases, additional diagnostic techniques should integrate conventional FNA cytology (i.e., galectin-3 immunostaining).

Although FNA cytology is a useful diagnostic tool for the assessment of thyroid nodules, it has some technical and conceptual limitations. For example, diagnosis of micropapillary carcinoma by FNA is almost impossible because of the small size of the lesion. On the other hand, other interpretative problems due in part to the quality of cytological preparation (e.g., inadequate sampling) but also to the skill of the cytopathologist result in a certain number of false-negative and false-positive results.

As a general consideration, it should be kept in mind that FNA cytology of thyroid lesions has an important diagnostic potential in a specific setting. For this reason, and in order to optimize the clinical management of thyroid nodules, a careful clinical and endocrinological evaluation that includes thyroid ultrasound scanning should always be performed before considering a thyroid FNA. When these guidelines are followed, a significant number of patients will not have any suspicious nodules to be considered for cytological analysis. These patients should be reassured and followed by the referring physician.

## See Also the Following Articles

Graves' Disease • Thyroid Carcinoma • Thyroid Function Tests • Thyroid Imaging • Thyroiditis, Infectious • Thyroiditis, Subacute • Thyroid Nodule

## Further Reading

- Bartolazzi, A., Gasbarri, A., Papotti, M., Bussolati, G., Lucante, T., Khan, A., Inohara, H., Marandino, F., Orlandi, F., Nardi, F., Vecchione, A., Larsson, O., and the Thyroid Cancer Study Group (2001). Application of an immunodiagnostic method for improving the preoperative diagnosis of nodular thyroid lesions. An International multicentre study. *Lancet* **357**, 1644–1650.
- Hamburger, J. I., and Kaplan, M. M. (1996). Evaluation of thyroid nodules by needle biopsy. In "Werner and Ingbar's Thyroid" (L. E. Braverman and R. D. Utiger, eds.), pp. 447–460. Lippincott–Raven, New York.
- Kini, S. R. (1987). The thyroid. In "Guides to Clinical Aspiration Biopsy" (T. S. Kline, ed.). Igaku Shoin, New York.
- Rausch, P., Nowels, K., and Jeffrey, R. B., Jr. (2001). Ultrasonographically guided thyroid biopsy: A review with emphasis on technique. *J. Ultrasound Med.* **20**, 79–85.
- Rosai, J., Carcangiu, M. L., and De Lellis, R. A. (1992). "Atlas of Tumor Pathology: Tumors of the Thyroid Gland," 3rd series, Fascicle 5. Armed Force Institute of Pathology, Washington, DC.

the free fraction can enter the peripheral cell. If the free concentration falls, the pituitary will be stimulated to secrete thyroid-stimulating hormone (TSH) so as to restore adequate free T4 and T3 concentrations. This will occur when free T4 concentrations change due to altered TBG levels. The kinetics of the protein changes is slow (half-life = 6 h for transthyretin and 4 days for TBG).

For a long time, it was believed that the changes in serum concentrations of TBG would not affect thyroid hormone economy. Recently, it has been shown that in treated but severely hypothyroid patients, pregnancy or estrogen treatment increases the demands of T4 treatment and serum TSH values increase significantly. This is the first evidence that the increased total T4 pool under estrogen treatment cannot be achieved solely by decreasing the metabolic rate and metabolism of T4. Therefore, under physiological circumstances, there is a discrete adaptation of thyroidal secretion to ensure an unchanged thyroid function. The therapeutic adaptations are necessary in severely hypothyroid patients, whereas it is rarely necessary to increase T4 treatment in partially hypothyroid patients.

TBG binds T3 with a lower affinity than it does T4, with the affinity constant for T3 being  $5 \times 10^{-8}$  M, and its serum levels are less affected by changes in serum proteins. Despite this, and despite the fact that T3 is the active thyroid hormone, other factors markedly reduce the diagnostic value of T3:

1. T3 is mainly an intracellular hormone, and the circulating pool represents no more than 15 to 20% of its whole body pool.
2. T3's production is mainly nonthyroidal and is markedly influenced by the metabolic state of the organism.
3. T3's production is controlled by monodeiodination, and its degradation also by conjugation.

Monodeiodination plays a crucial role in the control of serum T3 levels. Three enzymes—deiodinase types I, II, and III—have been characterized. They are selenoproteins and, hence, very rare mammalian proteins. However, they differ markedly concerning their functional characteristics. Quantitatively, the most important enzyme is deiodinase type I, which is found mainly in the liver and kidney but in small amounts is ubiquitously present. It is localized at the inner plasma membrane. Under physiological conditions, approximately 30% of T4 is converted by this enzyme to T3. The expression and activity of this enzyme are stimulated by T3; for example, in hyperthyroidism the activity is very high, whereas in hypothyroidism it is very low. However, many other nonthyroidal factors

affect its activity. In any catabolic state—malnutrition, starvation, severe infections, or metabolic or malignant diseases—its activity is decreased and this is reflected in decreased serum T3 levels. Another example is drug interferences such as glucocorticoids and amiodarone, which are potent inhibitors of deiodinases. Stimulation of deiodinase type I can be observed with overfeeding and/or a high-carbohydrate intake.

The activity profile of deiodinase type II is very different. It has a very low affinity constant ( $K_m$ ). The enzyme that is found in the pituitary and brain is most active when T4 is absent. It is inhibited with increasing T4 concentrations. It is directly implicated in the control of serum TSH. Both enzymes attack the outer ring of T4. The third enzyme, deiodinase type III, deiodinates only at the inner ring, a function that can be shared under particular conditions with deiodinase type I. Interestingly, deiodinase type III is localized at the outer cell membrane, and by degrading T4, it refuses its entry into the cell.

Thyroidal production of T3 is minimal compared with that of T4. Without iodine deficiency, T3 represents less than 5% of the secreted T4. There are at least two conditions where this percentage can increase: in iodine deficiency and in hyperthyroidism. In hyperthyroidism, the thyroidal deiodinase is strongly stimulated so that some of the T4 is converted during the process of secretion to T3. In iodine deficiency, this mechanism is also operating, but the poorly iodinated thyroglobulin (Tg) also has a high percentage of T3 that can go up to 50% that of T4. This contributes to a higher proportion of T3 in serum, and the T3-to-T4 ratio can increase from the physiological ratio of approximately 2% to 4% in hyperthyroidism and to much higher values in severe iodine deficiency. If selenium deficiency is associated with iodine deficiency, the increased ratio of T3 to T4 is not seen because the lack of selenium affects the activity of the deiodinases. Clinically, the increased T3-to-T4 ratio was of diagnostic significance for T3 toxicosis that is now diagnosed more adequately by a suppressed serum TSH in the presence of a still normal free T4. For practical purposes, one can conclude that the serum levels of T3 are influenced by many factors other than thyroidal secretion and, therefore, are of lower diagnostic value than free T4 values.

rT3 is the inactive counterpart of T3 and has been advocated as a diagnostic tool, but today none of them remains valid. Nevertheless, it is an interesting parameter that gives a good mirror image of T4-to-T3 conversion. If deiodinase type I is inhibited, serum rT3 levels increase. There are two reasons for this. First, rT3 needs deiodinase type I to be degraded and

accumulates if the enzyme is low. Second, there may be some shift of T4-to-T3 conversion to rT3.

For practical purposes, one can conclude that the serum levels of T3 are influenced by many factors other than thyroidal secretion and, therefore, are of lower diagnostic value than free T4 values.

### Methods of Serum Free T4 and T3 Measures

As mentioned previously, free T4 represents an extremely small fraction of the total T4. Today, for routine purposes, no fully satisfactory free T4 measure is available.

The best but cumbersome methods use ultrafiltration or equilibrium dialysis. The latter method was the first reliable one; however, it can yield falsely elevated values if FFAs are artificially increased by the presence of *in vitro* heparin or by repeated freezing and thawing. Both methods cannot be used in routine work.

Two different approaches for measuring the free T4 indirectly are available. One is still widely used in the United States: the free T4 or T3 index test, which is based on the fact that the product of total serum T4 times the result of the T3 resin uptake gives an arbitrary index that roughly correlates with free T4 values. The T3 resin uptake methods are based on the following principle: because the serum protein binding of T4 is mainly dependent on TBG, an increase of it will shift the equilibrium between TBG and free T4 in favor of TBG. If  $^{125}\text{I}$  T3 or T4 is added, a higher percentage will be bound to TBG, yet in any case the binding to TBG would be so strong and the free fraction of T3 or T4 would be so small that one could not measure it. Therefore, one adds a competitor, such as resins or antibodies, which binds in normal serum 15 to 30% of the  $^{125}\text{I}$  T3. Artifactual results with T3 resin uptake methods are frequent and similar to the ones found using commercial free T4 methods.

The so-called direct measures of free T4 and T3 methods are based on radioimmunoassays (RIAs) or chemiluminescent assays. Yet the free T4 concentrations are far too low to be measured directly. Therefore, commercial companies have developed T4 analogues that do not bind to TBG and are distributed only within the free hormone fraction. In addition, this fraction is artificially increased by competitors of T4 for TBG. These methods have the advantage over the free T4 index in that they yield absolute concentrations of free T4 and their interpretation is easy. However, because of technical limitations, they are also prone to artifactual values, particularly in systemic

illness where unknown competitors for T4 binding to TBG have been described. It follows that some commercial methods will have a tendency of too low values in severe illnesses and with decreased serum protein concentrations. Some of these problems can be overcome by more cumbersome techniques (two-step methods), but even these methods are not free of artifactual results. The latter tend to give too high values in severe illness. Nevertheless, in ambulatory medicine, these commercial methods are satisfactory and render total T4 and T3 methods unnecessary. In addition, the clinician should rely first on serum TSH, with serum free T4 and T3 levels being useful adjuncts for the refinement of diagnosis and treatment and for excluding some extremely rare thyroidal conditions.

These are TSHomas (pituitary tumors secreting TSH) and thyroid hormone resistance. In TSHomas, serum TSH values are normal or slightly increased, yet both serum free T4 and T3 are increased and are associated with clinical signs of hyperthyroidism. In thyroid hormone resistance, serum TSH can be normal, low, or slightly increased in the presence of increased free T4 and T3. However, the clinical picture is composite, with the heart often being the sole organ with signs of hyperthyroidism.

In most instances, serum free T4 and T3 levels are confirmatory. Yet relying only on serum TSH for diagnosis might lead to misleading conclusions because there are rare diseases where serum TSH will not help or may be misleading. In the presence of a high clinical suspicion of thyroid dysfunction or abnormal serum TSH, free T4 and T3 are necessary. Total T4 and total or free T3 are indicated only when the results of serum TSH and free T4 values are discordant. Among thyroid function tests, serum TSH is the first choice.

### Serum TSH

Measurement of serum TSH values is the firstline diagnostic test. Its sensitivity is very high, whereas its specificity is slightly lower (90%) because drugs and severe illness can decrease serum TSH levels in euthyroid patients. The predominant role of serum TSH as a diagnostic parameter is due to its being the most sensitive parameter of peripheral thyroid hormone action. It increases exponentially when serum free T4 decreases. It differs from the clinically more relevant effects of T3 on heart, liver, brain, and other tissues for the following reasons:

1. The intracellular regulation of thyroid hormone metabolism in the pituitary and brain is not encountered in peripheral tissues.

2. The action of thyroid hormones in the pituitary is inhibitory. In the periphery, it was thought that the action of T<sub>3</sub> was mainly stimulatory, yet recent studies (including gene screening) indicate that gene expression is more often down-regulated than stimulated by T<sub>3</sub>.

3. The role of the hypothalamus for TSH secretion by thyrotropin-releasing hormone (TRH), somatostatin, dopamine, and others adds complexity of coordination between T<sub>3</sub> and these other factors.

#### **Clinical Use of Serum TSH Measures**

Serum TSH is measured with immunometric assays using chemiluminescent probes. Progress has been constant in this field, and the lower detection limit of the newest assays is 0.002 mU/L or less (fourth-generation immunometric assays). At 0.02 mU/L, most immunometric assays have a coefficient of variation of 20%, whereas at higher TSH values, this decreases to 5% or less. A coefficient of variation of 20% is the upper limit of what is clinically acceptable.

The clinician should be aware that these assays measure only the immunological, and not the biological, properties of TSH. This is no problem when diagnosing hyperthyroidism, but there is good evidence of discordance between these two parameters in primary and hypothalamic hypothyroidism in the fetus and (possibly) in severely ill patients.

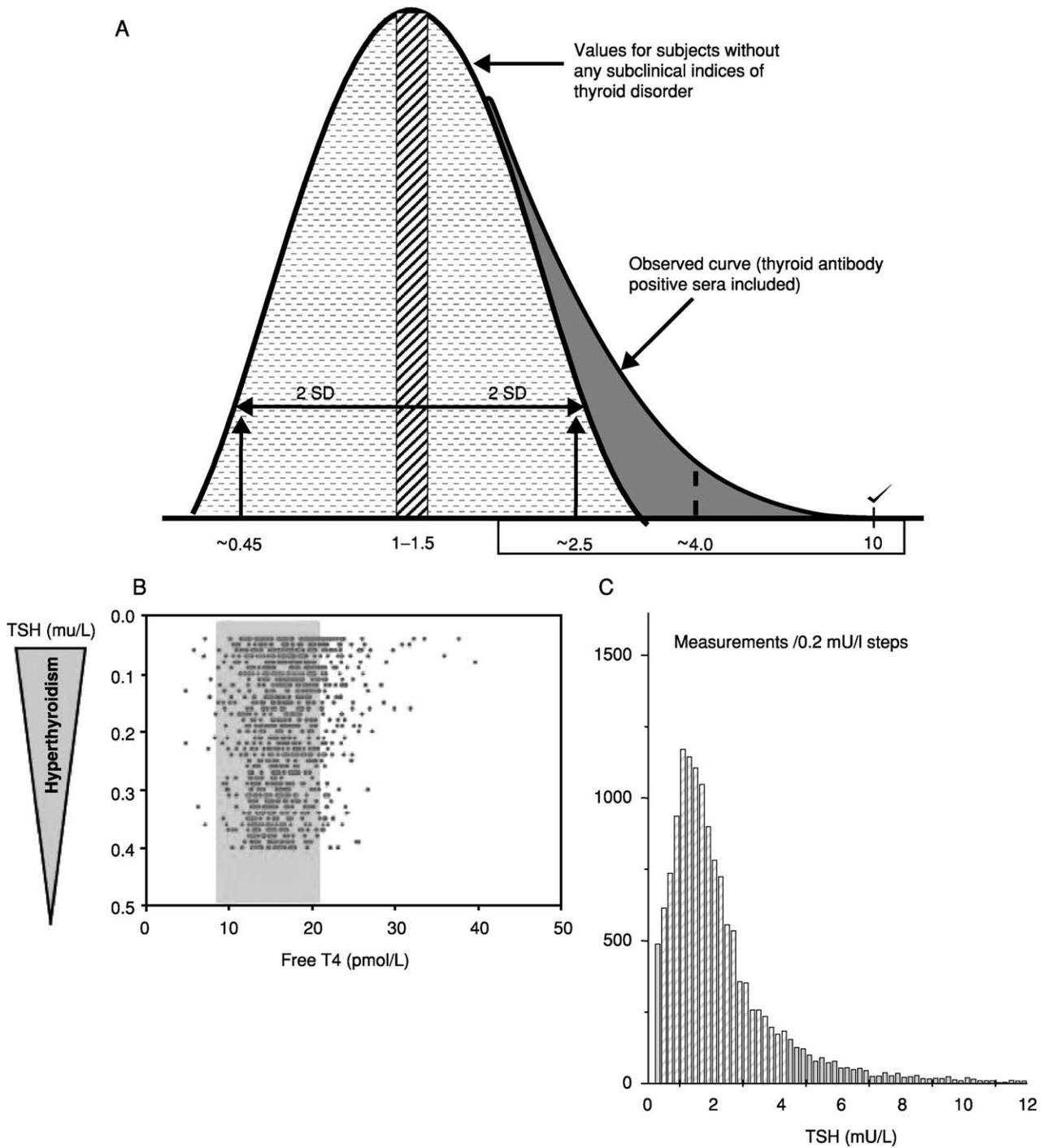
It also appears that antibodies recognize variable degrees of TSH glycosylation that occur in several thyroid pathologies. Different results in the range of serum TSH from 3 to 7 mU/L have been reported and are possibly related to such differences in antibodies.

TSH is secreted in a pulsatile manner and has a circadian rhythm. TSH spikes do not exceed 0.5 mU/ml. Serum TSH levels are highest during the evening (at 23:00 h), that is, during the first hours of sleep. The peak is displaced in night workers. In young individuals, the mean serum TSH levels are in the morning ( $0.9 \pm 0.3$  mU/L) and at 23:00 h ( $1.9 \pm 0.6$  mU/L). In the elderly, serum TSH levels tend to be slightly lower ( $0.7 \pm 0.6$  and  $1.3 \pm 0.9$  mU/L, respectively). These fluctuations are rarely critical for the clinical evaluation of thyroid function. The normal range extends from 0.4 to 4.0 mU/L. Recent studies tend to narrow the normal range to 0.4 to 2.5 mU/L, with a median of 1.0 to 1.5 mU/L and with the African American population having a slightly higher mean serum TSH than the Caucasian population (Fig. 1). This rigorous definition is based on 120 individuals who were tested for absence of goiter, antithyroid antibodies, and family history of thyroid disease. From the clinical point of view, a serum TSH of 0.4 mU/L or less

needs investigation. One study suggested that serum TSH values of 2.5 to 4 mU/L already represent an increased risk of developing overt thyroid disease, with the annual incidence of clinically manifest disease being approximately 5%. The concomitant presence of thyroid autoantibodies probably increases this risk. The significance of low serum TSH depends on the degree of serum TSH suppression that can be classified into three different degrees:

1. *Serum TSH 0.1 to 0.4 mU/L (for some authors, to 0.6 mU/L).* This is frequently seen in subclinical hyperthyroidism. Serum free T<sub>4</sub> values are normal and serum free T<sub>3</sub> values are only rarely increased. Such serum TSH values are not fully specific for subclinical hypothyroidism because they can also be found in severe nonthyroidal illness with or without glucocorticoid or dopaminergic treatment. In these latter situations, serum free T<sub>4</sub> and T<sub>3</sub> levels tend to be decreased or in the lower normal range. The differential diagnosis can be established by clinical observation and repetitive measures. If found repeatedly, this slight decrease of serum TSH is compatible with partial thyroid autonomy, with the cause being a hot nodule, a multinodular goiter, or euthyroid Graves' disease. Thyroid scintigraphy can be helpful in these situations. Recent iodine contamination may play an important role here and can be the origin of transient subclinical hyperthyroidism. Rarely, such serum TSH levels can be found in pituitary or hypothalamic insufficiency, but in these conditions the changes in thyroid function are mostly of secondary importance, with other pituitary insufficiencies being more pronounced.

2. *Serum TSH 0.05 to 0.1 mU/L (for some authors, 0.02 to 0.1 or 0.2 mU/L).* The differential diagnosis is similar to the previous one. In addition, serum free T<sub>4</sub> and/or T<sub>3</sub> levels tend to be at the upper limit or slightly increased. Some authors consider that even a small increase in free T<sub>3</sub> stands for overt and not subclinical hyperthyroidism, whereas others (including this author) prefer that the definition remain mainly clinical. Such serum TSH levels are the goal of T<sub>4</sub> treatment in patients with differentiated thyroid cancer. These patients present with serum free T<sub>4</sub> levels at the upper limit of normal, and in one-third of these patients free T<sub>4</sub> levels will be slightly increased. Serum T<sub>3</sub> levels are in the normal range. This can be explained by the absent thyroidal component of T<sub>3</sub> secretion. Low serum TSH levels are frequent in panhypopituitarism but not in hypothalamic disease. It is rather rare to find such low serum TSH levels in systemically ill patients unless they are treated with large doses of dopamine and/or glucocorticoids.



**Figure 1** Current TSH immunometric assays. (A) The area between 0.45 and 2.5 mU/l represents the values for a population (mean  $\pm$  2 SD) with documented absence of thyroid diseases. The bordering tails indicate that this population is not separated from the borderline pathological cases that represent for TSH levels above 2.5 mU/l in a majority of cases still euthyroid autoimmune thyroiditis (figure courtesy of Professor Carole Spencer). (B) Free T4 values in the subclinical hyperthyroid range as a function of decreasing serum TSH values are shown. Even though some free T4 values are above the normal range, the majority falls within the normal limits. Below TSH values of 0.1 mU/l, free T4 values tend to increase. Increased free T3 values are even less frequent despite partly suppressed serum TSH values (data not shown). (C) Consecutive serum TSH determinations (16,000) over 10 months in a diagnostic laboratory are shown. It includes known and unknown pathologies, the limits being arbitrarily set at 12 mU/l and 0.2 mU/l. It documents what is shown in panel A: one cannot separate three different populations (euthyroidism, subclinical hyper- or hypothyroidism). The free T4 values shown in panel B are extracted from the lower end of TSH values in panel C. The data in B and C were obtained by courtesy of UNILABS, SA, Geneva.

3. *Serum TSH less than 0.02 mU/L.* Such values are encountered only in endogenous hyperthyroidism and possibly in severe panhypopituitarism. Some assays will allow measuring levels as low as 0.002 mU/L. Therefore, even in frankly hyperthyroid patients, in the future, one might evaluate the severity of hyperthyroidism by serum TSH only. There are no data correlating the very low serum TSH levels with the clinical severity of hyperthyroidism. It is now well established that reactivation of TSH secretion during treatment for hyperthyroidism may lag behind the normalization of peripheral thyroid hormones and clinical manifestations by several weeks or even months. There is no clear explanation for this, but it is important to realize that in this situation serum TSH is not a valid parameter of euthyroidism.

#### **Use of TSH Measurements in the Case of Artifacts Free T4 Values**

Two diagnostic pitfalls can occur when serum TSH measurements are the only thyroid hormone parameter measured. Serum levels of thyroid hormones may be normal, slightly decreased, or increased in the case of thyroid hormone resistance with frankly increased serum free T4 and T3 levels.

In the case of a pituitary TSH-secreting adenoma, called a TSHoma, serum TSH levels are normal or moderately increased (up to 20 mU/L) despite clinical hyperthyroidism with increased serum free T4 and T3 levels.

Serum TSH measurements are essential for excluding thyroid dysfunction in two situations with apparently increased free T4 levels. One abnormality involves T4 binding to serum transport proteins. There can be a congenital high affinity and capacity binding of albumin or transthyretin for T4. T3 is much less affected. In addition, some neuroendocrine tumors secrete very large amounts of normal transthyretin (familial dysalbuminemic hyperthyroxinemia [FDH] and familial or acquired hyperthyroxinemia due to transthyretin excess). The affected patients are absolutely euthyroid because their free T4 values, measured by ultrafiltration or equilibrium dialysis, and T4 production rates are normal. Some commercial kits of free T4 measurements can yield artifactually increased very high free T4 values, even though newer versions of these kits avoid such errors. Few laboratories are equipped to identify these proteins. If this methodology is not available, differential diagnosis with pituitary tumors or thyroid hormone resistance should include serum T3, serum TSH  $\alpha$ -subunit, a TRH test, and possibly magnetic resonance imaging

(MRI) of the pituitary. Serum T3 and the  $\alpha$ -subunit are increased in TSHomas and in thyroid hormone resistance, but only exceptionally in dysalbuminemic hyperthyroxinemia. Basal serum TSH and the TRH test are normal in the case of abnormalities in serum transport proteins, whereas the response to TRH tends to be exaggerated in thyroid hormone resistance. In most but not all cases of TSHomas, there will be no response to TRH stimulation. Therefore, the value of this test is only indicative.

Another artifactual increase of free T4 values is encountered in the presence of autoantibodies to T4 or T3, even though most modern kits are no longer affected by such antibodies. Euthyroidism can be confirmed with a TRH test, yet in most countries TRH for clinical testing is no longer available.

#### **Artifacts of TSH Measures**

There is only one condition where serum TSH values can be artifactually increased, namely in the presence of antibodies against mouse immunoglobulin (IgG) in the patient's serum. Monoclonal mouse IgGs are part of all immunometric assays and, therefore, will be trapped by these antibodies. Recently, immunometric assays have overcome this problem. If necessary, the artifactual increase in serum TSH can be diagnosed by omitting anti-TSH antibodies in the assay and/or by a TRH test. It will show a complete nonresponse to TRH that may hint at this artifact or at a TSHoma.

The well-established overwhelming diagnostic power of serum TSH measurements can raise two questions. First, should one restrict the firstline diagnostic parameters of thyroid dysfunction to TSH only, thereby omitting free T4 measures? This might be justified for systematic screening, yet free T4 measurements are inexpensive and will be ordered mostly during further workup of the patient's thyroid disease. Second, should serum TSH be used systematically as a screening test of the adult female population over 50 or even over 35 years of age and be repeated at regular intervals (e.g., every 5 years)? There is already a precedent. The neonatal screening with TSH and the price of serum TSH measurements have decreased markedly and will probably decrease even further, making such a proposal a realistic issue.

#### **Special Techniques**

The main indications for the TRH test (200  $\mu$ g intravenously, blood sampling at 0, 15, 30, 60, and 120 min) are pituitary diseases. For primary thyroid dysfunction, it has lost all of its indications except those mentioned previously. In most countries, intravenous TRH is no longer available, with the nasal spray being

used instead. This way of testing represents a stronger and more prolonged TSH stimulation. Blood should be sampled before and 30 and 60 min after the TRH for measuring TSH or 3 h later for measuring an increase in serum T<sub>3</sub>.

Measuring the  $\alpha$ -subunit of TSH, the common subunit of all pituitary glycoproteins, is important in the case of TSHomas and obviously in gonadotropinomas as well.

## Thyroglobulin

Physiologically, Tg is released in small amounts from the thyroid. It is estimated that 1 g of tissue corresponds to 1  $\mu$ g/L serum Tg if serum TSH is in the normal range. Disruption of the normal thyroid structure results in leakage of Tg, and high serum levels are found in goitrous patients. In multinodular goiter, Tg levels can reach very high levels that overlap with those found in metastatic thyroid cancer patients. Marked and transient increases can be seen after <sup>131</sup>I treatment and after thyroid surgery. High levels are also found in Graves' disease and in functioning (and occasionally even nonfunctioning) thyroid adenomas.

The measurement of serum Tg levels has benefited from the advances in immunometric assays. Today, an assay should have a functional sensitivity (with a coefficient of variance of 20%) of less than 1.0 to 0.5  $\mu$ g/L, and interference of Tg antibodies should be kept to a minimum. In practice, there are considerable differences among the commercial kits, and each laboratory should determine the sensitivity of its assay. Therefore, two additional controls are recommended:

- Tg antibodies should be measured, and some authors recommend not using serum concentrations of Tg if anti-Tg antibodies are present.
- Another control consists in adding pure Tg to a separate sample of the patient's serum. A recovery of 80 to 120% of the added Tg excludes major interference. Adding small amounts of external Tg (1  $\mu$ g/L rather than 50  $\mu$ g/L) is recommended. Some authors consider that the recovery test does not guarantee absence of interference and so reject this approach. It is hoped that technical improvements of immunometric assays will allow most of these problems to be minimized.

## Indications

Tg measurements are useful in neonatology where low levels are found in thyroid agenesis or dysgenesis and where high levels are found in thyroid hormone

resistance and in some rare inborn errors of thyroid hormone synthesis. They are a valuable adjunct to scintigraphy and thyroid ultrasound.

Low Tg values are an excellent means to differentiate between thyrotoxicosis factitia and endogenous hyperthyroidism. In the presence of a normal-sized or even moderately increased thyroid, the ingested L-thyroxine will suppress Tg levels to values of less than 10  $\mu$ g/L. Yet in the presence of very large goiters, Tg levels will often not be suppressible.

Tg values can be of value for following the evolution of hyperthyroidism and during remission. In addition, in subacute thyroiditis, an increase in Tg levels has been documented. The measurement of Tg values in these clinical situations is not frequently used.

Tg measures are not useful in the differential diagnosis between benign nodular goiter and thyroid carcinoma because Tg levels tend to be increased and overlap in both conditions. In an occasional patient with an unidentified pulmonary or bone metastasis, a very high Tg level may help to establish the origin of the tumor.

The greatest merit of Tg measurements comes from the surveillance of differentiated follicular or papillary thyroid cancer because it allows easy detection of recurrences (Table I). Tg levels are a function of the tumor mass and tumor differentiation. Tg levels tend to be lower in tumors that are more aggressive and are absent in undifferentiated cancers.

The surveillance of well-differentiated thyroid cancers with Tg values has become the first choice of approach. It is crucial not to rely only on basal Tg values and to consider values after TSH stimulation as well. The use of recombinant TSH has greatly facilitated the surveillance because its use is equally valuable than inducing severe hypothyroidism.

If Tg values are more than 2  $\mu$ g/L, the presence of thyroid tissue is certain, indicating the presence of neoplastic tissue in the case of complete surgical or <sup>131</sup>I ablation. The changes in its concentration over time parallel the evolution of the tumor. Tg levels are particularly important for diagnosing absence of disease. For this purpose, one cannot rely on nonmeasurable Tg levels (<0.5  $\mu$ g/L) under T<sub>4</sub> treatment. If the patient is rendered hypothyroid, or if the patient receives recombinant TSH and serum Tg values do not exceed the critical limit of 2  $\mu$ g/L, the patient can be considered cured. In these situations, thyroid scintigraphy becomes obsolete.

Recently, a complementary method for determining the presence of tumor tissue has been advocated in cases of interference in serum measurements of Tg by anti-Tg antibodies. The presence of circulating

Table 1 Diagnostic Use of Serum Thyroglobulin Levels in Well-Differentiated Thyroid Cancer

Thyroglobulin value		Thyroglobulin antibodies	<sup>131</sup> I scintigraphy	Loco-regional tumor	Distant metastasis	"Cure"
TSH <0.1 mU/L	TSH >30 mU/L					
	<2 µg/L	Negative	Negative	Excluded	Excluded	Most likely
<2 µg/L		Negative	—	Not excluded	Unlikely	Not proven
2 to 10 µg/L		Negative	Positive or negative	Certain	Unlikely but possible	No
	2 to 10 µg/L	Negative	Positive or negative	Certain	Unlikely	No
>10 µg/L		Negative or positive	Mostly positive	Certain	Not excluded	No

thyroid follicular cells is obtained by using real-time quantitative reverse transcription–polymerase chain reaction (RT-PCR) based on the amplification of Tg mRNA. If normal thyroid tissue has been destroyed, the presence of Tg mRNA strongly supports recurrence of disease. The method is sensitive and independent of circulating anti-Tg antibodies, but it currently is limited to specialized research laboratories.

### Autoantibodies

For clinical use, three types of antibodies are of interest: the antithyroperoxidase (anti-TPO), the antithyroglobulin antibodies (anti-Tg), and the anti-TSH receptor antibodies (anti-TRAb). Among the anti-TRAb, one can distinguish blocking and stimulating antibodies. The recently discovered anti-sodium-iodide symporter antibodies do not seem to play a diagnostic role.

The thyroid antibodies belong to the IgG class, and only for anti-TPO antibodies have some complement fixing and cytotoxic activity been described in rare cases. The antigenic sites of the TPO have been well characterized; there are six of them. For anti-Tg and anti-TRAb, this is less well known because the antigenic sites are mainly conformational and not linear. Recently, it has been possible to produce functioning anti-TRAb in mice.

Indications to measure thyroid antibodies can serve diagnostic, follow-up, and screening purposes. TPO antibodies are the most sensitive antibodies for the diagnosis of any thyroid autoimmune process, and they are widely used in combination with the TSH for diagnosing the etiology of a thyroid disease. Unfortunately, the absolute values are not standardized and there are differences among the many commercial products. Their presence confers an increased risk of developing hypothyroidism and is an indication

for clinical follow-up. They are also useful in the follow-up of Graves' disease because their titer may reflect the activity of the disease. In Hashimoto's thyroiditis, the follow-up of these patients with measurements of antibodies may be confusing and is not generally recommended. There is no specific indication for Tg antibodies because they are identical with the ones enumerated previously. However, on rare occasions, the TPO antibodies may be negative, whereas Tg antibodies will reveal thyroid disease.

The TRAb are mostly not useful for the diagnosis of Graves' disease. Similarly to TPO antibodies, their titer may follow the activity of the disease. They can also be present in primary hypothyroidism and in postpartum thyroiditis; therefore, they are not diagnostic of Graves' disease.

There is one seldom but clear indication for measuring these antibodies, namely in pregnant patients with Graves' disease in whom the thyroid has been removed. In these cases, one has no clinical indication about the severity of the autoimmune process and, on rare occasions, placental transfer of a high titer of TRAb may cause neonatal hyperthyroidism.

Screening with anti-TPO antibodies is advocated during pregnancy because their presence indicates increased risk of postpartum thyroiditis.

### IN VIVO ISOTOPE IMAGING

Diagnostic *in vivo* isotopic imaging will yield information on thyroidal uptake of the isotope and functional morphology of the thyroid. Two isotopes, 123-iodine (<sup>123</sup>I) and Tc99m pertechnetate (Tc<sup>99m</sup>), are used nearly exclusively. There is still some indication for <sup>123</sup>I for whole body scintigraphy of patients with thyroid cancers. The place for other isotopes, such as thallium and gallium, has not been established. Thallium has been used for identifying thyroid

Table II Isotopes for Thyroid Imaging

	$^{123}\text{I}$	$^{131}\text{I}$	$\text{Tc}^{99\text{m}}$	F18 thioglucose
Half-life	13 h	8 days	6 h	110 min
Emission	Gamma	Beta, gamma	Gamma	Gamma
Radiation exposure (rad)/mCi	13	1300	0.13	
Diagnostic dose ( $\mu\text{Ci}$ )	300	50 (or 5 to 10,000)	2000	
Route	Oral	Oral	Intravenous	Intravenous
Absorption	Gastric	Gastric		
Particularities				Uptake by metabolically active cells
Maximal uptake	4 to 8 h	4 to 8 h	20 to 40 min	

metastases, and gallium 67 is taken up by inflammatory cells, particularly mast cells, and therefore can be used in some cases of destructive thyroiditis such as subacute thyroiditis (Table II).

### **123-Iodine**

The physical characteristics of  $^{123}\text{I}$  are given in Table II. It is close to an ideal isotope in that the irradiation of the thyroid is minimal. Fasting, and with intact gastric acidity the iodide, is taken up specifically by the gastric mucosa. Thyroidal uptake occurs very rapidly. In the case of concomitant food intake, the uptake is delayed and occurs by intestinal absorption.

Its uptake is dependent on the activity of the sodium-iodide symporter, formally called iodide trapping of the thyroid. Once transported into the cell, the free intrathyroidal iodide is immediately incorporated into Tg. Only under pathological conditions is the free iodide pool of the thyroid increased. This can be tested by the perchlorate discharge test. Once stored in the colloid as Tg, the labeled T4 will remain in the thyroid for a long time, depending on the total amount of stored colloid. In most conditions, the reserve in Tg is huge; therefore, the turnover of this iodide is slow, even though the most recently formed colloid is preferentially resorbed and secreted. As a rule of thumb, only 1 to 2% of the accumulated dose will be secreted per day. Therefore, the disappearance of the isotope will be mainly a function of its physical half-life (13 h for  $^{123}\text{I}$ ). During secretion, most of the  $^{123}\text{I}$  is secreted as T4 and to a lesser extent as T3, and only a minor part of iodide is lost.

Iodine contamination is one of the major technical problems resulting in invalid thyroid imaging. In individuals with a normal thyroid and a daily iodide intake of 200  $\mu\text{g}$  or more, a single dose of 30 mg iodide together with the tracer will reduce the uptake

to background levels. A similar effect can be obtained with half the dose of iodide given over several days. However, it should be noted that this applies for a completely normal thyroid. In areas of moderate iodine deficiency where small multinodular goiters are frequent, the suppression of  $^{123}\text{I}$  uptake by iodine needs higher and longer treatment with similar or higher doses. Accidental iodine contamination is a frequent event because large amounts of iodine containing contrast media are injected during axial tomography. These substances are highly water soluble and are eliminated within 4 to 6 weeks. Iodide is rapidly eliminated by the renal route (clearance of 50 ml/min), and in renal insufficiency they are correspondingly decreased.

Today, the major culprit of long-term iodine contamination is amiodarone. This substance and its biologically active metabolite, desethylamiodarone, have an approximate half-life of 4 to 6 weeks. However, iodine contamination from its degradation products can last for months or up to 1 to 2 years. Historically, other organic iodine compounds can be responsible for even longer lasting contaminations. For instance, dyes used for myelography give rise to lifelong contamination and can even pass the placenta with consequent low thyroid uptake in children.

Among the noniodinated substances interfering in thyroidal uptake, methimazole, propylthiouracil, and perchlorate should be mentioned, with the latter being a specific inhibitor of the iodide symporter. In severe nonthyroidal illness, thyroid function and thyroid uptake may be depressed. Drugs such as glucocorticoids (30–60 mg of prednisone) and dopamine and its analogues have multiple impacts, but the most significant one is a reduction of TSH secretion with a subsequent reduction in thyroid function. Older literature also mentions sulfonamides, but the more modern drugs in this class have not been reported to interfere.

### Tc<sup>99m</sup> Pertechnetate

Tc<sup>99m</sup> is close to ideal for studying the trapping of iodide by the thyroid. It is easily available and of low cost. Because it can only give information on anion trapping, the measurements are performed within 1 h after intravenous injection. The thyroidal uptake of Tc<sup>99m</sup> is approximately 10 times lower than that for <sup>123</sup>I (0.4–4.0%). The precise normal range is also dependent on iodine intake; therefore, it must be validated for each laboratory. As compared with <sup>123</sup>I, Tc<sup>99m</sup> may occasionally show uptake of cold nodules that may erroneously be taken for functioning nodules.

It should be noted that on many occasions thyroid imaging with isotopes has been replaced by ultrasonography. The indications for scintigraphy could be summarized as follows. In hyperthyroidism, Tc<sup>99m</sup> is as valid as <sup>123</sup>I for the differential diagnosis of nodular or diffuse goiter. If the thyroid is diffusely enlarged without concomitant exophthalmos or pretibial myxedema, the positive uptake will allow excluding silent and/or subacute thyroiditis and will help in establishing a diagnosis of thyrotoxicosis factitia and/or iodine contamination. In the case of multinodular goiter with a moderate decrease of serum TSH (<0.6 mU/L), the scintigraphy may confirm autonomous regions. In the case of very painful thyroid and inflammatory symptoms, an absent thyroid uptake may confirm the diagnosis of subacute thyroiditis. In the case of an isolated euthyroid thyroid nodule, the uptake will identify a cold nodule, even though Tc<sup>99m</sup> may give erroneous results and the predominant position of this investigation has been superseded by cytology.

In most cases, ultrasound evaluation of the thyroid gland is as useful as or better than scintigraphy. Nevertheless, the two examinations can be complementary. Scintigraphy will allow identifying an inactive nodule of the thyroid that the ultrasound may reveal to be cystic, mixed, or solid. In multinodular goiter, the scintigraphy will clearly give the patchy appearance of the functioning and (possibly) autonomous tissue. However, its value in describing inactive or cold nodules in this situation is limited.

### The Role of <sup>131</sup>I-iodine

The disadvantage of the <sup>131</sup>I isotope is its rather large delivered dose of irradiation to the patient. In thyroid cancer patients, <sup>131</sup>I has kept its place for diagnostic procedures, even though the use of Tg is reducing its indications. During the follow-up of these patients, whole body scintigraphies are obtained with <sup>131</sup>I

during intense TSH stimulation that is classically obtained by stopping T4 or T3 substitution or, more recently, by injecting recombinant human TSH.

Hypothyroidism is obtained by switching 7 weeks before <sup>131</sup>I treatment to T3 substitution that has to be stopped 12 to 14 days before giving the dose. Serum TSH levels should increase to more than 30 mU/L, and the efficiency of the whole body scan or treatment can be increased by concomitant regimen recommendations such as avoidance of iodized salt and iodide-rich bread, eggs, and fish. A small dose of diuretics (thiazides) can also be recommended to decrease the circulating iodide pool. Large diagnostic doses of <sup>131</sup>I are used (5–10 mCi) because the uptake of the cancerous cell is much lower than normal and is mostly markedly below 0.5% of the administered dose. The whole body scintigraphy is performed 3 to 4 days after administration of the radioactive iodide. A <sup>131</sup>I scintigraphy or therapy should not be repeated before 8 to 12 weeks because it induces a stunting effect (transiently abolished uptake).

This technique is very sensitive for detecting small metastases and in most cases correlates very well with the presence of serum Tg levels. However, it has been reported that there can be some dissociation and that serum Tg levels can be increased even in the absence of residual <sup>131</sup>I iodide uptake. It is thought that these patients have metastasis and should benefit from a more complete workup with subsequent surgery and/or <sup>131</sup>I treatment.

### Rare and/or Experimental Methods for Thyroid Imaging

The perchlorate discharge test can be performed only with <sup>123</sup>I or <sup>131</sup>I. It is a rarely used test that was described initially for detecting inherited defects of intrathyroidal iodine metabolism (absence of sodium-iodide symporter, absence of Tg, and dehalogenase defect). Its diagnostic value was for acquired small iodine organification defects that can be seen in autoimmune thyroiditis and Graves' disease treated with <sup>131</sup>I. Today, this test has no place in routine diagnostic procedures.

The methodology is as follows. One should measure the uptake and give 1 g of perchlorate 2 h after giving <sup>123</sup>I. One should measure the uptake again 1 h later. A decrease of more than 30% is significant. The test can be made more sensitive by giving 500 µg of potassium iodide (KI) together with the <sup>123</sup>I.

Positron emission tomography with fluoro-deoxyglucose is a very promising technique for detecting metabolically active tissue. The active principle,

deoxyglucose, can be taken up by cells but cannot be metabolized. Malignant cancer cells have a high glucose metabolism. Initially, it was thought that this test would yield the best results in poorly differentiated, and hence metabolically very active, tumors. However, experience has shown that even differentiated thyroid cancers can occasionally be identified with this technique.

### See Also the Following Articles

Antithyroid Drugs • Drug Effects and Thyroid Function • Hypothyroidism, Diagnosis of • Thyroglobulin • Thyroid Fine Needle Aspiration Cytology • Thyroid Hormone Metabolism • Thyrotoxicosis: Diagnosis • TSH (Thyroid-Stimulating Hormone; Thyrotropin)

### Further Reading

- Ayala, A. R., Danese, M. D., and Ladenson, P. W. (2000). When to treat mild hypothyroidism. *Endocrinol. Metab. Clinics North America* **29**, 399–415.
- Hollowell, J. G., Staehling, N. W., Flanders, W. D., *et al.* (2002). Serum TSH, T4, and thyroid antibodies in the United States population (1988 to 1994): National Health and Nutrition Examination Survey (NHANES III). *J. Clin. Endocrinol. Metab.* **87**, 489–495.
- Kobayashi, H., Sakurai, A., Katai, M., and Hashizume, K. (1999). Autosomally transmitted low concentration of thyroxine-binding globulin. *Thyroid* **9**, 159–163.
- Kohrle, J. (1999). Local activation and inactivation of thyroid hormones: The deiodinase family. *Mol. Cell Endocrinol.* **151**, 103–119.
- Losa, M., Giovanelli, M., Persani, L., Mortini, P., Faglia, G., and Beck-Peccoz, P. (1996). Criteria of cure and follow-up of central hyperthyroidism due to thyrotropin-secreting pituitary adenomas. *J. Clin. Endocrinol. Metab.* **81**, 3084–3090.
- Mandel, S. J., Larsen, P. R., Seely, E. W., and Brent, G. A. (1990). Increased need for thyroxine during pregnancy in women with primary hypothyroidism. *N. Engl. J. Med.* **323**, 91–96.
- Meier, C. A., and Burger, A. G. (2000). Effects of drugs and other substances on thyroid hormone synthesis and metabolism. In “The Thyroid” (L. E. Braverman and R. D. Utiger, eds.), pp. 265–280. Lippincott Williams & Wilkins, Philadelphia.
- Mori, Y., Jing, P., Kayama, M., *et al.* (1999). Gene amplification as a common cause of inherited thyroxine-binding globulin excess: Analysis of one familial and two sporadic cases. *Endocrinol. J.* **46**, 613–619.
- Nebert, D. W., and Russell, D. W. (2002). Clinical importance of the cytochromes P450. *Lancet* **360**, 1155–1162.
- Savagner, F., Rodien, P., Reynier, P., Rohmer, V., Bigorgne, J.-C., and Malthiery, Y. (2002). Analysis of Tg transcripts by real-time RT-PCR in the blood of thyroid cancer patients. *J. Clin. Endocrinol. Metab.* **87**, 635–685.
- Spencer, C. A. (2000). Serum thyroglobulin measurements: Clinical utility and technical limitations in the management of patients with differentiated thyroid carcinomas. *Endocrinol. Pract.* **6**, 481–484.
- Spencer, C. A., Takeuchi, M., and Kazarosyan, M. (1996). Current status and performance goals for serum thyrotropin (TSH) assays. *Clin. Chem.* **42**, 140–145.
- Spencer, C. A., Takeuchi, M., Kazarosyan, M., *et al.* (1998). Serum thyroglobulin autoantibodies: Prevalence, influence on serum thyroglobulin measurement, and prognostic significance in patients with differentiated thyroid carcinoma. *J. Clin. Endocrinol. Metab.* **83**, 1121–1127.
- Stockigt, J. R. (2001). Free thyroid hormone measurement: A critical appraisal. *Endocrinol. Metab. Clinics North America* **30**, 265–289.
- Vanderpump, M. P., Tunbridge, W. M., French, J. M., *et al.* (1995). The incidence of thyroid disorders in the community: A twenty-year follow-up of the Whickham Survey. *Clin. Endocrinol. (Oxf)* **43**, 55–68.
- Wang, W. P., Larson, S. M., Fazzari, M., *et al.* (2000). Prognostic value of [18F]fluorodeoxyglucose positron emission tomographic scanning in patients with thyroid cancer. *J. Clin. Endocrinol. Metab.* **85**, 1107–1113.
- Ward, G., McKinnon, L., Badrick, T., and Hickman, P. E. (1997). Heterophilic antibodies remain a problem for the immunoassay laboratory. *Am. J. Clin. Pathol.* **108**, 417–421.
- Weiss, R. E., and Refetoff, S. (2000). Resistance to thyroid hormone. *Rev. Endocrinol. Metab. Disorders* **1**(1–2), 97–108.

midline of the floor of the primitive pharynx, in the presumptive sublingual region. At E 9.5, the thyroid primordium buds from the primitive pharynx and begins to migrate caudally, forming an endodermal-lined diverticulum that descends in the tissue in front of the neck. At this stage, the migrating thyroid primordium is still connected to the pharyngeal floor by a narrow channel, the thyroglossal duct. At E 11.5, the thyroglossal duct disappears and the thyroid primordium loses all connections with the floor of the pharynx and reaches its final destination, in front of the trachea, by E 13.5. Then, the thyroid expands and follicular organization appears. By E 14.0, the differentiation process of TFCs begins and thyroid-specific genes are expressed. Finally, 2 days later, thyroxine is detected in the fetal thyroid.

## MOLECULAR ASPECTS OF THYROID GLAND DEVELOPMENT

During the past few years, it has become possible to identify a number of genes expressed in various stages of thyroid development and their role(s), as deduced by the phenotypes obtained in mouse models where such genes have been mutated (inactivated) by gene targeting.

The discovery that the transcription factors TTF-1 (thyroid transcription factor-1), Foxe1 (forkhead box e1, formerly called TTF-2), Pax8 (paired box gene 8), and Hhex (hematopoietically expressed homeobox) are expressed not only in mature thyroid cells but also in their precursors offered a useful tool for the exploration of the genetic basis of the developmental process of the thyroid gland.

At E 8.5, in the primitive pharynx, the endodermal cells fated to become TFCs are univocally specified by the expression of TTF-1, Foxe1, Pax8, and Hhex. These factors are also present in other embryonic tissues, but all four are coexpressed only in the presumptive thyroid bud from the moment a thickening composed by proliferating cells appears in the midline of the floor of the primitive pharynx. When the thyroid diverticulum forms and begins its migration, the thyroid primordium still expresses TTF-1, Foxe1, Pax8, and Hhex, and the simultaneous presence of these four factors will remain for the rest of the life as a hallmark of differentiated TFCs.

The expression of TTF-1, Foxe1, Pax8, and Hhex is necessary but not sufficient to guarantee the morphogenesis of the thyroid. Other genes also are required for the correct development of the gland. Among these, Tshr (thyroid-stimulating hormone receptor) seems to be involved in the final steps of differentiation of TFCs (Table I).

### TTF-1

TTF-1 was first identified as a nuclear factor able to bind to specific sequences that are present in both *Tg* and *TPO* genes. TTF-1 is a transcription factor that recognizes and binds to specific DNA sequences via a 60-amino acid DNA-binding domain called homeo-domain. The protein is encoded by a single gene, *Titf1* in mice and *TTF1* in humans, located on chromosomes 12 and 14q13, respectively.

During embryonic life, TTF-1 is expressed in the developing thyroid, lungs, diencephalon, and posterior pituitary. Studies in *Titf1* knockout mice have demonstrated that this factor is absolutely necessary for the

Table I Timing and Molecular Phenotype of Thyroid Development in Humans and Mice

Embryonic stage		Histological feature	Gene expression <sup>a</sup>				Hormonogenesis
Mice (day)	Humans (week)		TTF-1 Pax8	Foxe1 Hhex	Tg TPO	Tshr NIS	
8.5	3.0	Thyroid anlage	+		–	–	–
9.5	3.5	Thyroid primordium migration initiates	+		–	–	–
11.5	4.0–5.5	Thyroglossal duct disappears	+		–	–	–
13.5	6.5–7.0	Thyroid reaches final destination					
14.5	8.5	Expansion and functional differentiation	+		+	–	–
16.5	10.0	Onset folliculogenesis	+		+	+	+

<sup>a</sup>The expression of the indicated genes has only been studied in the mouse embryo.

correct development of the thyroid. *Titf1* null mice display a perinatal lethal phenotype characterized by severe defects in both lung and forebrain development as well as a lack of thyroid and posterior pituitary. At an early stage of development, in the absence of TTF-1, the thyroid primordium forms in its correct position, but already by E 10.0 the thyroid primordium appears to be smaller than that in wild-type embryos and soon undergoes degeneration, probably due to an apoptotic process. Hence, TTF-1 is dispensable for the initial commitment of thyroid cells but is required for the survival and subsequent differentiation of these cells. Because in the absence of TTF-1 thyroid cells disappear before the gland accomplishes its final differentiation, it is not possible to assess the role of TTF-1 *in vivo* in the adult thyroid. However, experiments performed in cultured differentiated thyroid cells have demonstrated that TTF-1 is relevant for the expression of thyroid-specific genes. The role of TTF-1 in embryonic life is being investigated. TTF-1 is present in organs derived from the ventral wall of the anterior foregut (thyroid and lungs) but is absent in the esophagus, which derives from the dorsal wall of the gut. Furthermore, in the absence of TTF-1, there is no septation between the trachea and esophagus. Hence, a possible role of TTF-1 could be to contribute to the specification of the anterior part of the gut. However, we do not know which genes TTF-1 controls during thyroid morphogenesis. Studies on the phenotype of *Titf1* null embryos indicate that this transcription factor regulates the expression of some signaling molecules. *Fgf8* and TGF-related peptide growth factor BMP4 are down-regulated in the diencephalon and lungs, respectively, in the absence of TTF-1. It is possible that in the developing thyroid as well, TTF-1 could control the expression of signaling molecules required for the survival of thyroid cells.

### Pax8

Pax8 is a transcription factor characterized by the presence of a 128-amino acid DNA-binding domain (paired domain) that recognizes and binds to a single site present in *Tg* and *TPO* promoters. The gene encoding Pax8 (called *Pax8* in mice and *PAX8* in humans) is located on chromosome 2 in both species. During embryonic life, Pax8 is expressed since E 8.5 in the kidneys, brain, and thyroid primordium. In *Pax8* knockout mice, the thyroid is severely affected, follicular cells cannot be detected, and the mice die within a few weeks. In the absence of Pax8, at an early stage of embryogenesis, the thyroid primordium

forms, buds from the gut, and begins its migration. However the thyroid primordium is nearly undetectable by E 12.0. Hence, like TTF-1, Pax8 is required for the survival of thyroid precursor cells.

Experiments in cultured cells have demonstrated that Pax8 can activate transcription of thyroid-specific genes at their chromosomal locus. The coexpression of Pax8 and TTF-1 in thyroid cells only has suggested that these factors can cooperate in the stimulation of thyroid genes. Indeed, studies have provided us with the demonstration that Pax8 and TTF-1 directly interact *in vivo* in differentiated thyroid cells. Our knowledge of the function of Pax8 in the developing thyroid is still scarce. It has been demonstrated that in the absence of Pax8, the expression of both *Foxe1* and *Hhex* is strongly reduced in the precursors of follicular cells. Pax8 could then have a specific upper role in the regulatory pathway controlling the development of the thyroid.

### Foxe1

*Foxe1* is a transcription factor characterized by the presence of a forkhead domain as a DNA-binding domain. It is encoded by the *Foxe1* gene in mice and the *FOXE1* gene in humans, located on chromosomes 4 and 9q22, respectively. Early during development, at E 8.5, *Foxe1* is expressed along the whole foregut, at variance with TTF-1 and Pax8, whose presence in the gut is restricted to the thyroid anlage. *Foxe1* knockout mice display a perinatal lethal phenotype characterized by a severe cleft palate and athyreosis or ectopia of the thyroid. Analysis of thyroid development shows that in the absence of *Foxe1* at E 8.5, the thyroid anlage is specified, but at E 9.5, thyroid precursor cells are still on the floor of the pharynx, showing a clear defect of migration. At E 15.0, the *Foxe1* knockout embryos exhibit either a small sublingual thyroid remnant or no thyroid gland at all. Hence, *Foxe1* is absolutely necessary to promote migration of TFC precursors. Furthermore, *Foxe1* also could be implicated in the control of the survival of thyroid cells, as shown by the absence of thyroid in many *Foxe1* knockout mice.

### Hhex

*Hhex* is a homeodomain containing the transcription factor encoded by the *Hhex* gene in mice and the *HHEX* gene in humans, located on chromosomes 19 and 10q23, respectively. During embryonic life, *Hhex* is present in the developing thyroid and in several organs derived from the foregut endoderm. In *Hhex* null embryos at E 8.5, the thyroid anlage is visible, but

already at E 9.5 the thyroid primordium is absent or hypoplastic and the expression of TTF-1, Foxe1, and Pax8 is down-regulated.

It is possible that Hhex is implicated in the regulation of the expression of TTF-1, Foxe1, and Pax8 in the thyroid primordium. Hhex could then be required to maintain the expression of these transcription factors in the thyroid.

### Tshr

Tshr is a protein that belongs to the superfamily of G protein-coupled receptors encoded by the *Tshr* gene in mice and the *TSHR* gene in humans, located on chromosomes 12 and 14q31, respectively. Tshr is detected in the developing thyroid after the completion of the migration of the primordium, during the same stage at which Tg appears and before the first evidence of follicular organization in the gland.

In adults, thyroid-stimulating hormone (TSH) is the main physiological agent implicated in the regulation of the main functions of the thyroid and exerts its cellular effects by binding to Tshr. Inactivating mutations in this receptor cause severe hypothyroidism during postnatal life. During the early stages of thyroid development, a functional Tshr is not required but is necessary for the final step of follicular cell differentiation. Indeed, in the absence of a correct TSH/Tshr signaling, the expression of both TPO and NIS in TFCs is strongly down-regulated.

## MOLECULAR GENETICS OF THYROID DYSGENESIS

In 85% of cases, permanent congenital hypothyroidism detected in newborns is due to impaired development of the thyroid. Defects in the mechanisms that allow the growth, survival, migration, or differentiation of the thyroid primordium can result in thyroid dysgenesis (TD). This term indicates an ectopic or hypoplastic thyroid (or both) as well as the absence of the gland.

It has been demonstrated that in some patients, TD is due to mutations in genes involved in thyroid development and already identified in mouse models such as *TITF1*, *PAX8*, *FOXE1*, and *TSHR*. These results confirm that the products of these genes are required for a correct thyroid morphogenesis (Table II).

### *TITF1*

Some patients affected by a syndrome characterized by severe neurological disturbances, respiratory distress, and thyroid alterations have been found to carry heterozygous mutations within the *TITF1* gene. When tested *in vitro*, the various mutations described encode for proteins that do not display either functional activity or a dominant negative effect on the wild-type form. These data indicate that a reduced expression of TTF-1 is not compatible with the normal development of lungs, brain, and thyroid in humans. On the contrary, mice carrying only a

Table II Mouse and Human Phenotypes Following Inactivation of Some Genes Normally Expressed during Thyroid Development

Protein	Gene		Tissue expression	Phenotype		Inheritance in humans
	Mouse	Human		Null mice	Patients carrying inactivating mutations	
TTF-1	<i>Titf1</i>	<i>TITF1</i>	Thyroid Lung Brain Posterior pituitary	Thyroid and pituitary absent; severe defects in lung and diencephalon	Mild thyroid hypoplasia; respiratory distress; ataxia	Dominant
Pax8	<i>Pax8</i>	<i>PAX8</i>	Thyroid Kidney	Thyroid absent	Thyroid hypoplasia	Dominant
Foxe1	<i>Foxe1</i>	<i>FOXE1</i>	Thyroid Esophagus Palate Hair follicles Anterior pituitary	Thyroid absent or ectopic; cleft palate	Athyreosis; cleft palate; spike hair; choane atresia	Recessive
Tshr	<i>Tshr</i>	<i>TSHR</i>	Thyroid	Thyroid hypoplasia	Thyroid hypoplasia	Recessive

functional allele of *Titf1* show a very mild, possible strain-dependent phenotype.

### PAX8

Mutations in *PAX8* have been described in sporadic and familial cases of congenital hypothyroidism with TD. The phenotype of the patients ranges from mild hypothyroidism to severe hypoplasia of the thyroid. As in the case of TTF-1, all affected individuals are heterozygous from the mutated allele, and in the familial cases transmission is autosomal dominant. This dominant effect of *PAX8* mutations could be due to a gene dose requirement (haploinsufficiency). In these patients, the single functioning allele is not able to produce a sufficient amount of Pax8 to support the normal development of the gland.

### FOXE1

Homozygous mutations in the *FOXE1* gene have been described in two pairs of siblings affected by Bamfort syndrome. This syndrome is characterized by cleft palate, bilateral choanal atresia, spiky hair, and congenital hypothyroidism with TD. The patients have been found to carry a homozygous missense mutation in the *FOXE1* gene, and because of this, the mutant protein was unable to bind DNA. All of the patients show athyreosis, whereas in mice the absence of Foxe1 causes either athyreosis or defects in thyroid migration. Because *FOXE1* mutations have been identified in only four patients, it is not known whether this discrepancy between humans and mice is significant or not.

### TSHR

Mutations in the *TSHR* gene have been identified in patients affected by congenital hypothyroidism with thyroid hypoplasia and compensatory increased TSH secretion. The disease is inherited as an autosomal recessive trait, and the phenotype ranges from asymptomatic hyperthyrotropinemia to congenital hypothyroidism with severe hypoplasia of the thyroid. The different penetrance of the phenotype could be due to the amount of residual activity of the receptor as well as to the individual genetic background.

### CONCLUSION

Although early in development TTF-1, Foxe1, Pax8, and Hhex are expressed in the thyroid primordium, none of them is essential for the initial steps of thyroid

morphogenesis. Genes required for thyroid anlage specification have not yet been identified. It is possible that the identification of the genes controlling TTF-1, Foxe1, Pax8, and Hhex expression can give us information on how thyroid precursor cells begin their differentiation. Mutations in genes involved during the initial steps of thyroid morphogenesis could be responsible for thyroid agenesis (i.e., absence of the formation of a thyroid anlage), a phenotype not found in any of the available animal models.

TTF-1, Foxe1, Pax8, and Hhex are transcription factors whose role is to control the expression of target genes that ultimately actuate a specific developmental program. The identification of these genes is still a matter of study. A number of cases of congenital hypothyroidism with TD could be due to mutations in genes that are targets of the transcription factors TTF-1, Foxe1, Pax8, and Hhex.

Most genes expressed in the developing thyroid are still expressed in the adult gland. The creation of animal models with a thyroid-specific, conditional knockout of these genes will be an invaluable tool in elucidating their role in a physiological context.

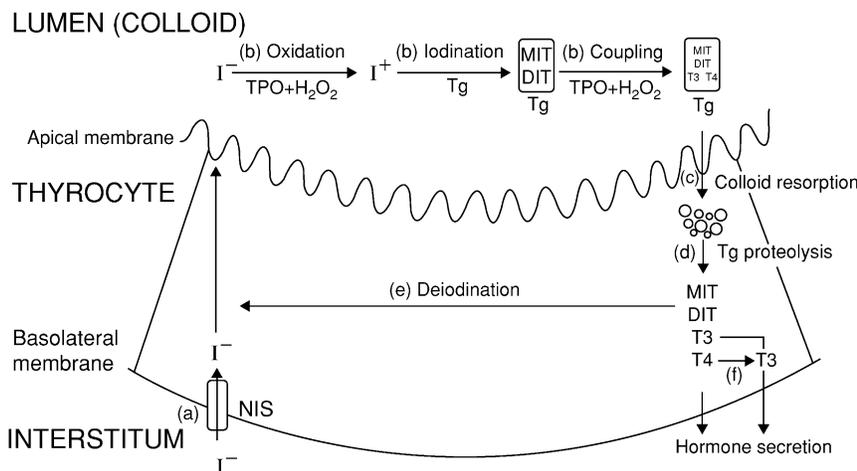
### See Also the Following Articles

Adrenal Cortex, Development • G Protein-Coupled Receptors • Hypothyroidism, Congenital • Thyroglobulin • TSH Receptor (Thyrotropin Receptor)

### Further Reading

- Damante, G., Tell, G., and Di Lauro, R. (2001). A unique combination of transcription factors controls differentiation of thyroid cells. *Prog. Nucleic Acid Res. Mol. Biol.* **66**, 307–356.
- De Felice, M., Ovitt, C., Biffali, E., Rodriguez-Mallon, A., Arra, C., Anastassiadis, K., Macchia, P. E., Mattei, M. G., Mariano, A., Schoeler, H., Macchia, V., and Di Lauro, R. (1998). A mouse model for hereditary thyroid dysgenesis and cleft palate. *Nat. Genet.* **19**, 395–398.
- Di Lauro, R., and De Felice, M. (2001). Anatomy and development of thyroid. In "Endocrinology IV" (L. J. De Groot and J. L. Jamison, eds.), pp. 1268–1277. W. B. Saunders, Philadelphia.
- Gillam, M. P., and Kopp, P. (2001). Genetic regulation of thyroid development. *Curr. Opin. Pediatr.* **13**, 358–363.
- Kimura, S., Hara, Y., Pineau, T., Fernandez-Salguero, P., Fox, C., Ward, J., and Gonzalez, F. (1996). The T/ebp null mouse: Thyroid-specific enhancer-binding protein is essential for the organogenesis of the thyroid, lung, ventral forebrain, and pituitary. *Genes Dev.* **10**, 60–69.
- Macchia, P. E., De Felice, M., and Di Lauro, R. (1999). Molecular genetics of congenital hypothyroidism. *Curr. Opin. Genet. Dev.* **9**, 289–294.
- Mansouri, A., Chowdhury, K., and Gruss, P. (1998). Follicular cells of the thyroid gland require Pax8 gene function. *Nat. Genet.* **19**, 87–90.
- Martinez Barbera, J. P., Clements, M., Thomas, P., Rodriguez, T., Meloy, D., Kiuoussis, D., and Beddington, R. S. (2000). The

- homeobox gene Hex is required in definitive endodermal tissues for normal forebrain, liver, and thyroid formation. *Development* **127**, 2433–2445.
- Postiglione, M. P., Parlato, R., Rodriguez-Mallon, A., Rosica, A., Mithbaokar, P., Maresca, M., Mariani, R. C., Davies, T. F., Zannini, M. S., De Felice, M., and Di Lauro, R. (2002). Role of the thyroid-stimulating hormone receptor signaling in development and differentiation of the thyroid gland. *Proc. Natl. Acad. Sci. USA* **99**, 5462–5467.
- Van Vliet, G. (2003). Development of the thyroid gland: Lessons from congenitally hypothyroid mice and men. *Clin. Genet.* **63**, 445–455.



**Figure 1** Schematic diagram of thyroid hormone biosynthesis and release. Subsequent metabolic steps are (a) iodide transport via the  $Na^+/I^-$  symporter (NIS) inhibited by  $ClO_4^-$  and  $SCN^-$ ; (b) oxidation of  $I^-$  to  $I^+$  and iodination of tyrosine residues in thyroglobulin (Tg), and coupling of mono-iodotyrosine (MIT) and diiodotyrosine (DIT) to thyroxine ( $T_4$ ) or triiodothyronine ( $T_3$ ), catalyzed by thyroid peroxidase (TPO) and inhibited by propylthiouracil and methimazole; (c) colloid resorption, inhibited by lithium and  $I^-$ ; (d) proteolysis of Tg, inhibited by  $I^-$ ; (e) deiodination of MIT and DIT; and (f) deiodination of  $T_4$ , inhibited by propylthiouracil.

or  $T_4$ ). Approximately 150  $\mu\text{g}$  of inorganic iodide is required daily for normal thyroid hormone biosynthesis. Iodine is an essential but rare element; it is mainly supplied in food, either naturally (especially in sea fish) or artificially, by the iodization of salt, bread, or milk. The thyroid gland accumulates iodide by an active energy-requiring transport mechanism: the  $Na^+/I^-$  symporter (NIS) is able to concentrate iodide from the serum against a large concentration gradient. Other proteins involved in thyroid hormone biosynthesis (schematically represented in Fig. 1) are thyroid peroxidase (TPO) and thyroglobulin. TPO (carbohydrate content 10%) contains four glycosylation sites that are important for the tertiary structure at the active site, and a prosthetic (heme) group; its main form is TPO-1 (103 kDa), but alternative splicing (by which exon 10 is spliced out) gives rise to the variant TPO-2 of 97 kDa, which is inactive with respect to iodination. Thyroglobulin is a large homodimeric glycoprotein (carbohydrate content 10%) containing 134 tyrosyl residues, of which 40 are available for iodination. Tyrosine residues 5, 1290, 2533, and 2746 are located in the amino terminus of the thyroglobulin molecule and are preferentially iodinated; these are known as the hormonogenic sites.

Once iodide is taken up in the thyrocyte, it rapidly moves to the apical surface and into the lumen. Thyroid hormone biosynthesis occurs within the thyroglobulin molecule at the apical surface. It involves two oxidative reactions, both catalyzed by

TPO located in the apical plasma membrane. The first is iodination of the tyrosyl residues in thyroglobulin, resulting in monoiodotyrosine and diiodotyrosine. The iodination requires prior peroxidation of  $I^-$  to form its more reactive form  $I^+$  (or hypoiodite). The second is the coupling of iodotyrosines in the thyroglobulin: coupling of a donor diiodotyrosine residue to an acceptor diiodotyrosine results in  $T_4$  and that of a donor monoiodotyrosine to an acceptor diiodotyrosine in 3,5,3'-triiodothyronine ( $T_3$ ) (Fig. 2.) The hormonogenic sites are only partially used in the biosynthesis: under normal circumstances (at an iodine content of thyroglobulin of 0.5%, i.e., 26 atoms iodine per 660 kDa molecule of thyroglobulin), each thyroglobulin molecule contains 5 monoiodotyrosine, 4.5 diiodotyrosine, 2.5  $T_4$ , and 0.7  $T_3$  molecules. The thyroglobulin is stored in the colloid.

Release of  $T_4$  and  $T_3$  from the gland first involves resorption of stored thyroglobulin from the colloid. In a process called micropinocytosis, or coated vesicle-dependent endocytosis, there is an invagination of the apical plasma membrane into small vesicles, producing colloid droplets. The vesicles are transported to the basolateral surface of the thyrocyte via endosomes; internalized thyroglobulin that is poorly iodinated is, however, recycled back to the follicular lumen. Colloid droplets with highly iodinated thyroglobulin fuse with lysosomes. Intralysosomal peptidases catalyze the degradation of thyroglobulin, releasing  $T_4$ ,  $T_3$ , monoiodotyrosine, and diiodotyrosine. Monoiodotyrosine

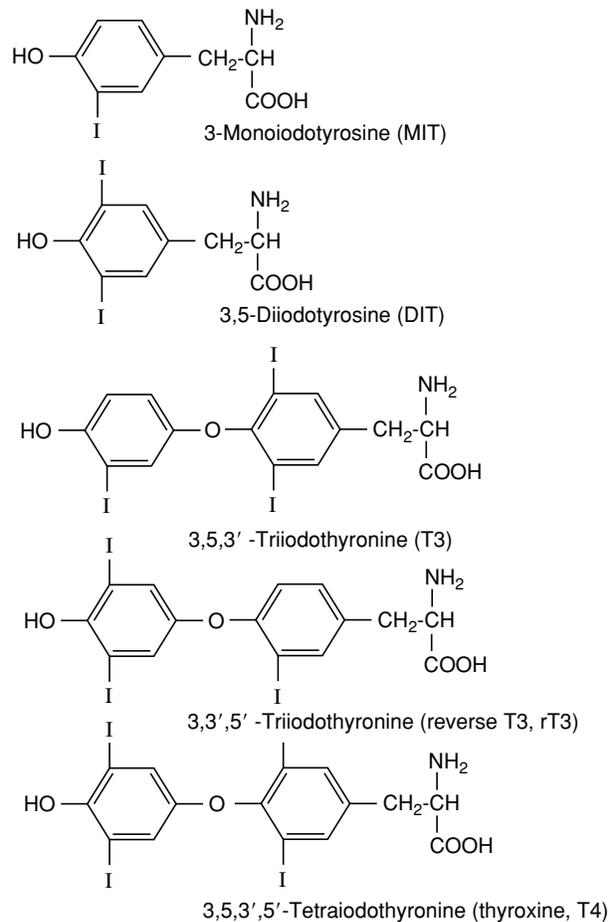


Figure 2 Structures of some iodotyrosines and iodothyronines.

and diiodotyrosine are deiodinated by an NADPH-dependent dehalogenase, and the liberated iodide is largely re-utilized for hormone synthesis, being only partly lost from the thyroid gland. The thyrocyte also contains 5'-deiodinase activity (type I), converting a small amount of hydrolyzed T<sub>4</sub> into T<sub>3</sub>.

Thyroid hormone (and a small amount of thyroglobulin) is secreted from the basolateral surface of the thyrocyte into the circulation. Thyroxine is the main secretory product of the thyroid gland (~130 mol/day). Thyroidal T<sub>3</sub> secretion is much lower (~10 mol/day), providing only 20% of the daily T<sub>3</sub> production. Secretion of the inactive iodothyronine rT<sub>3</sub> (reverse triiodothyronine) is still smaller, because the conditions for its formation within the thyroglobulin molecule (requiring coupling of a donor diiodotyrosine residue to an acceptor monoiodotyrosine) are less favorable.

The iodide uptake in the thyrocyte, the iodide organification (i.e., the incorporation of iodine in tyrosines), and the formation of iodothyronines, the resorption of colloid, and the release of T<sub>4</sub>, and T<sub>3</sub> are all stimulated by TSH. Specific transcription factors have been identified that bind to the promoter region of thyroglobulin and TPO genes: thyroid-transcription factors TTF-1 and TTF-2 (now called Foxe<sub>1</sub>, Forkhead box e1) and PAX-8. They induce via transactivation expression of these genes, but they may also play a role in ontogenesis: during embryonic development they are expressed before thyroglobulin, TPO, and TSH receptor.

## REGULATORY MECHANISMS

The activity of the thyroid gland is tightly regulated by feedback and feedforward control at the pituitary and hypothalamus. After being secreted into the bloodstream, thyroid hormone circulates predominantly bound to serum thyroid hormone binding proteins. Upon arrival at extrathyroidal target tissues, thyroid hormones enter cells via specific uptake mechanisms. Once inside the cell, thyroid hormones undergo a variety of metabolic reactions, most importantly deiodination. Thyroid hormone binds to specific nuclear binding sites, resulting in modulation of gene transcription.

## See Also the Following Articles

Graves' Disease • Hypothalamus–Pituitary–Thyroid Axis • Thyroglobulin • Thyroid Hormone Action • Thyroid Hormone-Binding Proteins • Thyroid Hormone Metabolism • Thyroid Hormone Receptors • TSH (Thyroid-Stimulating Hormone; Thyrotropin)

## Further Reading

- Capen, C. C. (2000). Comparative anatomy and physiology. In "The Thyroid" (L. E. Braverman and R. D. Utiger, eds.), 8th ed., pp. 20–43. Lippincott Williams & Wilkins, Philadelphia.
- Carasco, N. (2000). Thyroid iodide transport: The Na<sup>+</sup>/I<sup>-</sup> symporter (NIS). In "The Thyroid" (L. E. Braverman and R. D. Utiger, eds.), 8th ed., pp. 52–61. Lippincott Williams & Wilkins, Philadelphia.
- Dunn, J. T. (1999). Thyroid hormone and secretion. In "Thyroid Disease Manager." <http://www.thyroidmanager.org>
- Taurog, A. M. (2000). Hormone synthesis: Thyroid iodide metabolism. In "The Thyroid" (L. E. Braverman and R. D. Utiger, eds.), 8th ed., pp. 61–85. Lippincott Williams & Wilkins, Philadelphia.

hypothalamus–pituitary–thyroid axis, suppressing the production of thyroid-stimulating hormone (TSH or thyrotropin) in the anterior pituitary as well as that of thyrotropin-releasing hormone (TRH) in the hypothalamus. These signals lead to reduced hormone production by the thyroid. The central regulation of thyroid activity is therefore aimed at maintaining certain levels of thyroid hormone to ensure proper physiological function of virtually every tissue. Consequently, the general involvement of thyroid hormone becomes apparent only in disease states that lead to reduced or increased plasma hormone levels. Because of its diverse effects and essential role in normal physiology, thyroid hormone has been likened to a vitamin.

Perhaps not surprising, the mechanisms by which thyroid hormone exerts its myriad effects are complex and certainly not completely understood. A large number of processes are directly affected by thyroid hormone, but the full physiological effect is often the result of additional, secondary effects. As appears to be the case for various steroid hormones, nonnuclear actions may contribute to the full spectrum of direct effects of thyroid hormone, but evidence for this is relatively sparse. In contrast, following the seminal work of Tata and Oppenheimer in the 1960s and 1970s it was generally accepted by the early 1980s that regulation of gene expression, mediated by nuclear thyroid hormone receptors (TRs), is the principal mechanism of action of the hormone. The independent discoveries of multiple  $T_3$ -binding nuclear receptors in 1986 opened the way to a detailed analysis of transcriptional regulation as the molecular basis of thyroid hormone action. This mode of action proves to be particularly complex, and together with an increasing amount of information on the role of local thyroid hormone metabolism in different tissues, it is becoming clear that thyroid hormone action can be modulated at the level of individual tissues throughout life.

## EVOLUTION OF THYROID HORMONE ACTION

Information on the origins and earliest effects of thyroid hormone has increased our understanding of the effects in higher organisms. A brief overview of the evolutionary development of thyroid hormone action and its major effects is therefore given, followed by a description of the mechanisms of action and the tissue-specific effects of thyroid hormone in mammals, including man.

## Synthesis of Thyroid Hormone in Invertebrates

Spontaneous iodination of tyrosine residues in proteins occurs in the iodine-rich marine environment. Subsequent coupling of two iodinated tyrosines and digestion of proteins may release iodothyronines, including  $T_4$  and  $T_3$ , and these thyroid hormones are indeed found in significant amounts in marine invertebrates. At least in some species of jellyfish, organification of iodine to form  $T_4$  actually appears to play a role in the timing of the asexual reproductive cycle in a process called strobilation, which may be likened to metamorphosis in amphibians.  $T_4$  production is therefore an ancient biochemical process with possibly a physiological role in the most primitive invertebrates, predating the evolution of specialized thyroid cells. Such cells are found in the immediate precursors of vertebrates, the protochordates (sea squirts). These cells actively form  $T_4$ -containing proteins but are not yet organized in the typical thyroid follicle structure. Furthermore, release and subsequent absorption of  $T_4$  appear to be dependent on digestion of the secreted proteins in the gut.

## Synthesis of Thyroid Hormone in Vertebrates

Follicular thyroid cells, in which thyroid cells enclose a follicular lumen containing iodinated proteins (thyroglobulin), are found in the most primitive living vertebrates, the sea lampreys. At this point in evolution, thyroid cells have acquired a protein-digesting enzyme that allows the release of  $T_4$ , and in adult lampreys the gland has become endocrine. Interestingly, during the larval stage of this organism, which lasts up to 5 years and constitutes the better part of its life, the thyroid cells are still nonfollicular and  $T_4$  secretion is exocrine followed by absorption. Upon metamorphosis, these epithelial cells form thyroid follicles and secreted  $T_4$  is then directly absorbed by the vascular system. Therefore, the development of the lamprey seems to recapitulate the evolutionary acquisition of the endocrine thyroid gland.

All vertebrates contain follicular thyroid cells, although the localization and organization into a discrete gland vary greatly, from scattered follicles in the subpharyngeal area in lampreys to follicles located mostly in the kidneys in fish, a disc-shaped gland next to the heart in turtles, two widely separated globules on either side of the trachea in birds, and a

bilobed gland connected by an isthmus straddling the trachea in lizards, as is also the typical structure of the thyroid gland in all mammals. Despite these differences, the circulating levels of free  $T_3$  and  $T_4$  are within one order of magnitude throughout the vertebrate lineage, with the exception of some species of fish.

### Modes of Action and Effects of Thyroid Hormone in Invertebrates

The presence of  $T_4$  or  $T_3$ , particularly in marine organisms, is not proof of a physiological function given the spontaneous generation of iodothyronines under certain conditions. Nevertheless, iodothyronines appear to have effects in invertebrates as mentioned previously, although direct evidence for these effects is limited. A stronger case for a physiological role for thyroid hormone in primitive organisms is perhaps made by the presence of the specialized  $T_4$ -producing cells in lampreys. This has interesting implications for the primordial mode of action of thyroid hormone. Phylogenetic analysis of the large family of DNA-binding nuclear receptors, of which the TRs are members, shows that the common ancestral proteins were DNA-binding factors without ligand-binding properties. The capacity to selectively bind hormones and other regulatory compounds evolved relatively late. In the case of the thyroid receptors, it appears to have occurred during the development of the vertebrate lineage. The earliest effects of thyroid hormone must then be assumed to be extranuclear, and this mode of action may still be relevant in higher organisms.

### Modes of Action and Effects of Thyroid Hormone in Vertebrates

Development of the large diversity of effects in vertebrates is linked to the adaptive progression of tissue distribution of the  $T_3$  receptors (TRs) and mutations within the receptors allowing interactions with other DNA-binding proteins or accessory factors. Genes may have developed thyroid hormone responsiveness by acquiring the specific sequence in their promoter region that is required for binding of TRs. In contrast, the part of the receptor that recognizes this six-base pair sequence is highly conserved in this family of receptors and therefore not a factor in the diversification of  $T_3$  action. Further development of (variable) tissue responsiveness is probably also linked to the evolving expression patterns of the thyroid

hormone-metabolizing deiodinases, which are present throughout the vertebrate lineage.

### Growth and Development

The receptor-mediated effects of thyroid hormone on growth and development are evident in all vertebrates. The most striking of these is the transition of the larval amphibian from an aquatic vegetarian organism to a terrestrial carnivore. This experiment can be done with a salamander by simply adding  $T_4$  to the aquarium water. The ensuing metamorphosis leads to radical changes in the appearance of the animal, including replacement of the fin-like tail and appendages with legs and claws and a round tail; replacement of a smooth skin with a thicker keratinized skin; the loss of gills and acquisition of lungs; development of eyelids; complete restructuring of the mouth, tongue, and intestinal tract to allow a mostly carnivorous diet; and numerous biochemical changes to accommodate a different respiratory and metabolic physiology. The increase in thyroid activity and thyroid hormone levels at metamorphic climax in amphibians triggers and sustains the process. Cells in all tissues undergoing metamorphic changes express TRs, and thyroid hormone induces or represses the expression of many genes. In addition, responsiveness of tissues is actually increased by  $T_3$ -dependent stimulation of the expression of TRs, as is the case in the restructuring of the tail. A surge in thyroid hormone levels is similarly critical in early growth and differentiation in fish, reptiles, and birds. All these effects have their counterparts in the fetal and perinatal development in mammals, particularly with respect to brain development. In humans, the critical involvement of thyroid hormone is illustrated by the striking and mostly irreversible effects of congenital hypothyroidism on growth, metabolism, reproduction, and mental development (cretinism). The surge in  $T_3$  and  $T_4$  levels reaches its peak in humans during the first 2 months after birth, coinciding with the phase of maximal cortical growth. This growth spurt starts during the third trimester when fetal thyroid activity increases and the dependence on maternal thyroid hormone wanes. However, thyroid hormone is required for central nervous system development beginning in the 10th week of fetal life, coinciding with the appearance of TRs in neuronal tissue. Relatively low levels of thyroid hormone are essential throughout intrauterine development to ensure proper timing and progression of the various phases of brain development. Studies in rats show that  $T_4$  levels are critical because neurons preferentially take up this iodothyronine and then convert it to  $T_3$ .

### Metabolism

The basal metabolic rate (BMR), which is the sum of all energy-consuming processes of an individual when completely at rest, is reduced in humans by 40% in hypothyroidism, whereas it is increased by 50% in hyperthyroidism. Thyroid hormone has essentially the same stimulatory effect on metabolism and concomitant  $O_2$  consumption in other mammals and birds. It was previously thought that this effect played a crucial role in the evolutionary development of endothermy since metabolism in cold-blooded animals appeared unresponsive to thyroid hormone. However, this effect proved dependent on body temperature and the BMR of fish, amphibians, and reptiles is similarly responsive to  $T_3$ , albeit the extent of the effect is somewhat less than in mammals. Thyroid hormone does not appear to play a major role in thermoregulation in warm-blooded animals, and the effect on energy turnover, and hence heat production, is considered to be a second-order phenomenon.

The  $T_3$ -dependent stimulation of the BMR is the sum of increased energy consumption and substrate metabolism. The major energy-consuming processes in all tissues at rest are maintenance of the mitochondrial proton ( $H^+$ ) gradient; maintenance of the gradients of calcium ( $Ca^{2+}$ ), sodium ( $Na^+$ ), and potassium ( $K^+$ ) over various cellular membranes; and synthesis of DNA, RNA, and proteins. The effect on the BMR is not primarily the result of the stimulation of a single process, as has long been maintained for the effect of thyroid hormone on the sodium–potassium pump ( $Na^+K^+ATPase$ ). Cellular ion homeostasis is nevertheless an important aspect of the effect of thyroid hormone on the BMR. This homeostasis involves the energy-dependent transport of  $Ca^{2+}$ ,  $Na^+$ , and  $K^+$  ions against gradients across the plasma membrane, as well as across intracellular membrane systems in the case of  $Ca^{2+}$ , to maintain their critical cytosolic concentrations. These processes account for a considerable part of the metabolic rate in mammals due to the passive and facilitated leak of ions through membranes. The expression and activity of the major ion-pumping ATPases are stimulated by thyroid hormone, as are the passive-leak characteristics of membranes. The proliferation of some intracellular membranes is also increased by thyroid hormone, and together these effects result in increased ion fluxes and ATP consumption. Proliferation of mitochondria as well as their capacity for oxidative phosphorylation are also stimulated by thyroid hormone, ensuring sufficient ATP production to meet the greater demand due to increased ion cycling and other processes.

To fuel the extra ATP turnover, thyroid hormone increases the rate of glucose production by the liver and increases the availability of the necessary substrates—amino acids, glycerol, and free fatty acids from proteins and fat. Thyroid hormone stimulates the breakdown of protein and fat, but at the same time it stimulates the synthesis of these compounds. This cycling of substrates, which consumes a considerable amount of energy, appears futile, but it allows for the regulated liberation of the substrates mentioned previously.

### Which Iodothyronines Are Active in Mammals?

#### $T_4$ and $T_3$

The thyroid produces mainly  $T_4$ , whereas 80% of circulating  $T_3$  results from peripheral enzymatic deiodination of  $T_4$ , predominantly in the liver. The protein-bound and free plasma levels of  $T_4$  in all species that have been studied are substantially higher than those of  $T_3$ . In humans, the difference is approximately 40-fold for total hormone and 5-fold for free hormone. Nevertheless, as noted previously,  $T_3$  is considered to be the active form of thyroid hormone in mammals. Although  $T_4$  can affect the expression of thyroid hormone-responsive genes, the affinity of the thyroid hormone receptors for  $T_4$  is considerably lower compared to that for  $T_3$ . In some thyroid hormone effects, such as the effect on neurons and the down-regulation of pituitary TSH production,  $T_3$  cannot substitute for  $T_4$  and  $T_4$  appears to be the active form. In the latter case, as in neurons, the genes encoding the TSH $\alpha$  and TSH $\beta$  subunits are indeed responsive to  $T_3$ , but the pituitary cellular  $T_3$  content is determined by  $T_4$  uptake and subsequent deiodination rather than by the plasma  $T_3$  levels.

#### Other Iodothyronines

Progressive deiodination of  $T_4$  and  $T_3$  by the three known deiodinases gives rise to a number of different triiodothyronines and diiodothyronines. Of these, reverse  $T_3$  (3,3',5'-triiodothyronine) and 3,5- $T_2$  (3,5-diiodothyronine) have also been shown to elicit  $T_3$ -like effects. Because the affinity of the TRs for these compounds is negligible and the effects do not require *de novo* protein synthesis, the mechanism of action must be extranuclear. In the case of stimulation of mitochondrial activity by 3,5- $T_2$ , it was indeed shown to involve a direct interaction of this iodothyronine with mitochondrial enzymes.

Enzymatic removal of the amino group (deamination) of  $T_4$  and  $T_3$  gives 3,5,3',5'-tetraiodothyroacetic acid (Tetrac) and 3,5,3'-triiodothyroacetic acid (Triac), respectively. Tetrac can be deiodinated to Triac, and Triac has an even higher affinity for TRs than  $T_3$ . When administered at concentrations comparable to those of  $T_4$  and  $T_3$ , these compounds show  $T_3$ -like effects.

Although it cannot be ruled out that in certain circumstances cellular concentrations of these metabolites may reach levels high enough to elicit significant effects, they are generally considered to be of little or no physiological relevance based on the known low plasma concentrations.

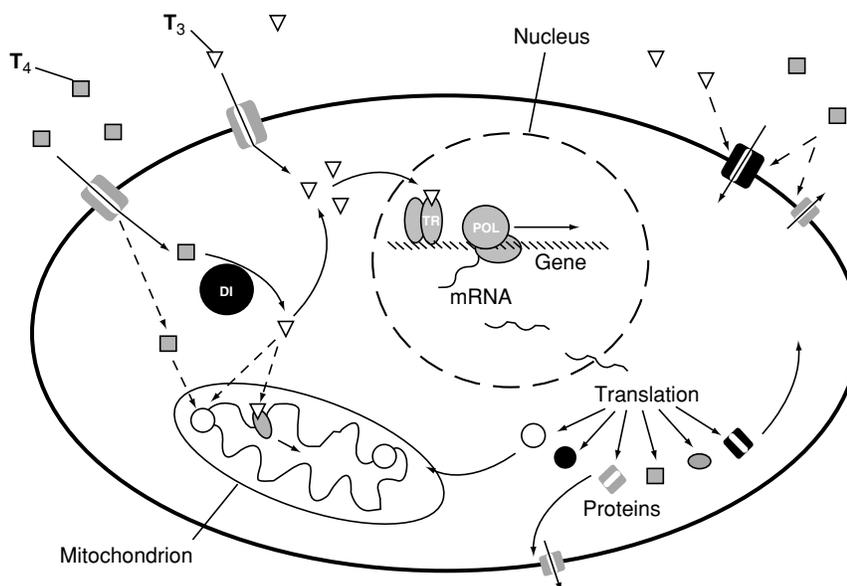
## MECHANISMS OF THYROID HORMONE ACTION

In the following sections, the mechanisms of the primary actions of thyroid hormone are presented (i.e., those mediated by  $T_3$  receptors and those that may involve direct interaction with a target). Additional actions may include effects on the stability of specific proteins or the messenger RNAs (mRNAs) encoding these proteins. However, there are few examples, and

the mechanism is not understood. Figure 1 summarizes the modes of action that are discussed here.

### Extranuclear Action

Stimulation by  $T_4$  and  $T_3$  of the activity of the plasma membrane  $Ca^{2+}$ -pump ( $Ca^{2+}$ -ATPase) in red blood cells, which do not have a nucleus, is the best documented example of an extranuclear thyroid hormone effect. The *in vitro* effects correlate with increased activity of this enzyme in erythrocytes taken from hyperthyroid patients and decreased activity in those from hypothyroid patients. *In vitro* and *in vivo* studies also show that  $T_3$  and  $T_4$  stimulate glucose uptake in a variety of tissues, increase heart rate, and induce vasodilation, with effects appearing within minutes. Similarly, in various cultured cell types, the addition of physiological amounts of  $T_3$  or  $T_4$  almost instantly stimulates  $Ca^{2+}$  influx or the activity of proteins involved in cellular signaling processes. These effects, however, are transient and it is not clear whether they are relevant under normal conditions. The molecular mechanism underlying the mostly membrane-associated rapid effects is also unclear. There is evidence of a direct interaction of  $T_3$  or  $T_4$  with membrane-bound proteins, thereby altering their properties. Based on a large body



**Figure 1** Modes of action of thyroid hormone.  $T_4$  and  $T_3$  are actively taken up by the cell by specific transporters. Deiodinases (DI) can convert  $T_4$  to  $T_3$ , adding to the cellular  $T_3$  pool.  $T_3$  binds to its nuclear receptor (TR) on the promoter of a  $T_3$ -responsive gene and activates transcription of the gene by RNA polymerase (POL). Messenger RNAs (mRNA) are translated into protein constituents of the cell, including mitochondrial proteins. Extranuclear actions are indicated by dashed arrows.  $T_4$  and  $T_3$  may affect substrate uptake or ion fluxes across the outer membrane by direct interaction with the transport proteins or its lipid environment. The activity of mitochondrial enzymes may be similarly affected. Transcription and expression of genes encoded by mitochondrial DNA may be modulated by  $T_3$  through specific TRs in mitochondria.

of data, an alternative hypothesis proposes that  $T_3$  and  $T_4$  affect membrane properties in a less specific way by becoming part of it. The physicochemical properties of the molecules are such that the lipophilic, iodinated phenolic part inserts readily in the outer leaflet of the lipid bilayer, whereas the hydrophilic amino and carboxyl groups prevent the molecule from crossing the membrane (it is for this reason that specific thyroid hormone transporters are required for cellular uptake). This partitioning in the membrane at normal levels of  $T_3$  and  $T_4$  reduces its fluidity, and this will affect the activity of membrane-associated enzymes.

Such a mechanism may also account for the immediate effects of  $T_3$  and  $T_4$  on many mitochondrial membrane-bound enzymes, although this is a matter of debate. Irrespectively, the number of mitochondria, mitochondrial membrane density, and mitochondrial activity generally increase with increasing thyroid activity. This may in part be a secondary response to greater ATP demand. However, the presence of high-affinity thyroid hormone-binding proteins in the inner mitochondrial membrane in hormone-responsive tissues, but not in refractory tissues such as testis and spleen, suggests a unique pathway of thyroid hormone action in mitochondria. These proteins are known to be truncated forms of the TR- $\alpha$  protein and indeed act as  $T_3$ -dependent transcription factors of genes encoded in the mitochondrial genome. Apart from this relatively rapid receptor-mediated action, several nuclear-encoded mitochondrial enzymes are regulated via nuclear TRs.

Changes in thyroid status also lead to alterations in the lipid composition and properties of membranes, most notably affecting the saturation of fatty-acyl chains of the phospholipids. Such changes have also been shown to affect membrane-associated enzymes. These lipid alterations are thought to be related to the effects of thyroid hormone on the expression of lipid desaturases. Any subsequent effects on membrane-associated processes are therefore secondary to a nuclear-mediated action of the hormone.

### Receptor-Mediated Nuclear Action

There are two known genes for thyroid hormone receptors (i.e., TR- $\alpha$  and TR- $\beta$ ). Multiple receptor isoforms and related proteins are expressed through alternative splicing of mRNAs. Of these, TR- $\alpha_1$ , TR- $\beta_1$ , TR- $\beta_2$ , and TR- $\beta_3$  are bona fide TRs in that they bind  $T_3$  and confer transcriptional regulation of target genes. Other forms, such as TR- $\alpha_2$  (DNA binding but not hormone binding) and the  $\Delta$ TR- $\alpha$ / $\Delta$ TR- $\beta$  forms (hormone binding but not DNA binding), can act

as dominant negative factors interfering with TR action. These factors may increase the capacity for tissue-specific fine-tuning of thyroid hormone effects. Some *in vivo* data support such a role, but the precise physiological relevance in different tissues is unknown.

Studies of receptor knockout mice lacking either functional TR- $\alpha$  or TR- $\beta$  receptors indicate a large degree of redundancy between both isoforms. The molecular mechanism of transcription regulation is considered identical for the TR- $\alpha$  and TR- $\beta$  isoforms, which is supported by *in vitro* analyses. Some isoform-specific effects are observed in knockout mice as well as in cases of TR- $\beta$  mutations in humans, including cardiac pacemaking (TR- $\alpha_1$ ), development of hearing (TR- $\beta_1$ ), and thyroid hormone feedback regulation (TR- $\beta_2$ ). These dependencies reflect tissue-restricted expression of TRs rather than gene-specific regulatory mechanisms.

### Regulation of Gene Transcription

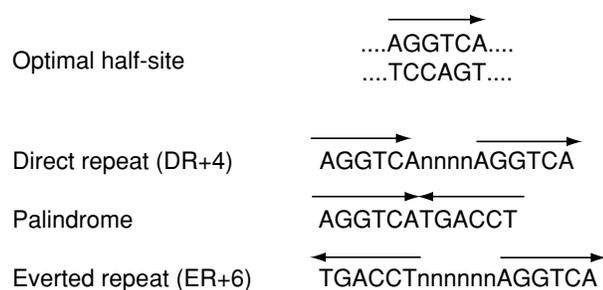
Transcription of a gene's DNA into mRNA requires recruitment of the RNA polymerase complex to general transcription factors that are already bound to the promoter region of the DNA, immediately upstream of the coding sequence. The rate of expression of a gene is basically determined by the ease with which these complexes are formed. Promoter activity and transcription rate are therefore terms that describe the frequency with which transcription of a gene is initiated following assembly of the complex. Gene transcription is primarily modulated by factors that either facilitate and stabilize the assembly of the transcription machinery or disrupt it. The RNA polymerase complex that transcribes the gene is exceptionally large, containing more than 100 different proteins. In contrast, nuclear DNA in its compact form is not readily accessible for such large complexes. Every 200-base pair stretch of DNA is wrapped around a disc-shaped histone protein complex forming a string of so-called nucleosomes that compact the DNA further by combining with other proteins. The resulting chromatin structure, however, is dynamic and can be made accessible to large DNA-binding complexes by modification of the histone proteins. Addition of acetyl groups to histones relaxes the chromatin structure, whereas deacetylation closes it. Two classes of enzymes are responsible for this reversible chemical modification: histone acetyl transferases (HATs) and histone deacetylases (HDACs). These proteins are key factors in the regulation of gene expression because they can be recruited by transcription factors, including nuclear hormone receptors.

### Thyroid Hormone Response Elements

The TRs belong to the large family of ligand-dependent transcription factors, which includes receptors for steroid hormones, retinoic acid and related retinoids, and vitamin D. Specific DNA sequences (response elements) are required in the promoters of genes to allow binding of these receptors, making the gene hormone responsive. Receptors in this family recognize the six-base pair element AGGTCA and interaction of two receptors (dimerization) is typically required for functional activity. The DNA motif to which each receptor binds is consequently referred to as a half-site, and the spacing and orientation of the half-sites determine receptor specificity of the response element. In the case of TRs, the optimal arrangement of half-sites in a thyroid hormone response element (TRE) is either a direct repeat spaced by four base pairs, a palindrome, or an everted repeat with a six-base pair gap (Fig. 2).

Deviations from the optimal half-site sequence and, to a lesser extent, from the optimal spacing between them are common in naturally occurring TREs. This variation is responsible for large differences between TREs in TR-binding affinity and stimulation of promoter activity. The relaxed constraints placed on TRE structures are evident in Fig. 3, which shows several natural TREs. TREs often consist of three or four half-sites allowing for binding of additional TRs, which further increases promoter activation. In addition, a promoter may contain several TREs contributing to the overall  $T_3$  responsiveness.

TRs readily form heterodimers on TREs with the retinoid X receptor (RXR) with greater binding affinity than that of TR homodimers. Transcription activation by  $T_3$  is also higher, requiring no ligand for RXR. RXRs are also considered the natural partner for TRs



**Figure 2** The optimal half-site for TR binding is depicted as in double-stranded DNA. Only the top strand of DNA is depicted in the three TREs that confer high  $T_3$  responsiveness to a promoter. The arrows indicate the orientation of half-sites in a TRE, and “n” can be G, A, T, or C. The hexamer sequence and optimal half-site arrangements are derived from studies of synthetic TREs.

*in vivo* because of the ubiquitous presence of these proteins.

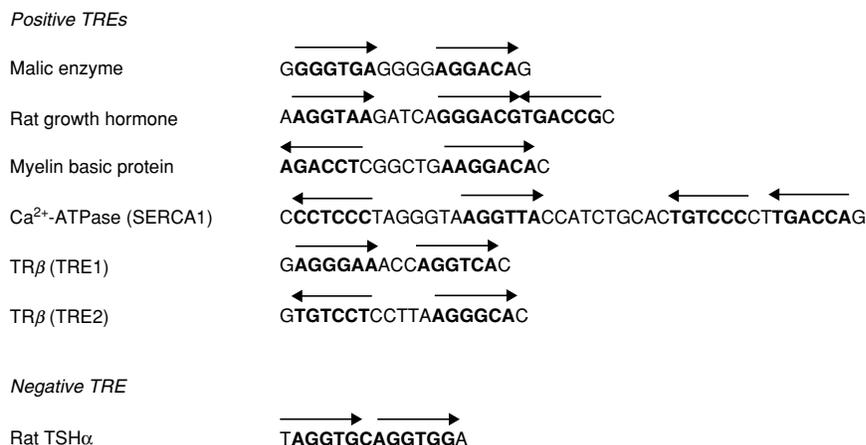
Interactions between different hormone receptors within the family are possible given their similar structure and DNA-binding properties. *In vitro* studies support this, showing, for instance, that estrogen receptors can interfere with TR action on  $T_3$ -responsive promoters, whereas TRs can bind to and block estrogen response elements. This so-called crosstalk may be relevant in tissues in which both hormone signaling routes are active, such as the hypothalamus in this example, but *in vivo* data are lacking.

### Stimulation of Gene Transcription by $T_3$

One aspect of TRs that is unique in the nuclear receptor family is the fact that TRs and TR-RXR heterodimers are bound to TREs in the unliganded state, whereas the other receptors require ligand for DNA binding. Furthermore, binding of unliganded TRs results in a substantial repression of basal promoter activity. Activation of transcription is induced by binding of  $T_3$  to the TR. This not only relieves the repression but also further stimulates promoter activity. Figure 4 summarizes the current model of the *in vivo* molecular mechanism underlying this action of  $T_3$ . The essential classes of factors involved are depicted, but different proteins exist within each class, some of which show tissue-specific expression. Transcription is repressed in the absence of  $T_3$  because the TR recruits a histone deacetylase to the promoter by means of a ubiquitous corepressor protein that serves as an adapter, having binding sites for both TR and the HDAC protein. Binding of  $T_3$  induces a marked conformational change in the TR, which disrupts binding of the corepressor complex. Instead, binding of coactivator proteins is now possible. Several such proteins form a complex with histone acetylase activity. This in turn opens up the local chromatin structure, allowing the assembly of an active transcription complex. Additional direct interactions of the TRs with basal transcription factors appear to further facilitate this process.

### Inhibition of Gene Transcription by $T_3$

The mechanism of  $T_3$ -induced repression of transcription, such as in the case of the TSH- $\alpha$ , TSH- $\beta$ , and TRH genes, is not understood. Corepressors and coactivators appear to be equally involved in this mode of regulation but with exactly opposite effects compared to the positively regulated genes (i.e., unliganded TRs stimulate transcription, which is abolished after  $T_3$  binding). Although TRE-like binding sites or even single half-sites have been implicated in



**Figure 3** A selection of natural TREs illustrating the variation in half-site sequence and spacing between half-sites. All indicated half-sites are functional in these TREs and the invariable central two Gs (Cs) are essential for TR binding. TREs from the rat growth hormone gene and the Ca<sup>2+</sup>-ATPase gene from rat skeletal muscle are examples of composite response elements. Both TREs from the human TRβ promoter contribute to the T<sub>3</sub> responsiveness of this gene. The TRE from the rat TSHα gene is an example of a half-site structure conferring a negative T<sub>3</sub> response.

some down-regulated promoters, a consensus negative TRE has not been defined.

Many genes are expected to be directly regulated by T<sub>3</sub> based on the responsiveness of expression under various conditions. The list of promoters that are shown to contain functional TREs is continuously growing. A selection of such genes are presented in [Table I](#) together with a short description of the function of each gene.

## EFFECTS OF THYROID HORMONE ON TARGET TISSUES

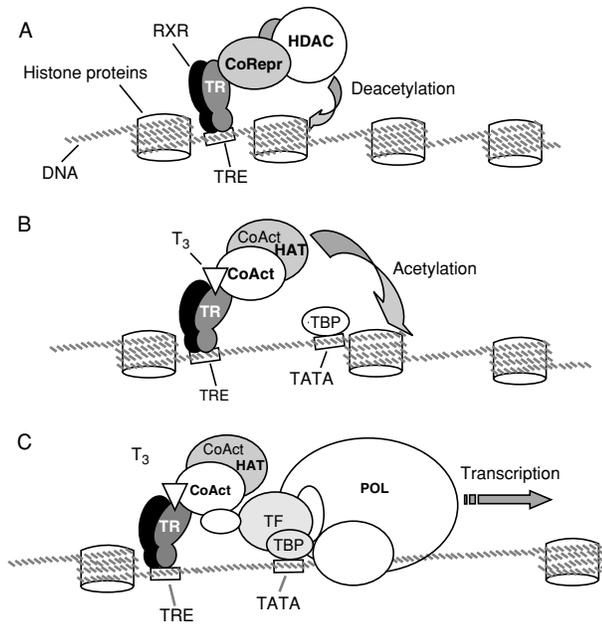
The large number of different processes, factors, and mechanisms summarized in the previous sections provide an explanation for the variety of tissue-specific actions of thyroid hormone. The responsiveness of tissues with respect to transcription of specific genes is dependent on the mix of TRs and their numerous cofactors, the details of which are poorly understood. Furthermore, the rate of cellular T<sub>3</sub> and/or T<sub>4</sub> uptake and local activity of hormone-metabolizing deiodinases will determine a tissue's T<sub>3</sub> concentration. As noted previously, many of the physiological effects of altered thyroid hormone levels are secondary to a limited number of direct actions of T<sub>3</sub>, and a large-scale analysis of gene expression indicates that the current list of genes that are directly regulated by T<sub>3</sub> is far from complete.

The effects of thyroid hormone on some of the major responsive tissues, based on both clinical and animal

experimental data, are described in the following sections. The emphasis is on the effects for which the primary targets of thyroid hormone are known.

### Brain

The development of the central nervous system is severely impaired in the absence of sufficient levels of thyroid hormone. The incomplete maturation of the brain is evidenced by neurological defects and mental retardation in humans. The timing of connections between axons, dendrites, and their targets in the developing brain is critical for the formation of the neuronal network. Diminished axonal growth and reduced dendritic arborization are evident in cerebral cortex, hippocampus, and cerebellum and in the visual and auditory cortex of hypothyroid rats. The expression of many genes is dependent on T<sub>3</sub>, but only a few have been shown to be directly regulated, such as myelin basic protein and brain-derived neurotrophic factor. The action of T<sub>3</sub> on the other genes may in part be permissive, which means that T<sub>3</sub> is essential for the action of another factor. For example, the stimulation of expression of ornithine decarboxylase by growth hormone (GH) in the brain is absolutely dependent on thyroid hormone. This enzyme is essential in nucleic acid and protein synthesis, and thyroid hormone alone has no effect on its expression. Interestingly, in hypothyroidism all enzymes that are dependent on thyroid hormone eventually reach normal levels of expression. It therefore appears that



**Figure 4** Model of the T<sub>3</sub>-dependent stimulation of transcription in a T<sub>3</sub>-responsive gene. (A) Basal or repressed state in the absence of T<sub>3</sub>. DNA is in its compacted form wound around histone proteins (nucleosomes) and bound to other ubiquitous proteins (not depicted). TR-RXR dimers are bound to the TRE. TR forms a complex with a protein with histone deacetylase activity (HDAC), which maintains the compact structure. The interaction is through a corepressor protein (CoRepr). (B) Activation. The corepressor complex is replaced by coactivator proteins (CoAct) upon binding of T<sub>3</sub> to its receptor. Histone acetylase activity (HAT) in the CoAct proteins opens up the nucleosome structure. The TATA box-binding protein (TBP) is a general transcription factor that binds to the DNA sequence TATA at the transcription initiation site. (C) Transcription. Following assembly of multiple general transcription factors (TF), which is stabilized by proteins that interact with the TR complex, the RNA polymerase complex (POL) is recruited to the promoter and transcription ensues.

the timing of thyroid action is critical, and that there is a transient period of hormone responsiveness during the development of the brain.

## Bone

Development and growth of bone are critically dependent on thyroid hormone, as evidenced by the typical short stature of adults when neonatal hypothyroidism remains untreated. The cell types that are involved in bone formation (osteoblasts) and bone resorption (osteoclasts) are both stimulated by T<sub>3</sub>, with effects on several critical enzymes. In hyperthyroid patients, increased bone formation and resorption result in a net loss of bone thickness and increase in porosity, leading to greater risk of fractures. Little

is known about the primary targets and mechanism of action of T<sub>3</sub> in bone cells, mainly because these cells are difficult to study in culture. Thyroid hormone is required for the effects of GH on bone growth and development, which involve the production of insulin-like growth factor 1. Some of the effects of T<sub>3</sub> may therefore result from this permissive action of thyroid hormone. However, studies using TR knockout mice indicate that many of the effects of T<sub>3</sub> on bone metabolism are nuclear mediated and independent of GH.

## Heart

Cardiovascular effects are among the most prominent clinical manifestations of thyroid disease. The expression of key cardiac proteins is dependent on thyroid hormone, as evidenced by the profound changes in cardiac performance in the transition from hypothyroidism to hyperthyroidism. Thyroid hormone directly regulates the expression of genes involved in membrane depolarization, muscle contraction, Ca<sup>2+</sup> handling, and adrenergic signaling (Table I). These actions include positive and negative regulation of transcription. In addition, extranuclear effects on membrane ion fluxes contribute to the increase in heart rate and a reduction in vascular resistance. All these effects combine to increase the performance of the heart, with a concomitant increase in energy turnover. A key factor is the stimulation of the Ca<sup>2+</sup>-pump of the sarcoplasmic reticulum (SR), whereas its inhibitory protein phospholamban is reduced. The SR membrane system is the intracellular store for Ca<sup>2+</sup> required for each contraction. The higher Ca<sup>2+</sup>-pump activity increases the amount of Ca<sup>2+</sup> that can be released from the SR as well as the speed with which it can be taken up again during relaxation. The result is an increased contractile capacity of the heart. Together with the increase in heart rate, the hemodynamic load placed on the heart increases with higher T<sub>3</sub> levels. Such mechanical load is a major independent stimulus for growth of the heart and responsible for the progressive cardiac hypertrophy seen with increasing thyroid hormone levels. The combined effects of T<sub>3</sub> on cardiac function and vascular resistance result in an increase in cardiac output adapted to the higher overall metabolic demand made by the organism.

## Skeletal Muscle

The overall effect of thyroid hormone on skeletal muscle is similar to that described previously for the

Table I Genes Known to be Transcriptionally Regulated by Thyroid Hormone<sup>a</sup>

	Gene function
<b>Gene transcription stimulated by T<sub>3</sub></b>	
BDNF	Brain-derived neurotropic factor. Stimulation of expression of this brain-specific regulatory factor may in part explain the critical role of thyroid hormone in brain development.
Myelin basic protein β <sub>1</sub> -Adrenergic receptor	Major protein component of the insulating myelin sheaths surrounding axons. Stimulation of expression of this receptor accounts for the potentiating effect of thyroid hormone on the adrenergic responsiveness of many tissues, such as liver, fat, skeletal muscle, and heart.
TR-β <sub>1</sub>	Thyroid hormone receptor β <sub>1</sub> . Stimulation of expression of this TR is particularly important for the timing of brain development.
HCN2	Hyperpolarization-activated cyclic nucleotide-gated channel. Cardiac ion channel involved in pacemaker activity.
MHC-α	Myosin heavy-chain α. One of the two ATPase proteins responsible for cardiac contraction. MHC-α imparts faster contraction at higher energy consumption than MHC-β. MHC-β is the predominant isoform in human heart.
SERCA2a	Sarcoplasmic reticulum (SR) Ca <sup>2+</sup> -ATPase. Ion pump that regulates intracellular calcium, playing a key role in the contraction-relaxation cycle of the heart. SERCA2a is also expressed in skeletal muscle.
SERCA1	Isoform of SERCA2a but exclusively expressed in skeletal muscle. It can be expressed at much higher levels than SERCA2a and its activity may account for up to 50% of the total energy consumption of contracting muscle.
GLUT4	Insulin-responsive glucose transporter. Responsible for insulin-stimulated uptake of glucose, particularly in skeletal and cardiac muscle and adipose tissue.
PEPCK	Phospho-enol-pyruvate carboxy kinase is a key enzyme in gluconeogenesis in liver and skeletal muscle.
Malic enzyme	Malic enzyme, or malate dehydrogenase, catalyzes the last oxidation step in the citric acid cycle in mitochondria.
Cholesterol hydroxylase	Cholesterol 7α-hydroxylase is a key enzyme in the hepatic degradation of cholesterol into bile acids.
Na <sup>+</sup> /K <sup>+</sup> -ATPase	Ubiquitous ion pump in the plasma membrane of cells responsible for maintaining the gradients of sodium and potassium between the extracellular and intracellular space.
MyoD and myogenin	Two different transcription factors that drive the expression of many muscle-specific genes during the differentiation of skeletal muscle (myogenesis).
rGH	Rat growth hormone. The TRE in this pituitary gene was the first to be analyzed in detail. A similar TRE has not been identified in the human GH gene.
Deiodinase type 1	This enzyme is responsible for the peripheral production of T <sub>3</sub> through deiodination of T <sub>4</sub> . It is primarily expressed in liver and kidney.
UCP1	Stimulation of expression of uncoupling protein 1 is responsible for the T <sub>3</sub> -induced thermogenic response in brown adipose tissue. It uncouples mitochondrial respiration, thereby converting oxidative energy directly to heat.
<b>Gene transcription inhibited by T<sub>3</sub></b>	
MHC-β	Myosin heavy-chain β. Slower and more energy-efficient isoform of the two myosin ATPases responsible for cardiac contraction. The hypothyroid rodent heart expresses almost exclusively MHC-β, which is nearly completely replaced by MHC-α in hyperthyroidism. This shift occurs to some extent in the hyperthyroid human heart.
TSH-α and-β	Both anterior pituitary genes encoding the α- and β-subunits of thyroid-stimulating hormone (TSH or thyrotropin) are negatively regulated by thyroid hormone. This is part of the negative feedback loop regulating plasma thyroid hormone levels.
TRH	Thyrotropin-releasing hormone, synthesized in the hypothalamus, is the principal positive regulator of TSH synthesis. Repression of transcription of this gene is another aspect of the feedback inhibition of thyroid hormone production.

<sup>a</sup>One or more TREs have been identified in the promoters of these genes.

heart: that is, an increase in performance at the expense of a higher energy turnover. Stimulation of the rates of contraction and relaxation by thyroid hormone is a classic observation in thyroidology. Changes in expression of genes similar or identical to those in cardiac tissue underlie this effect, but multiple isoforms and cell-specific and innervation-dependent factors make for a less straightforward regulation by  $T_3$ . Generally,  $T_3$  induces a shift in the expression of myosin-ATPase proteins, which determine the rate of contraction, from the slow myosin heavy-chain form (MHC I) to faster MHC II forms. MHC I is identical to MHC- $\beta$  in heart and this gene is also negatively regulated in skeletal muscle by  $T_3$  at the promoter level. Direct regulation is not established for the multiple MHC II forms that are similar, but not identical, to cardiac MHC- $\alpha$ . As in heart, the relaxation rate is determined by the level of expression of the  $Ca^{2+}$ -pump of the SR. Muscle also expresses the SERCA2 form present in the heart, but  $T_3$  preferentially stimulates the expression of a second isoform (SERCA1). This protein is virtually identical to the SERCA2 form, but the promoter of the SERCA1 gene is completely different, with different TREs allowing much higher expression of the protein compared to the SERCA2 form. Proliferation of the SR membrane is also stimulated by  $T_3$ . The expression of the fast MHC isoforms and the high  $Ca^{2+}$ -pump levels increase the energy cost of contraction two- or threefold in the transition from hypothyroidism to hyperthyroidism. The economy of force production is consequently reduced by  $T_3$  since absolute force production is not affected by these changes.

The timing of expression of a series of fetal and adult MHC isoforms during development is dependent on thyroid hormone. Similar to brain, the adult expression patterns are ultimately obtained irrespective of the thyroid status. This is not the case for the SERCA1 gene, which is absolutely dependent on thyroid hormone. Effects of neonatal hypothyroidism on muscle development in rats are fully reversible by treatment with thyroid hormone.

## Liver

Plasma levels of cholesterol, particularly low-density lipoprotein (LDL) cholesterol, are increased in hypothyroid patients. This is the result of decreased LDL cholesterol uptake by the liver (due to reduced LDL receptor numbers) and reduced hepatic metabolism of cholesterol. These effects can be reversed by treatment with thyroid hormone. The enzyme cholesterol 7 $\alpha$ -hydroxylase is the rate-limiting step in the degradation

of cholesterol into bile acids, and its expression is stimulated by  $T_3$  at the promoter level. Whether the expression of LDL receptors is also directly stimulated by  $T_3$  is not known.

$T_3$  stimulates the expression of a number of enzymes involved in lipogenesis, including malic enzyme, glucose-6-phosphate dehydrogenase, and fatty acid synthase. Malic enzyme is particularly responsive and the promoter of this gene contains a strong TRE. Interestingly, the same enzyme is expressed in brain, but it is unresponsive to  $T_3$ . In liver, several signal routes involved in energy metabolism, such as insulin signaling, converge on the promoter of this gene and interact with the  $T_3$  signal. For instance, the stimulatory effect of  $T_3$  on malic enzyme expression is increased 10-fold when a normal diet is replaced by a high-carbohydrate/fat-free (lipogenic) diet.

Synthesis of glucose (gluconeogenesis) from precursors such as amino acids and glycerol is an important liver function.  $T_3$  stimulates the expression of key enzymes in this process, including pyruvate carboxylase, glucose-6-phosphatase, and phospho-enol-pyruvate carboxy kinase (PEPCK). A TRE has been identified in the promoter of the PEPCK gene.

In addition,  $T_3$  stimulates the expression of the deiodinase type I gene, which converts  $T_4$  to  $T_3$ . This enzyme is responsible for most of the circulating  $T_3$  in the body and its promoter contains several TREs.

## Adipose Tissue

Differentiation of preadipocyte cells into mature white adipose tissue is dependent on thyroid hormone, and the same key enzymes involved in lipid metabolism in liver are influenced by  $T_3$  in fat tissue. High levels of  $T_3$  shift the balance between fat synthesis and breakdown to net lipolysis, resulting in generalized fat mobilization and loss of body fat stores. Although brown adipose tissue (BAT) is virtually absent in adults, as it is in most larger mammals, newborns have appreciable amounts of this type of fat tissue. It gets its color from the cytochromes in mitochondria, which are particularly abundant in these fat cells. In rodents, BAT is the site of the thyroid hormone-dependent extra heat production in response to prolonged cold exposure. Circulating thyroid hormone levels do increase, but this is not essential for cold adaptation and the action of the hormone appears to be permissive. Surprisingly, however, the adipocytes greatly increase their  $T_3$  content by increasing the expression of deiodinase type II. Like type I, this enzyme converts  $T_4$  to  $T_3$  and its expression is stimulated by the higher levels of

circulating adrenergic hormones triggered by the cold stress. The increase in  $T_3$  in the brown adipocyte in turn induces the transcription and expression of a protein called thermogenin or uncoupling protein 1, which uncouples respiration in mitochondria. This means that the energy produced by the numerous mitochondria in these cells is not stored in the form of ATP but rather directly released as heat. This mechanism is equally important in maintaining body temperature in newborns and it is an example of how thyroid hormone action may be regulated at the level of individual tissues.

### See Also the Following Articles

Amiodarone and Thyroid • Drug Effects and Thyroid Function • Iodine • Resistance to Thyroid Hormone

(RTH) • Thyroid Hormone-Binding Proteins • Thyroid Hormone Metabolism • Thyroid Hormone Receptors

### Further Reading

- Hulbert, A. J. (2000). Thyroid hormones and their effects: A new perspective. *Biol. Rev.* **75**, 519–631.
- Klein, I., and Ojamaa, K. (2001). Mechanisms of disease: Thyroid hormone and the cardiovascular system. *N. Engl. J. Med.* **344**, 501–509.
- Morreale de Escobar, G. (2001). The role of thyroid hormone in fetal neurodevelopment. *J. Pediatr. Endocrinol. Metab.* **14**, 1453–1462.
- O'Shea, P. J., and Williams, G. R. (2002). Insight into the physiological actions of thyroid hormone receptors from genetically modified mice. *J. Endocrinol.* **175**, 553–570.
- Yen, P. M. (2001). Physiological and molecular basis of thyroid hormone action. *Physiol. Rev.* **81**, 1097–1142.
- Zhang, J., and Lazar, M. A. (2000). The mechanism of action of thyroid hormones. *Annu. Rev. Physiol.* **62**, 439–466.

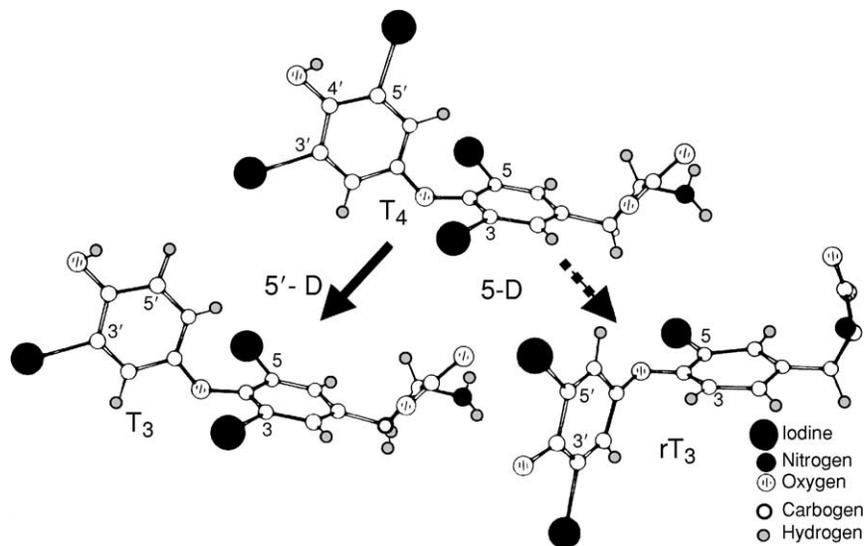
Deiodinase activity in the pituitary thyrotroph cells controls the negative feedback regulation of thyroid hormones on thyrotropin synthesis and secretion.

## INTRODUCTION

The roles of the essential trace element iodine and of the thyroid gland for development and normal function of vertebrate organisms, including humans, have been known for a long time, but it was not until 1915 that iodine containing prohormone thyroxine ( $T_4$ ) was isolated by Edward C. Kendall. Charles R. Harington and George Barger performed the structure analysis and chemical synthesis of  $T_4$  in 1927, and in 1952 Jack Gross and Rosalind Pitt-Rivers identified the bioactive thyroid hormone 3,3',5-triiodothyronine ( $T_3$ ), the "real thyroid hormone" (Fig. 1). Various enzymes metabolizing the amino acid-derived thyroid hormones have been characterized during the past 60 years. These enzymes generate a complex pattern and network of  $T_4$ -derived iodothyronine metabolites, some of which have specific biological functions that differ from those of the bioactive thyroid hormone  $T_3$ .  $T_3$  exerts its influence mainly by modulation of action of the ligand-dependent transcription factors, the nuclear  $T_3$  receptors (TRs). In addition to this key mechanism of action, several other biological effects of  $T_4$  and iodothyronines derived therefrom have been reported. However, these effects are less well characterized (e.g., activation of processes associated

with the plasma membrane of cells, modulation of the organization of the cytoskeleton, or direct influence of iodothyronines on mitochondrial function). Iodothyronines comprise iodinated metabolites of the prohormone  $T_4$  that have an intact diphenylether structure that is generated by coupling of two iodinated tyrosines residues of thyroglobulin during biosynthesis of the hormone in the thyroid gland. The only source of  $T_4$  is the thyroid gland, whereas the other iodinated metabolites are formed intracellularly in various organs and tissues.  $T_4$  is a highly hydrophobic, lipid-soluble compound that circulates in the blood bound to three main transport or distribution proteins: thyroxine-binding globulin (TBG), transthyretin (TTR), and albumin.  $T_4$  and its metabolites are excreted mainly via the bile and the feces. Iodide liberated from the iodothyronine metabolites by enzymatic deiodination, together with the fraction of iodide taken up by food, is excreted into the urine. Urinary iodide concentrations provide an indicator of the iodine supply of the organism.

$T_4$  and the iodothyronines are derivatives of the amino acid tyrosine and carry ionic charges in solution. Therefore, efflux of  $T_4$  and  $T_3$  from the thyrocytes and uptake of iodothyronines into target cells require transport processes such as facilitated transport or active or passive carrier-mediated transport, which is catalyzed by specific proteins not completely characterized. Tissue- and development-specific expression of these transporters of iodothyronines contributes to



**Figure 1** Enzymatic deiodination of the prohormone  $T_4$  to the thyromimetically active  $T_3$  by type I or type II 5'-deiodinase and inactivation of  $T_4$  to  $rT_3$  by type III 5-deiodinase. 5'-D, 5'-deiodination (phenolic or outer-ring deiodination); 5-D, 5-deiodination (tyrosyl- or inner-ring deiodination). Numbers indicate substituent positions ( $n'$ ) at the phenolic (outer) ring or ( $n$ ) at the tyrosyl (inner) ring.

the bioavailability of thyroid hormones in cells and organs and provides a first level of control for metabolism of iodothyronines by intracellular enzymes that depend on substrate availability. No enzymes metabolizing thyroid hormones in the blood or in the interstitial space have been identified. The free hormone concentrations in serum and interstitial space are very low due to the binding of  $T_4$  and  $T_3$  by the high-affinity distribution proteins TBG, TTR, and albumin. The other iodothyronine metabolites exhibit lower affinity for these proteins than  $T_4$  and  $T_3$  and thus might be more available to metabolic transformations.

## ENZYMATIC DEIODINATION OF THYROID HORMONES

### Sequential Monodeiodination Both Activates and Inactivates Thyroid Hormones

Reductive deiodination of the prohormone  $T_4$  generates two iodothyronine metabolites with different biological function (Fig. 1). Deiodination in the 5' position of the phenolic (outer) ring of  $T_4$  produces the thyromimetically active hormone  $T_3$ . Deiodination in the 5 position of the tyrosyl ring (inner ring) generates reverse  $T_3$  ( $rT_3$ ), a metabolite devoid of thyromimetic action at the nuclear  $T_3$  receptor but with the potential to act as a competitor to  $T_4$  in the 5'-deiodination reaction.  $rT_3$  circulates in the serum of normal persons in concentrations similar to those of  $T_3$ . Deiodinase enzymes are expressed in a development-, tissue-, and cell-specific manner and are regulated by different factors. All three deiodinase enzymes belong to the class of selenocysteine-containing proteins and are encoded by three different genes. With few exceptions, selenium availability is not a limiting factor in humans for the expression of the three deiodinase selenoenzymes, which rank very high in tissue-specific selenium supply among the selenoproteins. However, *in vitro* and animal studies indicate selenium-dependent regulation of the three deiodinases. Only during severe selenium deficiency, under protein-free diets due to metabolic diseases (phenylketonuria), long-term total parenteral nutrition, or malabsorption syndromes may effects of selenium deficiency on the expression of deiodinase enzymes and  $T_3$  formation be observed in children and adults.

### Two 5'-Deiodinases Are Involved in Activating Thyroid Hormone Metabolism

5'-Deiodination at the phenolic ring is catalyzed by two enzymes, the type I and the type II 5'-deiodinase (5'-DI and 5'-DII, respectively), which differ in various aspects as indicated in Table I. 5'-DI is less selective in substrate specificity than 5'-DII, which preferentially deiodinates  $T_4$  with a low-nanomolar Michaelis–Menten constant ( $K_m$ ) for this substrate. 5'-DII produces  $T_3$  mainly for the local supply in tissues expressing this enzyme, which renders these tissues independent from circulating  $T_3$ . 5'-DII does not contribute significantly to circulating plasma  $T_3$ , which is mainly provided by 5'-DI action on  $T_4$  in liver, kidney, and thyroid tissues expressing high 5'-DI activity in normal healthy subjects. The extent to which thyroid 5'-DI activity contributes to  $T_3$  secreted by the thyroid gland, which also produces  $T_3$  apart from  $T_4$  during thyroid hormone synthesis especially under conditions of iodine deficiency and in Graves' autoimmune thyroid disease, is unclear. In the latter constellation, 5'-DII activity also contributes to thyroid  $T_3$  production from  $T_4$  because 5'-DII expression in Graves' disease is stimulated by thyroid-stimulating autoantibodies, similar to thyrotropin (TSH) stimulation of 5'-DII in autonomous adenoma. 5'-DI also participates in the inactivation of  $T_4$  by deiodination at the tyrosyl ring, generating  $rT_3$  and degrading  $T_3$  to 3,3'- $T_2$ . Sulfated iodothyronines are also substrates of 5'-DI.

### Type III 5-Deiodinase Inactivates Thyroid Hormones

The third deiodinase, 5-deiodinase (5-DIII), removes iodine from the 5 position of the tyrosyl ring (inner-ring deiodination) (Fig. 1). 5-DIII is responsible for the majority of  $rT_3$  formation and  $T_3$  degradation and is considered the main thyroid hormone-inactivating enzyme. Expression of 5-DIII occurs in several tissues and cells that do not respond to thyroid hormone action. Thus, 5-DIII can be considered the one component of the "gatekeepers" of thyroid hormone action that prevents  $T_3$  action from occurring at an inappropriate time, location, or concentration. Because 5-DIII can act on both  $T_4$  (forming  $rT_3$ ) and  $T_3$  (forming 3,3'- $T_2$ ), it can perform the gatekeeping function in cell types that are exposed to  $T_4$  and/or  $T_3$  and thus might act independently of the expression pattern of cell type-specific thyroid hormone

Table I Properties of the Deiodinase Enzymes<sup>a</sup>

	Type I 5'-deiodinase (5'-DI)	Type II 5'-deiodinase (5'-DII)	Type III 5-deiodinase (5-DIII)
Function	Systemic > local T <sub>3</sub> production; degradation of rT <sub>3</sub> and sulfated iodothyronines	Local > systemic T <sub>3</sub> production	Inactivation of T <sub>4</sub> and T <sub>3</sub>
Expression	Liver, kidney, thyroid, pituitary; at low levels in heart and many other tissues; loss of expression in tumors	(Hypothyroid) pituitary, brain, brown adipose tissue, skin, placenta; thymus, pineal, and hardierian gland; glial cells and tanyocytes; overexpression in mesothelioma cell line	Placenta, brain; many other tissues, except normal adult pituitary and adult healthy liver, thyroid, kidney; reexpression in pathological organs (liver, pituitary); overexpression in hemangioma
Cosubstrate	DTT or DTE <i>in vitro</i> (K <sub>m</sub> , 2–5 mM); not glutathione or thioredoxin <i>in vivo</i>	DTT or DTE <i>in vitro</i> (K <sub>m</sub> , 5–10 mM); higher concentrations than for 5'-DI	DTT or DTE <i>in vitro</i> (K <sub>m</sub> , 10–20 mM); higher concentrations than for 5'-DI
Subcellular location	Endoplasmatic reticulum in liver; inner plasma membrane in kidney and thyroid	Inner plasma membrane; p29 subunit associated with F-actin respectively perinuclear vesicles	Endoplasmatic reticulum
Human gene location	1p32–p33	14q24.3	14q32
Essential amino acid residues	Histidine, selenocysteine, cysteine, phenylalanine	Selenocysteine	Selenocysteine
Enzyme stimulation or induction	T <sub>3</sub> , retinoids; TSH and cAMP in thyroid only; testosterone in liver; selenium, carbohydrate	cAMP; FGF; β-adrenergic agonists, nicotine; phorbolsters via PKC; ANP and CNP via cGMP in glial cells	T <sub>3</sub> , FGF, EGF; selenium
Inhibition	PTU, iodoacetate, aurothioglucose, iopanoate, dexamethasone	T <sub>4</sub> , rT <sub>3</sub> , iopanoate	Iopanoate

<sup>a</sup>Abbreviations used: ANP, atrial natriuretic peptide; CNP, C-type natriuretic peptide; DTE, dithioerythreitol; DTT, dithiothreitol; EGF, epidermal growth factor; FGF, fibroblast growth factor; PKC, protein kinase C.

transporters, restricting the uptake of these hormones. On the other hand, the development-, tissue-, and cell-specific expression pattern of the two 5'-deiodinase enzymes provides a system of systemic or local production of the thyromimetically active nuclear receptor ligand T<sub>3</sub>. Expression of the deiodinase enzymes in the thyrotroph cells of the anterior pituitary controls the local production and levels of T<sub>3</sub>, which suppresses gene expression and secretion of TSH. Therefore, deiodinases play a key role in negative feedback regulation of TSH by circulating thyroid hormone levels.

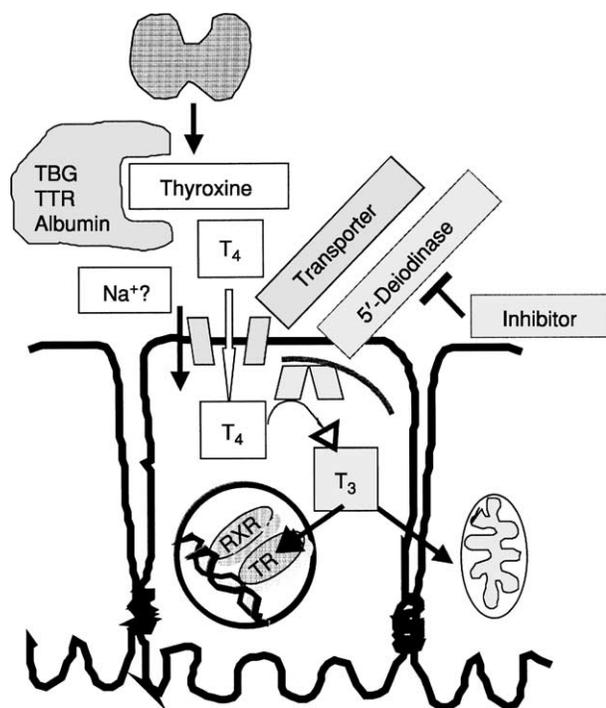
## INTRACELLULAR TARGETING AND COMPARTMENTALIZATION OF ACTIVE AND INACTIVE IODOTHYRONINES

### Subcellular Location of Deiodinases

Because all three deiodinase enzymes are located inside the cell with active sites exposed toward the cytosol (Fig. 2), their action also depends on the

availability of the main substrate T<sub>4</sub>, which reaches the cytosol via cell-type-specific T<sub>4</sub> carriers. T<sub>3</sub> entering the cells via T<sub>3</sub> transporters may therefore bypass the control exerted by the intracellular deiodinase network and gain direct access to the nuclear T<sub>3</sub> receptors. There is evidence that in some cells compartmentalization or distinction occurs for the two different T<sub>3</sub> pools. T<sub>3</sub> locally originating from the two 5'-deiodinases and T<sub>3</sub> taken up by plasma membrane T<sub>3</sub> transporters might be handled differently and might reach the nuclear T<sub>3</sub> receptors through distinct intracellular routes. Moreover, T<sub>3</sub> formed by the high activity of 5'-DI in liver, kidney, or thyroid might not be targeted to the nuclear T<sub>3</sub> receptor of the same cells but rather channeled toward efflux and export into circulation and thus reach (other) target cells only after its transient circulation in the blood pool of thyroid hormones.

The determination of the detailed subcellular location and extensive research on the intracellular compartmentalized fate of iodothyronines are required to clearly understand the molecular steps of thyroid hormone economy and action at the cellular level.



**Figure 2** Schematic representation of thyroid hormone synthesis by the thyroid; transport and distribution by the serum-binding proteins, thyroxine-binding protein (TBG) and transthyretin (TTR); cellular uptake by thyroid hormone transporters; intracellular activation of the prohormone T<sub>4</sub> to the thyromimetically active hormone T<sub>3</sub>; binding of T<sub>3</sub> to nuclear or mitochondrial T<sub>3</sub> receptors; and modulation of transcription of T<sub>3</sub>-regulated genes. RXR, retinoic acid receptor; TR, T<sub>3</sub> receptor.

5'-DI activity has been shown to be associated with the inner leaflet of the basolateral plasma membrane in proximal kidney tubular epithelial cells. The subcellular location of 5'-DI in the thyroid epithelial cells is not unequivocally clear, and in liver 5'-DI has been shown to be located in the rough and smooth endoplasmic reticulum with orientation toward cytosol, a finding that does not exclude an additional location at a lower concentration in the inner leaflet of the plasma membrane. 5'-DII activity has been demonstrated to be associated with the inner leaflet of the plasma membrane of several cell types. Type III 5D activity is also associated with cellular membranes. All three deiodinase enzymes are important membrane enzymes whose activity depends on membrane integration. Studies on the subcellular location of the three deiodinases using transient or stable overexpression of these selenoproteins, mainly in cell types that normally do not express deiodinase activity, have provided information on the different locations of these proteins, and data on the inactivation of the deiodinases by the ubiquitin-associated proteasome

protein degradation system have also been presented. However, these experiments should be interpreted with caution because overexpression models were used, an artificial cellular environment was chosen, and mainly marker enzymes or proteins were determined that do not correspond to classical and established integral membrane protein markers of rough and smooth endoplasmic reticulum, cytosolic- or luminal-oriented microsomal proteins, plasma membrane-associated proteins, apical or basolateral plasma membrane protein markers, and intra- or extracellularly oriented integral plasma membrane proteins, which have not been systematically studied. In addition, the proteasome protein degradation system has not been studied in detail for integral membrane proteins and is not well understood for cell lines overexpressing proteins outside of the specific cellular context.

### Iodothyronine Deiodinases Participate in the Control of Cell-Type-Specific Production and Action of Thyroid Hormone

It has long been known that thyroid hormone action, mainly exerted by the thyromimetically active T<sub>3</sub>, depends on the cellular expression pattern of T<sub>3</sub> receptors, which mediate thyroid hormone action on gene expression. The cloning of several isoforms of (nuclear) thyroid hormone receptors, encoded by two different genes (TR- $\alpha$  and TR- $\beta$ ), which occur in different splice forms that bind either both ligand and DNA (TR- $\alpha_1$ , TR- $\beta_1$ , TR- $\beta_2$ , and several variants) or only DNA (TR- $\alpha_2$ ), added to the complexity of the regulation of thyroid hormone action. This complexity has become even more complicated by the demonstration that in many instances the cellular location of the T<sub>3</sub> receptor differs from that of T<sub>3</sub> formation by the 5'-DI or 5'-DII enzymes. Two examples are the distinct location of 5'-DII-forming T<sub>3</sub> and the T<sub>3</sub> receptor-mediating T<sub>3</sub> action in the development of the cochlea of the inner ear of the mouse or the development of the frog eye. Whereas T<sub>3</sub> receptors are located in the sensory epithelium of the inner ear, the 5'-DII enzyme producing the active ligand T<sub>3</sub> is located in the surrounding connective tissue cells in a development-dependent manner. This implies that T<sub>3</sub> produced from T<sub>4</sub> must be released from the connective tissue cells and taken up by the sensory epithelial cells. Therefore, defects in both the time-dependent and the spatial organization of this network of ligand availability might lead to disturbed hearing function,

known to be affected by (fetal or congenital) hypothyroidism. In the development of the retina of the frog eye, a close correlation between the expression of  $T_3$  receptors and 5-DIII is reported. Local overexpression of 5-DIII, removing the active ligand  $T_3$ , disturbs the development of the eye. Compartmentalization of the  $T_3$  receptors, the ligand-activating 5'-deiodinase, and the ligand-inactivating 5-DIII occurs at the implantation site of the embryo in the uteroplacental unit and the fetal epithelium. Again, the expression patterns of the three deiodinase enzyme forms appear strictly associated with different cell types in a time- and space-dependent manner, thus allowing for strictly controlled ligand availability to the developing fetus. In the latter case, the deiodinases, especially the  $T_3$ - and  $T_4$ -inactivating enzyme 5-DIII, are essential for the control of thyroid hormone transfer from the maternal to the fetal compartment because it is well-known that excessive exposure of the fetal tissues and the fetal brain to thyroid hormone leads to irreversible brain damage such as occurs in uncontrolled maternal hyperthyroidism during pregnancy.

### Amphibian Metamorphosis Is a Paradigm of Spatial and Temporal Control of Thyroid Hormone Expression

The previous examples of spatial and temporal control of the active ligand  $T_3$  relate to the developmental program of expression of  $T_3$  receptors and deiodinase enzymes in amphibian metamorphosis that is under the central control of thyroid hormone. Although the thyroid gland provides the prohormone  $T_4$  and a small amount of circulating  $T_3$  (which might be needed to prime the metamorphic events), the local production and removal of the active ligand  $T_3$  are under the control of the deiodinase network, which is expressed in a cell-type-specific manner. Whether  $T_4$  and  $T_3$  transporters also contribute to precursor or ligand availability remains to be determined. Several cell-type-specific transporters for thyroid hormones have been identified. Disruption of these organized networks of control of ligand availability for  $T_3$  receptors might occur during various diseases, affecting either the cellular organization of tissues or the expression of the three deiodinase enzymes.

### ABERRANT AND INAPPROPRIATE EXPRESSION OF DEIODINASES

The concept of the control of availability of the thyromimetically active ligand  $T_3$  by appropriate expression of thyroid hormone deiodinases has gained

support as a result of the identification of several pathological constellations in which aberrant or inappropriate expression of deiodinase enzymes has been observed. The first example is the overexpression of 5-DIII in some infantile hemangiomas leading to extremely severe hypothyroidism that in some cases cannot be treated with excessive doses of administered  $T_4$ —a syndrome called consumptive hypothyroidism. The opposite situation is found for some mesothelioma tumors, in which overexpression of 5'-DII leads to elevated  $T_3$  production. For 5'-DI, no clear-cut alterations of expression have been described in humans. However, several animal experimental models with altered expression or regulation of 5'-DI activity have been described. Decreased (hepatic) expression of 5'-DI leads to constellations characterized by normal to elevated serum  $T_4$ , decreased or normal serum  $T_3$ , and elevated  $rT_3$ . The latter parameter might be the leading sign of decreased expression of 5'-DI because it is almost exclusively degraded by 5'-DI, whereas its production occurs via 5-DIII and, to a minor extent, 5'-DI. Therefore, the ratio of serum  $T_3/rT_3$  might be the most sensitive parameter of altered 5'-DI activity. No evidence of overexpression of 5'-DI has been reported in humans.

Two polymorphisms have been reported in the 5'-DI gene associated with altered plasma thyroid hormone levels and ratios. In combination with a further polymorphism of the TSH receptor gene, clear-cut alterations of the setpoint of the pituitary feedback control of thyroid hormone secretion and homeostasis are apparent. A polymorphism in the 5'-DII gene leading to an amino acid exchange in the enzyme has been found to be associated with obesity in diabetic patients, but no functional alterations of characteristics of the variant 5'-DII enzyme have been found.

### LOW $T_3$ SYNDROME OR EUTHYROID SICK SYNDROME

A long-standing clinically relevant conundrum is the low  $T_3$  syndrome or the euthyroid sick syndrome, characterized by low serum levels of  $T_3$ , elevated  $rT_3$ , and normal, elevated, or, in later stages, decreased serum  $T_4$  without appropriate elevation of serum TSH levels. This peculiar hormone constellation has been explained by decreased hepatic  $T_3$  formation by 5'-DI, normal or decreased thyroid hormone synthesis, and lack of pituitary stimulation of TSH secretion (Table II). Many facets of regulation of thyroid homeostasis contribute to this syndrome, which is

**Table II Pathogenesis of Low Serum Levels of T<sub>3</sub> and T<sub>4</sub> in the Euthyroid Sick Syndrome of Nonthyroidal Illness**


---

Low T <sub>3</sub> state
Decreased hepatic Type I 5'-deiodinase activity
Increased hepatic Type III 5-deiodinase activity
Decreased concentrations of serum thyroid hormone-binding proteins
Circulating inhibitors of binding of T <sub>3</sub> to serum proteins
Decreased tissue uptake of T <sub>3</sub> by peripheral tissues
Decreased TSH secretion
Hypercortisolism
Elevated cytokines (IL-1, IL-6, and TNF- $\alpha$ )
Interference of drugs with serum binding, tissue uptake, and 5'-deiodination (dopamine, glucocorticoids, propranolol, and uremic compounds)
Low T <sub>4</sub> state
Decreased serum concentration of T <sub>4</sub> -binding proteins
Circulating inhibitors of binding of T <sub>4</sub> to serum proteins
Decreased tissue uptake of T <sub>4</sub> by peripheral tissues
Decreased TSH secretion
Hypercortisolism
Elevated cytokines (IL-1, IL-6, and TNF- $\alpha$ )
Interference of drugs with serum binding, tissue uptake, and 5'-deiodination (glucocorticoids, dopamine, amiodarone, salicylates, and uremic compounds)

---

IL, interleukin; TSH, thyroid-stimulating hormone; TNF- $\alpha$ , tumor necrosis factor- $\alpha$ .

observed during carbohydrate starvation, acute and chronic illness, surgical intervention, trauma, infection, sepsis, or after administration of certain drugs that inhibit (hepatic) 5'-DI activity. In addition to decreased hepatic and, in some cases, impaired renal and thyroidal 5'-DI activity, the remarkable interruption of the normal negative feedback of thyroid hormones at the hypothalamic and pituitary level is a hallmark of this syndrome. Experiments using various animal models, including knockout mice models, suggest that elevated levels of proinflammatory cytokines (interleukins-1 and -6, tumor necrosis factor- $\alpha$ , and interferon- $\gamma$ ) in these conditions contribute to the inhibition of deiodinase activity and the interruption of the negative feedback at the central level. In addition, both thyroid hormone binding in serum and cellular uptake appear to be changed in these conditions. If recovery or causal treatment of the underlying disease are not possible, the condition may worsen because T<sub>4</sub> production decreases (low T<sub>4</sub> syndrome), a condition with a bad prognosis and high mortality rate. Attempted treatment or substitution with thyroid hormones (T<sub>4</sub> or T<sub>3</sub>) were unsuccessful and even worsened the situation by generating a negative nitrogen balance or an even more catabolic constellation.

Important new information on the pathogenesis of this syndrome has become available. In a clinical study of intensive care patients, from a systematic analysis

of the activities of the three deiodinases in biopsies of postmortem tissues (liver and skeletal muscle) and corresponding serum levels of thyroid hormones, together with other routine clinical chemistry parameters, it became obvious that in critically ill patients a reexpression of hepatic 5-DIII occurs, which is similar to the pattern in fetal human liver. This finding may explain the rapid decrease in circulating T<sub>3</sub> and increase in rT<sub>3</sub> in combination with decreased hepatic 5'-DI activity. Lowest 5'-DI and highest 5-DIII activities were observed in patients who died from cardiovascular complications, suggesting that poor tissue perfusion alters the normal expression levels of these two deiodinases. No significant expression of 5'-DII activity was found in both kidney and muscle. In addition, successful disruption of this vicious cycle of the euthyroid sickness syndrome has been demonstrated in intensive care patients who received a combined stimulation by the hypothalamic-releasing hormones thyroliberin (TRH) and growth hormone-releasing peptide, whereas administration of either component alone was not successful in restoring the growth hormone-thyroid hormone axis in severe illness. Pulsatile secretion of the pituitary hormones growth hormone (GH), TSH, and prolactin was reamplified by this combination of releasing factors, which also substantially increased circulating levels of insulin-like growth factor-1, T<sub>4</sub>, and T<sub>3</sub> while

avoiding an increase in  $rT_3$ . Apparently, interference at the hypothalamic level of regulation enabled this metabolic improvement and the restoration of normal feedback loops. Whether this treatment regimen can be used for all variants of the euthyroid sick syndrome remains to be established. Independent support for this hypothesis derives from observations in children and adult patients treated with GH. This regimen led to decreased  $T_4$  and  $rT_3$  and increased serum  $T_3$  levels independent of alterations in TSH and suggests stimulation of hepatic 5'-DI activity. This interpretation is supported by several studies of animal models indicating increased hepatic 5'-DI expression and activity after GH treatment. No data are available on a GH-dependent decrease in hepatic 5-DIII, which is normally not expressed in rodent liver.

### TISSUE-SPECIFIC EXPRESSION PATTERNS OF DEIODINASE ENZYMES AND HOMEOSTASIS OF NET $T_3$ FORMATION

The observations of hepatic reexpression of 5-DIII and decreased activity of 5'-DI in severe illness suggest that the developmental pattern of expression of deiodinases can be reverted under certain conditions or that an alteration of the cellular composition of a given tissue results in marked alterations in the net balance of homeostasis of  $T_3$  formation. Decreased expression of 5'-DI without evidence of a concomitant increase in 5-DIII expression is also found in tumors of the thyroid, kidney, prostate, and testis. In pituitary adenoma, decreased expression of 5'-DI is associated with elevated 5-DIII activity, whereas the normal pituitary shows only low levels of this enzyme. Depending on the type of pituitary tumor, 5'-DI expression is increased or reduced. Treatment of human cell lines derived from tumors expressing deiodinase activity with agents modulating DNA methylation or histone acetylation or administration of retinoic acid, an agent known to induce cell differentiation, leads to reexpression of 5'-DI activity, even in cell lines that have lost 5'-DI expression. These observations indicate that several regulatory factors involved in development and differentiation exert marked influence on the production and homeostasis of  $T_3$ , which is essential for maintenance of the differentiated state and the metabolic function of various cell types, especially epithelial cells. Therefore, adequate production of thyromimetically active  $T_3$  seems to be a key parameter of the normal cell and functional tissue, which is under a complex network

of control by potent factors involved in development, differentiation, and maintenance of function of a vertebrate organism.

### METABOLISM OF THYROID HORMONES AT THE AMINO ACID SIDE CHAIN

Thyroid hormones and iodothyronines derived therefrom have an alanine amino acid side chain that can undergo oxidative decarboxylation and deamination, as found for other amino acids. Although no *in vivo* experimental evidence is available on decarboxylation of iodothyronines to yield metabolites with an amine side chain, oxidative decarboxylation of  $T_4$  and  $T_3$  seems to occur because relevant circulating levels of both tetraiodothyro-acetic acid (Tetrac) and the corresponding triiodothyro-acetic acid (Triac) can be measured in serum. The enzymes catalyzing these reactions have not been characterized in detail. However, the metabolites Tetrac and Triac are of potential relevance because Tetrac is a good substrate for the 5'-deiodinases and Triac is a potent ligand for nuclear  $T_3$  receptors. Whether Triac reaches the nuclear  $T_3$  receptor target *in vivo* and modulates the expression of  $T_3$ -regulated genes is unclear. Triac has a short biological half-life compared to that of  $T_3$  and is used in pharmacological doses for treatment of thyroid hormone resistance, for which it efficiently suppresses elevated TSH but does not lead to hyperthyroid conditions in heart, liver, and other organs due to its rapid degradation and lack of accumulation in tissues.

### CONJUGATION OF THYROID HORMONES BY GLUCURONIDASES AND SULFOTRANSFERASES

#### Glucuronidation

Multiple UDP-glucuronyltransferases appear to be involved in the glucuronidation of thyroid hormones. The 4'-OH group of iodothyronines is rather acidic, depending on the iodine substitution pattern of the phenolic ring, and Tetrac and Triac are better substrates for glucuronidation than their parent compounds  $T_4$  and  $T_3$ . Glucuronidated iodothyronines show increased solubility in water and are rapidly excreted via the bile, but they may also undergo enterohepatic circulation. Drugs and agents interfering with UDP-glucuronyltransferases might therefore affect thyroid hormone metabolism and elimination

as indicated by findings on the effects of several halogenated aromatic compounds (PCBs and dioxins) in animal models. Compounds that increase  $T_4$  glucuronidation decrease serum levels of  $T_4$  (e.g., phenobarbital and several halogenated aromatic compounds).

### Sulfation

The second important pathway of thyroid hormone conjugation of the phenolic hydroxyl group is catalyzed by sulfotransferases, which belong to a large superfamily of enzymes with tissue- and development-specific expression patterns. Sulfation accelerates the deiodination of different iodothyronines by the type I deiodinase and initiates irreversible hormone degradation.  $T_3$  sulfate found during pregnancy in fetal and maternal circulation might function as a reservoir from which active  $T_3$  is recovered by tissue sulfatase activity.  $T_2$  sulfate, also found in fetal circulation, might represent an elimination form of excess thyroid hormone secreted toward the maternal circulation when fetal type I deiodinase activity is still low. Significant sulfotransferase activity has been found in fetal tissues. Relevant sulfatase activities that might liberate  $T_3$  from circulating  $T_3$ -sulfate have been observed in (maternal) placenta and liver during pregnancy as well as in other tissues.

### DRUGS INTERFERING WITH THYROID HORMONE METABOLISM

Several drugs and agents interfere with thyroid hormone metabolism, especially the deiodination reactions. Many phenolic aromatic, mono- or polycyclic agents interfere with iodothyronine binding and deiodination by the deiodinases. With few exceptions, these agents specifically interfere with one or more of the deiodinase enzymes. Several drugs routinely used in clinics interfere with  $5'$ -DI. Most efficient inhibition is observed for some iodinated oral cholecystographic X-ray contrast agents (ipodate or iopanoic acid), the antiarrhythmic drug amiodarone and its metabolites, the synthetic glucocorticoid dexamethasone, some antiphlogistic and anti-inflammatory agents, and the large group of plant secondary phenolic metabolites (flavonoids, isoflavonoids, aurones, and chalcones). The latter compounds are contained in significant amounts in daily consumed plant-derived food, but it is unclear whether their consumption contributes to goitrogenesis under conditions of inadequate iodide supply. Many of these deiodinase-inhibiting

compounds lead to a constellation of serum thyroid hormones reminiscent of the low  $T_3$  syndrome: normal or elevated  $T_4$ , low  $T_3$ , elevated  $rT_3$ , and normal TSH. The main effect of these compounds is inhibition of  $5'$ -DI. Some agents are potent inhibitors of all three deiodinases (e.g., iopanoic acid). No selective or isoenzyme-specific inhibitors of  $5'$ -DII and 5-DIII are known. The antithyroid drug 6-*n*-propyl-2-thiouracil is a selective inhibitor of  $5'$ -DI in most species. However, its clinical use as an antithyroid drug is based mainly on its inhibition of thyroid hormone synthesis catalyzed by the thyroperoxidase and less so via inhibition of  $5'$ -DI activity. The compound aurothioglucose is a potent inhibitor of  $5'$ -DI but also inhibits the other deiodinases at higher concentrations due to the presence of a selenocysteine residue in their active sites. The synthetic flavonoid F21388 derived from natural compounds is a potent inhibitor of  $5'$ -DI and  $5'$ -DII and has been shown to modify transplacental thyroid hormone transfer and thyroid hormone transport to the fetal brain. The development of isoenzyme-specific inhibitors may provide the possibility for tissue- and cell-specific interference of thyroid hormone action in a manner similar to that of the steroid field, for which selective inhibitors of steroid activation (e.g., aromatase inhibitors) and inactivation are already in clinical use.

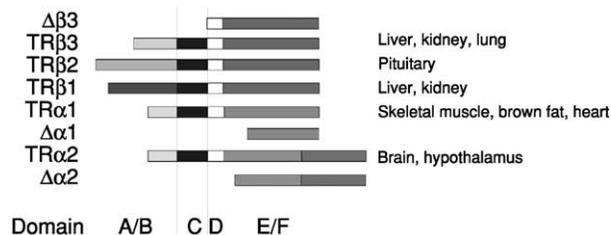
### See Also the Following Articles

Amiodarone and Thyroid • Drug Effects and Thyroid Function • Hypothalamic Hypothyroidism • Iodine • Resistance to Thyroid Hormone (RTH) • Thyroid Hormone Action • Thyroid Hormone-Binding Proteins • Thyroid Hormone Receptors • Thyroid Peroxidase

### Further Reading

- Bianco, A. C., Salvatore, D., Gereben, B., Berry, M. J., and Larsen, P. R. (2002). Biochemistry, cellular and molecular biology, and physiological roles of the iodothyronine selenodeiodinases. *Endocr. Rev.* **23**, 38–89.
- Hennemann, G., Docter, R., Friesema, E. C., De Jong, M., Krenning, E. P., and Visser, T. J. (2001). Plasma membrane transport of thyroid hormones and its role in thyroid hormone metabolism and bioavailability. *Endocr. Rev.* **22**, 451–476.
- Köhrle, J. (2000). The selenoenzyme family of deiodinase isozymes controls local thyroid hormone availability. *Rev. Endocr. Metab. Dis.* **1**, 49–58.
- Köhrle, J. (2002). Iodothyronine deiodinases. *Methods Enzymol.* **347**, 125–167.
- Leonard, J. L., and Köhrle, J. (2000). Intracellular pathways of thyroid hormone metabolism. In "The Thyroid" (L. E. Braverman and R. D. Utiger, eds.), Vol. 8, pp. 136–173. Lippincott, Philadelphia.

- Papanicolaou, D. A. (2000). Euthyroid sick syndrome and the role of cytokines. *Rev. Endocr. Metab. Dis.* **1**, 43–48.
- Peeters, R. P., van Toor, H., Klootwijk, W., de Rijke, Y. B., Kuiper, G. G. J. M., Uitterlinden, A. G., and Visser, T. J. (2003). Polymorphisms in thyroid hormone pathway genes are associated with plasma TSH and iodothyronine levels in healthy subjects. *J. Clin. Endocrinol. Metab.* **88**, 2880–2888.
- Peeters, R. P., Wouters, P. J., Kaptein, E., van Toor, H., Visser, T. J., and Van den Berghe, G. (2003). Reduced activation and increased inactivation of thyroid hormone in tissues of critically ill patients. *J. Clin. Endocrinol. Metab.* **88**, 3202–3211.
- Tata, J. R. (1999). Amphibian metamorphosis as a model for studying the developmental actions of thyroid hormone. *Biochimie* **81**, 359–366.
- Van den Berghe, G., Wouters, P., Weekers, F., Mohan, S., Baxter, R. C., Veldhuis, J. D., Bowers, C. Y., and Bouillon, R. (1999). Reactivation of pituitary hormone release and metabolic improvement by infusion of growth hormone-releasing peptide and thyrotropin-releasing hormone in patients with protracted critical illness. *J. Clin. Endocrinol. Metab.* **84**, 1311–1323.
- Visser, T. J., Kaptein, E., Glatt, H., Bartsch, I., Hagen, M., and Coughtrie, M. W. H. (1998). Characterization of thyroid hormone sulfotransferases. *Chem. Biol. Interact.* **109**, 279–291.



**Figure 1** Schematic representation of the four thyroid hormone receptor (TR) isoforms. The receptor isoforms are shown with the tissues where their expression is most prominent. The differences between the isoforms are indicated by different shading symbolizing the different domains. The schematic domain structure is indicated at the bottom, with A/B signifying the N-terminal receptor domain, C the DNA binding domain, D the hinge region, and E/F the ligand binding domain.

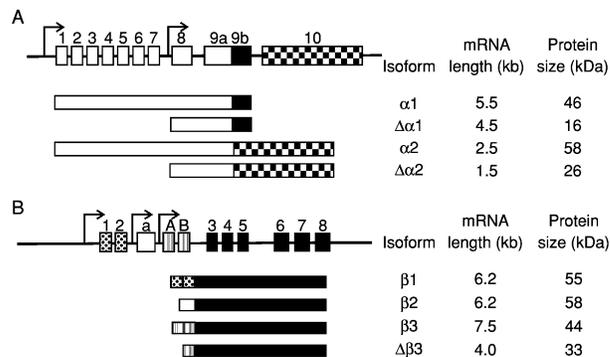
has led to a better understanding of resistance to thyroid hormone, the possible mechanism of action of amiodarone, and the development of thyroid hormone agonist and antagonists.

## TR ISOFORMS

### Gene Structure

TRs are encoded by two different gene loci, TR- $\alpha$  [locus *c-erbA- $\alpha$*  on chromosome 17q11.2 (human), 10 (rat), and 11 (mouse)] and TR- $\beta$  [locus *c-erbA- $\beta$*  on chromosome 3p24.3 (human), 15 (rat), and 14 (mouse)]. The TR- $\alpha$  gene encodes for TR- $\alpha_1$  and its splice variant TR- $\alpha_2$ . Two further isoforms arise from the TR- $\alpha$  gene, namely TR- $\Delta\alpha_1$  and TR- $\Delta\alpha_2$ . The TR- $\alpha$  gene contains 10 exons, and the alternative splice site for TR- $\alpha_2$  is located on exon 9 (Fig. 2). The TR- $\alpha_2$  splice variant, which is only found in mammals, has no intact ligand-binding domain (LBD) but can act as a constitutive repressor of TR action. An internal promoter located on intron 7 is responsible for the formation of TR- $\Delta\alpha_1$  and TR- $\Delta\alpha_2$  variants, which are unable to bind hormone and DNA and can act as constitutive repressors of the other TRs.

The TR- $\beta$  locus encodes the isoforms TR- $\beta_1$ , TR- $\beta_2$ , TR- $\beta_3$ , and a truncated receptor TR- $\Delta\beta_3$  (unable to bind thyroid hormone), which are generated via alternative splicing and/or differential promoter usage (Fig. 2). They differ only in the amino-terminal part of the receptor (Fig. 1). The TR- $\beta$  locus contains 11 exons, of which exons 3–8 are common to all TR- $\beta$ s and also show high homology with the TR- $\alpha$  isoforms. Exons a, A, and B were discovered later. Alternative splicing of exons 1 and 2 results in the



**Figure 2** (A) Schematic representation of the transcripts generated by the *c-erbA $\alpha$*  locus. Exons common to the TR- $\alpha_1$  and TR- $\alpha_2$  isoforms are depicted as white rectangles. The black rectangles show the regions specific to the TR- $\alpha_1$  isoforms and the black blocked rectangles show the regions specific to the TR- $\alpha_2$  isoforms. The arrows indicate the different promoters used to generate the different TR- $\alpha$  isoforms. (B) Schematic representation of the transcripts generated by the *c-erbA $\beta$*  locus. Exon 3–8, common to all TR- $\beta$  isoforms, are shown as black rectangles. Exons 1 and 2, common to the TR- $\beta_1$  isoform, are shown as dotted rectangles. Exon a of the TR- $\beta_2$  isoform is depicted as a white rectangle, and exons A and B of the TR- $\beta_3$  and B of TR- $\Delta\beta_3$  are shown as gray rectangles. The arrows indicate the different promoters used to generate the different TR- $\beta$  isoforms.

N-terminal part of TR- $\beta_1$ , exon a for TR- $\beta_2$ , and exons A and B for TR- $\beta_3$  and its  $\Delta$  variant.

### Protein Structure

The TRs have a domain organization similar to that of the other members of the nuclear receptor superfamily. There are four domains: the A/B domain (the amino-terminal part, which differs in TRs), the C domain (which contains the DNA binding domain), the D domain (hinge region), and the E/F domain (LBD) (Fig. 1). Each of these domains plays a different role in the interaction with proteins and DNA. Generally, it is believed that the A/B domain, together with the E domain, is involved in the interaction with coactivator proteins. DNA binding of the TRs is regulated by the C and E domains, and corepressor interaction sites are located mostly in the D and E domains. The protein sequence is highly conserved between rodents and humans.

The A/B domains of the TR- $\alpha$  and TR- $\beta$  isoforms differ in size and sequence. The protein sequence of the amino-terminal part of TR- $\alpha_1$  and TR- $\alpha_2$  is the same, whereas it differs between TR- $\beta$  isoforms. There is a high homology between the different isoforms of rat, human, and mouse, although the TR- $\beta_2$  isoform in rats is 38 amino acids longer than those in

mouse and human. TR- $\beta_3$  and TR- $\Delta\beta_3$  have only been found in rats.

The role of the A/B domain in transcriptional activation is poorly understood due to the fact that its activity is highly dependent on the species or cell types being studied. Some studies have shown that transcriptional activity is lost after deletion of the N-terminal amino acids of the TRs, suggesting that the most likely function of this domain is in regulation of the interaction of the TR with other proteins, such as transcription factors or coactivators.

The C domain possesses two zinc fingers (Fig. 3), each containing four cysteines coordinated by a zinc ion and four functional sites called P-, D-, T-, and A-boxes. The first zinc finger is important in the specific association with the TRE sequences. In the presence of DNA, the receptor dimerizes, followed by placement of the first zinc finger of the DNA-binding domain of each monomer in direct contact with the DNA major groove. The second zinc finger of each monomer is stabilized by direct interaction between

the this zinc finger and the TRE. The P-box is also important in sequence-specific recognition of response elements of other receptors, such as ER, RXR, RAR, and VDR. The D-box is involved in distinguishing the spacing between half-sites of hormone response elements. The T- and A-boxes are involved in heterodimerization. Because of their critical role, changes in the zinc fingers or boxes result in abrogation of DNA-binding and transcriptional activity.

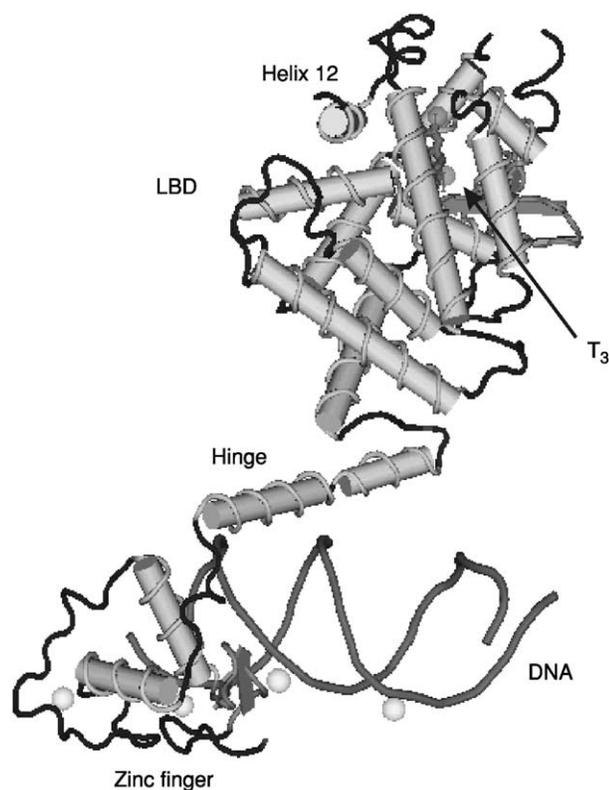
The hinge region (D domain) links the C domain to the LBD (Fig. 3). It interacts with corepressors since mutation of this domain in the TR- $\beta_1$  receptor abolishes the basal repression. This indicates either that the hinge region of the unliganded TR can provide a contact surface with corepressor or that the hinge region has an allosteric effect on these interactions since the interaction with corepressors is also associated with sequences in the LBD.

TRs possess motifs called nuclear localization signals (NLSs). The consensus sequences mediating nuclear translocation consist of two clusters of two or three basic amino acids separated by two nonbasic amino acids [(K/R)(K/R)XX(K/R)(K/R)(K/R)]. For TR- $\beta_1$ , the KRLAKRK (amino acids 184–190) in domain D mediates the T<sub>3</sub>-induced nuclear translocation. The TR- $\alpha$  isoforms also possess several NLSs in the D and E domains, and this receptor shows nuclear-cytoplasmic shuttling.

#### Ligand-Binding Domain (E Domain)

The LBD or E domain is required for nuclear localization, homo- and/or heterodimerization, and, most important, ligand binding and the ligand-induced change in transcriptional activity. In the case of the thyroid hormone receptor, the ligand T<sub>3</sub> is tightly packed inside the binding pocket (Fig. 3). The LBD in its inactive form is able to bind to so-called corepressor molecules that help the receptor keep the gene silent. The major event upon ligand binding is that helix 12, which is composed of the final amino acids of the receptor protein, moves into a different position, thus releasing the corepressor. It then forms a different surface that will allow binding of the coactivator, which helps the receptor in gene activation. Molecular studies have established that an LXXLL amino acid motif within the coactivators mediates the interaction with ligand-activated nuclear receptors.

The LBD domain also harbors important sequences such as the heptad repeats. These repeats form hydrophobic interfaces for TR homo- and heterodimerization. One of these conserved regions, the ninth heptad, is important for TR homodimerization



**Figure 3** Impression of the combination of the three-dimensional (3D) structures of the TR- $\beta$  DNA and ligand binding domain (from the NCBI-MMDB database, structure Nos. 1BSX and 2NLL). It can be seen that T<sub>3</sub> is tightly packed inside the molecule. The zinc fingers, hinge region, ligand binding domain (LBD), and helix 12 are indicated.

and RXR-TR heterodimerization since mutation of the ninth heptad decreases TR homo- and heterodimerization. For TR- $\alpha_2$  the dimerization region is probably located in the DNA-binding domain (note that TR- $\alpha_2$  has no ninth heptad).

Although sequence comparisons show great similarity between species and between TR- $\beta$  and TR- $\alpha$ , an intriguing difference exists in the ligand-binding capacity for the T<sub>3</sub> analogue TRIAC between these isoforms. TR- $\beta_1$  binds TRIAC 1.5-fold stronger than does TR- $\alpha_1$ . There are only a few amino acid alterations between the two receptor isoforms in the hormone-binding pocket. However, these changes allow a subtle change in the position of one of the arginine residues (Arg<sup>228</sup> for TR- $\alpha_1$ , and Arg<sup>282</sup> for TR- $\beta_1$ ), which explains the increased binding in TR- $\beta_1$ .

## DISTRIBUTION AND FUNCTION OF THYROID HORMONE RECEPTORS

After the discovery of TR-encoding genes, developmental and tissue-specific patterns of TR expression were reported based on their mRNA. TR- $\alpha_1$ , TR- $\alpha_2$ , and TR- $\beta_1$  are expressed in almost all tissues. From studies of frogs, it appears that the TR- $\alpha$  isoforms are predominant during prenatal development, followed by expression of TR- $\beta$  isoforms; it has been suggested that this may hold true for other species as well. In adult organisms, TR- $\alpha$  isoform mRNAs are highly expressed in skeletal muscle and the TR- $\Delta\alpha$  mRNAs are expressed in the intestine. TR- $\beta_1$  mRNA is most highly expressed in the liver, kidney, and brain. TR- $\beta_2$  mRNA expression is limited to certain hypothalamic areas, the pituitary, and the retina. TR- $\beta_3$  and TR- $\Delta\beta_3$  mRNA expression is restricted to the liver, spleen, heart, and lung. Even when different receptor isoforms are present, one is the major isoform in a particular tissue (Table I).

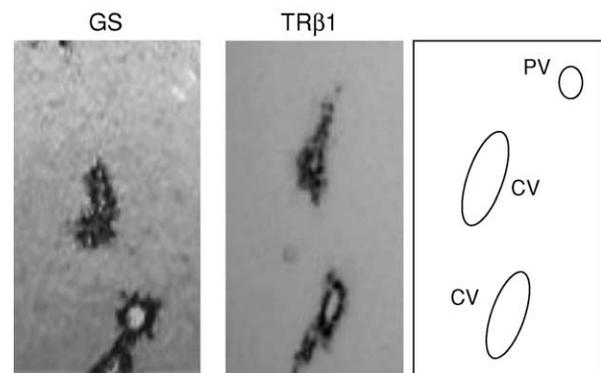
Other regulatory possibilities arise from the fact that many T<sub>3</sub>-dependent genes are not distributed homogeneously in a tissue. In the liver, T<sub>3</sub>-dependent genes, such as *PEPCK* and *Spot14*, are expressed around the portal vein or around the central vein, respectively (zonal distribution). Studies have shown that the TR isoforms in liver are not homogeneously expressed but are also expressed in a zonal fashion. From these studies, it appears that TR- $\beta_1$  is expressed in only a few cell layers around the central vein of the liver lobule (Fig. 4). Interestingly, the *Spot14* gene, which is expressed in the same subpopulation of cells around the central vein, is exclusively dependent on the TR- $\beta_1$  isoform for its regulation by thyroid

**Table I** Isoform-Selective TR Functions in Different Organs

Thyroid hormone target	Receptor isoforms mediating actions
Hepatic gene regulation	TR- $\beta$ > TR- $\alpha$
Bone development	TR- $\alpha$ > TR- $\beta$
TSH suppression	TR- $\beta$ > TR- $\alpha$
Ligand-independent TSH elevation	TR- $\alpha$ > TR- $\beta$
Cochlear development and function	TR- $\beta$
Maturation of small intestine	TR- $\alpha$ = TR- $\beta$
Cardiac gene expression	TR- $\alpha$ > TR- $\beta$
Heart rate	TR- $\alpha$
Retinal development	TR- $\beta$
Growth	TR- $\alpha$ > TR- $\beta$
Immune function	TR- $\alpha$ > TR- $\beta$
Temperature regulation	TR- $\alpha$ > TR- $\beta$

hormone. The TR- $\alpha_1$  and - $\alpha_2$  isoforms are expressed in a wider area but are still clustered around the central vein. These findings indicate that when studying TR isoform-dependent effects, it is important to consider these effects not only *in vitro* but also *in vivo* in relation to the local expression of the different TR isoforms.

The elucidation of the TR isoform-dependent effects was greatly helped by the advent of mice devoid of one or more specific isoforms. The TR- $\alpha_1$  knockout is fertile and shows a mild hypothyroid phenotype, reduced body temperature, and reduced heart rate. Selective ablation of TR- $\alpha_2$  results in overexpression of TR- $\alpha_1$ , low levels of thyroid hormones,



**Figure 4** Zonal expression of TR- $\beta_1$  in rat liver. Rat liver slices were incubated with polyclonal anti-TR- $\beta_1$ . Monoclonal antiglutamine synthetase was used as a control to stain the central veins on a consecutive slice. It can be clearly seen that TR- $\beta_1$  is expressed in the same subset of cells that express GS. A schematic drawing depicting the position of the central (CV) and portal (PV) veins in the slices is shown on the right. Magnification,  $\times 100$ .

and normal levels of TSH. Interestingly, the phenotype of TR- $\alpha_2$  mutant mice also shows signs of hyperthyroidism, such as decreased body weight, elevated heart rate, and increased body temperature. These data suggest that the balance between TR- $\alpha_1$  and TR- $\alpha_2$  may provide an additional level of adjustment of hormone responsiveness in certain tissues.

In mice missing the full-length TR- $\alpha_1$  and TR- $\alpha_2$  (but that have the  $\Delta\alpha$  isoforms), the thyroid gland develops abnormally, there is arrested maturation of the intestine and reduced bone growth, and the mice die within a few weeks after birth. Interestingly, when all TR- $\alpha$  isoforms are deleted (both full-length and  $\Delta$ ), the phenotype is less severe. The selective inactivation of the TR- $\beta$  gene results in thyroid hyperplasia, increased serum thyroid hormones and TSH, impaired T<sub>3</sub>-dependent regulation of cholesterol metabolism, and defects in cochlear function (similar to the resistance to thyroid hormone syndrome in humans).

The TR knockout animal models emphasize two important features. First, the mice without all known isoforms are still viable. The existence of an unknown receptor isoform is a possibility that has not been clarified. Second, the deletion of a particular isoform can be partly compensated by other receptors. However, certain genes or processes will be influenced by the deletion of a particular isoform, as seen in the case of genes involved in lipid metabolism or inner ear development.

## TR AGONISTS AND ANTAGONISTS

Solving the X-ray structure has allowed the development of TR isoform-specific agonists and antagonists. As for the knockouts discussed previously, it is not an all-or-nothing phenomenon but it is a matter of preferential stimulation or inhibition. One of the first isoform-specific agonists to be synthesized was GC-1, which showed a clear preference for TR- $\beta_1$  both *in vitro* and *in vivo*. The X-ray structure suggested that compounds with a 5'-aryl extension could act as antagonists because they would interfere with the proper folding of helix 12 (the lid on the box). However, many compounds based on GC-1 that have large extensions act as agonists with the exception of GC-14, which acts as a partial TR- $\beta_1$  antagonist. This can probably be explained by the fact that the side chains are not rigid enough and therefore allow helix 12 (the lid) to assume its proper position. Other synthetic ligands have been developed that behave as antagonists (DIBRT and NH-3) or partial antagonists (NH-4). Of these, NH-3 appears to be the first

high-affinity TR antagonist that also inhibits TR action in a *Xenopus* development model. It has been shown to block both coactivator and corepressor binding. The latter is strange since all known receptor antagonists promote the binding of these corepressors because although they block the proper positioning of helix 12, they leave the corepressor binding site intact. This means that as a result of NH-3 binding, helix 12 assumes a position that precludes both coactivator and corepressor binding. In light of the fact that the TRs are also able to bind without hormone and thus inhibit gene expression, this may be a beneficial property.

## CONCLUSION

Solving the structure has allowed the design and understanding of the mechanism of action of antagonists and agonists. With the current knowledge of receptor structure, it can be envisaged that it may be possible to design receptor agonists that will correct the receptor defect in thyroid hormone resistance. For instance, it has been shown that the shift of 0.3 Å of helix 6 of the TR due to mutation of alanine-317 to threonine is the cause of decreased T<sub>3</sub> binding. If an agonist were found that could “live” with this small shift and thus activate the receptor, patients harboring this particular mutation could be treated. Furthermore, the fact that TRs are not homogeneously expressed in target tissues and the design of novel agonists and antagonists open up exciting possibilities for a directed interference in specific cellular processes.

## See Also the Following Articles

Amiodarone and Thyroid • Drug Effects and Thyroid Function • Iodine • Resistance to Thyroid Hormone (RTH) • Thyroid Hormone Action • Thyroid Hormone-Binding Proteins • Thyroid Hormone Metabolism

## Further Reading

- Cheng, S. Y. (2000). Multiple mechanisms for regulation of the transcriptional activity of thyroid hormone receptors. *Rev. Endocr. Metab. Dis.* **1**, 9–18.
- Flamant, F., and Samarut, J. (2003). Thyroid hormone receptors: Lessons from knockout and knock-in mutant mice. *Trends Endocrinol. Metab.* **14**, 85–90.
- Jungermann, K., and Kietzmann, T. (1996). Zonation of parenchymal and nonparenchymal metabolism in liver. *Annu. Rev. Nutr.* **16**, 179–203.
- Lazar, M. A. (1993). Thyroid hormone receptors: Multiple forms, multiple possibilities. *Endocr. Rev.* **14**, 184–193.
- O’Shea, P. J., and Williams, G. R. (2002). Insight into the physiological actions of thyroid hormone receptors from genetically modified mice. *J. Endocrinol.* **175**, 553–570.

- Sachs, L. M., Damjanovski, S., Jones, P. L., Li, Q., Amano, T., Ueda, S., Shi, Y. B., and Ishizuya-Oka, A. (2000). Dual functions of thyroid hormone receptors during *Xenopus* development. *Comp. Biochem. Physiol. B* **126**, 199–211.
- Wagner, R. L., Apriletti, J. W., McGrath, M. E., West, B. L., Baxter, J. D., and Fletterick, R. J. (1995). A structural role for hormone in the thyroid hormone receptor. *Nature* **378**, 690–697.
- Webb, P., Nguyen, N. H., Chiellini, G., Yoshihara, H. A., Cunha Lima, S. T., Apriletti, J. W., Ribeiro, R. C., Marimuthu, A., West, B. L., Goede, P., Mellstrom, K., Nilsson, S., Kushner, P. J., Fletterick, R. J., Scanlan, T. S., and Baxter, J. D. (2002). Design of thyroid hormone receptor antagonists from first principles. *J. Steroid Biochem. Mol. Biol.* **83**, 59–73.
- Yen, P. M. (2001). Physiological and molecular basis of thyroid hormone action. *Physiol. Rev.* **81**, 1097–1142.
- Zandieh Doulabi, B., Platvoet-ter Schiphorst, M., van Beeren, H. C., Labruyere, W. T., Lamers, W. H., Fliers, E., Bakker, O., and Wiersinga, W. M. (2002). TR(beta)1 protein is preferentially expressed in the pericentral zone of rat liver and exhibits marked diurnal variation. *Endocrinology* **143**, 979–984.

Table I Human Thyroid Hormone-Binding Proteins

Proteins	Bound hormone (%)	
	T <sub>4</sub>	T <sub>3</sub>
<b>Major THBPs<sup>a</sup></b>		
Thyroxine-binding globulin	66	75
Transthyretin	11	10
Albumin	20	9
<b>Minor THBPs</b>		
Lipoproteins	3	6
Immunoglobulin M	?	?
Immunoglobulin G	?	?

<sup>a</sup>THBPs, thyroid hormone-binding proteins.

oligosaccharide chains with an average of 10 sialic acid residues. Carbohydrates affect the half-life of TBG in serum since deglycosylation is associated with rapid clearance of the protein by the liver. In addition, carbohydrate removal slightly reduces TBG immunoreactivity and T<sub>4</sub>-binding activity and decreases its stability. TBG has only one binding site for iodothyronines, and it binds T<sub>4</sub> with a higher affinity than T<sub>3</sub>. The very high affinity of TBG for thyroid hormones explains why TBG, although present in serum at a much lower concentration than TTR or HSA, carries approximately 65% of T<sub>4</sub> and 75% of T<sub>3</sub> (Table I).

TBG concentration in normal adult human serum ranges from 12 to 20 mg/liter (Table II), with a maximal T<sub>4</sub>-binding capacity of 0.14–0.25 mg T<sub>4</sub>/liter. TBG is detectable in the 12-week-old fetus; concentrations in newborns are higher than in adults, decline until midadulthood, and increase thereafter.

### Transthyretin

TTR is a 56-kDa protein composed of four identical subunits, each containing 127 amino acids; it does not contain carbohydrates (Table II). It has two identical thyroid hormone binding sites, but normally only one of them is occupied. The normal serum TTR concentration is 250 mg/liter (Table II), corresponding to maximal binding capacity of 2 mg T<sub>4</sub>/liter. TTR binds approximately 10% of T<sub>4</sub> and 10% of T<sub>3</sub>. In addition to thyroid hormone, TTR also binds retinol-binding protein and is therefore involved in vitamin A transport. Synthesis of TTR occurs mostly in the liver, but the protein is also expressed in the pancreatic islet cells, the retina, and the epithelial cells of choroid plexus in both rats and humans. TTR synthesized in the choroid plexus may play an important role in brain develop-

Table II Characteristics of the Major Thyroid Hormone-Binding Proteins<sup>a</sup>

Characteristic	Protein		
	TBG	TTR	Albumin
Molecular weight (kDa)	54	56	66
Structure	Monomer	Tetramer	Monomer
Carbohydrates (%)	20	0	0
Association constant (M <sup>-1</sup> )			
T <sub>4</sub>	1 × 10 <sup>10</sup>	2 × 10 <sup>5</sup>	1.5 × 10 <sup>6</sup>
T <sub>3</sub>	1 × 10 <sup>9</sup>	1 × 10 <sup>6</sup>	2 × 10 <sup>5</sup>
Serum concentration (mg/liter)	12–20	250	40,000
Half-life (days)	5	2	15
Gene location (chromosome)	Xq22.2	18	4 (humans) 5 (mice)
Site of synthesis	Liver	Liver Choroid plexus Retina Pancreas	Liver

<sup>a</sup>Abbreviations used: TBG, thyroxine-binding globulin; TTR, transthyretin.

ment because it may help maintain the appropriate T<sub>4</sub> concentration in the central nervous system and favor its uniform distribution in different areas of the central nervous system.

The TTR gene exists in a single copy located on chromosome 18 and is composed of four exons spanning 7.3 kilobase pairs. The 5'-flanking region has a highly conserved DNA sequence among species, suggesting a crucial role in the regulation of TTR gene expression.

### Albumin

Albumin is a 66-kDa protein composed of 585 amino acids (Table II). It has a relatively strong binding site for thyroid hormone and several additional sites with much lower affinity. It does not contain carbohydrates. Its serum concentration is very high (40 g/liter), and the percentage of thyroid hormone bound to albumin is approximately 20% of T<sub>4</sub> and 10% of T<sub>3</sub>.

The human albumin gene consists of a single copy and is located on the long arm of chromosome 4, linked to vitamin D-binding α<sub>2</sub>-globulin, whereas in mice the gene is located on chromosome 5, close to the α-fetoprotein gene. There is 90% homology between

the human albumin gene and the corresponding gene in rodents.

### Lipoproteins

Lipoproteins are complex molecules composed of a protein moiety (apolipoproteins) and a lipid (both polar and nonpolar) moiety. They bind approximately 3% of T<sub>4</sub> and 6% of T<sub>3</sub>. High-density lipoproteins are the major lipoprotein plasma carriers of thyroid hormones through a specific interaction with their apolipoproteins (A-I, A-II, A-IV, C-I, C-II, C-III, and E). These apolipoproteins have a single thyroid hormone binding site encoded by exon 3 (exon 2 for apolipoprotein A-IV) of the respective gene. The thyroid hormone binding site on apolipoproteins is distinct from the apolipoprotein portion that binds to cell lipoprotein receptors. The physiological role of thyroid hormone binding to lipoproteins remains to be defined, but lipoproteins may facilitate enterohepatic circulation, transplacental passage, and central nervous system distribution of thyroid hormones, and they may be involved in thyroid hormone delivery to target tissues with cell surface receptors for apolipoproteins.

## VARIATIONS IN THYROID HORMONE-BINDING PROTEINS

### Acquired Variations

#### TBG

Many drugs and pathophysiologic conditions (Table III) are associated with changes in serum TBG concentration related to variations in either TBG synthesis or metabolic clearance rate. Hyperthyroidism and hypothyroidism cause a slight decrease and increase, respectively, in serum TBG levels due to an effect on liver synthesis of the protein. Pregnancy and estrogen therapy cause an increase in serum TBG concentration. This appears to be related to the longer half-life of TBG in the circulation because of estrogen-induced increased sialylation of the protein. Serum TBG values are also increased in patients with acute or chronic hepatitis and in a significant proportion of cases of hepatocarcinoma. Whereas in hepatitis the increase in TBG is probably the consequence of TBG release from damaged liver cells, in hepatocarcinoma the underlying mechanism may be increased liver synthesis of TBG. Patients with nephrotic syndrome have a reduced TBG concentration due to massive renal protein loss. Losses of TBG through peritoneal membrane are likely to account for the

Table III Acquired Thyroxine-Binding Globulin (TBG) Variations

Condition/drug	Serum TBG concentration
Hyperthyroidism	Decreased
Hypothyroidism	Increased
Pregnancy	Increased
Acute and chronic hepatitis	Increased
Hepatocellular carcinoma	Increased
Nephrotic syndrome	Decreased
Chronic renal failure	Decreased
AIDS	Increased
Diabetic ketoacidosis	Decreased
Starvation	Decreased
Cushing's syndrome	Decreased
Acromegaly	Decreased
Oat cell carcinoma	Increased
Estrogens	Increased
Androgens	Decreased
Anabolic steroid	Decreased
Glucocorticoids	Decreased
Perphenazine	Increased
5-Fluorouracil	Increased
Heroin, methadone	Increased
Clofibrate	Increased
L-Asparaginase	Decreased
Interleukin-6	Decreased

decrease in TBG concentration observed in patients with chronic renal failure undergoing regular peritoneal dialysis. Serum TBG (but not CBG) levels are increased in AIDS patients, possibly due to associated hepatitis or to a specific enhancement of TBG hepatic synthesis. Patients with diabetic ketoacidosis often have decreased serum TBG levels, which might be related to the lack of stimulation of liver protein synthesis by insulin. Starvation or extreme protein-calorie malnutrition cause a decrease in serum TBG concentration likely related to decreased hepatic synthesis of the protein. These effects, as well as the decrease in TBG that occurs in severe terminal illness, may be mediated by inhibition of TBG synthesis caused by interleukin-6. Minor variations in serum TBG concentration have been reported in several other pathophysiologic conditions (Table III).

In addition to estrogens, other drugs cause an increase in serum TBG concentration, including 5-fluorouracil, clofibrate, heroin, and methadone (Table III). Conversely, administration of androgens, anabolic steroids, glucocorticoids, and L-asparaginase has been associated with decreased TBG levels in the circulation (Table III).

**TTR**

Serum TTR concentration is often decreased in patients with severe nonthyroidal illness, particularly during protein-calorie malnutrition, nephrotic syndrome, liver diseases, and cystic fibrosis (Table IV). In such circumstances, serum TTR levels decrease, whereas TBG and albumin concentrations remain normal. Both decreased liver synthesis of TTR (possibly mediated by interleukin-6) and its accelerated degradation contribute to these changes. TTR may be increased in patients with pancreatic endocrine tumors (insulinomas or glucagonomas) or gastrointestinal carcinoids, probably due to TTR synthesis by the neoplasm. TTR levels are increased in the central nervous system but not in the serum of patients with endogenous depression or with Parkinson's disease (after adrenal medullary autotransplantation). These changes probably reflect an increased TTR synthesis by the choroid plexus. Many drugs affect serum TTR concentration, and the effect is often the converse of that on TBG. Thus, estrogens decrease serum TTR concentration and androgens, anabolic steroids, and glucocorticoids increase serum TTR concentration (Table IV). Although the underlying mechanisms are not completely understood, variations in TTR synthesis likely contribute to these changes.

**Albumin**

Albumin concentration is decreased in many acute and chronic nonthyroidal illnesses. These variations occur concomitantly and are always associated with the previously mentioned similar changes in serum TBG and TTR concentrations.

**Table IV Acquired Transthyretin (TTR) Variations**

Condition/drug	Serum TTR concentration
Protein-calorie malnutrition	Decreased
Nephrotic syndrome	Decreased
Liver diseases	Decreased
Cystic fibrosis	Decreased
Insulinoma	Increased
Glucagonoma	Increased
Gastrointestinal carcinoids	Increased
Depression	Increased <sup>a</sup>
Parkinson's disease	Increased <sup>a</sup>
Estrogens	Decreased
Androgens	Increased
Anabolic steroids	Increased
Glucocorticoids	Increased

<sup>a</sup>TTR concentration is increased in cerebrospinal fluid but not in serum.

**Inherited Variations****TBG**

Familial forms of TBG deficiency and TBG excess, both inherited as X-linked traits, exist. These defects involve the TBG gene rather than the rate of TBG disposal. Complete TBG deficiency, partial TBG deficiency, and TBG excess are distinguished according to serum TBG levels in hemizygous subjects. When TBG deficiency is complete, affected males have no detectable TBG in serum, whereas carrier females have half the normal serum TBG levels. In partial TBG deficiency, serum TBG concentration in heterozygous females is usually higher than half the normal value. In the presence of excess TBG, serum concentration of the protein is usually two- to fourfold higher than normal.

Complete TBG deficiency occurs in approximately 1 in 15,000 newborn males. Eleven TBG variants account for complete TBG deficiency. In most cases, a single nucleotide substitution, a frameshift due to nucleotide deletion, or multiple nucleotide deletions are the mechanisms leading to early termination of translation and truncation of the TBG molecule. Mutations may also occur outside the coding region of the TBG gene. In a family with complete TBG deficiency, no mutations were detected either in the coding or in the promoter regions of the gene.

Partial TBG deficiency occurs in 1 in 4000 newborns. Six different TBG variants cause variable degrees of decreases in serum TBG concentration. Some of these variants are unstable, have a reduced binding affinity for T<sub>4</sub> and T<sub>3</sub>, or show an abnormal migration pattern on isoelectric focusing. A Japanese family with partial TBG deficiency has been reported with normal thyroid hormone-binding affinity, normal isoelectric focusing pattern, normal heat stability, and no mutations in the TBG gene coding region. In this family, the hereditary transmission appeared to be autosomal dominant.

Inherited TBG excess is a rare condition, occurring in approximately 1 in 25,000–30,000 newborns. The pathophysiological basis of TBG excess has been shown to be TBG gene amplification (duplication and triplication), whereas no mutations in the coding and promoting regions have been detected.

**TTR**

Many TTR variants characterized by single amino acid substitutions have been described, most in patients with familial amyloidotic polyneuropathy, amyloidotic cardiomyopathy, or senile systemic amyloidosis. Some of these TTR variants have a reduced binding

affinity for thyroid hormone. A different TTR variant characterized by an increased affinity for  $T_4$  is responsible for a pattern of euthyroid hyperthyroxinemia (i.e., TTR-associated hyperthyroxinemia).

### Albumin

A well-characterized inherited albumin variation transmitted as an autosomal dominant trait is familial dysalbuminemic hyperthyroxinemia (FDH), which is characterized by the presence in serum of an albumin variant with increased affinity for thyroid hormones. In many cases, the albumin variant has increased affinity for  $T_4$  only; in other instances, an increased affinity for  $T_3$  and/or reverse  $T_3$  is also present. Three different single nucleotide substitutions have been identified as the molecular basis for the increased albumin affinity for thyroid hormone.

The inherited absence of albumin (analbuminemia) and the polymorphism called bisalbuminemia have negligible effects on thyroid hormone transport because the decrease in albumin levels is partially compensated for by a slight increase in TBG and TTR levels.

## EFFECTS OF VARIATIONS IN THYROID HORMONE-BINDING PROTEINS ON THYROID FUNCTION TESTS

Variations in THBP concentration or affinity profoundly affect serum total thyroid hormone concentrations. This is particularly true for TBG because it has a major role in thyroid hormone binding. Accordingly, a decrease or an increase in serum TBG concentration lead to a decrease or an increase, respectively, in serum total thyroid hormone levels. Although the latter changes are similar to those found in hypothyroidism and hyperthyroidism, respectively, they do not reflect thyroid hypofunction or hyperfunction because they are not associated with variations in the metabolically active, free (unbound) thyroid hormone fraction. Similar considerations are tenable for FDH and TTR-associated hyperthyroxinemia. Therefore, TBG excess, FDH, and TTR-associated hyperthyroxinemia are among the most important causes of euthyroid hyperthyroxinemia. The latter may be independent of THBP variations and caused by drugs (e.g., amiodarone, propranolol, iodinated contrast agents, and L-thyroxine), resistance to thyroid hormones, or the acute phase of some psychiatric disorders (Table V).

Thus, should serum total thyroid hormone measurement provide results that are in contrast with the clinical picture, a THBP abnormality should be suspected and searched for. The correct definition

**Table V Causes of Euthyroid Hyperthyroxinemia**

TBG excess
Familial dysalbuminemic hyperthyroxinemia
Transthyretin-associated hyperthyroxinemia
Amiodarone
Propranolol
Iodinated contrast agents
L-Thyroxine
Resistance to thyroid hormone
Acute phase of psychiatric disorders

of thyroid status requires measurement of serum free thyroid hormones and thyrotropin concentrations. This approach is particularly useful when THBPs (e.g., TBG excess or FDH) coexist with thyroid disorders, such as Graves' disease or Hashimoto's thyroiditis. In these circumstances, serum total thyroid hormone levels may be normal in hypothyroid patients, whereas the increased levels of hyperthyroid patients may not easily be distinguished from the increased concentrations due to THBP abnormalities.

Because serum free thyroid hormone determination is crucial for the assessment of thyroid status and to avoid inappropriate treatment for hyperthyroidism or hypothyroidism, it is essential to select methods for free thyroid hormone measurement that are not affected by the abnormal THBP concentration or affinity. The two-step methods in which free hormone is first separated from protein-bound hormone by dialysis, ultrafiltration, column adsorption chromatography, or immunoabsorption provide the most reliable results. In fact, in the second step (immunoassay) the tracer is not in contact with THBP, thus preventing interaction between the two and the consequent artifactual results.

## CONCLUSION

THBPs exert functions that are important for thyroid physiology. They provide a buffering action, preventing abrupt changes in serum thyroid hormone levels; function as a storage system for thyroid hormones; and are involved in targeted delivery of thyroid hormone at the tissue level, thus facilitating thyroid hormone cellular distribution. TBG is the major THBP in serum since it binds approximately two-thirds to three-fourths of  $T_4$  and  $T_3$ . Both inherited and acquired variations of the major THBPs (TBG, TTR, and albumin) have been demonstrated. These variations do not modify thyroid status but do affect the results of serum total thyroid hormone measurement and

may lead to incorrect diagnosis and inappropriate treatment for hyperthyroidism or hypothyroidism. Thus, for a correct definition of thyroid status, determination of free  $T_4$  and  $T_3$  by assays that are not influenced by THBPs is required.

### See Also the Following Articles

Resistance to Thyroid Hormone (RTH) • Thyroid Hormone Action • Thyroid Hormone Metabolism • Thyroid Hormone Receptors

### Further Reading

- Bartalena, L. (1990). Recent achievements in studies on thyroid hormone-binding proteins. *Endocr. Rev.* **11**, 47–74.
- Bartalena, L., and Robbins, J. (1993). Thyroid hormone transport proteins. In "Clinics in Laboratory Medicine: Pathophysiology of Thyroid Disease" (G. C. Klee, ed.), pp. 583–598. Saunders, Philadelphia.
- Bartalena, L., Bogazzi, F., Brogioni, S., Burelli, A., Scarcello, G., and Martino, E. (1996). Measurement of serum free thyroid hormone concentrations: An essential tool for the diagnosis of thyroid dysfunction. *Hormone Res.* **45**, 142–147.
- Benvenga, S., and Robbins, J. (1993). Lipoprotein–thyroid hormone interactions. *Trends Endocrinol. Metab.* **4**, 194–198.
- Refetoff, S. (1990). Inherited thyroxine-binding globulin (TBG) abnormalities in man. *Endocr. Rev.* **10**, 275–293.
- Robbins, J. (2000). Thyroid hormone transport proteins and the physiology of hormone binding. In "Werner's and Ingbar's The Thyroid" (L. E. Braverman and R. D. Utiger, eds.), 8th ed., pp. 105–120. Lippincott Williams & Wilkins, Philadelphia.
- Schussler, G. C. (2000). The thyroxine-binding proteins. *Thyroid* **10**, 141–149.

Table I Nuclides of Iodine with Clinical Relevance

Nuclide of iodine (I)	Half-life	Dose	Clinical uses
$^{127}\text{I}$	Nonradioactive		Necessary for life; can be imaged by fluorescent scanning
$^{123}\text{I}$	13 h	100–300 $\mu\text{Ci}$ 3.7–7.4 MBq	For measurement of uptake, scanning of thyroid
$^{123}\text{I}$	13 h	1–5 mCi 37–185 MBq	For imaging thyroid cancer after surgical thyroidectomy
$^{131}\text{I}$	8 days	1–10 mCi 37–370 MBq	For diagnostic images of thyroid cancer after surgical thyroidectomy
$^{131}\text{I}$	8 days	30–>200 mCi 1.1–7.4 MBq	For imaging thyroid cancer days after radioiodine treatment
$^{124}\text{I}$	4 days	5 mCi 185 MBq	Positron emitter for imaging; provides functional and volumetric information

tetrafosmin,  $^{18}\text{F}$  fluorodeoxyglucose (FDG),  $^{111}\text{In}$  octreotide, and  $^{99\text{m}}\text{Tc}$  demercaptosuccinate (DMSA). These are shown in Table II and their roles in imaging thyroid cancer are described later.

## INSTRUMENTS

The optimal instrument is a gamma camera fitted with a pinhole collimator. This provides high-resolution scans and allows anterior and oblique views to be produced. Tomographic images, or single photon emission computed tomography, provide better resolution but do not provide additional relevant information. For whole-body imaging, a dual-headed whole-body camera is recommended. For imaging positrons, dedicated positron emission tomography

(PET) cameras have better resolution than hybrid PET–gamma cameras.

## NORMAL THYROID GLAND

The adult thyroid gland has an average weight of 15–20 g and appears as two pear-shaped lobes connected by an isthmus. The appearance on scan is variable, with some degree of asymmetry being common (Fig. 1). Iodide that is not trapped by the thyroid gland is primarily excreted in the urine, with small amounts trapped by the salivary glands, stomach, choroid plexus, and lactating breast tissue. Iodine trapped by nonthyroidal tissues is not organified. Recent intake of iodine-rich foods and drugs decreases radioiodine trapping by the thyroid gland

Table II Properties of Radionuclides and Radiopharmaceuticals Used in the Diagnosis of Thyroid Disorders

Radionuclide	Usual administered dose	Main clinical use
Thallium 201	3–5 $\mu\text{Ci}$	Whole-body scan for thyroid cancer, usually when Tg positive and $^{131}\text{I}$ scan negative
$^{99\text{m}}\text{Tc}$ -sestamibi	111–185 MBq	Has been used to differentiate malignant from benign nodules
	20–25 mCi	Whole-body scan for thyroid cancer as above
$^{99\text{m}}\text{Tc}$ -tetrafosmin	740–925 MBq	Has been used to differentiate malignant from benign nodules
	20–25 mCi	Whole-body scan for thyroid cancer as above
$^{111}\text{In}$ -octreoscan	740–925 MBq	Has been used to differentiate malignant from benign nodules
	1–10 mCi	Minimal value for differentiated thyroid cancer
$^{18}\text{F}$ -fluorodeoxyglucose	37–370 MBq for scan	Moderately useful in medullary cancer
	15 mCi	Whole-body scan for thyroid cancer useful in Tg-positive, $^{131}\text{I}$ -negative patients
	555 MBq	Has been used to differentiate malignant from benign nodules Useful in medullary cancer
$^{99\text{m}}\text{TcV}$ dimercaptosuccinate (pentavalent DMSA)	15–20 mCi	Useful for medullary cancer but not available in the United States
	555–740 MBq	



**Figure 1** A normal thyroid scan. The imaging was completed 24 h after the oral ingestion of 7.4 MBq  $^{123}\text{I}$ . There is homogeneous uptake in both lobes, and the isthmus shows less uptake. The normal 24-h uptake is approximately 10–30% when the dietary iodine intake is approximately 500  $\mu\text{g}$  daily.

and lowers the calculated uptake value. Similarly, administration of intravenous radiographic contrast agents impairs thyroidal iodine trapping.

## NONTOXIC GOITER

The most common cause of goiter worldwide is iodine deficiency. In iodine-deficient regions, the uptake of radioiodine is increased. Imaging of a diffuse (endemic) goiter in regions of low iodine intake reveals a uniformly enlarged gland with a relatively homogeneous pattern of uptake. The most common form of goiter in the United States is Hashimoto's disease. The scan is seldom required for Hashimoto's disease, but when obtained the appearance can vary from a diffusely enlarged gland with normal uptake to that similar to Graves' disease, patchy uptake, or a gland with significantly reduced uptake. With increasing numbers of emigrants to the United States from iodine-deficient regions, multinodular goiter is becoming more common, especially in women and those of advanced age. Imaging typically demonstrates a markedly heterogeneous radioiodine distribution due to the presence of multiple nodules with varying degrees of function. This condition can progress to thyrotoxic nodular goiter if one or more of the nodules enlarge and develop autonomous function.

## HYPERTHYROIDISM (THYROTOXICOSIS)

Thyrotoxicosis refers to the effects of excess thyroid hormone; hyperthyroidism implies that the excess hormones are produced and secreted by the thyroid. In the setting of thyrotoxicosis, diagnostic radioiodine studies are primarily used to differentiate low-uptake thyrotoxic conditions (e.g., silent thyroiditis), which are self-limited without treatment, from high-uptake conditions that persist unless treated (Table III). Uptake measurement of radioiodine can help determine optimal treatment doses of  $^{131}\text{I}$  in the latter case.

### Thyrotoxicosis with High Uptake of Radioiodine (Hyperthyroidism)

#### Graves' Disease

Graves' hyperthyroidism is caused by autoantibodies to the receptor for TSH (TSI). These autoantibodies cause continuous production of thyroid hormone, and the thyroid is unresponsive to normal inhibitory feedback mechanisms (i.e., it is nonsuppressible). Imaging of the thyroid in Graves' disease reveals a diffusely enlarged gland with uniformly increased accumulation of tracer throughout both lobes (Fig. 2). The 24-h uptake value is elevated, often in the range of 60–80%. Visualization of the pyramidal lobe is more

**Table III** Causes of Thyrotoxicosis

Thyrotoxicosis with high uptake	Thyrotoxicosis with low uptake
Graves' disease	Excess thyroid hormone Thyrotoxicosis factitia Thyrotoxicosis medicamentosa Hamburger thyrotoxicosis
Single toxic adenoma	
Single hot nodule	Thyroiditis
Functioning nodule	Subacute thyroiditis De Quervain's
Toxic multinodular goiter	Silent thyroiditis Postpartum thyroiditis
Functioning pituitary tumor secreting TSH	
	Excess iodine Contrast agents
Pregnancy-associated tumor	Medications, amiodarone
Hydatidiform mole	
Choriocarcinoma	
Struma ovarii	



**Figure 2** A thyroid scan 24h after ingestion of 7.4MBq  $^{123}\text{I}$  in Graves' disease. The uptake is increased (usually >40%) and the gland is larger and plumper.

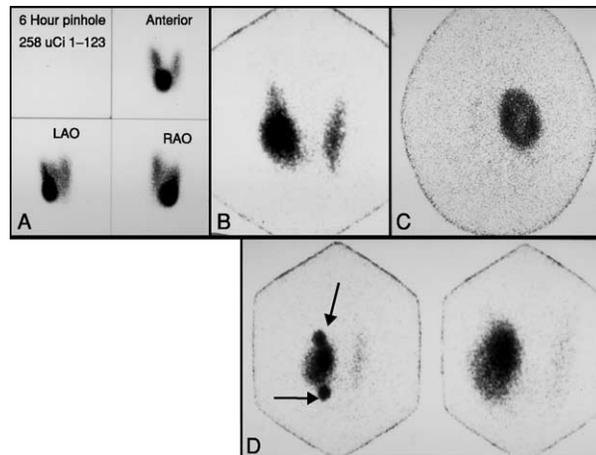
common in cases of autoimmune thyroid disease, perhaps due to increased stimulation of otherwise minimally functioning remnant tissue, and it is seen in more than 50% of patients.

#### **Toxic Multinodular Goiter**

Toxic multinodular goiters are more common in older patients and in those from areas of endemic iodine deficiency. Unlike euthyroid multinodular glands, these are accompanied by symptoms and signs of hyperthyroidism due to the unchecked, autonomous overproduction of thyroid hormone by one or more of the nodules. Scans reveal heterogeneous uptake of tracer by nodules with varying degrees of functionality. At least one nodule demonstrates increased uptake, accounting for the toxic aspect of this condition. Because the uptake tends to be less elevated in this condition than in Graves' disease, and because the condition is more resistant to radiation than Graves' disease, higher doses of therapeutic radioiodine are required to achieve successful ablation.

#### **Solitary Toxic Adenoma**

In single toxic adenoma ("hot" nodule), imaging reveals high uptake within this nodule, with little or no uptake throughout the remainder of the gland due to suppression by low TSH (i.e., the remainder of the gland remains sensitive to the normal feedback mechanisms). Scans of toxic adenomas can also reveal

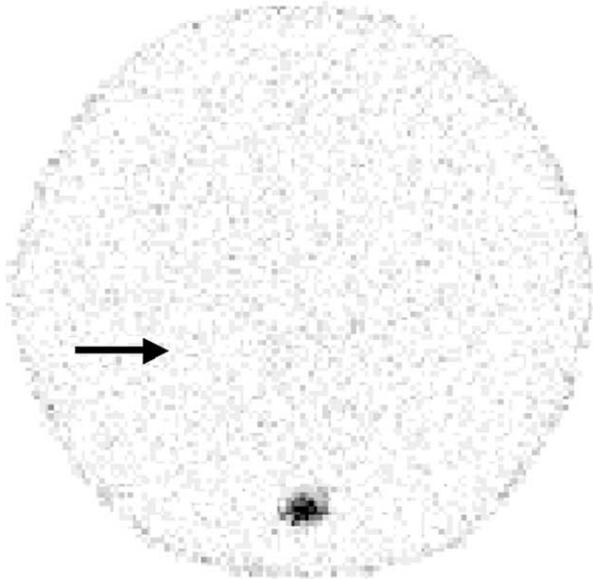


**Figure 3** Examples of functioning hot nodules (A) in a euthyroid patient, (B) in a mildly hyperthyroid patient, and (C and D) in hyperthyroid patients. In C and D, there is suppression of normal thyroid. Figure 3C shows central degeneration, and Fig. 3D shows markers where edges of the nodule are palpated.

"cold" (i.e., relatively photopenic) regions within the otherwise intense uptake of the adenoma due to necrosis (degeneration). **Figure 3** shows the range of scan findings in functioning thyroid nodules. The uptake in a solitary toxic adenoma is often only mildly elevated or near the upper limit of normal. A functioning euthyroid nodule is likely benign, and the rate of progression to thyrotoxicosis is usually slow but increases when the nodule is large (>3 cm) and the patient is advanced in age.

#### **Thyrotoxicosis with Low Uptake of Radioiodine**

Thyroiditis is a general term for a group of disparate conditions, several of which have thyrotoxic symptoms and signs and biochemical findings despite low iodine uptake by the thyroid gland. The excess thyroid hormones result from uncontrolled release of previously stored hormone from disrupted follicles. Imaging reveals minimal trapping of radioiodine by the thyroid gland, and uptake measurements are depressed (**Fig. 4**). Included in this category are subacute thyroiditis, silent thyroiditis, postpartum thyroiditis, and martial arts thyroiditis, which is also called traumatic thyroiditis. These conditions are usually transient and self-limited. Acute thyroiditis (thyroid abscess) also shows reduced uptake. Thyrotoxicosis with low uptake can also result from oversupply of thyroid hormone by exogenous sources. This can be intentional (thyrotoxicosis medicamentosa) in patients with thyroid cancer



**Figure 4** A thyroid scan in a patient with subacute thyroiditis. There is no uptake by the thyroid. A radioactive marker over the sternal notch shows the expected site of the thyroid (arrow).

or factitious when the patient conceals ingesting the medication. Patients who need thyroid hormone commonly take slightly more than a physiological dose because it makes them feel better. This is not factitious thyrotoxicosis because both patient and physician recognize the deception. Thyroid imaging reveals a pattern similar to that seen in thyroiditis, with minimal uptake of radioiodine and a markedly suppressed uptake value. Low thyroglobulin is an indicator of factitious thyrotoxicosis.

In addition, hyperthyroidism can be secondary to ectopic overproduction of thyroid hormone (as in functioning thyroid metastases, trophoblastic disease, struma ovarii, or teratomas). Scanning outside the thyroid determines the source of thyroid hormone production. Excess intake of iodine can cause hyperthyroidism (Jod Basedow effect). This is more common in patients who have been chronically iodine deficient and then are exposed to excess iodine.

## THYROID NODULES

Thyroid imaging using radiopharmaceuticals has a limited role in patients with a solitary thyroid nodule. Scintigraphy can be used to determine the functional status of such nodules, recognizing that nonfunctioning (cold) nodules have a greater likelihood of being malignant than do functioning (hot) nodules. The vast majority of nodules imaged, however, are nonfunctional (approximately 90%), and 10–15% of these

are malignant. Thus, the positive predictive value of scintiscan for thyroid cancer is low (on the order of 15%). This raises concern about the cost-effectiveness of scintigraphy. Because of the low positive predictive value of thyroid scan to diagnose cancer, patients with thyroid nodules who are euthyroid are better evaluated by fine needle aspiration (FNA). In contrast, when patients have a nodule and are hyperthyroid, scintigraphic imaging can be performed first to discriminate between a benign functioning nodule and a nonfunctioning nodule in the setting of Graves' disease. The latter case is relatively uncommon, but it warrants further evaluation by FNA. This discussion regarding the relative malignant potential of hot and cold nodules applies to adults only and does not hold for children, in whom functional nodules have a relatively high likelihood of malignancy.  $^{123}\text{I}$  is preferred to  $^{99\text{m}}\text{TcO}^4$  because it provides more physiologic information and predicts response to therapeutic  $^{131}\text{I}$  for functional nodules. There are disparate results for  $^{123}\text{I}$  and  $^{99\text{m}}\text{TcO}^4$ . A cold  $^{123}\text{I}$  nodule that traps pertechnetate has a significant probability of being malignant.

## ECTOPIC THYROID TISSUE

The thyroid begins its embryologic development in the posterior oral cavity and subsequently migrates downward toward its final destination in the neck base. This migration may be arrested, leaving thyroid tissue in ectopic locations ranging from the base of the tongue to the pericardium. Ectopic thyroid tissue is usually hypofunctional, and the resultant elevation in TSH stimulates the ectopic tissue to grow larger over time. When ectopic thyroid tissue is present (above the neck base), a normal cervical thyroid gland is almost always absent. Diagnosis should be made using a combination of thyroid function tests and radioiodine scintigraphy. The embryologic descent is marked by the thyroglossal duct, and cysts can develop along this route. Thyroglossal duct cysts generally do not take up radioiodine and are thus better confirmed with ultrasound or other anatomic imaging studies. The pyramidal lobe is a normal remnant of the thyroglossal duct tract and can rarely be seen scintigraphically in normal scans and in approximately two-thirds of scans in autoimmune hyperthyroidism.

## THYROID CANCER

After thyroidectomy for thyroid carcinoma, radioablation with  $^{131}\text{I}$  can be used to eradicate any

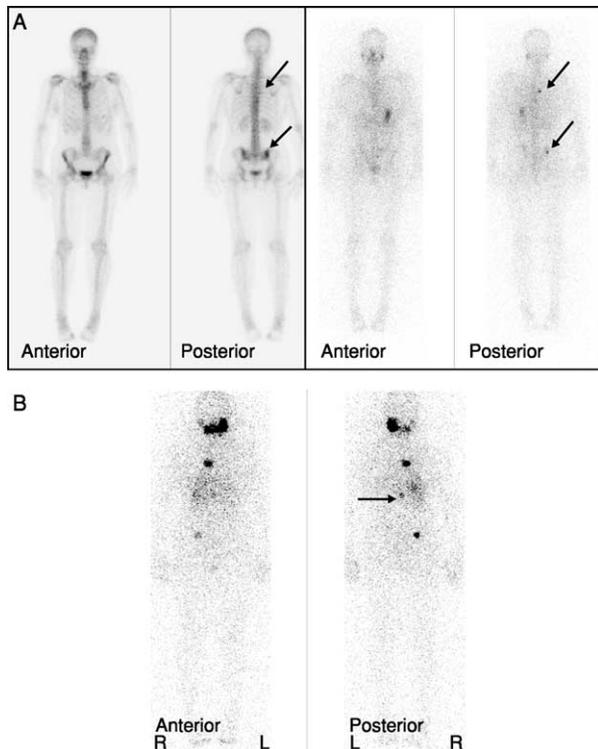
remaining thyroid tissue within the thyroid bed or metastases to lymph nodes or distant sites. Whole-body imaging has an important role in the postoperative management of the thyroid cancer patient. A diagnostic scan prior to treatment is useful to determine the dose of therapeutic  $^{131}\text{I}$  to be prescribed based on uptake in the thyroid region and the presence or absence of uptake in lymph nodes or other metastatic sites. Opinions vary regarding the necessity of the diagnostic scan prior to the first radioablation, but follow-up using the diagnostic scan is well established. Imaging of the patient following ablation allows the physician to verify uptake of the radiopharmaceutical by the thyroid tissue and to identify any areas of uptake not seen on the diagnostic scan. Periodic follow-up imaging is performed to monitor for recurrence of thyroid cancer; the general guidelines proposed by some authorities recommend annual  $^{131}\text{I}$  imaging until two consecutively negative studies are found. A TSH level  $>25\text{--}30\text{ mU/liter}$  facilitates the detection of small amounts of thyroid tissue. We advise using a level of  $50\text{ mU/liter}$  or higher. Levo-thyroxine is withdrawn for 4 weeks; alternatively, triiodothyronine ( $\text{T}_3$ ), which has a shorter half-life, is substituted for 4 weeks, allowing time for  $\text{T}_4$  to be metabolized, and then  $\text{T}_3$  is discontinued for 2 weeks. The introduction of recombinant human TSH (rhTSH) has made it possible to image (and treat) with  $^{131}\text{I}$ , without rendering the patient hypothyroid. Studies show that rhTSH is almost equivalent to endogenous TSH stimulation for determining the presence or absence of cancer, provided scan and serum thyroglobulin values are obtained. Peak serum TSH levels using rhTSH can be higher (the mean value in our experience with  $>100$  patients is  $140\ \mu\text{m/liter}$ ) than those obtained after conventional withdrawal of thyroid hormone, but the time of stimulation is shorter. Patients prefer the rhTSH protocol because hypothyroidism is avoided. Since rhTSH has been studied extensively only in the diagnostic setting and patients scanned after surgery frequently require  $^{131}\text{I}$  therapy, it may be prudent to reserve rhTSH for follow-up when it is anticipated that the scan will be negative. There are reports of the value of rhTSH in therapy. In any situation in which prolonged hypothyroidism and sustained elevation of TSH would be disadvantageous, rhTSH should be considered (e.g., when the metastases are in confined anatomic spaces, such as the spinal cord, and expansion of these could cause clinical problems).

The plasma inorganic iodine level is an important factor with regard to the amount of radioiodine that is trapped by the thyroid. Decreasing the intake of

iodine to  $30\text{--}50\ \mu\text{g/day}$  over 7–14 days increases the uptake two or three times, thus theoretically increasing the effectiveness of radioablation. A low-iodine diet is recommended for 2 weeks prior to radioiodine scanning (details of a low-iodine diet are available at [www.Thyca.org](http://www.Thyca.org)). Diagnostic whole-body scanning is conducted 2–4 days after administration of  $37\text{--}370\text{ MBq }^{131}\text{I}$ .  $^{131}\text{I}$  is used because its long half-life enables imaging after 48–96 h or more. Anterior and posterior whole-body images and spot scans of the neck with corresponding uptake measurements are obtained. Normal thyroid traps significantly more iodine than metastases; therefore, when there is a normal remnant it might have to be ablated prior to treating metastases. The importance of a skilled thyroid surgeon is emphasized. Lymph node metastases are usually in the lateral neck and less commonly in the mediastinum. Pulmonary metastases can be focal or diffuse, whereas skeletal lesions are focal in nature. The sensitivity of diagnostic  $^{131}\text{I}$  scan for papillary cancer and follicular cancer has been reported to be 45–80%. The sensitivity of posttherapeutic  $^{131}\text{I}$  scans is higher. Controversy exists regarding whether use of  $^{131}\text{I}$  for diagnostic scan can cause “stunning,” which is the inability of the thyroid tissue to take up a therapeutic dose of  $^{131}\text{I}$  secondary to radiation on the thyroid tissue by the diagnostic dose. Some investigators have not found the stunning effect after administration of 74 or 185 MBq of  $^{131}\text{I}$ . Stunning appears to occur when larger diagnostic doses are prescribed and when there is a delay between testing and treatment. Reasons for lack of iodine uptake by cancers include genetic changes in the  $\text{Na}^+/\text{I}^-$  symporter, Hurthle cell types, and poorly differentiated follicular and papillary carcinomas. Retinoic acid has been used to promote redifferentiation and induce  $^{131}\text{I}$  uptake in thyroid cancers with previously  $^{131}\text{I}$ -negative papillary, follicular, and mixed cell-type tumors. These studies have shown mixed results, but overall this strategy appears to have minimal clinical impact.

### Diagnostic Scanning with $^{123}\text{I}$

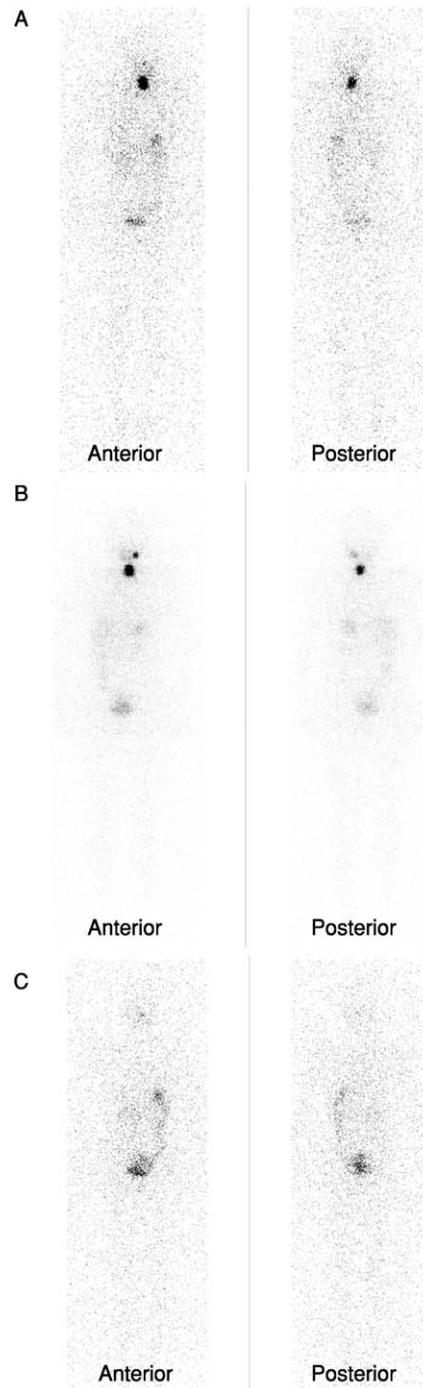
$^{123}\text{I}$  emits gamma rays at lower energies than does  $^{131}\text{I}$ , and it does not emit beta particles and is unlikely to induce thyroid stunning. At least one study has demonstrated a higher rate of ablation after  $^{131}\text{I}$  treatment when  $^{123}\text{I}$  was used in the diagnostic scan, indicating the possibility that  $^{123}\text{I}$  may replace  $^{131}\text{I}$  for whole-body scintigraphy (Fig. 5). A dose of  $74\text{ MBq }^{123}\text{I}$  has been shown to have the same overall effectiveness in diagnostic imaging as a dose of  $74\text{ MBq }^{131}\text{I}$ .



**Figure 5** (A, right) Anterior and posterior whole-body scan 24 h after a dose of 74 MBq  $^{123}\text{I}$  in a patient with thyroid cancer who had thyroidectomy and prior  $^{123}\text{I}$  therapy. There are two metastases in the posterior image (arrows). These are seen faintly in a rib and sacroiliac joint on bone scan (left, arrows). (B) A whole-body scan 10 days after a therapeutic dose of 7.4 GBq  $^{123}\text{I}$ . The two lesions seen on diagnostic scan are shown. In addition, there is a faint lesion in the low thoracic spine that was not imaged with  $^{123}\text{I}$  (arrow). There is significant uptake in the salivary glands. Liver uptake is due to metabolism of radioactive thyroid hormones.

### Posttherapy Scanning

A posttherapy  $^{131}\text{I}$  scan is usually performed 5–7 days following radioiodine ablation. **Figure 6** shows whole-body  $^{131}\text{I}$  scans—first a diagnostic scan followed by a posttherapy scan and a second diagnostic scan 12 months later to determine whether  $^{131}\text{I}$  treatment was successful. There is a higher sensitivity and clearer delineation of lesions using posttherapy scans due to the higher dose of  $^{131}\text{I}$ . Some authors report that as many as one-third of patients had metastases to the lymph nodes and lungs seen on posttherapeutic study that were not seen on the diagnostic scan. In fact, some clinicians perform only a posttherapeutic scan. We do not recommend this approach because in our experience, the diagnostic scan determines how much therapeutic  $^{131}\text{I}$  to prescribe and the posttherapy scan seldom shows additional clinically relevant information.



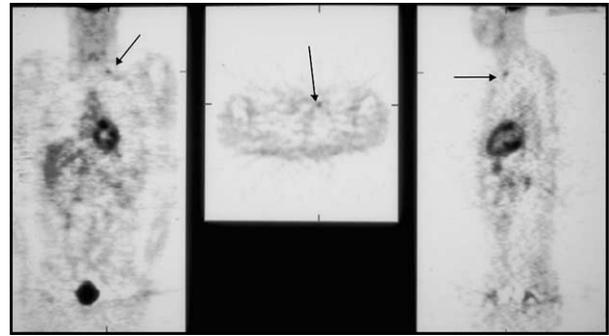
**Figure 6** (A) A diagnostic scan made 72 h after a dose of 74 MBq  $^{123}\text{I}$ . There is uptake in the thyroid bed. Physiological uptake is present in the stomach and intestines. (B) Whole-body scan in the same patient 1 week after 3.7 GBq  $^{123}\text{I}$  therapy. There is uptake in the thyroid bed and a left cervical node, and there is also uptake in the liver and gut. There is no stunning. (C) Whole-body scan in the same patient 1 year after I-131 treatment. The scan was made 48 h after a dose of 74 MBq  $^{123}\text{I}$ . RhTSH was used to stimulate uptake. There is no evidence of disease. There is physiological uptake in the gut and nasopharynx.

### False-Positive $^{131}\text{I}$ Scans

Physiologic uptake is seen in the salivary glands, nasal mucosa, gastric mucosa, small bowel, and colon. Excretion can also be identified in the bladder and bowel. These should not be confused with thyroid metastases. However, contamination by secretions or excretions can be misinterpreted as metastases. Other false positives have been reported in sinusitis, dental disease, tracheostomy, bronchiectasis, thymus, gallbladder, Meckel's and Zenker's diverticulum, psoriatic plaque, rheumatoid arthritis, hiatal hernia, and achalasia. Nonthyroidal cancers that have been mistaken as recurrent thyroid disease or metastases include salivary adenocarcinoma, meningioma, lung cancer, ovarian cancer, breast cancer, teratoma, neurilemoma, and gastric adenocarcinoma. Labeled thyroid hormone may also collect diffusely in the liver in the postablation scan, after sufficient production by residual thyroid tissue and concentration by hepatocytes.

### Other Radionuclide Scanning Techniques

Other scanning procedures that do not depend on iodine trapping can be employed for cancers that do not trap iodine. PET appears to be the first choice. When used in patients who are Tg positive and  $^{131}\text{I}$  scan negative, PET has a sensitivity of approximately 60–80%. False-positive results may be due to uptake of FDG in the tense cervical muscles of anxious patients and in brown fat. Thallous-201 chloride is taken up by all types of thyroid cancer. The maximal cancer to background ratios are obtained 10–15 min after injection. Anterior and posterior whole body images are obtained. Neither a low-iodine diet nor thyroid hormone cessation are necessary. The sensitivities range from 45 to 94%. In one study, the detection rate of recurrent or metastatic thyroid carcinoma using  $^{201}\text{Tl}$  was similar to that of FDG PET, and the two modalities are mostly concordant as well as complementary to  $^{131}\text{I}$  scintigraphy. However, FDG PET is capable of providing better image quality. Sestamibi has been used to identify thyroid cancers and metastases; anterior and posterior whole-body images are obtained 10–20 min after injection. More than 90% of the tracer is found in the inner mitochondrial matrix. The sensitivity is 70–90%. FDG PET was found to be more sensitive in the detection of recurrent thyroid cancer than  $^{99\text{m}}\text{Tc}$  sestamibi.  $^{99\text{m}}\text{Tc}$  tetrafosmin has properties similar to those of sestamibi. This radiopharmaceutical is used most often in the detection of local recurrence and cervical lymph node metastases. The sensitivity of tetrafosmin scanning for



**Figure 7** PET scan 1 h after intravenous injection of 555 MBq  $^{18}\text{F}$  fluorodeoxyglucose. The patient had an elevated Tg but negative  $^{123}\text{I}$  diagnostic and three negative posttherapy scans. Abnormal uptake is seen in the left supraclavicular node. This node was removed surgically and Tg became undetectable.

the detection of metastases has been reported to be 70–90%.  $^{111}\text{In}$  octreotide, an analogue of somatostatin, has been most useful in imaging residual or metastatic medullary thyroid cancer since these neuroendocrine tumors express somatostatin receptors.  $^{111}\text{In}$  octreotide scans have occasionally been useful in imaging differentiated thyroid carcinomas, especially in the case of Hurthle cell carcinoma. **Figure 7** shows a positive PET scan in a patient who had an elevated Tg but negative diagnostic and posttherapy scans. Specificities are higher with Tg  $\geq 5$ –10  $\mu\text{g}/\text{liter}$  and have been correlated with increasing thyroglobulin levels. Misinterpretation of tense or active muscles in the neck and larynx has led to false positives; therefore, it is important that the patient remain relaxed during the procedure and avoid speaking or chewing. Incidental focal uptake seen in the thyroid during PET performed for other indications is highly suspicious for primary thyroid cancer.

### IMAGING MEDULLARY THYROID CANCER

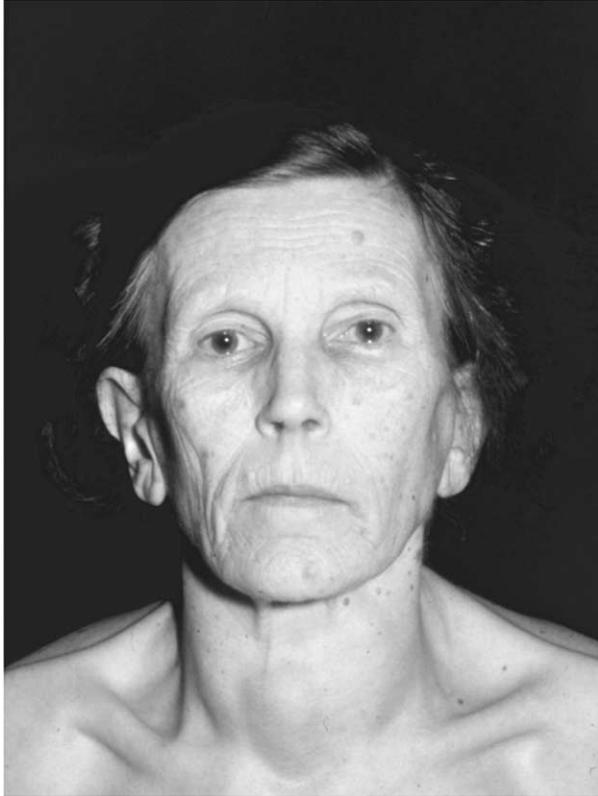
When calcitonin remains high following surgery, residual medullary cancer is present. Noninvasive methods of detecting and imaging medullary cancer include  $^{201}\text{Tl}$ ,  $^{99\text{m}}\text{Tc}$  sestamibi,  $^{111}\text{In}$  octreotide, radioiodinated meta-iodobenzylguanidine,  $^{131}\text{I}$  anti-CEA, and  $^{99\text{m}}\text{Tc}$ -labeled DMSA. PET is superior to computed tomography or magnetic resonance imaging in detecting metastatic and recurrent medullary thyroid cancer, and sensitivities as high as 76% have been reported. Patients with rapidly rising serum calcitonin levels during the first postoperative year benefit most from a PET scan.

## See Also the Following Articles

Graves' Disease • Nontoxic Goiter • Thyroid Carcinoma • Thyroid Fine Needle Aspiration Cytology • Thyroid Nodule • Thyrotoxicosis: Diagnosis • Toxic Adenoma • Toxic Multinodular Goiter

## Further Reading

- Gulzar, Z., Jana, S., Young, I., *et al.* (2001). Neck and whole-body scanning with 5-mCi dose of  $^{123}\text{I}$  as diagnostic tracer in patients with well-differentiated thyroid cancer. *Endocr. Practice* **7**, 244–249.
- Haugen, B. R., Pacini, F., Reiners, C., *et al.* (1999). A comparison of recombinant human thyrotropin and thyroid hormone withdrawal for the detection of thyroid remnant or cancer. *J. Clin. Endocrinol. Metab.* **84**, 3877–3885.
- McDougall, I. R. (1995). Whole body scintigraphy with radioiodine-131. A comprehensive list of false positives with some examples. *Clin. Nucl. Med.* **20**, 869–875.
- McDougall, I. R. (1997).  $^{131}\text{I}$  treatment of  $^{131}\text{I}$  negative whole body scan and positive thyroglobulin in differentiated thyroid carcinoma: What is being treated? *Thyroid* **7**, 669–672.
- McDougall, I. R. (1997). 74MBq radioiodine  $^{131}\text{I}$  does not prevent the uptake of therapeutic doses of  $^{131}\text{I}$  (i.e., it does not cause stunning in differentiated thyroid cancer). *Nucl. Med. Commun.* **18**, 505–512.
- McDougall, I. R., Davidson, J., and Segall, G. M. (2001). Positron emission tomography and the thyroid with an emphasis on thyroid cancer. *Nucl. Med. Commun.* **22**, 485–492.
- Pellegriti, G., Scollo, C., Giuffrida, D., Vigneri, R., Squatrito, S., and Pezzino, V. (2001). Usefulness of recombinant human thyrotropin in the radiometabolic treatment of selected patients with thyroid cancer. *Thyroid* **11**, 1025–1030.
- Schlumberger, M. J. (1998). Papillary and follicular thyroid carcinoma. *N. Engl. J. Med.* **338**, 297–306.
- Shankar, L. K., Yamamoto, A. J., Alavi, A., and Mandel, S. J. (2002). Comparison of  $^{123}\text{I}$  scintigraphy at 5 and 24 hours in patients with differentiated thyroid cancer. *J. Nucl. Med.* **43**, 72–76.
- Siddiqi, A., Foley, R. R., Britton, K. E., *et al.* (2001). The role of  $^{123}\text{I}$ -diagnostic imaging in the follow-up of patients with differentiated thyroid carcinoma as compared to  $^{131}\text{I}$ -scanning: Avoidance of negative therapeutic uptake due to stunning. *Clin. Endocrinol.* **55**, 515–521.
- Singer, P. A., Cooper, D. S., Daniels, G. H., *et al.* (1996). Treatment guidelines for patients with thyroid nodules and well-differentiated thyroid cancer. *Arch. Intern. Med.* **156**, 2165–2172.



**Figure 1** Patient with a clinically solitary thyroid nodule located in the midline of the anterior neck.

role of specific candidate genes in the etiology of nodular thyroid disease have not provided a clear picture, mostly because too small and too few families have been studied. Although single genes may play a role in certain families, it is thought that genetic heterogeneity (i.e., no single gene is either necessary or sufficient for disease development) is highly likely.

The following sequence of events appears to lead to the development of nodular thyroid disease: First, iodine deficiency or other goitrogenic factors induce thyroid hyperplasia. Second, due to increased proliferation during this stage, mutagenesis is increased. In the case of hot or toxic nodules, these mutations confer constitutive activation of the cyclic AMP cascade (e.g., thyroid-stimulating hormone receptor and  $G_s\alpha$  protein mutations). This eventually leads to stimulation of iodine uptake and metabolism, thyroid hormone synthesis and release, and hyperthyroidism. In the case of cold thyroid nodules, a similar mechanism, but with mutations in genes that favor dedifferentiation (e.g., *ras* oncogene), is suggested. These latter mutations initiate growth but not function of the affected thyroid cells.

## EPIDEMIOLOGY

Unfortunately, our knowledge is hampered by a lack of population-based longitudinal studies using sensitive diagnostic imaging (e.g., ultrasound) allowing distinction between uninodular and multinodular disease and morphologic as well as functional characterization. Despite these shortcomings, there is a clear pattern of increased thyroid nodularity with decreasing iodine intake. The fact that nodules exist in the face of iodine sufficiency and even iodine excess emphasizes the importance of other environmental etiologic factors acting in concert with genetic factors.

In the Whickham survey (United Kingdom), a solitary thyroid nodule was present in 5.3% of women and 0.8% of men (6.6:1 ratio). Size and function of the nodules were not indicated. In the Framingham, Massachusetts, study, a solitary thyroid nodule was present in 6.4% of women and 1.6% of men. Both investigations used clinical evaluation (palpation) and were performed in an iodine-sufficient area. If ultrasound is used, the prevalence of thyroid nodules >10 mm is usually 20–30%, increasing with age and in areas with insufficient iodine intake. In autopsy studies, 50% or more have either single or multiple thyroid nodules. If investigated using isotope scintigraphy, approximately 10–15% of all nodules are autonomously functioning (taking up the isotope, hot or toxic), whereas 85–90% are nonfunctioning (cold, no isotope uptake). The incidence of clinical disease is estimated to be 0.1% by palpation, corresponding to a lifetime risk of 5–10%.

## NATURAL HISTORY

The natural history with respect to growth and function varies and is difficult to predict in a given patient since no specific growth parameters exist. Therefore, it is difficult to decide whether a patient can be monitored without treatment or should be offered treatment before the nodule grows any more.

In the Framingham survey, new nodules appeared with an incidence of 1 per 1000 individuals per year, resulting in an estimated lifetime risk for developing a nodule of 5–10%. After exclusion of the minority of patients who have rapid growth and symptoms and clinical suspicion of malignancy, and who are therefore offered treatment, nodules on average do not change significantly over time. Nodules that increase in size are predominantly solid and carry a higher risk of harboring thyroid malignancy than those predominantly cystic, which are more prone to decrease in size or even disappear. In many patients, ultrasound

will identify additional nodules not evident at clinical investigation. Given time, many patients will be classified as having multinodular goiter.

In the subgroup of hot nodules, the rate of evolution into a toxic nodule is approximately 4% annually. The risk is closely related to nodule size. If the nodule is >3 cm, the risk is 20% within 6 years, whereas the risk is only 2–5% if nodule size is <2.5 cm.

## DIAGNOSIS

### Clinical Evaluation

There is no clear-cut relation between thyroid nodule size, morphology, and function, on the one hand, and the complaints of the individual patient, on the other hand. The majority of patients have few or no clinical symptoms. Therefore, given normal thyroid function and exclusion of malignancy, many need no treatment. A rough management algorithm for the majority of patients with a thyroid nodule is shown in Fig. 2.

### Manifestations

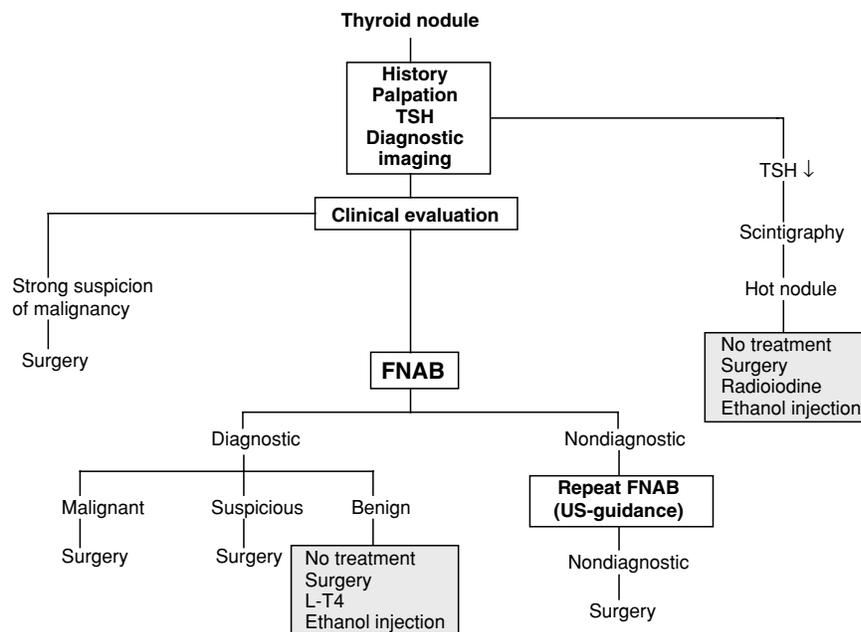
When present, the most important symptoms and signs are caused by compression of structures in the neck or, rarely, in the upper thoracic cavity. In

addition to various degrees of neck disfigurement, which by itself can merit treatment, the symptoms are related to compression of the trachea or esophagus. The symptoms of tracheal compression are dyspnea, stridor, cough, and choking sensation, but respiratory distress is rare unless the nodule extends into the thoracic cavity. An acute exacerbation may be caused by hemorrhage into a nodule or by upper respiratory infections causing endotracheal swelling. Complaints due to esophageal compression are less common, as is vocal cord paralysis caused by stretching and/or compression of the recurrent laryngeal nerves.

A high proportion of functioning nodules cause slight or overt hyperthyroidism due to excessive secretion of thyroid hormones. This condition, with or without pressure symptoms or cosmetic complaints, may by itself merit treatment.

### Clinical Examination

The evaluation of a patient with a thyroid nodule comprises a careful history and physical examination focusing on inspection of the neck, including regional lymph nodes, the upper thorax, and palpation of the thyroid. This clinical evaluation should preferably be



**Figure 2** Management algorithm outlining a cost-effective evaluation and treatment of the solitary thyroid nodule. In case of strong suspicion of malignancy, surgery is advised irrespective of fine needle aspiration biopsy (FNAB) results. In case of a nondiagnostic result, repeat FNAB yields a satisfactory aspirate in 50% of patients. FNAB guided by ultrasonography (US-guidance) allows sampling from the periphery of a solid nodule or solid part of a mixed solid–cystic nodule, increasing the sufficiency rate. The options in case of a diagnostic FNAB include those for both solid and cystic nodules. In case of recurrent cysts, the possibilities are reaspiration, surgery, or ethanol injection. The shaded boxes indicate treatment options.

done with the patient swallowing gulps of water and the head tilted slightly backward. Observer variation is very high, and the specificity and sensitivity of the diagnosis of a thyroid nodule are low. Detection of nodules depends on their size, morphology, and location within the thyroid parenchyma; the anatomy of the patient's neck; and, most important, the training of the physician. The patient is usually unaware of the presence of a nodule smaller than 1.5–2 cm in diameter, and a nodule of 1.0 cm or less usually escapes detection by the physician.

### Assessment of Risk of Malignancy

A family history of benign goiter suggests a benign disorder but is not proof thereof. Medullary thyroid carcinoma or even papillary or follicular thyroid carcinoma in the family should raise suspicion. The risk of harboring cancer is highest in the young and the old, and the risk is higher in men than in women. Head or neck irradiation during childhood for a number of benign conditions leads to clinically evident thyroid abnormality in 10–40% of these individuals 5–40 years later. Thyroid carcinomas, mainly papillary carcinomas, are seen in 30% of those with thyroid abnormality. The importance of fallout radiation is tragically evidenced by the epidemic of childhood papillary thyroid cancer seen in Belarus and Ukraine after the Chernobyl nuclear accident. Rapid nodule growth (weeks to months) and symptoms of local invasion, such as pain, dysphagia, hoarseness, or dyspnea, suggest a carcinoma; however, only a minority of patients have these symptoms. Growth during thyroid hormone therapy is particularly worrisome. Sudden growth is most likely a thyroid cyst or hemorrhage into a previously undetected nodule. A hard and immobile nodule is suggestive of thyroid carcinoma, as is same-sided lymphadenopathy. Virtually all patients with thyroid carcinoma have normal thyroid function, as do most patients with benign thyroid nodules. On the other hand, thyroid hyperfunction normally rules out thyroid malignancy. In case of a high clinical suspicion of malignancy (Table I), thyroid surgery should be recommended, irrespective of a benign needle biopsy, since the likelihood of malignancy is very high.

### Laboratory Investigations

The only relevant biochemical test that is routinely needed is serum thyrotropin (TSH), which if normal indicates normal thyroid function. Subnormal serum TSH values should lead to determination of free

**Table I The Most Important Clinical Factors Increasing the Likelihood of Thyroid Malignancy in a Patient with Normal Thyroid Function and a Solitary Thyroid Nodule**

Family history of thyroid malignancy
Rapid nodule growth (especially during levothyroxine therapy)
Very firm or hard nodule
Fixation to adjacent structures
Vocal cord paralysis (evidenced by laryngoscopy)
Regional lymphadenopathy (enlarged regional lymph nodes)
Detection of distant metastases
Age <20 or >60 years
Male sex
History of head and neck irradiation
Diameter of nodule >4 cm and partially cystic
Compression symptoms (dysphagia, hoarseness, dyspnea)

thyroxine ( $T_4$ ) and free triiodothyronine ( $T_3$ ) in serum. If serum TSH is decreased on repeat examination, treatment of this hypermetabolic state should be offered independent of whether free  $T_4$  and/or free  $T_3$  are elevated, especially in the elderly. Isotope scintigraphy is recommended and will most likely demonstrate a functioning nodule. Most patients have normal serum TSH, including those with thyroid malignancy. Elevated serum TSH with or without decreased serum free  $T_4$  suggests that the patient has chronic autoimmune thyroiditis (Hashimoto's thyroiditis). This can be verified by demonstrating thyroid autoantibodies in serum.

Thyroid autoantibodies against thyroid peroxidase or thyroglobulin do not aid in the differentiation between malignant and benign nodules. However, they are markers of an increased risk of developing hypothyroidism (Hashimoto's thyroiditis) and hyperthyroidism (Graves' disease) spontaneously or secondary to surgery or treatment with radioactive iodine.

Calcitonin, a hormone produced by the parafollicular C cells of the thyroid gland, is the only clinically relevant biochemical marker of medullary thyroid carcinoma, which accounts for approximately 5% of all thyroid carcinomas. Basal or pentagastrin-stimulated serum calcitonin measurement is more sensitive than thyroid biopsy in detecting medullary thyroid carcinoma. However, there is no consensus on its routine use in patients with thyroid nodules.

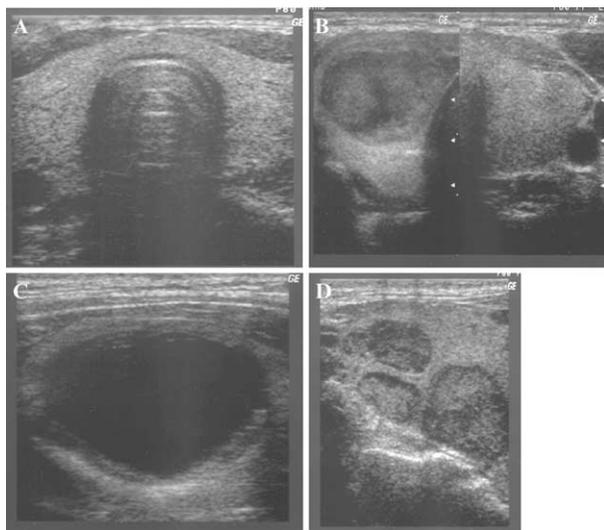
### Diagnostic Imaging

Neck palpation is very imprecise with regard to the determination of thyroid nodule morphology and size. For this reason, imaging methods are increasingly

used, although no imaging method can accurately differentiate benign and malignant nodules.

### Ultrasonography

This very sensitive technique with a high resolution has had a dramatic impact on clinical practice. When used in patients with a goiter, it has been shown to alter management in more than half of these patients. The increasing and widespread use, whether initially or during follow-up, is related to high availability, low cost, little discomfort to the patient, and its nonionizing nature. It allows determination of total thyroid volume, individual nodule size and echogenicity, and morphology of extranodular tissue and the evaluation of regional lymph nodes. Color-flow Doppler provides additional information regarding regional blood flow and nodule vascularity (Fig. 3). It distinguishes solid from cystic lesions and aids in the performance of accurate biopsies, punctures, and therapeutic procedures, such as percutaneous ethanol injection and laser therapy. Although there is no ultrasonographic pattern, alone or in combination with other techniques that may be considered specific for thyroid malignancy, characteristics such as hypoechogenicity, microcalcifications, and increased nodular flow are all predictive of malignancy to some extent. However, fine needle aspiration biopsy, preferably guided by ultrasonography, is far more accurate for this distinction.



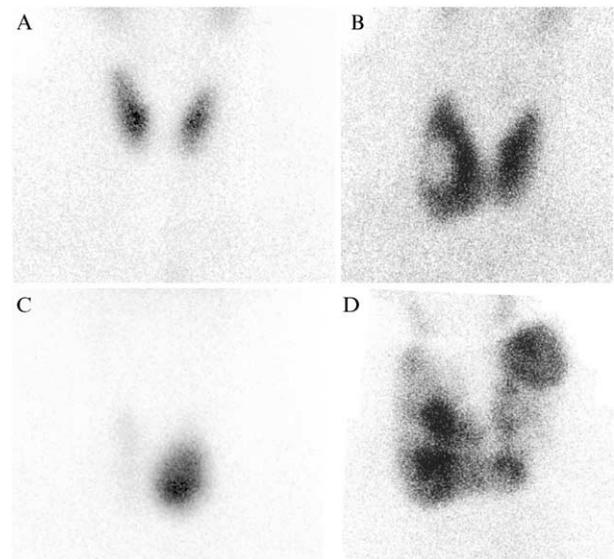
**Figure 3** Various appearances (morphological patterns) of a thyroid nodule using ultrasonography. (A) Normal thyroid tissue. (B) Solitary solid hypoechoic (dark) thyroid nodule in the right thyroid lobe (left) surrounded by normal thyroid tissue (medium gray). (C) Solitary cyst (black area) surrounded by normal thyroid tissue. (D) Multiple nodules with varying echogenicity (multinodular gland).

### Scintigraphy

Although the resolution of isotope imaging can be enhanced to 5–10 mm by tomography, this resolution is still far below that of ultrasonography. Therefore, scintigraphy is not so much used for evaluation of morphology as for evaluation of the regional uptake of the isotope and thereby the determination of functionality of the thyroid nodules (Fig. 4). Nodules with uptake by scintigraphy (hot or toxic) almost never harbor clinically important malignancy, although rare exceptions do exist. In an unselected population of patients with a thyroid nodule, 80–90% of the nodules were nonfunctioning (cold). The a priori risk of malignancy is probably no higher than 5% for such a nodule. Scintigraphy is inaccurate in estimating thyroid and nodule size as well as in diagnosing malignancy.

### Computed Tomography and Magnetic Resonance imaging

Computed tomography (CT) and magnetic resonance imaging (MRI) are expensive, time-consuming, and not readily available for imaging thyroid nodules. Their major strength is their ability to diagnose and assess the extent of substernal/intrathoracic thyroid tissue much more precisely than any other method. Both methods are well suited for visualizing the trachea and demonstrating narrowing of the tracheal



**Figure 4** Various appearances (morphological patterns) of a thyroid nodule using scintigraphy. (A) Normal uptake in two thyroid lobes. (B) Solitary nonfunctioning (cold) nodule in the right thyroid lobe (left). (C) Solitary functioning (hot or toxic) nodule in the left thyroid lobe (right). (D) Multiple nodules with varying degrees of isotope uptake (multinodular gland).

area or a decrease in its volume. However, this measure correlates poorly with lung function.

### Fine Needle Aspiration Biopsy

Fine needle aspiration biopsy (FNAB) provides the most direct and specific information about a thyroid nodule and is used by virtually all thyroid specialists in the initial evaluation of a patient with a solitary thyroid nodule or a dominant nodule in a multinodular goiter. It is without complications, inexpensive, and easy to learn to perform. Use of FNAB has reduced the number of thyroid surgeries by approximately 50%, doubled the surgical yield of thyroid cancer, and reduced the overall cost of medical care for these patients by 25%.

The technique involves the use of a 5- to 20-ml plastic syringe with a 22- to 27-gauge needle. The skin is cleaned with alcohol, and sometimes skin infiltration with 1 or 2 ml of 1% lidocaine is used. The needle attached to the syringe is inserted perpendicular to the anterior surface of the neck. Negative pressure is applied, and as soon as bloody fluid in the hub of the needle appears, pressure is released and the needle withdrawn. No fluid should enter the syringe. If the nodule is a cyst or partly cystic, the aspiration should be followed by FNAB of any residual solid component. Investigation of the cyst sediment rarely aids in the diagnosis of malignancy. After withdrawal, the needle is detached and the specimen is evacuated onto a slide. The specimen should be smeared immediately. Often, air drying is used and a number of staining methods are available.

Diagnostically useful FNAB specimens are obtained in approximately 80% of the cases and rebiopsy typically reduces the number of insufficient samples by half. The number of sufficient samples increases with operator experience, use of ultrasound guidance, the number of aspirations, when the nodule is solid, and with increasing cytopathologist experience, but it is highly dependent on the criteria used for adequacy of a sample. The relative distribution of FNAB results is given in Table II.

Diagnostic accuracy of FNAB at large depends on the classification of the 10–15% of suspicious lesions, of which 15–25% are malignant. If regarded negative, sensitivity will decrease and specificity will increase. If regarded positive, the converse is true. Patients with suspicious, malignant, and nondiagnostic FNAB results (after reaspiration) should be operated on (Fig. 2). If this strategy is followed, the risk of postponing the diagnosis of malignancy in the approximately

**Table II Etiology of Thyroid Nodules and the Relative Distribution of Fine Needle Aspiration Biopsy Results<sup>a</sup>**

<b>Etiology</b>	<b>Distribution (%)</b>
<b>Benign (no evidence of malignancy)</b>	70
Colloid nodule	
Cyst	
Thyroiditis (acute, subacute, or chronic)	
<b>Suspicious</b>	10
Follicular neoplasia	
<b>Malignant</b>	4
Follicular carcinoma	
Papillary carcinoma	
Medullary carcinoma (C-cell carcinoma)	
Undifferentiated carcinoma (anaplastic)	
Lymphoma	
Metastasis	
<b>Nondiagnostic (insufficient)</b>	16 <sup>b</sup>

<sup>a</sup>Data are representative of the author's institution.

<sup>b</sup>The number of nondiagnostic results can be halved by rebiopsy.

70% of cases in which nonsurgical therapy is an option can be reduced to 1%. Repeat FNAB during follow-up of nodules left untreated will virtually eliminate the risk of overlooking thyroid malignancy. Neither elaborate classification systems for suspicious FNAB findings nor the use of large-needle biopsy increase diagnostic accuracy. Attempts to include biochemical analysis of thyroid cyst fluid or immunodetection of various candidate molecules, such as thyroid peroxidase or lectin-related molecules, in the evaluation of thyroid cytology are still in the experimental stage.

### TREATMENT

There is no ideal treatment for the thyroid nodule. The optimal therapy varies depending on the size and morphology of the nodule and whether it is functioning (Fig. 2; Table III). Although nodules 1–1.5 cm or larger should undergo FNAB, treatment is often not necessary once malignancy has been ruled out. In the subcentimetric nodule, FNAB need not routinely be performed and treatment is rarely indicated.

### Levothyroxine Therapy

Although on the decline, it is still common practice to use thyroid suppression with levothyroxine (L-T4) in the management of solid thyroid nodules in the euthyroid patient. The aim is to shrink existing nodules, considered to be a favorable sign indicating that the

Table III Advantages and Disadvantages of the Treatment Options for the Solitary Thyroid Nodule

Treatment	Advantages	Disadvantages
Levothyroxine	Outpatient Low cost May slow nodule growth Possibly prevents new nodule formation	Low efficacy Lifelong treatment Regrowth after cessation Adverse effects (bone and heart) Not feasible with TSH suppressed
Surgery <sup>a</sup>	Prompt relief of symptoms Nodule ablation Definite diagnosis	Inpatient High cost Anesthesiological risk Surgical risk Vocal cord paralysis Hypoparathyroidism Hypothyroidism Bleeding and infection Scar
Radioiodine <sup>b</sup>	Outpatient Low cost	40% size reduction Contraceptives needed in fertile women Side effects Radiation thyroiditis Graves' disease Hypothyroidism Long-term cancer risk unknown
Ethanol injection	Outpatient Relatively low cost Thyroid function preserved	Repeat injections needed Low efficacy in large nodules Operator dependency Side effects <sup>c</sup> Pain Transient dysphonia Thyroiditis Extranodular fibrosis Complicates subsequent cytological interpretation

<sup>a</sup>In this case, unilateral operation limiting the risk of side effects.

<sup>b</sup>It can only be used in the nodule with uptake, whether thyroid function is increased or not.

<sup>c</sup>Except for various degrees of pain, side effects are rare.

nodule is benign. TSH suppression seems most beneficial in the subgroup of patients with small, solid nodules. Approximately 20% of solitary solid nodules actually regress as a result of L-T4 therapy, and cessation of therapy leads to rapid regrowth. On average, long-term therapy is without significant nodule-reducing effect. Growth can be suppressed or slowed, and the formation of new nodules may be prevented. However, this necessitates that serum TSH is suppressed to subnormal values, which may have adverse effects. This degree of TSH suppression, called mild or subclinical hyperthyroidism, is associated with an increased risk of atrial fibrillation, other cardiac side effects, and reduced bone density, potentially leading to osteoporosis. It is without effect in the cystic nodule

and in patients with spontaneously low serum TSH with or without elevated thyroid hormone levels. For these reasons, its use is questionable; at most, it can be used in younger patients with small nodules, in whom treatment is least necessary.

### Surgery

When there is malignant or suspicious cytologic features and/or symptoms due to the nodule, surgery is often recommended, especially for younger patients and in cases in which there are large nodules. The preferred operation is a unilateral removal of the affected lobe. The frequency of complications decreases with increasing experience and specialist

training and is generally low. Complications include temporary and permanent unilateral vocal cord paralysis (1–2 and 0.5–1.0%, respectively), temporary and permanent hypocalcemia (1.0 and 0.5%, respectively), and wound hematomas and infections (0.5 and 0.3%, respectively). The risk of complications increases with the extent of operation. In the patient with normal thyroid function postoperatively, there is no indication for routine L-T<sub>4</sub> treatment since this does not seem to hinder thyroid growth in the long term, at least in iodine-sufficient regions.

Although an option, surgery is rarely used in the hyperthyroid patient with a toxic nodule. Radioactive iodine treatment is the preferred treatment.

### Radioactive Iodine

If the patient has hyperthyroidism (toxic nodule), antithyroid drugs (propylthiouracil or methimazole) can normalize thyroid function but disease recurrence is the rule when medication is stopped. With the exception of a few patients who have a large nodule, in which case surgery may be indicated, radioactive iodine is the treatment of choice. This is also the case for the clinically euthyroid patient with a functioning (hot) nodule without hyperthyroidism, in whom treatment may be dictated by the nodule size, which may cause compression or cosmetic disturbances. In addition, radioactive iodine treatment is used to prevent hyperthyroidism (annual risk of approximately 4%).

A cure rate (i.e., normalization of thyroid function and the appearance on a thyroid isotope scintigram) of 75% is seen, and the nodule shrinks 30–40% following a single dose of radioactive iodine. Side effects are few, with rare cases of radiation thyroiditis and transition to Graves' disease. The risk of hypothyroidism is approximately 10% after 5 years and unrelated to the dose of radioactivity. The long-term risk of malignancy is unknown but considered negligible.

Radioactive iodine has no effect in the nonfunctioning (cold) thyroid nodule, whether solid or cystic. In the future, the possibility of stimulation with recombinant human TSH before radioactive iodine treatment may lead to an increased iodine uptake and also an effect in the solid, cold nodule.

### Ethanol Injection

Ethanol (70–100%) causes local small vessel thrombosis and coagulative necrosis, leading to fibrosis and permanent tissue ablation. It has been used in both

autonomously functioning thyroid nodules and nonfunctioning thyroid nodules, whether solid or cystic. If multiple injections are used, complete cure (normal serum TSH and isotope scintigraphy) can be achieved in 60–70% of patients with toxic nodules and 70–80% with hot nodules. A single ethanol instillation (after aspiration) in thyroid cysts reduces recurrence to approximately 20% compared to approximately 50% after aspiration alone. In solitary solid, nonfunctioning thyroid nodules, approximately 50% of patients are relieved of their clinical symptoms based on a 50% nodule volume reduction. Additional injections have little effect. It is an option for patients who do not wish to undergo radioiodine treatment or surgery. However, it often necessitates repeat treatment to obtain complete cure. The long-term effects are unknown, and the treatment is not devoid of side effects. The procedure, often used in Italian centers, is not a routine option, should still be classified as experimental, and requires the special technical skill that can be obtained only at a center familiar with interventional ultrasound.

Ultrasound-guided interstitial laser photocoagulation for solid solitary, benign, nonfunctioning thyroid nodules was recently introduced. One treatment lasting approximately 10 min resulted in a nodule reduction of approximately 40% and significant reduction of pressure symptoms. These results are similar to those obtained using ethanol therapy. The fact that the spread of energy with a laser (thermal destruction) can be controlled, as opposed to chemical destruction by injection of ethanol, may favor laser therapy in the long term. This treatment option is experimental.

### Acknowledgments

This work was supported by the Agnes and Knut Mørk Foundation, the A. P. Møller Relief Foundation, and the Novo Nordisk Foundation.

### See Also the Following Articles

Iodine Deficiency • Iodine, Radioactive • Smoking and the Thyroid • Thyroid Carcinoma • Thyroid Disease, Genetic Factors in • Thyroid Disorders in the Elderly • Thyroidectomy • Thyroid Fine Needle Aspiration Cytology • Thyroid Imaging • Toxic Multinodular Goiter

### Further Reading

Bennedbæk, F. N., and Hegedüs, L. (2000). Management of the solitary thyroid nodule. Results of a North American survey. *J. Clin. Endocrinol. Metab.* **85**, 2493–2498.

- Bennedbæk, F. N., Perrild, H., and Hegedüs, L. (1999). Diagnosis and treatment of the solitary thyroid nodule. Results of a European survey. *Clin. Endocrinol.* **50**, 357–363.
- Døssing, H., Bennedbæk, F. N., and Hegedüs, L. (2002). Benign solitary solid cold nodules: US-guided interstitial laser photocoagulation—initial experience. *Radiology* **225**, 53–57.
- Eszlinger, M., Krohn, K., and Paschke, R. (2003). Complementary DNA expression array analysis suggests a lower expression of signal transduction proteins and receptors in cold and hot thyroid nodules. *J. Clin. Endocrinol. Metab.* **86**, 4834–4842.
- Hegedüs, L., and Bennedbæk, F. N. (2002). Management of the single thyroid nodule. In “Oxford Textbook of Endocrinology” (J. A. H. Wass and S. M. Shalet, eds.). Oxford Univ. Press, Oxford.
- Hegedüs, L., Bonnema, S. B., and Bennedbæk, F. N. (2003). Management of simple nodular goiter: Current status and future perspectives. *Endocr. Rev.* **24**, 102–132.
- Krohn, K., and Paschke, R. (2001). Progress in understanding the etiology of thyroid autonomy. *J. Clin. Endocrinol. Metab.* **86**, 3336–3345.
- Ridgway, E. C. (2000). Clinical evaluation of solitary thyroid nodules. In “Werner and Ingbar’s The Thyroid: A Fundamental and Clinical Text” (L. E. Braverman and R. D. Utiger, eds.), pp. 949–956. Lippincott Williams & Wilkins, Philadelphia.

reactions: Iodotyrosine is formed at the active site of TPO through the reaction of iodide radicals with tyrosine radicals. In particular, TPO catalyzes two-electron oxidation of tyrosine and iodide and forms an iodinating intermediate ( $\text{TPO-I}^+$ ) that reacts with tyrosine and produces monoiodothyrosine (3-iodotyrosine; MIT) with TPO. As another possibility, Taurog proposed a reaction between oxidized TPO and  $\text{I}^-$  to produce hypoiodite ( $\text{OI}^-$ ), which also involves a two-electron change.

### Oxidative Coupling by TPO

TPO catalyzes two-electron oxidation of iodide and tyrosine to form MIT; further reaction of the MIT radical with iodide (one-electron oxidation) gives diiodotyrosine (3,5'-diiodotyrosine; DIT) through formation of DIT radicals. No matter the precise nature of the iodinating species, it is clear that iodide is oxidized by  $\text{H}_2\text{O}_2$  and TPO and transferred to the tyrosyl group within the Tg peptide chain. Not all of Tg's tyrosyls are equally accessible to iodination. Isolated from the thyroid, Tg rarely contains more than 1% iodine, or approximately 56 atoms of iodine per 660-kDa molecule of Tg. The molecule has approximately 134 tyrosyl residues among its two identical chains; thus, at most, only one-third of the tyrosyls are iodinated. After DIT and MIT formation, two residues of DIT couple to make L-thyroxine (tetraiodothyronine;  $\text{T}_4$ ) or one DIT and one MIT make L-triiodothyronine (triiodothyronine;  $\text{T}_3$ ), all still within the Tg molecule. In this reaction, further oxidation mediated by TPO and  $\text{H}_2\text{O}_2$  produces an iodophenyl free radical, leaving  $\text{T}_4$  or  $\text{T}_3$  at the acceptor site and dehydroalanine at the donor position.

### Regulation of Thyroid Hormone Synthesis by TPO

From the kinetic data on the iodinating and coupling reactions of free tyrosines catalyzed by TPO, it has been concluded that the mechanism of the enzyme action fits the preferential formation of  $\text{T}_4$ , even though the formation started from free tyrosine. The native structure of Tg plays an important role in the preferential formation of  $\text{T}_4$ . This is supported by the fact that a specific peptide structure of Tg is involved in the biosynthesis of thyroid hormone: The hormonogenic tyrosine residues are iodinated in rigid sequential order, resulting in the formation of DIT derivatives via MIT, which subsequently undergo one-electron oxidation to form  $\text{T}_4$ . It is important to note

that thyroid hormones formed in Tg are prevented from further oxidation by TPO, whereas free iodothyronines are readily oxidized by TPO.

### GENE STRUCTURE OF TPO

The cDNAs encoding TPO have been isolated in man, pig, rat, and mouse. Kimura *et al.* cloned two different cDNAs of hTPO: one of 3048 nucleotides [base pairs (bps)] called *TPO1* and the second (*TPO2*) of 2877 bps. *TPO1* coded for a protein of 933 residues and a molecular mass of 103 kDa, whereas *TPO2* was identical except that it lacked exon 10 and had 1 bp change, coding for a protein of 96 kDa. Both forms occur in normal and abnormal human thyroid tissue. *TPO2* appears enzymatically inactive because it does not bind heme, degrades rapidly, and failed to reach the cell surface in experiments with stable cell lines. There are different degradative pathways for the two forms. The TPO gene resides on chromosome 2p13, spans more than 150 kbps, and has 17 exons. It contains domains similar to those of acetylcholinesterase, low-density lipoprotein receptor, and insulin-like growth factor receptor. Several types of mutations of the TPO gene cause diminished iodide organification. TPO shares with NIS, Tg, and the TSH receptor the regulation of its gene expression by thyroid-specific transcription factors (TTFs), such as *TTF-1*, *TTF-2*, and *Pax-8*. Tg and TPO have the same binding sites for *TTF-1*, *TTF-2*, and *Pax-8* in their promoters, and the genes for both have *TTF-1* binding sites in their enhancer regions. TPO is synthesized on polysomes and transported to the Golgi, where it undergoes glycosylation; it is then packaged into exocytotic vesicles along with Tg. These vesicles fuse with the apical membrane in a process stimulated by TSH, and TPO is then found in the membrane associated with microvilli. Yokoyama and Taurog suggest that the C-terminal portion of the TPO molecule is in the cell cytoplasm, that the portion from residues 845 to 870 is in the apical membrane, and that the remainder, including residues 1–844, lies in the thyroid follicular lumen. TPO enzymatic activity is restricted to the apical membrane, but most of the thyroid's total amount of TPO is intracellular at the endoplasmic reticulum and perinuclear membrane. This intracellular protein is inactive due to improper folding and contains only high-mannose-type carbohydrate units; in contrast, the membrane TPO has complex carbohydrate units, essential for enzymatic activity. Chronic TSH stimulation increases the amount of TPO and its concentration at the apical membrane.

## MUTATIONS OF THE TPO GENE

Defective organification of iodine is due to abnormalities of Tg and TPO synthesis or H<sub>2</sub>O<sub>2</sub> production. Organification defect in iodine caused by abnormal H<sub>2</sub>O<sub>2</sub> production is rare, so abnormal Tg and TPO synthesis is thought to be the major cause of defective organification of iodine. The prevalence of neonatal hypothyroidism is approximately 1/4000, one-fifth of which is caused by genetically determined thyroid dyshormonogenesis. Because defective TPO activity is one of the two major causes of defective organification of iodine, approximately 1 in 40,000 newborns has hypothyroidism due to defective TPO. The results of clinical observations, however, suggest that compensatory hyperplasia of the thyroid tissue partially compensates hypothyroidism when TPO activity is borderline. Therefore, more than 1 in 40,000 newborns may have milder forms of hypothyroidism caused by defective TPO. Defects of TPO are both quantitative and qualitative; the latter include impaired binding to heme, impaired binding to Tg or iodine substrates, abnormal localization in thyrocyte, and abnormal susceptibility to inhibition. More than 110 cases of hereditary defective organification of iodine ascribable to defective TPO have been reported, and the hereditary form is autosomal recessive. Pedigree maps of these families show consanguineous marriage in many cases. The TPO gene mutations responsible for congenital goitrous hypothyroidism have been identified, causing either abnormal TPO with low or absent enzymatic activity or complete absence of TPO protein formation. It has been shown that Pendred's syndrome (an autosomal recessive disease characterized by defective iodine organification with goiter and congenital neurosensory deafness) is due to mutations of the pendrin gene. The decreased organification activity of this condition may be due to abnormal pendrin-TPO interactions of the apical membrane of thyroid follicular cells.

## TPO AS AUTOANTIGEN

TPO, Tg, and TSH receptor are the major thyroid autoantigens identified at the biochemical and molecular levels. In particular, TPO was identified in 1985 as the "thyroid microsomal antigen" reacting with thyroid microsomal autoantibodies, first described in the late 1950s. Anti-TPO autoantibodies (TPO-Ab) are detected together with anti-Tg (Tg-Ab) and anti-TSH receptor (TR-Ab) autoantibodies in sera of patients with autoimmune thyroid diseases (AITD); the majority of TPO-Ab are IgG<sub>1</sub> or IgG<sub>4</sub>.

Epitopes recognized by autoantibodies have been extensively studied using synthetic peptides, recombinant DNA molecules, and recombinant Fabs specific for TPO. B-cell epitopes are generally conformational, whereas T-cell epitopes are short, linear peptide fragments. In studies of B-cell epitopes with recombinant TPO, it has been found that many B-cell epitopes are located on the latter half of the COOH end of the TPO molecule and that the region between amino acids 590 and 767 contains an epitopic hotspot; approximately 80% of TPO-Ab recognize two conformational epitopes. TPO-Ab are detected in almost all patients with Hashimoto's thyroiditis and in the majority of those with Graves' disease. Although TPO-Ab mediate *in vitro* complement-dependent lysis and antibody-dependent, cell-mediated cytolysis of thyroid cells, they are probably devoid of cytolytic activity *in vivo* due to the inaccessibility of the apical membrane of the thyroid follicular cell. Some, but not all, TPO-Ab are able to inhibit TPO enzymatic activity *in vitro*, but the relevance of this phenomenon *in vivo* is probably minimal. Taken together, these data strongly support the concept that serum TPO-Ab are markers of thyroid autoimmunity but are not directly involved in thyroid cell damage.

Sera of AITD patients also contain autoantibodies that recognize cross-reacting epitope(s) between Tg and TPO (TGPO-Ab). The exact nature of TGPO-Ab remains to be clarified. Results of primary lymphocyte cultures with synthetic peptides have produced a degree of consensus regarding the TPO T-cell epitopes recognized in patients. Three regions of TPO protein probably contain T-cell epitopes: amino acid residues 110–250, 414–589, and 841–901. Amino acid residues 119–126 of TPO have been predicted from algorithms to be a T-cell epitope common to both TPO and Tg, and they have been proven to aid in the activation of thyroiditogenic cells. Porcine TPO can induce thyroiditis in mice. Interestingly, murine strains with a high incidence of thyroiditis induced by porcine TPO are quite different from those developing thyroiditis after immunization with Tg.

## See Also the Following Articles

Thyroid Autoimmunity • Thyroid Hormone Action • Thyroid Hormone Metabolism

## Further Reading

De Vijlder, J. J. M., Dinsart, C., Libert, F., Geurts van Kessel, A., Bikker, H., Bolhuis, P. A., and Vassart, G. (1988). Regional localization of the gene for thyroid peroxidase to human chromosome 2pter p12. *Cytogenet. Cell Genet.* **47**, 170–172.

- Dunn, J. T., and Dunn, A. D. (2001). Update on intrathyroidal iodine metabolism. *Thyroid* **11**(5), 407–414.
- Kimura, S., Kotani, T., McBride, O. W., Umeki, K., Hirai, K., Nakayama, T., and Ohtaki, S. (1987). Human thyroid peroxidase: Complete cDNA and protein sequence, chromosome mapping, and identification of two alternately spliced mRNAs. *Proc. Natl. Acad. Sci. USA* **84**, 5555–5559.
- Kopp, P. (1999). Pendred's syndrome: Identification of the genetic defect a century after its recognition. *Thyroid* **9**(1), 65–69.
- Mariotti, S., Caturegli, P., Piccolo, P., Barbesino, G., and Pinchera, A. (1990). Antithyroid peroxidase autoantibodies in thyroid diseases. *J. Clin. Endocrinol. Metab.* **71**, 661–669.
- McLachlan, S. M., and Rapoport, B. (1995). Genetic and epitopic analysis of thyroid peroxidase (TPO) autoantibodies: Markers of the human thyroid autoimmune response. *Clin. Exp. Immunol.* **101**, 200–206.
- Ohtaki, S., Nakagawa, H., Nakamura, M., and Kotani, T. (1996). Thyroid peroxidase: Experimental and clinical integration. *Endocr. J.* **43**(1), 1–14.
- Rapoport, B., and McLachlan, S. M. (2001). Thyroid autoimmunity. *J. Clin. Invest.* **108**, 1253–1259.
- Taurog, A. M. (2000). Hormone synthesis. Thyroid iodine metabolism. In "The Thyroid. A Fundamental and Clinical Text" (L. E. Braverman and R. D. Utiger, eds.), 8th ed. Lippincott Williams & Wilkins, Philadelphia.
- Yokoyama, N., and Taurog, A. (1988). Porcine thyroid peroxidase: Relationship between the native enzyme and an active, highly purified tryptic fragment. *Mol. Endocrinol.* **2**, 838–844.



## Thyroid-Stimulating Hormone

see TSH

amiodarone, iopanoic acid, propranolol, and glucocorticoids), or decreased TSH secretion (e.g., dopamine and its agonists, octreotide and glucocorticoids). Hypothyroidism and thyrotoxicosis may develop during therapy with iodine-containing drugs, and subclinical or overt hypothyroidism may be observed with long-term lithium therapy. Cytokines (e.g., interferon- $\alpha$  or interleukin-2) may precipitate hypothyroidism, thyrotoxicosis, or the biphasic pattern of silent thyroiditis, especially in the presence of preexisting thyroid autoimmunity. In hypothyroid patients taking L-T4, drugs such as ferrous sulfate, cholestiramine, cholestipol, and soybean formulations may interfere with L-T4 absorption.

## THYROID DISEASES IN THE ELDERLY

### Hypothyroidism

The prevalence of hypothyroidism in the elderly is high (0.5–6% for overt and 4–15% for subclinical hypothyroidism). Autoimmune thyroiditis is the main cause, but iatrogenic hypothyroidism (radioiodine administration, thyroid surgery, head and neck radiation, and antithyroid drugs) is also common. Excess iodine from amiodarone or iodinated radiographic contrast agents may induce hypothyroidism, preferentially in glands with preexisting autoimmunity. Hypothyroidism in the elderly develops insidiously and often lacks classic clinical features, and some manifestations may be erroneously attributed to “normal” aging or age-associated diseases. Unexplained increases in serum cholesterol, macrocytic anemia, severe constipation, congestive heart failure with restrictive cardiomyopathy, and subtle neurologic signs are the most common manifestations. Severe depression, lethargy, memory loss, apathy, and, rarely, psychosis or irreversible dementia may be observed. Elderly patients are more susceptible to myxedema coma, which may be precipitated by intercurrent NTI or cold exposure. The diagnosis of primary hypothyroidism is based on increased serum TSH, although the nonspecific effects of NTI and/or drugs must be taken into account. Decreased thyroid hormone concentrations may be observed in both hypothyroidism and NTI, but low FT4 is more frequent in thyroid failure. Anti-thyroid antibody tests help to differentiate autoimmune from nonautoimmune causes of hypothyroidism.

Subclinical hypothyroidism (increased serum TSH with normal FT4 concentration) occurs frequently in the elderly. Like overt thyroid failure, most cases are due to autoimmune thyroiditis or to previous treatment of hyperthyroidism. Progression to overt

hypothyroidism occurs in 2–18% of cases per year, with the highest percentages observed for subjects with high serum thyroid antibody titers. Subclinical hypothyroidism is associated with a slight but significant increase in serum lipids.

Therapy should be initiated with low doses of L-T4 (12.5–25  $\mu\text{g}/\text{day}$ ) followed by incremental increases (12.5–25  $\mu\text{g}/\text{day}$  every 4–8 weeks) until full replacement (1.1 or 1.2  $\mu\text{g}/\text{kg}$  body weight) after several months. Particular attention should be paid to patients with coronary disease to avoid angina or myocardial infarction. Indication for therapy in subclinical hypothyroidism is controversial. Serum TSH  $>10$  mU/liter and/or high thyroid antibody titers favor treatment, particularly when hypercholesterolemia and hypothyroid symptoms are present. Careful consideration of potential adverse reactions is always required.

### Hyperthyroidism

The prevalence of hyperthyroidism in the elderly is 0.5–2%, mostly due to Graves' disease, toxic nodular goiter (TNG), and toxic adenoma. The relative frequency of TNG is higher than in young patients, especially in areas of iodine deficiency. Iodine-induced hyperthyroidism occurs frequently in elderly patients using iodine-containing medications or radiographic contrast media, and amiodarone-induced thyrotoxicosis is the most frequent form. Elderly hyperthyroid patients frequently display few signs and symptoms, hence the terms apathetic or masked hyperthyroidism. Eye signs are often lacking, but Graves' ophthalmopathy, when present, is usually worse. Tachycardia is less common, but a high prevalence (25–35%) of atrial fibrillation is found in thyrotoxic elderly male patients. Weight loss is often associated with anorexia rather than increased appetite; muscle wasting and weakness are common, and the high risk of osteoporosis and bone fracture typical of old age is increased. Neuropsychiatric symptoms may be reported as primary manifestations. Elevated serum FT4 and/or FT3 and low TSH levels by sensitive assay establish the diagnosis. Serum TSH may be low in euthyroid severely ill patients with coexistent NTI, but undetectable rather than low serum TSH by third-generation assays strongly suggests hyperthyroidism. However, discrimination between the two conditions may not be possible without concomitant FT3 and FT4 assay.

Subclinical hyperthyroidism (low TSH and normal FT3 and FT4) must be distinguished from other

causes of low serum TSH, such as NTI. The prevalence of low serum TSH among elderly people is high (1.5–12.5%), but the number of cases progressing to overt hyperthyroidism is low (2–10%). Subclinical hyperthyroidism may be due to “subclinical” Graves’ disease, autonomously functioning thyroid nodules, or excessive/inappropriate thyroid hormone therapy. Subclinical hyperthyroidism in the elderly is associated with increased risk of atrial fibrillation and mortality from cardiovascular diseases, as well as reduced bone density in postmenopausal women.

Long-term therapy with antithyroid drugs is not recommended for elderly patients because of the high relapse rate after withdrawal and the increased incidence of adverse effects. Radioiodine ( $^{131}\text{I}$ ) is the treatment of choice since it results in a definitive cure and avoids the risks of surgery. Euthyroidism should be restored with antithyroid drugs before  $^{131}\text{I}$  therapy, and control of heart rate with beta-blockers (or calcium channels blockers) should be obtained before and after therapy, as long as the patient remains thyrotoxic. Long-term follow-up of thyroid function is mandatory and eventual hypothyroidism must be corrected. Most cases of subclinical hyperthyroidism in the elderly require active treatment due to cardiovascular and bone complications.

### Nontoxic Goiter and Thyroid Carcinoma

An age-dependent increase in thyroid volume and nodularity has been documented by echography, reaching a prevalence after 60 years of age of approximately 50% in iodine-deficient areas. Because surgical risks are higher in the elderly, most nontoxic nodular goiters are managed conservatively, unless there is strong suspicion of malignancy or significant airway obstruction. In patients with contraindications to surgery, radioiodine has been used successfully, with

partial reduction in thyroid size and relief of pressure symptoms.

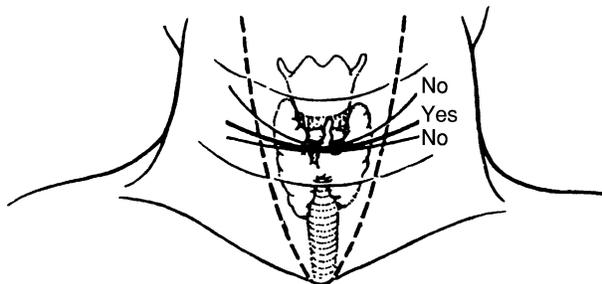
The ratio of papillary/follicular thyroid carcinoma is lower in the elderly (2:1) than in younger patients (3 or 4:1). Differentiated thyroid carcinoma is more aggressive in older patients, especially males. Anaplastic thyroid carcinoma is almost exclusively observed in patients older than 65 years of age, similar to other rare thyroid neoplasms, such as sarcomas and primary thyroid lymphomas. The therapeutic approach for differentiated thyroid cancer is similar to that followed for younger persons.

### See Also the Following Articles

Aging and Longevity of Human Populations • Graves’ Disease, Hyperthyroidism in • Graves’ Ophthalmopathy • Hyperthyroidism, Subclinical • Hypothyroidism, Treatment of • Nontoxic Goiter • Thyroid Autoimmunity • Thyroid Carcinoma • Toxic Adenoma • Toxic Multinodular Goiter

### Further Reading

- Chiovato, L., Mariotti, S., and Pinchera, A. (1997). Thyroid diseases in the elderly. *Baillieres Clin. Endocrinol. Metab.* **11**, 251–270.
- Mariotti, S., Franceschi, C., Cossarizza, A., and Pinchera, A. (1995). The aging thyroid. *Endocr. Rev.* **16**, 686–715.
- Mariotti, S., Chiovato, L., Franceschi, C., and Pinchera, A. (1998). Thyroid autoimmunity and aging. *Exp. Gerontol.* **33**, 535–541.
- Mokshagundam, S., and Barzel, U. S. (1993). Thyroid disease in the elderly. *J. Am. Geriatric Soc.* **41**, 1361–1369.
- Parle, J. V., Maisonneuve, P., Sheppard, M. C., Boyle, P., and Franklyn, J. A. (2001). Prediction of all-cause and cardiovascular mortality in elderly people from one low serum thyrotropin result: A 10-year cohort study. *Lancet* **358**, 861–865.
- Sawin, C. T., Geller, A., Wolf, P. A., Belanger, A. J., Baker, E., Bacharach, P., Wilson, P. W., Benjamin, E. J., and D’Agostino, R. B. (1994). Low serum thyrotropin concentrations as a risk factor for atrial fibrillation in older persons. *N. Engl. J. Med.* **331**, 1249–1252.



**Figure 1** The skin incision in a thyroidectomy is made 1 cm below the cricoid and follows normal skin lines. A 00 silk suture is pressed against the neck to mark the incision site. Reprinted from Clark (1985), with permission.

crease for optimal cosmetic effect. Thus, the incision is directly over the isthmus of the thyroid gland. The gentle curve of the Kocher incision is then marked using a taut silk suture (Fig. 1). The length of the incision is dependent on the existing pathology and the size of the patient's neck. It may vary from approximately 3 to 5 cm.

The skin incision is extended through subcutaneous fat and the platysma muscle. The superior flaps are created by dissection in an avascular plane just deep to the platysma muscle and superficial to the anterior jugular veins up to the level of the thyroid cartilage. The lower flap is mobilized in a similar fashion to the level of the suprasternal notch. The wound edges are protected with moistened drapes, and one or two self-retaining retractors are placed.

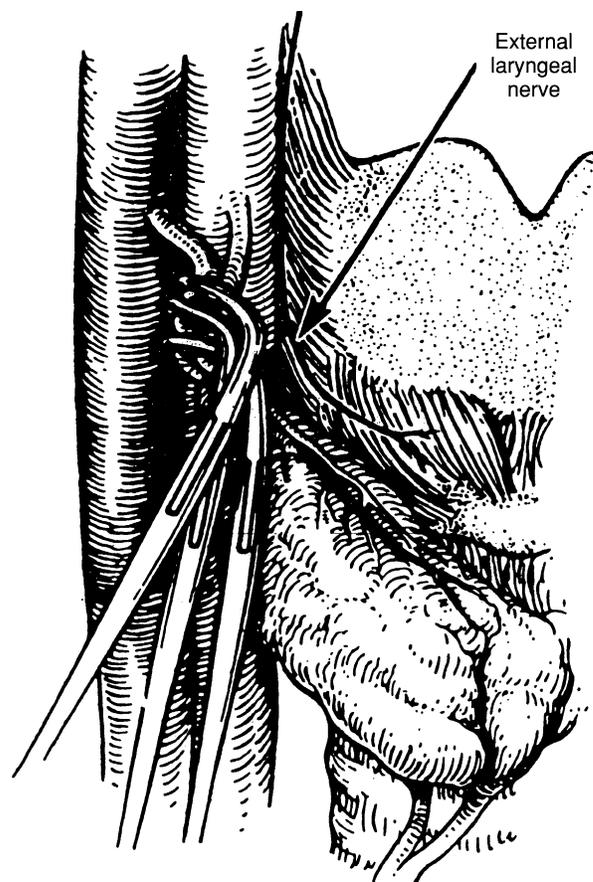
## MIDLINE DISSECTION

Excellent exposure is obtained by a midline incision through the median raphe of the superficial layer of the deep cervical fascia between the strap muscles. The midline is most easily identified low in the neck. The incision is extended superiorly to the thyroid cartilage and inferiorly to the suprasternal notch. Crossing veins in the lower neck may need to be ligated.

The sternohyoid muscle is dissected from the sternothyroid muscle laterally to provide better exposure of the thyroid gland. This can usually be done with blunt dissection. The sternothyroid muscle is then dissected from the underlying thyroid. The middle thyroid vein(s) is identified, divided, and ligated. This dissection is facilitated by medial retraction of the thyroid gland and lateral retraction of the strap muscles and carotid sheath.

Attention is turned to the superior pole. Retracting the thyroid in a caudal direction identifies the superior thyroid artery and veins. The tissue adjacent to the superior pole vessels can usually be swept from the thyroid with a peanut sponge. The space between the thyroid gland and cricothyroid muscle is opened, allowing the superior pole vessels to be skeletonized, triple clamped, ligated, and divided. The external branch of the superior laryngeal nerve is at risk of injury during this maneuver; there is a 10% reported injury rate. It has a variable course and is a thin nerve. The external branch of the superior laryngeal nerve is best preserved by not attempting to directly identify it but, rather, individually ligating the superior pole vessels close to the thyroid as opposed to mass ligation of the pedicle (Fig. 2).

The tissue posterior to the pole can now be easily swept from the thyroid gland by blunt dissection. The upper parathyroid gland is usually identified at



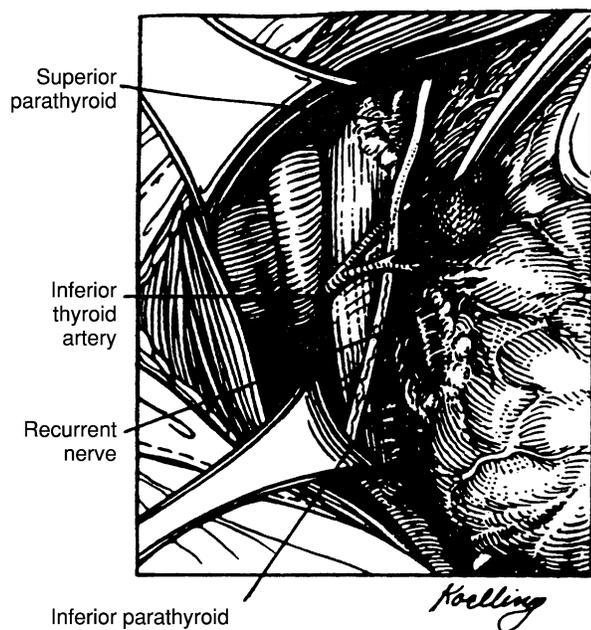
**Figure 2** The superior pole vessels are identified and individually ligated using three right-angle clamps placed with tips away from the external branch of the superior laryngeal nerve. Reprinted from Clark (1985), with permission.

the level of the cricoid cartilage, where the RLN enters the larynx posterior to the cricothyroid muscle. It should be preserved on its vascular pedicle. The area cephalad to the cricoid cartilage is considered relatively safe because the recurrent laryngeal nerve enters the cricothyroid muscle below the cricoid cartilage.

The pyramidal lobe, which is present in approximately 80% of patients, should be identified and mobilized to the level above the thyroid cartilage with the dissection plane immediately adjacent to this lobe. There are often numerous small vessels supplying the pyramidal lobe. When the pyramidal lobe is freed of all its lateral attachments, it is gently avulsed or divided from its superior attachment at the hyoid bone.

### IDENTIFICATION OF THE RECURRENT LARYNGEAL NERVE

Both right and left RLNs enter the larynx posterior to the cricothyroid muscle just above the cricoid cartilage. The right RLN takes a more oblique course in the neck and may pass anterior or posterior to the inferior thyroid artery (Fig. 3). In approximately 1% of patients, the right RLN is nonrecurrent and may



**Figure 3** The recurrent laryngeal nerve may run anterior, posterior, or between the branches of the inferior thyroid artery. The inferior parathyroid gland is usually identified within 1 cm of this junction. Reprinted from Clark (1985), with permission.

enter the thyroid from a superior or lateral direction. The left RLN almost always runs in the tracheoesophageal groove because of its deeper origin from within the thorax. Both recurrent nerves may branch before entering the larynx; this occurs more frequently on the left. Preservation of the medial branch is of utmost importance because it usually contains the motor fibers.

Retraction of the carotid sheath laterally and the thyroid medially and anteriorly places tension on the inferior thyroid artery, which helps to identify the RLN where it crosses the midportion of the thyroid gland. It is usually safe to identify the RLN low in the neck and then follow it to where it enters the cricothyroid muscle through the ligament of Berry. The tertiary branches of the inferior thyroid artery are individually ligated with fine ties and divided, mobilizing the thyroid lobe medially, away from the RLN.

The most difficult part of the dissection during a thyroidectomy usually involves the ligament of Berry. The ligament is situated at the posterior lateral portion to the thyroid gland just caudal to the cricoid cartilage. A small branch of the inferior thyroid artery traverses the ligament, as do one or more veins from the thyroid gland. These vessels are usually readily identified and ligated. Should bleeding occur, it should be controlled by pressure with no clamping until the RLN is identified. A small amount of persistent bleeding at the end of the case can be controlled with the placement of a small pledget of thrombin-soaked gel foam. A tubercle of Zuckerkandl may extend over the RLN at the ligament of Berry. In addition to having a consistent relationship with the location of the RLN, the upper parathyroid gland may be situated at the tip of this protruding portion of thyroid tissue.

### PRESERVATION OF THE PARATHYROID GLANDS

Eighty-five percent of people have four parathyroid glands usually situated immediately adjacent to the thyroid gland on the posterior lateral capsule. The upper parathyroid glands are most commonly lateral to the recurrent laryngeal nerve at the level of the ligament of Berry in a posterior position and are usually the easiest to preserve during thyroidectomy. The lower parathyroid glands are almost always situated anterior to the RLN and within 1 cm caudal to where the RLN crosses the inferior thyroid artery. If not observed here, they are usually in the thymus or

perithymic fat. Preservation of the blood supply to the parathyroid glands is best achieved by meticulous dissection from the thyroid capsule. Should it not be feasible to preserve the blood supply to a parathyroid gland, it should be biopsied to confirm it is parathyroid and autotransplanted into the contralateral sternocleidomastoid muscle.

In general, parathyroid glands should be preserved rather than transplanted if possible. The parathyroid glands should not be mobilized extensively because they may be devascularized. The thyroid gland should also be examined after its removal to ensure that a parathyroid gland has not been removed.

## REMOVAL OF SPECIMEN

Always keeping in mind the location of the RLN, the inferior thyroid vessels are dissected free, clamped, and ligated. The thyroid gland is retracted medially to expose the trachea. The thyroid lobe and isthmus are then easily dissected off the trachea with electrocautery. For a lobectomy, the isthmus is divided between two Colodny clamps on the side contralateral to the thyroid pathology, and the isthmus is suture ligated. The operative field is irrigated with warm saline and any bleeding is controlled.

The strap muscles and then platysma muscle are reapproximated with interrupted 4-0 absorbable suture, and the skin is closed with butterfly clips. The wound is dressed with sterile gauze bandages.

## FUTURE DIRECTIONS

Except for improved lighting and magnification, the fundamental approach to thyroid surgery, with

preservation of the RLN and parathyroid glands, has remained relatively unchanged during the past century. Minimally invasive surgery has evolved to include thyroid surgery. Endoscopic parathyroidectomy using CO<sub>2</sub> insufflation has been demonstrated to be technically feasible, and similar techniques have been applied to thyroidectomy. The development of highly sensitive nerve probes is under investigation. Its utility in avoiding injury to the external branch of the superior laryngeal nerve and preservation of the RLN in redo surgery has had promising preliminary results. Regardless of these new developments, the importance of adequate training, expertise, and knowledge of anatomy and techniques are fundamental to ensuring good outcomes.

## See Also the Following Articles

Medullary Thyroid Carcinoma • Nontoxic Goiter • Parathyroid Cancer • Parathyroid Glands, Pathology • Thyrotoxicosis, Treatment

## Further Reading

- Clark, O. H. (1985). "Endocrine Surgery of the Thyroid and Parathyroid Glands." Mosby, St. Louis.
- Clark, O. H., and Noguchi, S. (2000). "Thyroid Cancer—Diagnosis and Treatment." Quality Medical, St. Louis.
- Doherty, G. M., and Skogseid, B. (2001). "Surgical Endocrinology." Lippincott Williams & Wilkins, Philadelphia.
- Duh, Q.-Y., and Clark, O. H. (1997). "Textbook of Endocrine Surgery." Saunders, Philadelphia.
- Jossart, G. H., and Clark, O. H. (1996). Perioperative management and surgical technique. *Sci. Am.* 1-8.
- Wartofsky, L. (2000). "Thyroid Cancer. A Comprehensive Guide to Clinical Management." Humana Press, Totowa, NJ.

**Table I** Differential Diagnosis of the Painful Neck Mass

Nonthyroidal	
	Infected thyroglossal duct cyst
	Infected branchial cleft cyst
	Infected cystic hygroma
	Cervical adenitis
	Cellulitis of the anterior neck
Thyroidal	
	Subacute thyroiditis
	Infectious thyroiditis
	Acute hemorrhage into a cyst
	Acute hemorrhage into a benign or malignant nodule
	Rapidly enlarging thyroid carcinoma
	Painful Hashimoto's thyroiditis
	Radiation thyroiditis
	Globus hystericus

### Clinical Presentation and Diagnosis

Bacterial thyroiditis is often preceded by an upper respiratory infection, which may induce inflammation of the fistula and promote the transmission of pathogens to the thyroid, and it is more common in the late fall and late spring months. More than 90% of patients present with thyroidal pain, tenderness, fever, and local compression resulting in dysphagia and dysphonia, and signs and symptoms of systemic toxicity may be present. The thyroid is tender to palpation, with unilateral or bilateral lobar enlargement, and it is associated with erythema and warmth of the skin. Cervical lymphadenopathy is not a prominent feature unless there is a predisposing pharyngitis.

The differential diagnosis of bacterial thyroiditis can be divided into nonthyroidal and thyroidal causes (Table I). Essentially all of the nonthyroidal causes are

infectious in origin and present as discrete painful masses. Subacute thyroiditis is the most common cause of the painful thyroid and often results in both local and systemic symptoms similar to those seen in bacterial thyroiditis.

Thyroid function tests in the normal range are the most common finding in patients with bacterial thyroiditis (Table II), although thyrotoxicosis and hypothyroidism have been reported. Fine needle aspiration biopsy is the best laboratory test in the evaluation of infectious thyroiditis and is diagnostic in most cases, especially when tenderness is limited to a solitary nodule or a localized area and subacute thyroiditis has been ruled out. Gram stain and culture of the fine needle aspirate will reveal the causative organism in more than 90% of cases. Most imaging studies are adjunctive and are best reserved for patients for whom the diagnosis is unclear.

In the adult, *Staphylococcus aureus* and *Streptococcus pyogenes* are the offending pathogens in more than 80% of patients and are the sole pathogen in more than 70% of cases. In children,  $\alpha$ - and  $\beta$ -hemolytic *Streptococcus* and a variety of anaerobes account for ~70% of cases, whereas mixed pathogens are identified in >50% of cases.

### Management and Prognosis

Treatment of acute bacterial thyroiditis requires admission to the hospital, drainage of any abscess, and parenteral antimicrobial therapy aimed at the causative agent. If no organisms are seen on the gram stain, nafcillin and gentamycin or a third-generation cephalosporin is appropriate initial therapy in adults, whereas a second-generation cephalosporin or clindamycin is reasonable in children. Since a pyriform sinus fistula is the most

**Table II** Diagnostic Tests in the Evaluation of Infectious Thyroiditis

Test	Comments
Fine needle aspiration biopsy	Diagnostic in 90% of cases; test of choice; special stains considered necessary
Thyroid function tests	Usually normal; rare case reports of hypothyroidism and thyrotoxicosis
Erythrocyte sedimentation rate	Usually elevated; nonspecific test
Leukocyte count	Usually elevated; nonspecific test
Radionuclide imaging	Adjunctive test; radioiodine best; provides information regarding overall gland function
Neck radiograph	Adjunctive test; presence of gas indicates abscess with anaerobic organisms
Magnetic resonance imaging	Adjunctive test; helpful in identifying pyriform sinus fistulae
Ultrasonography	Adjunctive test; helpful in identifying pyriform sinus fistulae
Computed tomography	Adjunctive test; helpful in identifying pyriform sinus fistulae
Barium swallow	Adjunctive test; helpful in identifying pyriform sinus fistulae

common route of infection in bacterial thyroiditis, a barium swallow, computed tomography, or magnetic resonance imaging of the neck should be performed to search for communicating fistulae in most patients, especially children, with the first episode and all patients with recurrent episodes. Such fistulae must be surgically excised for definitive cure and prevention of recurrent infection.

Mortality from acute bacterial thyroiditis has markedly improved from the 20–25% reported in the early 1900s. A 1983 review by Berger *et al.* estimated an overall mortality of 8.6%. Reviews involving more than 100 patients failed to list mortality as a complication of acute bacterial thyroiditis. However, the mortality is close to 100% if the diagnosis is delayed and antimicrobial therapy is not initiated. In survivors, complete recovery is the norm, although there have been reports of transient hypothyroidism, vocal cord paralysis (which may also be transient), and recurrence of infection as sequelae of acute bacterial thyroiditis.

## FUNGAL INFECTIONS

Although rare, fungal infections of the thyroid are the next most common cause of infectious thyroiditis, comprising 15% of cases reported through 1980. The predominant offending organism is *Aspergillus* species, with at least 26 documented cases. Virtually all of the affected patients were immunocompromised and had disseminated disease; most cases of thyroidal infection were determined postmortem. Asymptomatic infection of the thyroid with *Pneumocystis carinii* is found in up to 20% of AIDS patients with disseminated *Pneumocystis* infection at autopsy. The diagnosis of *P. carinii* infection is made by performing Gomori's silver methenamine stain on specimens obtained by fine needle aspiration biopsy. Case reports of fungal infections of the thyroid have included *Coccidioides immitis*, *Histoplasma capsulatum*, *Candida albicans*, *Allescheria boydii*, and *Nocardia asteroides*.

## MYCOBACTERIAL INFECTIONS

The true incidence of infection of the thyroid with *Mycobacterium tuberculosis* is difficult to determine. Using strict pathological criteria, only 19 cases have been reported in the literature. Thyroidal tuberculosis is associated with disseminated or miliary disease and symptoms are usually present for months. Although at least three of the reported patients with tuberculous thyroiditis died and recurrent laryngeal

nerve paralysis has been described, resolution without sequelae usually follows appropriate antituberculous therapy. Infections with atypical mycobacteria, including *M. chelonii*, *M. intracellulare*, and *M. avian-intracellulare*, have also been described—the latter in patients with AIDS in the setting of widely disseminated disease. Although acid fast organisms have been found in the thyroid of individuals with disseminated *M. leprae*, symptomatic thyroid infection has not been described.

## PARASITIC INFECTIONS

Several parasitic agents have involved the thyroid on rare occasions. *Echinococcus granulosus* has been reported in the setting of a chronic goiter, with the diagnosis being made at the time of surgery. If echinococcal infection is suspected, biopsy of the lesion is contraindicated due to spillage and rupture of the cyst contents, and specific serologic testing should be performed. Surgical removal is the preferred mode of treatment, with antiparasitic agents useful as adjunctive therapy and for inoperable cases. Involvement of the thyroid with *Strongyloides stercoralis* has been described only in the setting of disseminated disease in immunocompromised patients. Mortality with disseminated *Strongyloides* infection is high due to both the infection and the immunocompromised status of the patient.

## SYPHILITIC INFECTION

Historically, secondary syphilis was frequently associated with pain and swelling of the thyroid that responded to antisiphilic therapy, which commonly included iodides. However, microbiologic evidence of syphilitic infection of the thyroid is lacking in most of these cases; thus, a direct relationship between treponemal infection and thyroid dysfunction cannot be determined. Indeed, only seven cases of gummata of the thyroid, presenting as painless nodules, have been reported in the world's literature.

## VIRAL INFECTIONS

The most common infectious organism found in the thyroid at postmortem examination in patients with AIDS is cytomegalovirus (CMV), occurring in the setting of disseminated CMV infection. However, symptomatic thyroidal infection with CMV has not been reported. Thyroiditis has been associated with mumps parotitis, although this is rare.

## CONCLUSION

Infectious thyroiditis is uncommon and the diagnosis often requires a high index of suspicion. A rational approach to such patients, including history, physical examination, laboratory evaluation, and fine needle aspiration biopsy, will allow the appropriate diagnosis to be made in the vast majority of cases.

## See Also the Following Articles

Pediatric HIV Infection and Hypothalamic–Pituitary–Adrenal Axis • Thyroid Carcinoma • Thyroid Fine Needle Aspiration Cytology • Thyroiditis, Postpartum • Thyroiditis, Subacute

## Further Reading

- Battan, R., Mariuz, P., Raviglione, M. C., Sabatini, M. T., Mullen, M. P., and Poretsky, L. (1991). *Pneumocystis carinii* infection of the thyroid in a hypothyroid patient with AIDS: Diagnosis by fine needle aspiration biopsy. *J. Clin. Endocrinol. Metab.* **72**, 724–726.
- Berger, S. A., Zonszein, J., Villamena, P., and Mittman, N. (1983). Infectious diseases of the thyroid gland. *Rev. Infect. Dis.* **5**(1), 108–122.
- Farwell, A. P. (2000). Infectious thyroiditis. In “The Thyroid” (L. E. Braverman and R. D. Utiger, eds.), 8th ed., pp. 1044–1050. Lippincott William & Wilkins, Philadelphia.
- Miyauchi, A., Matsuzuka, F., Kuma, K., and Takai, S. (1990). Piriform sinus fistula: An underlying abnormality common in patients with acute suppurative thyroiditis. *World J. Surg.* **14**(3), 400–405.
- Nonomura, N., Ikarashi, F., Fujisaki, T., and Nakano, Y. (1993). Surgical approach to pyriform sinus fistula. *Am. J. Otolaryngol.* **14**(2), 111–115.
- Park, S. W., Han, M. H., Sung, M. H., Kim, I. O., Kim, K. H., Chang, K. H., and Han, M. C. (2000). Neck infection associated with pyriform sinus fistula: Imaging findings. *Am. J. Neuroradiol.* **21**(5), 817–822.
- Rich, E. J., and Mendelman, P. M. (1987). Acute suppurative thyroiditis in pediatric patients. *Pediatr. Infect. Dis. J.* **6**(10), 936–940.
- Shah, S. S., and Baum, S. G. (2000). Diagnosis and management of infectious thyroiditis. *Curr. Infect. Dis. Rep.* **2**, 147–153.
- Singh, S. K., Agrawal, J. K., Kumar, M., and Shukla, H. S. (1994). Fine needle aspiration cytology in the management of acute suppurative thyroiditis. *Ear Nose Throat J.* **73**(6), 415–417.
- Takai, S.-I., Miyauchi, A., Matsuzuka, F., Kuma, K., and Kosaki, G. (1979). Internal fistula as a route of infection in acute suppurative thyroiditis. *Lancet* **1**, 751–752.

pregnancy, with postpartum increases in helper T cells and cytotoxic T cells.

The hormonal changes that occur during pregnancy have a profound effect on the production of immune tolerance. Increased corticotropin-releasing hormone production by the placenta stimulates the maternal adrenal glands to produce a state of hypercortisolism. There is activation of the sympathetic system with increased production of catecholamines. A moderate increase in 25-hydroxy vitamin D<sub>3</sub> and more significant increases in 1,25-dihydroxy vitamin D<sub>3</sub> also occur. Through intermediate mechanisms, these changes suppress proinflammatory cytokine formation and promote a shift from Th1 to Th2. The increase in plasma estrogen and progesterone levels may enhance this effect. These changes help to conserve the fetus and reduce chances of rejection.

Maternal immune mechanisms return to a normal nonpregnant state in the first few months to 1 year after delivery. However, there may be a rebound increase in some elements of the autoimmune reaction (e.g., antibody production or CD4:CD8 ratio), which may aggravate existing autoimmune disease or precipitate it for the first time in predisposed women.

### Women with PPT

The “immune rebound” hypothesis may explain some of the autoimmune mechanisms that form the basis of PPT. The behavior of antibodies to thyroglobulin (TgAb) and thyroid peroxidase (TPOAb) during pregnancy and the postpartum period has been well documented. This heightened immune response in the first few months of the postpartum period may be caused by several factors.

#### Microchimerism

There is a considerable influx of fetal cells into the maternal circulation at the time of delivery. These cells may persist for short or long periods of time in the maternal host. Evidence suggests that these “chimeric” cells may induce an immune reaction in the host by causing a breakdown in immune tolerance. One short-term effect may be the rebound immune enhancement seen in the postpartum period, which may cause an exacerbation or new onset of some autoimmune diseases.

#### T-Cell Changes

Evidence suggests that activation of both circulating and intrathyroidal T cells occurs in PPT. Both Walfish and Stagnaro-Green demonstrated the expression of MHC class II molecules and a higher percentage of increased CD4:CD8 ratios in subjects who developed

PPT. In a prospective study of TPOAb-positive women, Kuijpers showed a higher percentage of MHC II-expressing T cells in subjects who subsequently developed PPT compared to those who did not. However, Jansson’s group did not demonstrate a difference in circulating subsets of T lymphocytes in thyrotoxic and hypothyroid PPT subjects compared with normal controls. This group demonstrated a relative increase in intrathyroidal B cells and a relative decrease in CD8 cells (resulting in an increased CD4:CD8 ratio) in subjects with PPT.

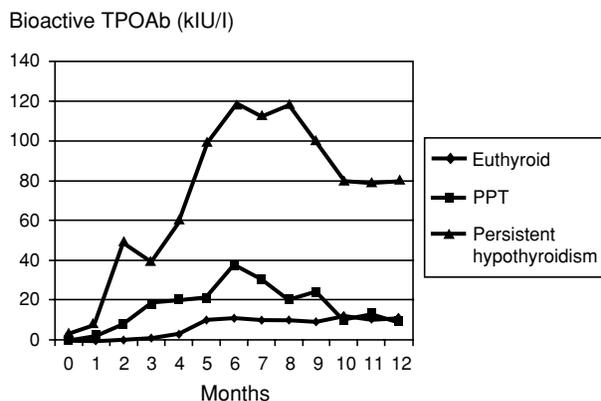
### PPT as an Immune-Mediated Disease

Several features of PPT point to the central importance of immune mechanisms in the pathogenesis of PPT (Table I). The majority of women who develop PPT are positive for serological markers of thyroid autoimmunity (i.e., TPOAb and TgAb). In our experience, all such women have TPOAb during early pregnancy. Evidence suggests that antibodies with a dual specificity for Tg and TPO may also be found in PPT at a higher prevalence than in normal control subjects. However, several investigators have reported PPT in TPOAb-negative women. In such women, the etiology of PPT is unclear. The histological changes occurring within the thyroid gland, with immune cell infiltration typical of autoimmune thyroid disease, give further credence to the immune pathogenesis of the disease. Biopsy of the thyroid gland in women with PPT shows lymphocytic infiltration and follicle formation reminiscent of those of Hashimoto’s thyroiditis. Several studies have shown HLA haplotype restrictions in PPT, which are commonly seen in autoimmune thyroiditis such as Hashimoto’s and Graves’ disease.

The subclasses of TPOAb that are able to activate the complement cascade (IgG<sub>1</sub>–IgG<sub>3</sub>) increase during

**Table I** Immunohistological Features of PPT

Presence of thyroid antibodies	TPOAb (majority)  Thyroglobulin antibodies and antibodies with dual specificity (minority)
Postpartum increase in TPOAb subclasses capable of activating complement	IgG <sub>1</sub> –IgG <sub>3</sub>
Histological changes	Lymphocyte infiltration and follicle formation within the thyroid gland
HLA haplotype restriction	Similar to autoimmune thyroiditis



**Figure 1** Biologically active TPOAb levels in antibody-positive women who were euthyroid or had transient or persistent postpartum hypothyroidism. Antibody concentrations returned to normal in those with PPT but remained elevated in subjects with persistent hypothyroidism.

the postpartum period (Fig. 1) and are associated with both phases of PPT. Jansson reported an increase in IgG<sub>1</sub> in hypothyroid PPT. Hall demonstrated an increase in IgG<sub>2</sub> and IgG<sub>3</sub> in biphasic PPT, with the increase in IgG<sub>3</sub> coinciding with the thyrotoxic phase. Briones-Urbina confirmed the IgG<sub>1</sub> and IgG<sub>2</sub> changes but found low IgG<sub>3</sub> levels. However, these investigators also confirmed that IgG<sub>4</sub>, which is incapable of influencing the complement cascade, remains relatively unchanged. This raises the interesting possibility of a pathogenetic role for these antibodies in PPT, perhaps through complement activation. A sublethal antibody-directed, complement-mediated attack on thyroid cells may result in increased secretion of thyroid hormones, producing the thyrotoxic phase of the disease. However, a more severe complement-mediated attack may produce damage to the follicular architecture of the gland and produce hypothyroidism. Conclusive evidence for such complement activation in PPT is lacking; studies have been unable to demonstrate terminal complement complexes (markers of complement activation) in TPOAb-positive women who developed PPT and who were followed weekly during the postpartum period.

## CLINICAL FEATURES AND MANAGEMENT OF PPT

### Prevalence

There is a wide variation in the reported worldwide prevalence of PPT. This variation may be explained by true geographic differences in prevalence (reflecting genetic heterogeneity and other factors) or by

methodological discrepancies in studies reported from different locations. Variability of factors such as diagnostic criteria, length of follow-up after delivery, frequency of postpartum blood sampling, and differences in hormone assay methodology may have contributed to this variation. An average prevalence of 5–7% is acceptable for unselected pregnant women from most iodine-replete populations.

### Predisposition to PPT

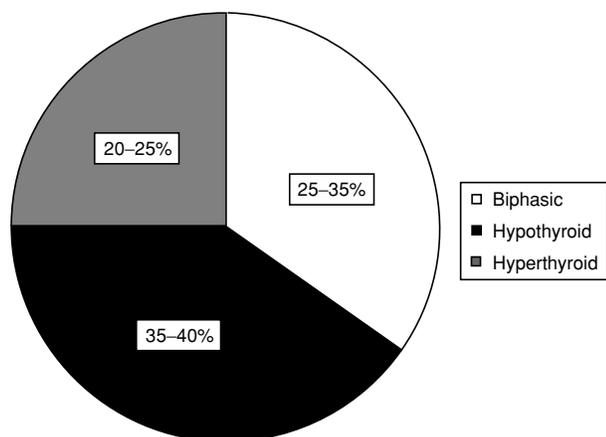
Women with TPOAb (and TgAb alone in less than 5%), type 1 diabetes mellitus, and previous PPT are at increased risk of developing PPT. We previously mentioned this increased risk in TPOAb-positive women. Studies in Cardiff show that approximately 50% of women with TPOAb during the early stages of pregnancy develop PPT. Other studies have shown this proportion to vary between 30 and 52%. Therefore, TPOAb is a marker of risk for the development of PPT but remains a weak predictor. There is a higher prevalence of PPT in subjects with type 1 diabetes mellitus. Gerstein followed up 40 of 51 pregnant subjects with type 1 diabetes, of whom 10 developed thyroid dysfunction (1 due to Graves' disease), and Alvarez-Marfany followed up 28 of 41 similar women, of whom 7 developed thyroiditis. Therefore, the incidence is approximately 25% in this group of women. However, the highest incidence of PPT is found in women who have had a previous episode of the disease. Sixty-nine percent of women who have TPOAb and had PPT following a previous pregnancy will develop a similar disease during the next pregnancy. However, only 25% who are TPOAb positive and remain euthyroid will develop thyroid dysfunction during the next pregnancy.

There is evidence of a role for environmental factors, such as the presence of a goiter and smoking and a family history of thyroid disease, in the causation of PPT. However, additional studies are needed to confirm this finding.

### Types of PPT and Clinical Features

Classically, PPT occurs after a full-term delivery. However, there are reports of postpartum thyroiditis occurring after early loss of pregnancy between 5 and 20 weeks of gestation.

PPT is typically a biphasic disease (Fig. 2). A transient thyrotoxic phase is followed by a period of recovery and then a hypothyroid phase. The thyrotoxic phase occurs at a median of 13 weeks postpartum and lasts 1 or 2 months; it is probably due to the release of



**Figure 2** Clinical types of PPT. The thyrotoxic phase usually precedes the hypothyroid phase.

performed thyroid hormone resulting from destruction of thyroid follicles. The hypothyroid phase that follows occurs at a median of 19 weeks postpartum, lasts longer (approximately 4–6 months), is accompanied by significant symptoms, and results from autoimmune follicular destruction and reduced hormone synthesis. Some women may require thyroxine replacement therapy during this phase. Rarely, the hypothyroid phase may precede the thyrotoxic phase. In some women, the two phases of PPT may occur independently of each other and either clinical or biochemical evidence of thyrotoxicosis or hypothyroidism alone develops during approximately the same postpartum periods as described previously. Significantly, as many as 30% of women who have TPOAb and PPT may develop permanent hypothyroidism requiring thyroxine replacement therapy by the end of the first postpartum year.

The symptoms of the thyrotoxic phase are mild and self-limiting. Fatigue, palpitations, weight loss, irritability, and heat intolerance are more commonly found in subjects with thyrotoxic PPT than in euthyroid postpartum women. They may also have tremor, nervousness, and psychological symptoms. The mild and nonspecific nature of these symptoms may cause diagnostic confusion, and they may be missed if a high index of suspicion is not maintained.

The hypothyroid phase lasts longer and may cause considerable morbidity. Fatigue, loss of concentration, constipation, muscle and joint pains, and stiffness are common complaints. Some of these symptoms may occur before the abnormalities in biochemical thyroid function become evident and also persist after euthyroidism is achieved. Studies from The Netherlands and United Kingdom have demonstrated a significant

excess of symptoms of minor and major depression. However, a study from Spain was unable to confirm this finding. The mechanism of depression in hypothyroid PPT is speculative but may be related to the reduced 5-hydroxytryptamine drive seen in this condition or to known cytokine release associated with this phase, affecting neurotransmission.

### Management of PPT

The temporal relationship to pregnancy and delivery, the presence of thyroid antibodies in the majority of women who develop symptoms, and the pattern and timing of biochemical thyroid dysfunction should alert clinicians to PPT in women at risk. Symptoms and the presence of a goiter are unhelpful in differentiating PPT from other causes of thyroid dysfunction in the postpartum period. The presence of eye signs in the thyrotoxic phase, however, favors a diagnosis of Graves' disease.

The thyrotoxic phase is distinguished from an exacerbation of quiescent or a new onset of Graves' disease relatively easily by radioiodine uptake scanning. Uptake is consistently low in PPT but high in Graves' disease. When postpartum thyrotoxicosis occurs in women with previously known Graves' disease, a low uptake confirms PPT (on the background of quiescent Graves' disease). However, care needs to be taken in the use of radioiodine as a diagnostic tool in nursing mothers. Technetium scans may be preferable. The presence of thyroid-stimulating hormone (TSH) receptor antibodies favors the diagnosis of Graves' disease. Thyroid ultrasonography, thyroglobulin estimation, and IL-6 measurement are of little practical value. Specific antithyroid drug therapy is not required in the thyrotoxic phase because symptoms and biochemical thyroid function return to normal in a few weeks. Occasionally, a beta-blocker may be indicated for symptom relief.

The hypothyroid phase usually follows the thyrotoxic phase and should be anticipated with periodic follow-up. An elevated TSH level at the appropriate time postpartum in women who are most likely TPOAb positive should alert clinicians to the diagnosis. This phase lasts longer and is associated with considerable morbidity. It may result in early permanent hypothyroidism, as described previously. Thyroxine therapy is indicated with a trial of withdrawal at 9–12 months. It is possible to withdraw thyroxine therapy in the majority of women at the end of this period, but a recurrence of symptoms associated with increased TSH levels on follow-up indicates the need for permanent replacement therapy.

## LONG-TERM OUTCOME FOLLOWING PPT

The long-term outcome of PPT has been examined in several studies (Table II). In our series, permanent hypothyroidism occurred as early as 9 months postpartum in approximately 30% of subjects who were TPOAb positive and had PPT. These women required thyroxine replacement therapy to maintain normal clinical and biochemical thyroid function. A review of long-term follow-up studies of PPT from geographically different locations reported a 12–61% prevalence of permanent hypothyroidism. The variability may in part be due to differences in the definition of PPT and long-term thyroid dysfunction, length of follow-up, and ascertainment. We followed 98 TPOAb-positive women (of whom 48 developed PPT) and 70 TPOAb-negative controls for 66–140 months. Forty-six percent of women who developed PPT were hypothyroid (some subclinically) at the end of the follow-up period compared to only 4% of women who were TPOAb positive but did not develop PPT and 1.4% of women who were TPOAb negative. The rate of conversion to hypothyroidism in women who were TPOAb positive and developed PPT was 7.1% per year, higher than that reported for women in community-based follow-up studies. Investigators from different areas of the world have confirmed a high prevalence of hypothyroidism at the end of variable follow-up periods in women who developed PPT. In a study of Japanese women with a similar length of follow-up after PPT (mean, 8.7 years), Tachi found a 29% prevalence of permanent hypothyroidism. In a Swedish study, Jansson found a 30% prevalence of hypothyroidism at 5 years. In Brazil, Barca found a 61% prevalence of hypothyroidism at the end of a 2-year

period of follow-up after PPT. The reason for this high prevalence of relatively early hypothyroidism is unclear.

These studies raise the interesting issue of the nature of thyroid damage following the initial episode of PPT. Several investigators have indicated the distinct possibility of a persistent but subtle abnormality (probably autoimmune in nature) of thyroid function and morphology in these women. Iodine perchlorate discharge tests were abnormal in 41% of Italian and 64% of Welsh women studied 3 and 7 years, respectively, following PPT, suggestive of a persistent organification defect. Furthermore, we found a significantly higher prevalence of thyroid ultrasound hypoechogenicity (due to autoimmune destruction) at 4–8 weeks postpartum in women who were TPOAb positive and developed PPT (45%) compared to antibody-positive women who did not develop PPT (17%) and to antibody-negative women (1.5%). There was a significantly higher prevalence of persistent abnormalities in the first group after 66–144 months of follow-up (although mildly reduced from the postpartum period), indicative of persistent autoimmune destruction following the initial episode of PPT. It seems likely that a persistent but low-grade immune destructive process severe enough to produce echogenic changes in the thyroid gland continues to occur in these women, although maximal damage occurs at the time of PPT.

These findings indicate the need for long-term follow-up of women who are TPOAb positive and who develop PPT. A relatively high incidence of early permanent hypothyroidism in these women and a higher than normal annual conversion rate to hypothyroidism (compared to that of TPOAb-positive

Table II Long-Term Follow-Up of PPT

Author	Follow-up (years)	No. of subjects	Thyroid dysfunction (includes subclinical hypothyroidism/abnormal TRH test)
Nikolai	3	25	6
Vargas	1	42	8
Tachi	5–16	44	10
Lervang	2	23	2
Jansson	5	50	13
Othman	2–4	43	10
Solomon	0.9–3.7	55	20
Kuijpers	2.5–3	14	6
Premawardhana	5–11.5	98	24
Lucas	3.3	42	5
Barca	2	49	30

women who do not develop PPT and women from community surveys), indicate the need for long-term follow-up. An increased TSH level with or without symptoms of hypothyroidism on annual (or more frequent) thyroid function testing is an indication for thyroxine replacement therapy.

## SCREENING FOR PPT

There is no consensus about screening for PPT. This relates as much to the absence of a highly sensitive and specific marker for prediction as to the lack of appreciation of the clinical problem and the long-term effects of PPT. The significant morbidity of PPT in the first postpartum year, the likelihood that approximately three-fourths of women who have PPT will have an episode in a future pregnancy, the high prevalence of long-term thyroid dysfunction following PPT, and the availability of effective treatment should make a screening strategy useful. Some authorities recommend a selective screening strategy aimed at only those women who have a high risk of developing PPT (i.e., women with type 1 diabetes mellitus and those who have had PPT in a previous pregnancy). However, a screening strategy should take into account the fact that PPT may occur in TPOAb-negative women. It is salutary to remember that the prevalence of diseases for which antenatal screening is currently recommended is considerably lower than that of PPT in women of childbearing age.

TPOAb has been proposed as a marker for the prediction of PPT. As mentioned previously, when measured in early pregnancy, TPOAb is present in approximately 10% of women. However, only approximately half of these develop PPT, raising questions about the sensitivity of TPOAb as a predictor. Ten studies have examined TPOAb as a predictor of PPT. No firm conclusions can be drawn from these studies for several reasons. Most investigators measured microsomal antibodies, whereas some measured antibodies to TPO (the specific microsomal antigen). The assay methods used and the timing of antibody measurement were variable. Whereas some investigators measured antibodies in the antepartum period, others measured them at delivery and in the postpartum period. These studies, however, showed that thyroid antibodies have a sensitivity of 0.45–0.89 and specificity of 0.91–0.98. Positive predictive value and relative risk were 0.31–0.78 and 20–50.7, respectively. A report confirmed the cost-effectiveness of a screening program. It is not known whether screening

will be improved by thyroglobulin estimation, measuring ultrasound thyroid volume, or assessment of complement activation.

## CONCLUSION

PPT is a common endocrine disorder affecting young women. The exact mechanisms of cellular damage have not been determined, although the autoimmune destructive nature of the disease has long been recognized. The immune perturbations of pregnancy and the postpartum period account for the modulation of thyroid autoimmunity, which is the hallmark of the disease, and the timing and nature of clinical and biochemical changes. Although the majority of patients have a short and self-limited illness, some have a prolonged and symptomatic disorder that requires specific therapy. The recognition of short- and long-term morbidity, and the need for permanent thyroxine supplementation in a significant minority of subjects following PPT, has raised the important but unsettled issue of screening for PPT, targeted perhaps at those at highest risk. We await the discovery of a sensitive and specific screening tool.

## See Also the Following Articles

Depression, Thyroid Function and • Thyroid Autoimmunity • Thyroiditis, Infectious • Thyroiditis, Subacute • Thyrotoxicosis, Overview of Causes

## Further Reading

- Gerstein, H. C. (1990). How common is postpartum thyroiditis? A methodologic overview of the literature. *Arch. Intern. Med.* **150**, 1397–1400.
- Lazarus, J. H., Premawardhana, L. D. K. E., and Parkes, A. B. (1998). Postpartum thyroiditis; 83–97. In "Endocrine Autoimmunity and Associated Conditions" (A. P. Weetman, ed.). Kluwer, London.
- Muller, A. F., Drexhage, H. A., and Berghout, A. (2001). Postpartum thyroiditis and autoimmune thyroiditis in women of childbearing age: Recent insights and consequences for antenatal and postnatal care. *Endocr. Rev.* **22**(5), 605–630.
- Premawardhana, L. D., Parkes, A. B., Ammari, F., John, R., Darke, C., Adams, H., and Lazarus, J. H. (2000). Postpartum thyroiditis and long term thyroid status: Prognostic influence of thyroid peroxidase antibodies and ultrasound echogenicity. *J. Clin. Endocrinol. Metab.* **85**, 71–75.
- Smallridge, R. C. (1999). Postpartum thyroid disease through the ages. *Thyroid* **9**, 671–673.
- Smallridge, R. C. (2000). Postpartum thyroid disease: A model of immunologic dysfunction. *Clin. Appl. Immunol. Rev.* **1**, 89–103.

**Table 1** Differential Diagnosis of Subacute Thyroiditis

Subacute thyroiditis
Acute hemorrhage into cyst or nodule
Infected thyroglossal duct cyst
Infected branchial cleft cyst
Acute pyogenic thyroiditis
Cellulitis of the anterior neck
Rapidly enlarging thyroid cancer
Painful Hashimoto's thyroiditis
Cervical adenitis
Radiation thyroiditis
Infected cystic hygroma
Rare infections (e.g., <i>Pneumocystis carinii</i> ) or other inflammatory disorders (e.g., amyloidosis)
Globus hystericus

all of whom were positive for HLA-Bw35. Two of the patients lived near each other but developed the disorder 1 year apart, and the third sibling lived several hundred miles away and had not been in contact with either of the other two when they had subacute thyroiditis. Thus, HLA-Bw35 appears to render individuals genetically susceptible to the development of the disorder.

An interesting cluster of subacute thyroiditis was reported in 12 patients in a single town in The Netherlands. Affected individuals bore clinical similarities to subacute thyroiditis, although only 1 patient was positive for HLA-Bw35, much less than would be expected. However, 5 of 11 patients tested positive for HLA-B15/62, a markedly greater frequency than expected. This suggests both genetic and clinical heterogeneity in subacute thyroiditis.

Although thyroid autoimmunity is not believed to play a role in subacute thyroiditis, autoimmune abnormalities associated with the disorder have been described. Antibodies directed against the thyroid-stimulating hormone (TSH) receptor have been reported. Also, a report of sensitization of T lymphocytes against thyroid antigen suggests the possibility of an autoimmune component, although it is believed that this is likely the result of released antigen during the active inflammation phase. Low titers of thyroid autoantibodies are sometimes detectable in patients with subacute thyroiditis.

## CLINICAL FEATURES

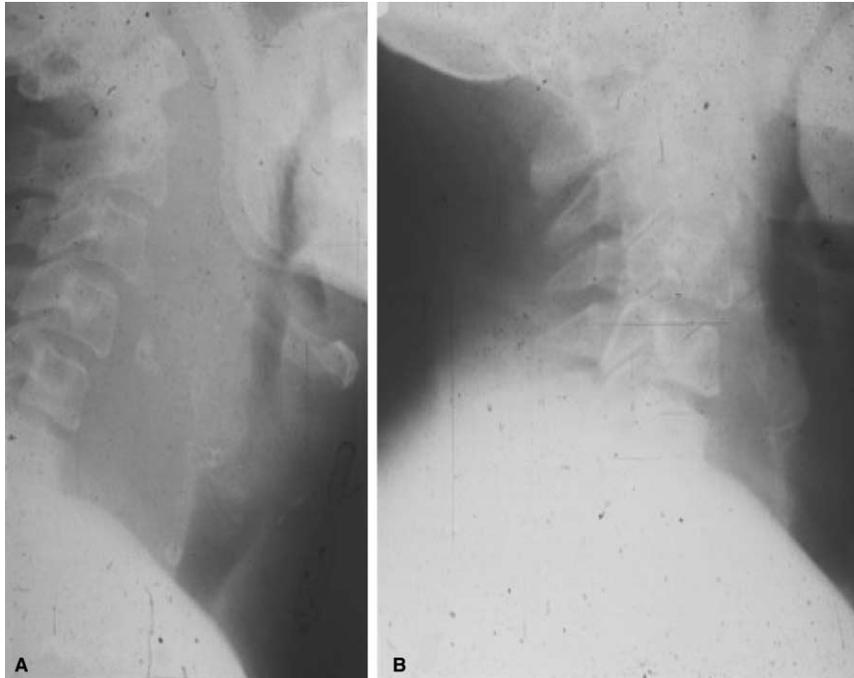
Subacute thyroiditis is more common in women (80% of cases) between the ages of 40 and 50 years. As mentioned previously, a viral prodrome is common,

as are symptoms of sore throat, weakness, low-grade fever, myalgias, and, frequently, dysphagia. Symptoms may develop gradually over a few weeks, although patients frequently describe the disorder as coming on within a few days. Anterior neck pain is usual and is generally more significant on one side. It frequently radiates to the ear, the mandible, the occiput, or even the upper chest. Commonly, the patient initially consults an ear, nose, and throat physician or a dentist because of pain in the area of the throat or mandible. Frequently, the pain will migrate from one side to the opposite thyroid lobe after a few weeks. Symptoms of hypermetabolism are frequent since thyrotoxicosis occurs in approximately 50% of affected individuals; symptoms may include diaphoresis, palpitations, tachycardia, and weight loss. Rarely, the clinical presentation of subacute thyroiditis may be so dramatic in onset and so pronounced in severity that obstructive symptoms may develop (Fig. 1).

Physical examination may disclose signs of hypermetabolism, with tachycardia, diaphoresis, and tremor. Palpation of the neck generally reveals a very firm to hard, exquisitely tender, ill-defined mass in one lobe of the thyroid gland, although there is frequently thickening and tenderness of the contralateral lobe as well. Tenderness may be so pronounced that patients may try to prevent palpation by the examiner. The overlying skin is occasionally erythematous.

## DIFFERENTIAL DIAGNOSIS

The differential diagnosis of subacute thyroiditis is shown in Table 1 and includes both thyroid and non-thyroid disorders. As mentioned previously, it is the most common thyroid etiology of anterior neck pain, with hemorrhage into a cyst or other benign nodule the second most common thyroidal cause of neck pain. Clinically, the two can readily be distinguished from each other since hemorrhage into a nodule is not accompanied by either viral symptoms or thyrotoxicosis. Moreover, hemorrhage is generally sudden in onset, and the mass tends to be very smooth on palpation. In addition, the radioactive iodine uptake in patients with hemorrhage is normal, and a radionuclide scan would reveal a filling defect. Pyogenic thyroiditis is characterized by features of an abscess, with a tender, palpable, fluctuant mass. In addition, the patient would likely have a high fever and a significantly higher white blood cell count than would the patient with subacute thyroiditis, and a radionuclide scan would reveal similar radionuclide findings as those for the patient with acute hemorrhage into a cyst. Patients with pyogenic thyroiditis are not thyrotoxic, nor are those with



**Figure 1** A 52-year-old man with severe subacute granulomatous thyroiditis with threatened airway obstruction. Note the pronounced retropharyngeal edema (A) and its dramatic resolution following intravenous administration of methylprednisolone (B). From Singer, P. A. (1991). Thyroiditis: Acute, subacute, and chronic. *Med. Clin. North Am.* 75, 61–77.

cellulitis of the anterior neck. Also, patients with cellulitis have features of infection and do not have discreet masses on palpation. The patient with an infected thyroglossal duct cyst or branchial cleft cyst may present with a painful mass, which may be fluctuant. Such lesions also may be distinguishable from subacute thyroiditis by virtue of their more superior (thyroglossal duct cyst) or lateral (branchial cleft cyst) locations. In addition, patients with such lesions are euthyroid.

Rapidly growing thyroid cancer is an unusual cause of anterior neck pain, but when it occurs patients complain more of an achy-type pain, often accompanied by symptoms of tracheo-esophageal pressure. Patients do not have features of acute inflammation, and on palpation the mass is nontender and usually rock hard. Painful Hashimoto's thyroiditis is also very uncommon and is accompanied by elevated titers of thyroid antibodies and, frequently, hypothyroidism.

Radioactive iodine administration, either for treatment of thyrotoxic Graves' disease or for remnant ablation in patients with differentiated thyroid cancer, may cause anterior neck pain and thyroid tenderness. The etiology of the inflammation is always obvious.

Several years ago, infection of the thyroid with *Pneumocystis carinii* was confused with subacute thyroiditis. We reported on three patients with biopsy-proven *P. carinii* infection who presented with anterior neck

pain and suppressed radioiodine uptake values; two of the patients were thyrotoxic. Other workers also reported similar cases at approximately the same time. What made most of the cases unique was the history of using inhaled aerosolized pentamidine for pneumocystis pneumonia prophylaxis, which may well have resulted in the pneumocystis organism seeking another primary target. Since pentamidine is no longer used for prophylaxis, no further cases of pneumocystis thyroiditis have been reported.

Thyroid amyloidosis associated with systemic amyloid has been reported, with tender goiters, low thyroid radioactive iodine uptake values, and elevated erythrocyte sedimentation rates but without thyrotoxicosis.

## LABORATORY FINDINGS

Biochemical thyrotoxicosis occurs in approximately half of patients with subacute thyroiditis. Serum  $T_4$  (free  $T_4$ ) and  $T_3$  (free  $T_3$ ) levels are elevated and serum TSH concentrations are suppressed. There is a disproportionate elevation of  $T_4$  relative to  $T_3$  since serum levels of thyroid hormones are due to "dumping" of preformed hormone into the circulation and therefore reflect intrathyroidal  $T_4$  and  $T_3$  content. This is in contrast to the thyroid hormone levels in

patients with Graves' disease, in which there is a disproportionate increase of serum  $T_3$  relative to serum  $T_4$ . In addition, impaired peripheral conversion of  $T_4$  to  $T_3$  due to the illness may contribute to the relatively lower  $T_3$  concentrations. In general, the thyrotoxicosis associated with subacute thyroiditis is mild or at most modest in its severity.

Radioactive iodine uptake is always suppressed, usually to less than 2% after 24 h. Indeed, if the radioactive iodine uptake is more than 5% after 24 h, the diagnosis of subacute thyroiditis is unlikely. The absent uptake of iodine by the thyroid is due to destruction of the iodine-trapping mechanism from the inflammatory process as well as from inhibition of TSH secretion by excess circulating thyroid hormone. It is important to perform a radioactive iodine uptake test in patients with suspected subacute thyroiditis in order to rule out other causes of anterior neck pain as well as other types of thyrotoxicosis.

The erythrocyte sedimentation rate is almost always  $>50$  mm/h. A normal sedimentation rate in a patient with a painful anterior neck mass places the diagnosis of subacute thyroiditis in question. A mild normochromic, normocytic anemia is frequently present, as is a slightly elevated total white blood cell count.

The serum thyroglobulin concentration is always elevated during the acute phase of subacute thyroiditis, reflecting the destruction of the thyroid follicular architecture and release of thyroglobulin into the circulation. However, the serum thyroglobulin is not recommended as a routine test in the evaluation of patients with suspected subacute thyroiditis since it is elevated in virtually all other thyroid disorders as well. It may be a helpful test when anterior neck discomfort is present, and when the diagnosis of subacute thyroiditis is in question, a normal serum thyroglobulin would cast serious doubt on the diagnosis.

Fine needle aspiration is not routinely recommended in the evaluation of suspected subacute thyroiditis but may be employed in patients in whom the diagnosis of a painful neck mass is not clear. The principal disorders to exclude in such circumstances include acute pyogenic thyroiditis and hemorrhage into either a malignant or benign thyroid nodule. Fine needle aspiration in subacute thyroiditis shows giant cells and pseudogranulomas.

Ultrasonography has been employed as an adjunct in the diagnosis of subacute thyroiditis; findings show diffuse areas of hypoechogenicity. However, ultrasound is not routinely used in the evaluation of patients with suspected subacute thyroiditis. Table II summarizes typical clinical and laboratory manifestations of subacute thyroiditis.

**Table II Clinical and Laboratory Characteristics of Subacute Thyroiditis**

<b>Symptoms</b>
Viral-type prodrome (fever, myalgias, sore throat)
Dysphagia
Anterior neck pain (frequently with radiation)
Symptoms of hypermetabolism (palpitations, tachycardia, weight loss)
<b>Signs</b>
Fever
Tachycardia
Tender, hard thyroid mass
<b>Laboratory findings<sup>a</sup></b>
Serum free $T_4$ elevated
Serum TSH suppressed
Erythrocyte sedimentation rate elevated
Thyroidal radioactive iodine uptake suppressed
Serum thyroglobulin elevated

<sup>a</sup>All of the tests on this list, except the thyroglobulin, should be included in the routine evaluation of subacute thyroiditis. The thyroglobulin is helpful when it is unclear if the neck pain is thyroid in origin.

## TREATMENT AND CLINICAL COURSE

Treatment of subacute thyroiditis is directed toward relief of pain and inflammation and control of thyrotoxic symptoms. Salicylates and nonsteroidal anti-inflammatory agents have been advocated by some physicians as preferred treatment, although they tend to be effective in only the mildest cases of the disorder. Most thyroid specialists recommend the use of glucocorticoids, which are effective at relieving pain within hours after oral administration. A divided dose of 30–60 mg of prednisone daily usually suffices. Lack of significant improvement within 24 h after initiation of steroids is uncommon since dramatic relief of pain usually occurs within hours. Indeed, the absence of rapid improvement would call into question the original diagnosis. The prednisone can begin to be tapered after approximately 1 week and discontinued within 3 or 4 weeks. Pain and swelling recur, often in the contralateral lobe, in approximately 20% of patients. When this occurs, prednisone should be resumed but with a lower dose. Approximately half of the original starting dose usually suffices. The tapering process can then begin anew.

Thyrotoxic symptoms may be controlled by the use of beta-blocking agents, with the dose depending on severity of symptoms. Propranolol in a dose of 10–40 mg three or four times per day or atenolol in a dose of 25–50 mg once daily is usually sufficient to control the hypermetabolic symptoms.

Thionamide agents are of no use in the management of subacute thyroiditis and are not recommended since the thyrotoxicosis results from release of preformed hormone rather than from increased synthesis. Therefore, drugs that inhibit thyroid hormone synthesis, such as thyroid-blocking agents, have no beneficial effect.

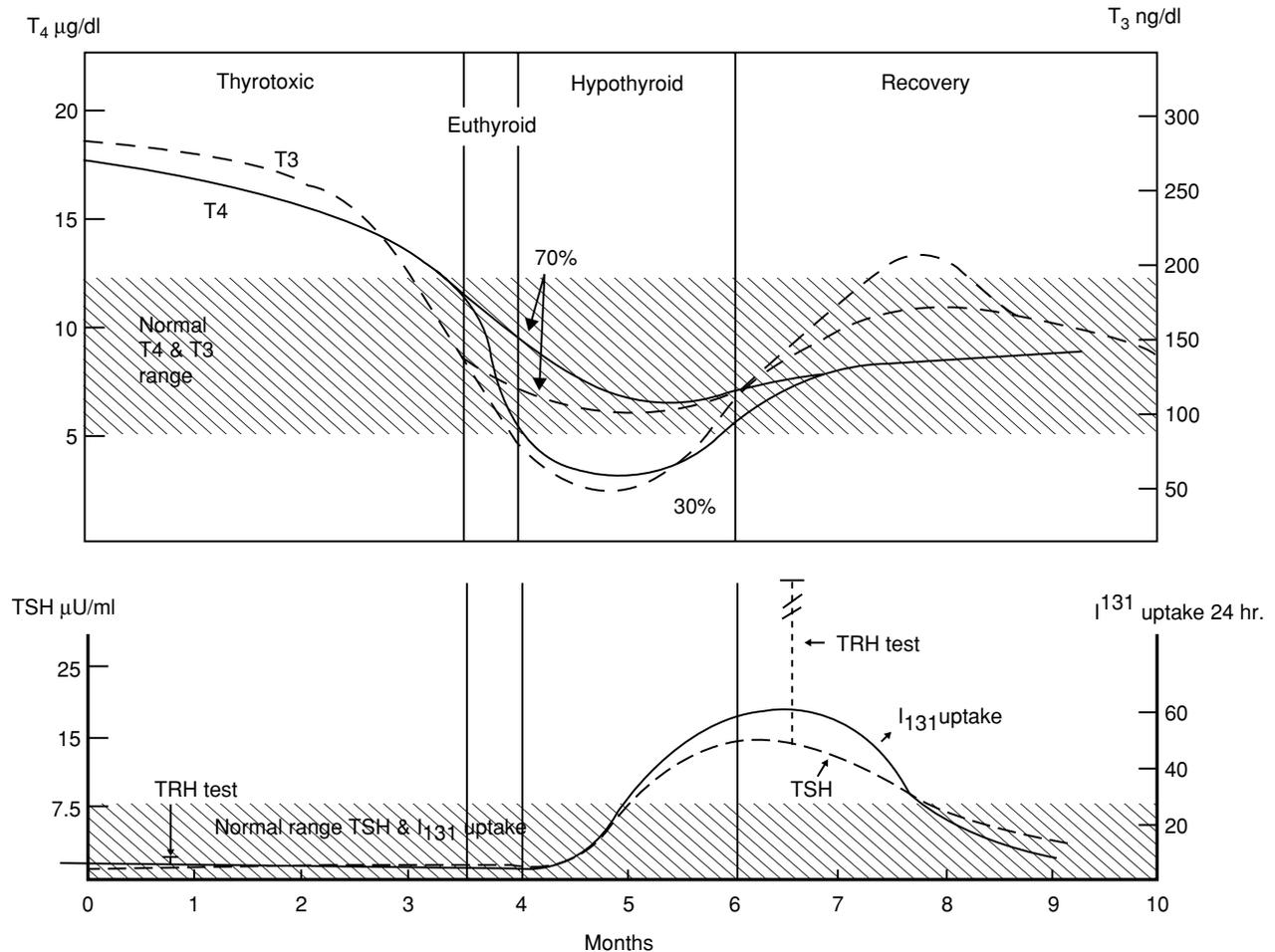
Sodium ipodate has been shown to be effective in correcting thyrotoxicosis more rapidly, although few physicians have reported using it. Cases have been reported in which radioactive iodine ablation, or even thyroidectomy, has been employed for patients with prolonged, disabling pain, but such circumstances are rare.

The clinical course of subacute thyroiditis generally follows four phases (Fig. 2). The initial or "acute" phase is characterized by pain, tenderness, and thyrotoxicosis, and it may last 2–12 weeks. It is during this phase, called

the thyrotoxic phase, that treatment with glucocorticoids is necessary. Following the thyrotoxic phase, a several-week interval of euthyroidism occurs. A hypothyroid period then follows, which may last from a few weeks to a few months. During the hypothyroid phase, levo-thyroxine therapy may be necessary. An asymptomatic recovery phase then follows, during which the thyroid is restored to normal function.

Not all patients with subacute thyroiditis progress through all four phases of the disorder since only approximately 50% of patients develop transient hypothyroidism. In most cases, the entire episode of subacute thyroiditis lasts no more than 6 months.

Late recurrences of subacute thyroiditis are uncommon but may occur, even years later. Typically, a repeat bout of subacute thyroiditis is milder than the original occurrence.



**Figure 2** Representation of changes observed during a typical bout of subacute thyroiditis. From Emerson, C. H., and Farwell, A. P. (2000). Sporadic silent thyroiditis, postpartum thyroiditis, and subacute thyroiditis. In "Werner & Ingbar's The Thyroid: A Fundamental and Clinical Text" (L. E. Braverman and R. D. Utiger, eds.), 8th ed., pp. 578–589. Lippincott Williams & Wilkins, Philadelphia.

Virtually all patients with subacute thyroiditis achieve restoration of normal thyroid function, although permanent hypothyroidism has been reported in 1–5% of patients. Moreover, it has been shown that patients who have suffered a bout of subacute thyroiditis may have subtle, permanent thyroid abnormalities. For example, patients have been shown to be sensitive to the inhibitory effects of exogenously administered iodides by exhibiting elevations in serum TSH concentrations, even years after having had subacute thyroiditis. Thus, perhaps patients with a history of subacute thyroiditis should be screened with serum TSH levels a few weeks after receiving exogenous iodides in pharmacologic quantities.

### Acknowledgment

I gratefully acknowledge the expert secretarial assistance of Elsa Ahumada.

### See Also the Following Articles

Thyroid Autoimmunity • Thyroid Disease, Genetic Factors in • Thyroid Fine Needle Aspiration Cytology • Thyroiditis, Infectious • Thyroiditis, Postpartum • Thyrotoxicosis, Overview of Causes

### Further Reading

De Bruin, T. W. A., Riekhoff, F. P. M., and DeBoer, J. J. (1990). An outbreak of thyrotoxicosis due to atypical subacute thyroiditis. *J. Clin. Endocrinol. Metab.* **70**, 396–402.

Emerson, C. H., and Farwell, A. P. (2000). Sporadic silent thyroiditis, postpartum thyroiditis, and subacute thyroiditis. *In*

“Werner & Ingbar’s The Thyroid: A Fundamental and Clinical Text” (L. E. Braverman and R. D. Utiger, eds.), 8th ed., pp. 578–589. Lippincott Williams & Wilkins, Philadelphia.

Farwell, A. P., and Braverman, L. E. (1996). Inflammatory thyroid disorders. *Otolaryngol. Clin. North Am.* **29**, 541.

Guttler, R., Singer, P. A., Axline, S. G., Greaves, T. S., and McGill, J. J. (1993). *Pneumocystis carinii* thyroiditis. *Arch. Intern. Med.* **153**, 393–396.

Iitaka, M., Momotani, N., Ishil, J., and Ito, K. (1996). Incidence of subacute thyroiditis recurrences after a prolonged latency: 24 year survey. *J. Clin. Endocrinol. Metab.* **81**, 466.

Kitaoka, H., Sakurada, T., Fukazawa, H., et al. (1985). An epidemiological and subacute thyroiditis in northern Japan. *Nippon Naibumpi Gakk. Zasshi* **61**, 554.

Ohsako, N., Tamai, H., Sudo, T., et al. (1995). Clinical characteristics of subacute thyroiditis classified according to human leukocyte antigen typing. *J. Clin. Endocrinol. Metab.* **80**, 3653.

Roti, E., Minelli, R., Gardini, L., et al. (1990). Iodine-induced hypothyroidism in euthyroid subjects with a previous episode of subacute thyroiditis. *J. Clin. Endocrinol. Metab.* **70**, 1581–1585.

Rubin, R. A., and Guay, A. T. (1991). Susceptibility to subacute thyroiditis is genetically influenced: Familial occurrence in identical twins. *Thyroid* **1**, 157.

Singer, P. A. (1991). Thyroiditis: Acute, subacute, and chronic. *Med. Clin. North Am.* **75**, 61–77.

Smallridge, R. C., De Keyser, F. M., Van Herle, A. J., et al. (1986). Thyroid iodine content and serum thyroglobulin: Cues to the natural history of destruction-induced thyroiditis. *J. Clin. Endocrinol. Metab.* **62**, 1213.

Volpe, R. (1981). Subacute thyroiditis [Review]. *Prog. Clin. Biol. Res.* **74**, 115.

Volpe, R. (1999). The management of subacute (de Quervain’s) thyroiditis [Review]. *Thyroid* **3**, 253.

Volpe, R., Row, V. V., and Ezrin, C. (1967). Circulating viral and thyroid antibodies in subacute thyroiditis. *J. Clin. Endocrinol. Metab.* **27**, 1275–1284.

Yamamoto, M., Saito, S., Sakurada, T., et al. (1987). Effect of prednisone and salicylate on serum thyroglobulin level in patients with subacute thyroiditis. *Clin. Endocrinol.* **27**, 339.

phosphatase (ALP), osteocalcin, and osteopontin. Thyroid hormone receptors are central to conferring  $T_3$  responsiveness to cells by binding to target genes either as homodimers or as heterodimers complexed with the *cis*-acting factor 9-*cis*RXR. Transcriptional regulation by thyroid hormones is mediated by ligand-dependent transcription factors called TRs (e.g., TR $\alpha$ 1, TR $\beta$ 1, TR $\beta$ 2). In osteoblasts, TRs are coexpressed with the *cis*- and *trans*-acting factors 9-*cis*RXR (RXR) and all-*trans* RAR (RAR) where they modify the regulation of endogenous gene expression (e.g., ALP, osteocalcin, osteopontin) by  $T_3$ .

The cellular osteoblast-like lineage hOb displays lower cytokine secretion, reduced immunostaining of TR- and  $T_3$ -binding sites, and decreased thyroid receptor function than does another osteoblast-like cell line, BMS; therefore, it seems that hOb cells play a lesser role and BMS cells play a greater role than previously envisaged in  $T_3$  regulation of bone remodeling. Although it is premature to extend these observations to the situation *in vivo*, this nonetheless highlights the potential importance of human bone marrow cells in future studies of  $T_3$  action on bone.

## Osteoblast/Osteoclast Function

### **Interference of Thyroid Hormones in Osteoblast/Osteoclast Function**

Thyroid hormones increase osteoclastic bone resorption by acting indirectly on osteoblasts; a direct response of osteoclasts to  $T_3$  is disputed, and most actions of  $T_3$  in bone are thought to be mediated via osteoblasts. Osteoclasts are the primary cell type that can resorb bone. They are highly motile, giant, multinucleated cells derived from hemopoietic tissue. The osteoclast precursor cells are closely related to cells of monocyte macrophage lineage. The formation of osteoclasts takes place only in the close vicinity of mineralized bone, and the multinucleated osteoclasts never appear in the circulation. The activity of osteoclasts is essential for the physiological resorption of bone during skeletal remodeling as well as for the maintenance of calcium homeostasis. Enhanced osteoclast activity is apparent in the excessive bone loss seen in a variety of pathological conditions such as thyrotoxicosis. Studies on the expression of the receptors for the stimulators of bone resorption, parathyroid hormone (PTH), and calcitriol [1,25(OH) $_2$  D $_3$ , vitamin D $_3$ ] reveal, surprisingly, that they are not present on osteoclasts or on their precursor cells but rather are expressed in bone-forming osteoblasts. The osteoblast/stromal cell ratio is crucial for the development and differentiation of the osteoclast via

a mechanism involving cell-to-cell contact. The osteoblastic stromal cells produce osteoclast-differentiating factors in response to a variety of stimuli that are essential for the formation of mature osteoclasts. One of these factors has been identified as RANKL (receptor activator of nuclear factor kappa B [NF- $\kappa$ B] ligand), which is a membrane protein expressed in the osteoblastic stromal cells and belongs to the tumor necrosis factor (TNF) family of growth factors. Osteoclast progenitors express RANK, the receptor for RANKL, at their cell surfaces, and the interaction between RANK and RANKL stimulates the development and differentiation of osteoclasts.  $T_3$  induces the expression of RANKL mRNA in primary osteoblastic cells (POB). This effect is amplified when cells are costimulated with calcitriol.

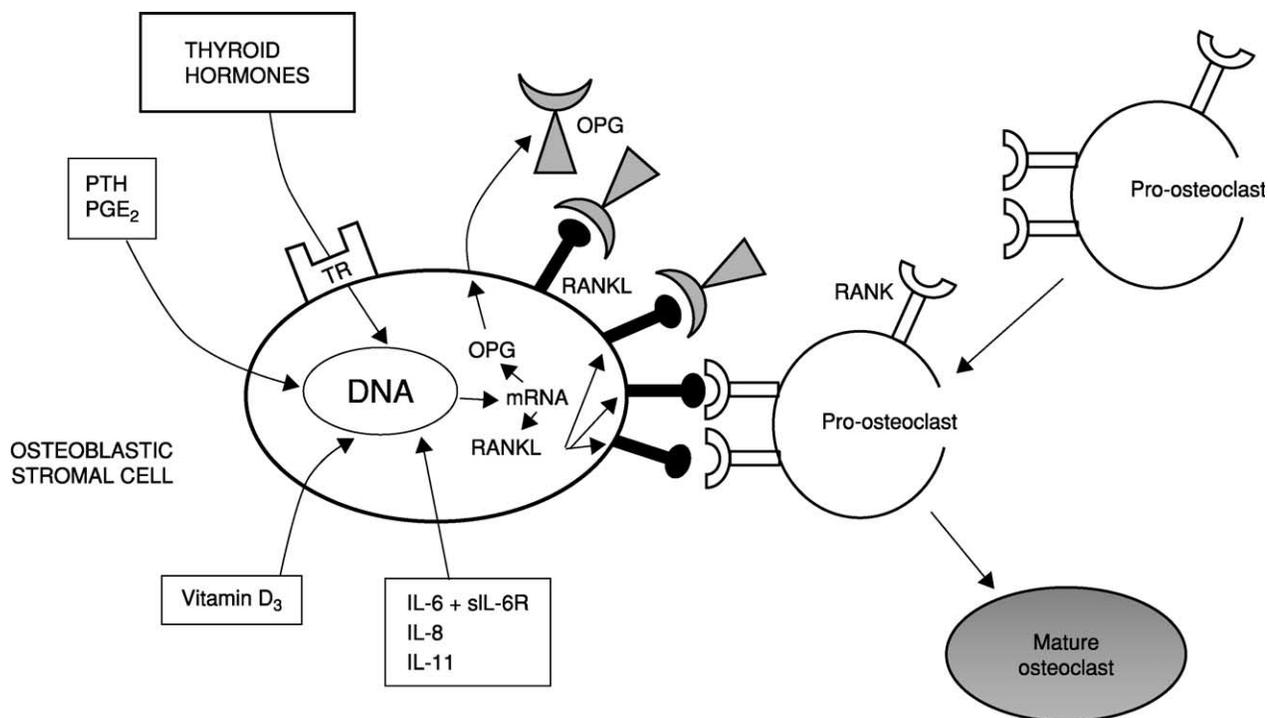
In addition to RANKL, the osteoblastic stromal cells produce osteoprotegerin, (OPG), which serves as a decoy receptor for RANKL, inhibiting osteoclastogenesis by preventing the interaction between the osteoblastic stromal cell and the osteoclast progenitor (Fig. 1).

The regulation of osteoclast apoptosis is a further mechanism by which osteoclast activity can be controlled. In fact, many resorptive agents are able to alter osteoclast life span and/or apoptotic rate. Osteoclast viability is increased by agents that stimulate resorption (e.g., macrophage colony-stimulating factor [M-CSF], interleukin-1 [IL-1]) that are produced by monocyte macrophage cell lines and also by the effect of thyroid hormones.

In conclusion, thyroid hormones, like other resorption agents, stimulate osteoclast activity in cocultures with osteoblasts but do not stimulate activity in highly purified osteoclast preparations. The primary effect of thyroid hormones is to stimulate osteoblastic production of downstream effectors that activate the osteoclast. These downstream effectors can act either in a paracrine fashion to directly activate osteoclast activity or in an autocrine fashion to further stimulate osteoblasts to produce the paracrine factors that directly activate the osteoclasts.

### **Local Bone Marrow Factors (Cytokines) Modulating Bone Remodeling in Thyrotoxic Bone Disease**

The release of cytokines by osteoblasts provides a mechanism by which osteoblasts could mediate the action of  $T_3$  on osteoclast resorption. In the microenvironment of the bone marrow, the cytokines IL1, M-CSF, and TNF can stimulate bone formation and resorption and play a critical role in regulating osteoclast formation and activity. Furthermore, it has been



**Figure 1** Schematic representation of the role of the RANK/RANKL/OPG system in osteoclast development. RANKL is expressed on the surface of the osteoblastic stromal cell and serves as a ligand for RANK, which is expressed on the surface of the osteoclast progenitor cell. The interaction between RANK and RANKL triggers the development of the mature osteoclast. OPG, produced by the osteoblastic stromal cell, serves as a decoy receptor for RANKL and inhibits osteoclastogenesis, preventing the interaction between the osteoblastic stromal cell and the osteoclast progenitor.

shown that cell cultures of human bone marrow stromal cells, containing osteoblast progenitor cells, release IL-6 and IL-8 in response to  $T_3$ . The former regulates osteoclast proliferation and recruitment, whereas the latter regulates osteoclast development and activity. In addition, IL-8 receptors have been identified in osteoclast cells. Other cytokines induced by thyroid hormones are also implicated in high-turnover bone resorption. Serum IL-6 and IL-8 are produced by a number of sources, including blood monocytes and bone tissue. IL-6 mRNA is present in thyroid follicles, but high levels of cytokines, independent of the thyroid gland, are also reported in patients with ablated thyroids. The soluble serum receptor for IL-6 (sIL-6R) regulates the biological activity of IL-6, and the levels of this receptor correlate well with those of serum thyroid hormones and may prove to be a better determinant of thyrotoxic bone resorption.

High levels of serum IL-6 have been demonstrated previously in thyrotoxicosis, including Graves' disease, and also in other conditions unrelated to autoimmunity such as toxic nodular goiter and iatrogenic

subclinical thyroid excess. The elevations in serum IL-6 and IL-8 in thyrotoxicosis seem to result from the chronic effect of thyroid hormone excess rather than from the accompanying autoimmune inflammatory response produced by Graves' thyroid or eye disease.

In hyperthyroidism, as in osteoporosis from estrogen deficiency, there is an inverse relationship between the IL-6 concentration and BMC values as well as between cytokines and deoxypyridinoline (Udpd) excretion. It is possible that the same mechanism, mediated by IL-6, induces osteopenia in two conditions with high bone turnover as in postmenopausal osteoporosis and thyrotoxicosis.

Serum levels of IL-6 and IL-8, as well as markers of bone remodeling, are elevated in untreated thyrotoxicosis but fall as thyroid hormone levels normalize with treatment. The normalization of Udpd precedes that of serum cytokine concentrations; although many studies have not found a correlation between serum cytokines and Udpd, the interaction of  $T_3$  with IL-6 and other cytokines should continue to be a focus in future investigations.

Evidence of a role for cytokines in  $T_3$ -associated bone resorption is lacking; this is due partly to an interference of their effects by a variety of other growth systemic factors such as PTH, calcitriol, and prostaglandin  $E_2$  ( $PGE_2$ ).

### **Systemic Factors Regulating Osteoblast/Osteoclast Function in Thyrotoxic Bone Disease**

Thyroid hormones, at concentrations approaching those that occur in thyrotoxicosis, also stimulate bone resorption and increase sensitivity to PTH. This causes a marked increase in bone resorption in bone organ cultures and stimulates osteoclast formation in both murine and human marrow cell cultures. Many of the calciotropic hormones and cytokines appear to act through a dual capacity to inhibit production of OPG and to stimulate the RANK system. The RANKL/RANK interaction is essential for the effect of PTH, calcitriol, and  $PGE_2$ . This is demonstrated by the ability of anti-RANKL antibodies to inhibit the bone resorption activity of these known inducers of bone resorption; in contrast, estrogen appears to inhibit RANKL production.

Studies have demonstrated that PTH regulates systemic levels of IL-6 in experimental animals. Furthermore, IL-6 has been shown to be an important mediator of the bone-resorbing activity of PTH *in vivo* and also plays a role in coupling PTH-induced bone resorption; IL-6 can stimulate proliferation of early osteoblast precursors, and PTH induces the differentiation and fusion of the precursors to form multinucleated osteoclasts.

The treatment of bone cell cultures with thyroid hormones produces a gradual increase in the concentration of  $PGE_2$ . This effect is abolished by indomethacine, which also reduces thyroid hormone-induced bone resorption.

In mouse bone cell cultures, the effect of calcitriol on the IL-6-dependent formation of osteoclast-like cells is also subject to modulation by  $T_3$ . The mechanism of interaction of these two hormones appears to involve the joint stimulation of the prostaglandin system.  $T_3$  alone does not induce osteoclast formation in cocultures of marrow cells with POB; rather, it enhances the calcitriol-induced osteoclast formation. In addition, calcitriol induces the expression of deiodinase, an enzyme that converts the prohormone 3,5,3',5'-tetraiodothyronine ( $T_4$ ) into its active form ( $T_3$ ). These data facilitate an understanding of the mechanism of osteoclast formation and suggest a novel interaction between thyroid hormones and calcitriol.

The addition of calcitriol to unfractionated bone cells produces a dose-dependent increase in osteoblast survival. Calcitriol also acts indirectly on osteoclasts via the production of IL-1 and IL-6 by osteoblasts, stimulating osteoclast resorption.

## **THYROTOXIC BONE DISEASE**

### **Biochemical Markers of Bone Remodeling in Thyrotoxic Bone Disease**

Thyroid hormones stimulate both osteogenesis and osteolysis to induce an acceleration in bone remodeling. Early biochemical studies attempted to quantify the calcium balance in hyperthyroidism. Mean serum calcium and phosphorus concentrations were found to be higher than those in normal controls. The relative hypercalcemia reduces serum PTH and calcitriol levels, both of which are negatively correlated with free  $T_4$  ( $FT_4$ ), but these levels are normalized when the hyperthyroidism is treated.

Biochemical markers of bone formation and bone resorption, such as osteocalcin, ALP, bone isoenzyme, and urinary collagen pyridinoline (Upyr) or Udpd cross-links, are elevated in thyrotoxic patients and indicate increased bone formation and osteoclastic bone resorption. The binding of  $T_3$  to its nuclear receptor in osteoblasts directly stimulates the osteoblasts to produce ALP, osteocalcin, and the propeptide of type I collagen. Furthermore, osteoblast activity mediates the  $T_3$  activation of osteoclasts, leading to bone resorption and the release of markers such as collagen pyridinoline and Udpd cross-links.

Total ALP and osteocalcin are elevated in 30% and 65 to 90% of thyrotoxic patients, respectively. Early studies found a significant correlation between ALP and  $T_4$  serum levels and osteoid volume. Conflicting data have been published on the potential correlation between osteocalcin and bone ALP and free  $T_3$  ( $FT_3$ ), so the precise relationship of osteocalcin to bone ALP remains unclear.

Novel markers related to collagen have been established as specific markers of bone resorption: the non-reducible cross-links of mature collagen, serum and urinary pyridinoline and deoxypyridinoline (Udpd/creatinine), and serum carboxyterminal telopeptide type I collagen. Urinary pyridinoline levels are raised in 99% of thyrotoxic patients but are less specific for bone than are urinary Udpd cross-links. Udpd accurately indicates thyrotoxic and subclinical advanced thyrotoxic bone resorption. Humoral markers of bone metabolism, such as bone ALP, osteocalcin, and Udpd, are more important to judge the early

effects of antithyroid treatment than is BMD, whose changes occur too slowly.

It is important to predict which patients are likely to suffer a long-term deficit in bone mineral mass so as to increase the efficacy of the therapy. Longitudinal and cross-sectional studies confirm that bone turnover is balanced within 2 to 3 weeks of therapy, and after 4 to 8 weeks of euthyroidism, a peak in the level of ALP occurs and Udpd levels return to within the normal range. However, continued increased serum ALP levels after 1 year of therapy may indicate continuing bone formation and, therefore, provide a marker for low bone mass, even in the presence of normalized Udpd excretion levels. This should be related to high levels of PTH; prior to treatment, endogenous PTH is slightly suppressed but increases significantly during the first year of therapy. The elevation of PTH seems to play a role in maintaining the high bone turnover rate despite the euthyroidism, conserving the bone calcium levels mediated by calcitriol, and increasing bone formation by an anabolic effect on osteoblasts.

Patients with Graves' disease exhibit suppressed serum thyroid-stimulating hormone (TSH) levels and often display elevated levels of TSH receptor antibody (TRAB). Cross-sectional studies indicate that markers of bone metabolism (e.g., ALP, Udpd, Upyr) are more strongly correlated with TRAB than with TSH in Graves' clinical and subclinical patients as well as in Graves' patients with normal TSH levels. These data could support the clinical usefulness of TRAB as a marker of bone metabolism in Graves' patients. Because TRAB correlates with neither FT<sub>3</sub> nor FT<sub>4</sub> but is closely correlated with biochemical markers of bone metabolism, TRAB might directly affect bone metabolism independently of thyroid function. This hypothesis is supported by a recent study demonstrating that osteoblasts possess functional TSH receptors.

### Bone Histomorphometry in Thyrotoxic Bone Disease

After the first report by von Recklinghausen, the histological bone changes in thyrotoxicosis were described as being similar to those of osteite fibrosa, osteoporosis, and osteomalacia.

More recently, histomorphometric analysis has shown that the bone hyperthyroid changes occurring in thyrotoxicosis are, in fact, specific and characterized by increased turnover in trabecular bone and increased remodeling and porosity in cortical bone. The predominant changes are related to bone

destruction where the osteoclast activity is increased and the osteolytic activity induces and surpasses the osteogenic activity.

In Graves' bone disease, bone mineralization is initially slightly greater than bone formation; however, subsequently, the equilibrium between the increase in bone mineralization and bone formation is restored. Osteoid production can even exceed the rate of mineralization; consequently, an osteomalacia-like, osteoid-rich pattern can be observed on histological analysis of bone biopsy specimens.

Because the rate of bone formation does not always equal that of bone resorption, trabecular thickness may decrease, giving rise to trabecular perforations and higher cortical porosity.

In conclusion, the thyrotoxic histomorphometric pattern is unique and characterized by increased osteoblast and osteoclast activity, giving rise to a net loss of bone volume. These changes are present in both cortical and trabecular bone, although they are more evident in cortical bone, whereas a reduction of the absolute bone volume occurs less often in cortical bone.

### Bone Mineral Density Markers in Thyrotoxic Bone Disease

Between 1970 and 1980, BMD was measured only in peripheral sites in patients with hyperthyroidism. However, a newer and more superior method, dual energy X-ray absorptiometry (DEXA), has been developed for evaluating bone mass, where measurements are taken in lumbar vertebrae or femoral neck, typical sites of osteoporosis-related morbidity. Longitudinal studies of thyrotoxic patients demonstrate that even a slight excess of thyroid hormones, regardless of the cause, is associated with decreased BMD in cortical and trabecular bone. In thyrotoxic patients, the baseline percentage BMD of vertebrae is 92.6% as compared with that of normal controls ( $0.91 \pm 0.03 \text{ g/cm}^2$  and  $0.85 \pm 0.2 \text{ g/cm}^2$ , respectively). After 12 months of antithyroid therapy, the mean lumbar spine BMD increases from an initial value of  $1.01 \text{ g/cm}^2$  to  $1.07 \text{ g/cm}^2$  and increases by 6.6% per year ( $P < 0.001$ ). The femoral neck BMC increases by 1.2% per year, and increases in femoral trochanter bone mineral of 3.2% per year have also been documented. No significant difference is observed between the BMDs of male and female patients. A significant reduction in BMD is seen in postmenopausal women affected by thyrotoxicosis or long-term treatment with L-T<sub>4</sub> suppressive therapy. Some authors, but

not others, have found a significant correlation between markers of bone formation and BMD. Conflicting results have also emerged regarding the correlation between BMD and the duration of hyperthyroidism, with some findings sustaining that percentage BMD is inversely correlated with levels of the TRAB. In addition, premature hair graying may be a weak marker for reduced BMD in women with a history of Graves' disease, although that is not the case in normal women. Hyperthyroid patients display a general reduction of bone mass in the axial skeleton that is only partially corrected, mainly at the femoral neck, after 5 to 7 months of biochemical euthyroidism. In children, severe osteopenia is observed on diagnosis of Graves' disease, with a preferential loss of cortical bone and the BMD reduced significantly, but this is rapidly corrected after 1 year of euthyroidism.

After surgical treatment for Graves' disease, thyrotoxicosis-associated bone loss in premenopausal women is fully restored, and subclinical hypothyroidism post-thyroidectomy may even result in higher BMD than in controls. In contrast, surgical thyroidectomy does not produce a similar result in postmenopausal women due to the influence of the menopause on bone resorption. T<sub>4</sub> therapy alone does not represent a significant risk factor for bone loss or, therefore, for osteoporotic fracture. Nonetheless, it is clear that there is a potential risk of bone loss in postmenopausal females.

## CLINICAL ASPECTS OF THYROTOXIC BONE DISEASE

Clinical studies have shown that hyperthyroidism is one of the major causes of secondary osteoporosis; however, unlike osteoporosis, it affects more than trabecular bone. The imbalance of the resorption/formation ratio reduces BMD and, therefore, constitutes a risk of thyrotoxicosis-associated osteoporotic fractures.

The etiology and duration of thyrotoxicosis do not seem to play a role in the severity of thyrotoxic bone disease, but the clinical manifestations of Graves' disease for bone may differ depending on the age of the patient, particularly for pre- and postmenopausal female patients.

Several humoral markers reflect osteogenesis. T<sub>3</sub> enhances the functional activity of mature osteoblasts but suppresses the differentiation of osteoprogenitor cells to osteoblasts. The marker of osteoblast activity at all stages (ALP) and the marker of mature osteoblast activity (osteocalcin) are increased in thyrotoxic bone disease.

The osteolytic humoral marker, Udpd cross-links/creatinine excretion, is a highly sensitive marker of increased bone metabolism in thyrotoxicosis, whereas there are only relatively small increases in osteocalcin and bone ALP determinants.

After 1 year of euthyroidism, thyrotoxicosis-associated bone loss may be reversible. The normalization of thyroid hormone levels with antithyroid drugs is followed by a significant increase in lumbar spine BMD, which is preceded by a significant attenuation of increased bone turnover. This recovery is sometimes incomplete in the lumbar spine and Ward's triangle after 5 to 7 months euthyroidism; therefore, it is important to identify hyperthyroid patients who are at risk for insufficient recovery.

Some longitudinal studies suggest that high ALP levels 1 year after the initiation of antithyroid therapy are associated with reduced BMD values and could predict poor restoration of BMD. During this phase, PTH rises and insulin-like growth factor-1 (IGF-1) bioactivity could be insufficient to restore bone mass despite euthyroidism. In addition, the use of antiresorptive drugs could be encouraged in these patients.

In postmenopausal thyrotoxic women, many factors in serum and the bone microenvironment unit interact to increase osteoclast activity and subsequent bone loss. This is due to the effect of increased thyroid hormone production together with a lack of skeletal protection by estrogen. The relative dehydroepiandrosterone (DHEA) and IGF-1 insufficiency that occurs during the postmenopausal period may constitute additional risk factors for developing enhanced bone loss. The high-turnover state of the early postmenopausal stage could predispose bone to the detrimental effects of hyperthyroidism, and postmenopausal women are most sensitive to accelerated bone loss from excessive T<sub>4</sub> therapy.

Conversely, other studies have demonstrated no significant difference in BMD expressed as a Z score between pre- and postmenopausal female hyperthyroid patients. This could suggest that the impact of thyrotoxicosis is great enough to surpass the effect of the menopause on bone mass, at least during the late postmenopausal period. The effect of thyroid hormones could aggravate the evident high bone turnover state during the early postmenopausal period. In contrast, during the late postmenopausal period, characterized by slow bone loss, thyroid hormone could produce a less detrimental effect on bone remodeling. This argues in favor of the early use of hormonal replacement therapy in early postmenopausal thyrotoxic women and, moreover, for the thyroid hormone replacement or suppressive therapy to be of

the smallest possible dose to produce the desired clinical effect.

Special discussion is required for the impact of thyrotoxicosis on bone metabolism in three specific situations: thyrotoxicosis in men, bone disease in subclinical hyperthyroidism, and thyrotoxicosis in children.

### Thyrotoxic Bone Disease in Men

There is little information on the effect of thyroid hormones on bone mass and the risk of fractures in men; however, it is clear that thyroid hormones have a smaller effect on bone in men than in women. Nonetheless, many experts believe that hyperthyroidism is one of the most important causes of osteoporotic fractures in men occurring after alcohol abuse, glucocorticoid excess, and hypogonadism.

Previous studies have shown that thyrotoxic patients display a lower radial BMD, compared with age-matched controls, and have a twofold increased risk of hip fractures.

In male patients with recent-onset Graves' disease, BMD values are reduced to levels similar to those reported in female hyperthyroid patients. As in women, this is more marked in cortical bone where there is a significant relationship with  $T_4$  levels. This is consistent with an improvement of BMD in Graves' patients on recovery from the hyperthyroidism and with partial recovery of bone mass after attainment of euthyroidism.

The presence of greater concentrations of total ALP and osteocalcin in men affected by Graves' disease suggests the beginning of the bone mass recovery period during the early months of effective treatment.

The effect of long-term suppressive  $T_4$  therapy for thyroid cancer on BMD is similar to that produced by Graves' disease. Nonetheless, the bone loss in Graves' disease is reversible to some extent, whereas the bone loss in patients on suppressive therapy is reasonably continuous. The prior administration of pamidronate could be used to prevent a thyroid hormone-induced increase in bone resorption.

The effect of excess endogenous and exogenous thyroid hormones is mildly deleterious in the axial bone mass in male patients. In another study, the same suppressive therapy caused only a small increase in the markers of bone metabolism without detectable changes in the BMD.

In hypothyroidism, thyroid hormone replacement therapy does not produce a difference in BMC, thereby excluding the possibility of a significant loss of cortical bone mineral by thyroid hormones.

Normal BMD is reported in male patients on  $T_4$  substitutive therapy.

Major evidence is emerging for the existence of male osteoporosis; therefore, male patients with a history of TSH suppression and thyrotoxicosis should be included in the preventive program of skeletal status assessment.

### Bone Disease in Subclinical Hyperthyroidism

Endogenous subclinical hyperthyroidism is identified by suppressed values of TSH in the presence of normal  $T_4$  levels. This is also observed during suppressive  $T_4$  treatment and is referred to as exogenous subclinical hyperthyroidism.

BMD and bone biochemical parameters may be influenced by the duration of subclinical hyperthyroidism and by the menopausal status of the patients. For subclinical thyrotoxicosis, it appears that the appendicular, rather than the axial, skeleton is more susceptible to minor thyroid hormone excess. In premenopausal women affected by the endogenous subclinical hyperthyroidism associated with Graves' disease or multinodular goiter, biochemical bone markers are not increased and BMD is not reduced. Conversely, long-term endogenous subclinical hyperthyroidism (2 years or more) may be a contributing factor for the development of osteoporosis and accelerated bone loss in postmenopausal women, mostly at sites where cortical bone predominates. A subtle increase in thyroid hormone, together with the lack of skeletal protection by estrogen and relative postmenopausal insufficiency of DHEA and IGF-1, is an additional risk factor for bone loss.

Antithyroid treatment of endogenous subclinical hyperthyroidism has a beneficial effect on bone loss in postmenopausal women. This is substantiated by significantly higher BMD values at the distal site of the forearm in treated patients, as compared with untreated patients, during the second year of treatment. The difference is small but cumulative over many years and could result in a decreased risk of fracture. The stabilization of the decline in BMD is remarkable considering the postmenopausal status. In premenopausal women affected by subclinical hyperthyroidism for Graves' disease, the antithyroid drug treatment produces a significant increase in bone mass and reduces the risk of secondary osteoporosis. It still appears to be important to achieve normal TSH levels in Graves' patients during the treatment so as to normalize their bone metabolism.

The T<sub>4</sub> treatment could produce exogenous subclinical hyperthyroidism. The T<sub>4</sub> dose that renders a patient euthyroid, as shown by normal TSH values, rarely produces an adverse effect.

The slightly suppressive T<sub>4</sub> administration can activate bone turnover but does not constitute a risk factor for bone loss or for osteoporosis in pre- and postmenopausal women with nontoxic goiter.

Two meta-analysis studies have shown a significant reduction in BMD, but only in postmenopausal women on long-term L-T<sub>4</sub> suppressive therapy; this was more marked in cortical bone than in trabecular bone. In postmenopausal women, the bone loss appears early during the suppressive treatment. Long-term suppressive doses of TSH produce BMD reduction at various skeletal sites and increase the risk of fractures. TSH suppressive doses should be prescribed only when appropriate and no longer than necessary so as to minimize the adverse effects of excessive doses of thyroid hormone on bone. The magnitude of bone loss, also during long-term therapy, depends on the serum levels of thyroid hormones as well as on the functional state of thyroid hormone receptors in bone tissue. However, other risk factors should be studied to prevent the possible loss of bone mass (e.g., age, weight, calcium intake in postmenopausal women, low physical activity in premenopausal women).

T<sub>4</sub> therapy represents a significant risk factor for BMD loss only in postmenopausal women with a previous history of thyrotoxicosis. In this condition, the smallest possible dose of thyroid hormone to produce the desired clinical effect must be used. Postmenopausal women taking both thyroid and estrogen hormones exhibit BMD values comparable to those observed in women taking only estrogen. The estrogen replacement therapy abolishes the reduction in femoral and vertebral BMD in postmenopausal women on L-T<sub>4</sub> therapy. The potential beneficial influence of estrogen replacement therapy suggests that estrogen administration should be encouraged in those patients. Anti-osteoclastic agents may also be appropriate as a preventive treatment in postmenopausal women at high risk for osteoporosis during T<sub>4</sub> therapy.

### Thyrotoxic Bone Disease in Children

Graves' disease is a rare condition during childhood and adolescence, with only 1 to 5% of all patients being children. The incidence of juvenile Graves' disease ranges from 0.1 in 100,000 patients in young children to 3.0 in 100,000 patients in adolescents. Only a few investigations in children and adolescents

have been published during the past decade or so. Accelerated growth and bone maturation is observed in prepubertal children; growth acceleration is present for several months before diagnosis, and the bone age is advanced by 1.5 to 2.5 years compared with the chronological age, irrespective of weight. During this phase, the bone maturation is affected by growth hormone (GH) and thyroid hormone, whereas at puberty, it is influenced mainly by sex hormones. This may explain the failure to observe similar growth acceleration and bone maturation in the pubertal patients for whom bone age corresponds to anagraphic age. In prepubertal children, the epiphysis is also significantly altered due to the exposition to high levels of thyroid hormones.

In growing children, untreated thyrotoxicosis and inappropriate T<sub>3</sub> replacement therapy can increase osteopenia in the short term but generally results in short-stature adults relative to predicted heights.

The severe osteopenia in children that is observed on diagnosis of Graves' disease is rapidly corrected after 1 to 2 years of treatment. Furthermore, antithyroid treatment dramatically reduces the bone resorption in pubertal girls and increases significantly both spinal and total body BMD, providing the physiological conditions to obtain optimal peak bone mass.

### OUTLINE OF THERAPY FOR THYROTOXIC BONE DISEASE

Experimental animal studies demonstrate that excess thyroid hormone induces cortical bone loss associated with high bone turnover that is higher in tibia than in vertebra. This is due to the effect of alendronate, evident in tibia but not in vertebra, that interferes with the recruitment of osteoclasts and increases bone volume by inhibiting osteoclast activity. In vertebra, the lack of effect of thyroid hormone and alendronate may be ascribed to a lower basal bone turnover.

The pamidronate is effective at preventing bone mineral loss in ovariectomized rats, both T<sub>4</sub> treated and untreated. This finding may have clinical relevance in estrogen-depleted patients for whom a treatment other than the reduction of T<sub>4</sub> administration would be desirable.

In normal men subjected to mild thyroid hormone excess for 8 days, the prior administration of pamidronate is useful in the prevention of thyroid hormone-induced increased bone resorption or induced osteopenia. Alendronate also produces an increase in BMD and a corresponding decrease in serum osteocalcin levels in both pre- and postmenopausal women.

A recent study assessed the effect of antiresorptive therapy with nasal calcitonin on bone metabolism in recently diagnosed hyperthyroid patients. The effect of exogenous calcitonin was greater in the patients than in normal controls.

The significant reduction in axial BMD in thyrotoxic patients was partially restored after attainment of the euthyroid state. Nonetheless, recovery was incomplete, with a 5% deficit compared with controls. The treatment with nasal calcitonin had no additional effect after attainment of the euthyroid state.

In postmenopausal women, estrogen replacement therapy is effective at preventing an increase in bone mineral metabolism in high-dose T<sub>4</sub> treatment as in thyrotoxicosis.

For the treatment of bone thyrotoxic disease in children, the antithyroid drug therapy dramatically reduces the bone resorption and increases significantly both spinal and total BMD, providing physiological conditions for the achievement of the peak bone mass.

In summary, published evidence indicates that TSH suppresser therapy for subclinical Graves' disease, and even minimally excessive thyroid hormone replacement or chronic suppression therapy, can be accompanied by a decrease in cortical bone mass. This raises important questions about the use of lifelong replacement therapy, especially for young patients. These data should also provoke caution in the use of traditional suppressive therapy for thyroid nodules. The correct interpretation of studies such as these requires keeping in mind the concept of remodeling space. Predictable changes in bone density will be observed simply by increasing or restricting the remodeling space. However, after some months, a new equilibrium will be achieved. To evaluate the clinical impact of such changes, it is necessary to carry out longer term studies than those that have been

reported, but for the present time, the smallest possible dose of thyroid hormones to achieve an euthyroid state or to suppress thyroid gland activity is advisable.

The critical effect of suppressive T<sub>4</sub> treatment on bone in thyroid cancer could be balanced by pamidronate administration. In these patients, it is also critical to assess bone mineral metabolism.

## Acknowledgement

I am indebted to Tracy Williams for her collaborative support.

## See Also the Following Articles

Bone Mass Measurement • Bone Turnover Markers • Hyperthyroidism, Subclinical • Interleukin-6 • Osteoporosis, Overview • Paget's Disease of Bone • Parathyroid Hormone (PTH) • Tumor Necrosis Factor (TNF)

## Further Reading

- Greenspan, S., and Greenspan, F. (1999). The effect of thyroid hormone on skeletal integrity. *Ann. Int. Med.* **130**, 750–758.
- Harper, K. D., and Weber, T. (1998). Secondary osteoporosis: Diagnostic consideration. *Endocrinol. Metab. Clinics North America* **27**, 325–348.
- Lerner, U. H. (2000). Osteoclast formation and resorption. *Matrix Biol.* **19**, 107–120.
- Mosekilde, L., Melsen, F., Bragger, J. P., Myhre-Jensen, O., and Schwartz Soerensen, N. (1977). Bone changes in hyperthyroidism: Interrelationship between bone morphometry, thyroid function, and calcium–phosphorus metabolism. *Acta Endocrinol.* **85**, 515–525.
- Motomura, K., and Brent, A. G. (1998). Mechanism of thyroid hormone action: Implications for the clinical manifestation of thyrotoxicosis. *Endocrinol. Metab. Clinics North America* **27**, 1–23.
- Pantazi, H., and Papapetrou, D. P. (2000). Changes in parameters of bone and mineral metabolism during therapy for hyperthyroidism. *J. Clin. Endocrinol. Metab.* **85**, 1099–1106.

2-h radioiodine uptake, although it should be feasible to obtain the result of serum T4 determination within a few hours on an emergency basis in most hospitals today. However, initiation of therapy should not be postponed when there is a high index of suspicion merely because one is awaiting laboratory confirmation of the diagnosis.

Other laboratory abnormalities often include modest hyperglycemia in the absence of diabetes mellitus. Moderate leukocytosis with a mild shift to the left is common even in the absence of infection. Increased serum calcium levels may be seen, perhaps due to both hemoconcentration and the known effects of thyroid hormone on bone resorption, but other serum electrolytes are usually normal. Hepatic dysfunction in thyrotoxic storm will result in elevated levels of serum lactate dehydrogenase, aspartate aminotransferase, and bilirubin.

Because serum cortisol levels are usually elevated in thyrotoxic individuals, a normal value may be interpreted as being inappropriately low. In view of the known coincidence of adrenal insufficiency with Graves' disease, one should maintain a reasonably high index of suspicion for this disorder, particularly if there is hypotension and suggestive electrolyte abnormalities. It would be prudent to obtain a serum sample for cortisol determination prior to the administration of corticosteroid. Even in the absence of adrenal insufficiency, adrenal reserve may be exceeded in thyrotoxic crisis due to the inability of the adrenal gland to meet the demand placed on it as a result of the accelerated turnover and disposal of glucocorticoids that occur in thyrotoxicosis.

## PATHOGENESIS

The precise pathogenesis underlying the precipitation of thyrotoxic storm is likely not to be the same in all cases and remains incompletely understood, although the magnitude of serum hormone levels per se does not appear to be critical. However, acute discharge of hormone resulting in rather sudden changes of its concentration, in the appropriate clinical setting, certainly can trigger crisis. This may be seen after <sup>131</sup>I therapy, withdrawal of propylthiouracil (PTU) therapy, vigorous palpation of the thyroid, or administration of lithium, stable iodine, or iodinated contrast dyes.

A possible interaction between the effects of excessive levels of circulating thyroid hormone and the catecholamines has been proposed. This is evidenced by the dramatic clinical improvement that follows the use

of agents that either deplete their tissue levels, such as reserpine, or block  $\beta$ -adrenergic receptors, such as propranolol. Although these agents are useful adjuncts to therapy, they should not be used by themselves because they might not prevent the occurrence of storm.

## TREATMENT

We believe that to avoid a fatal outcome, it is important to implement a four-pronged approach to the management of thyrotoxic storm. The relative importance for survival of each arm of therapy will vary in a given patient. First, specific antithyroid drugs must be used to decrease the increased thyroid production and release of thyroxine (T4) and T3. The second part of management consists of treatment intended to block the effects of the remaining, but excessive, circulating concentrations of free T4 and T3. The third component addresses any underlying precipitating illness such as infection or ketoacidosis. The final arm of therapy is composed of those several specific treatments that are directed against the underlying systemic decompensation that may be characterized by fever, congestive failure, shock, and the like. In view of the poor prognosis associated with incompletely treated thyrotoxic storm, no one component of this four-pronged therapeutic approach should be neglected.

### Therapy Directed against the Thyroid Gland

Inhibition of new synthesis of the thyroid hormones is achieved by administration of thionamide antithyroid drugs, such as PTU and methimazole (Tapazole), given orally. There are no available parenteral preparations of these compounds. Either methimazole or PTU may also be administered via the rectum if necessary. In view of the gravity of thyrotoxic storm, thionamide doses are much higher than those for otherwise uncomplicated thyrotoxicosis. Some experienced clinicians believe that PTU will provide more rapid clinical improvement because it has the additional advantage of inhibiting conversion of T4 to T3, a property not shared by methimazole. Separate treatment must be administered to inhibit the continuing release of T4 and T3 into the blood because thionamides act to reduce new hormone synthesis but have no effect on thyroidal secretion of preformed stores of hormone. Either inorganic iodine or lithium carbonate may be used for this purpose. Iodides may be given either orally as Lugol's solution or as a saturated solution of potassium iodide (eight

drops every 6 h). When iodine is administered together with full doses of antithyroid drugs, dramatic decreases in serum T<sub>4</sub> can be seen. The sequence of administration of these agents is extremely important. Use of iodine without prior thionamide dose is contraindicated because the iodine will provide extra substrate to enrich hormone stores within the gland, thereby generating the potential for further exaggeration of thyrotoxicosis.

In patients who may be allergic to iodine, lithium carbonate may be used as an alternative agent to inhibit hormonal release, although some caution has been raised in regard to its use in the setting of storm. This drug also may be used in thyrotoxic patients who are known to have serious toxic reactions to the thionamides.

### Therapy Directed against Ongoing Effects of Thyroid Hormone in the Periphery

For the purpose of acutely reducing the circulating hormones, peritoneal dialysis and plasmapheresis have been employed, as has experimental hemoperfusion through a resin bed or charcoal columns. Such aggressive management should be considered in severe cases.

Beta blockers are also commonly used. Propranolol is the agent most commonly used in the United States today. Large doses, such as 60 to 120 mg every 6 h, are used in crisis or impending crisis. Indeed, because of the more rapid metabolism of the drug in severe thyrotoxicosis, even larger oral doses, or preferably intravenous doses, should be given. Initial intravenous doses should be given cautiously, whereas the patient's cardiac rhythm is monitored continuously. Added benefits of  $\beta$ -adrenergic blockade in these patients include improvement in agitation, convulsions, psychotic behavior, tremor, diarrhea, fever, and diaphoresis.

### Therapy Directed against the Precipitating Illness

In most cases of thyrotoxic storm, therapy is not complete unless a diagnosis of the possible precipitating event is made and early treatment as indicated for that underlying illness is implemented.

It is important to be alert to the fact that conditions such as ketoacidosis, pulmonary thromboembolism, and stroke may underlie thyrotoxic crisis, particularly in the obtunded or psychotic patient, and require the same vigorous management ordinarily indicated. In the patient with thyrotoxic crisis in whom none of the latter precipitating factors is apparent, a diligent search for some focus of infection must be carried out. Empirical,

broad-spectrum antibiotic coverage may be warranted while awaiting results of cultures. In most patients who survive thyrotoxic crisis, clinical improvement is dramatic and demonstrable within 12 to 24 h.

### Therapy Directed against Systemic Decompensation

Fluid depletion caused by the hyperpyrexia, diaphoresis, vomiting, or diarrhea must be vigorously replaced to avoid vascular collapse. Fluid management must be individualized. Shock may be refractory to cautious fluid resuscitation in younger patients, whereas judicious replacement of fluids is necessary in elderly patients with congestive heart failure or other cardiac compromise. Intravenous fluids containing 10% dextrose in addition to electrolytes will allow repletion of the depleted hepatic glycogen. For fever, acetaminophen, rather than salicylates, is the preferred antipyretic because salicylates inhibit thyroid hormone binding and could increase free hormone, thereby transiently worsening the thyrotoxic crisis.

Vasopressor therapy may become necessary on a temporary basis to provide adequate hemodynamic support if hydration with intravenous fluid replacement is not effective. Stress dose glucocorticoids have been given on empirical grounds on the basis of postulated relative adrenal insufficiency. This approach has the added benefit of further blockade of peripheral conversion of T<sub>4</sub> to T<sub>3</sub>, and this is an additional justification for their use.

### See Also the Following Articles

Graves' Ophthalmopathy • Thyrotoxicosis: Diagnosis • Thyrotoxicosis Factitia • Thyrotoxicosis, Overview of Causes • Thyrotoxicosis, Systemic Manifestations • Thyrotoxicosis, Treatment

### Further Reading

- Aiello, D. P., DuPlessis, A. J., Pattishall, E. G., III, and Kulin, H. E. (1989). Thyroid storm presenting with coma and seizures. *Clin. Pediatr.* **28**, 571–574.
- Ashkar, F. S., Katims, R. B., Smoak, W. M., and Gilson, A. J. (1970). Thyroid storm treatment with blood exchange and plasmapheresis. *J. Am. Med. Assoc.* **214**, 1275–1279.
- Burman, K. D., Yeager, H. C., and Briggs, W. A. (1976). Resin hemoperfusion: A method of removing circulating thyroid hormones. *J. Clin. Endocrinol. Metab.* **42**, 70–78.
- Eriksson, M. A., Rubinfeld, S., Garber, A. J., and Kohler, P. O. (1977). Propranolol does not prevent thyroid storm. *N. Engl. J. Med.* **296**, 263–264.
- Feely, J., Forrest, A., and Gunn, A. (1980). Propranolol dosage in thyrotoxicosis. *J. Clin. Endocrinol. Metab.* **51**, 658–661.

Herrmann, J., Hilger, P., and Kruskemper, H. L. (1973). Plasmapheresis in the treatment of thyrotoxic crisis (measurement of half-concentration tissues for free and total T3 and T4). *Acta Endocrinol. (Copenh)* **173**(Suppl.), 22.

McDermott, M. T., Kidd, G. S., Dodson, L. E., and Hofeldt, F. D. (1983). Radioiodine-induced thyroid storm. *Am. J. Med.* **75**, 353–359.

Wartofsky, L., Ransil, B. J., and Ingbar, S. H. (1970). Inhibition by iodine of the release of thyroxine from the thyroid glands of patients with thyrotoxicosis. *J. Clin. Invest.* **49**, 78–86.

Yeung, S. C., Go, R., and Balasubramanyam, A. (1995). Rectal administration of iodide and propylthiouracil in the treatment of thyroid storm. *Thyroid* **5**, 403–405.



## Thyrotropin

see TSH (Thyroid-Stimulating Hormone)



## Thyrotropin Receptor

see TSH Receptor

**Table II** Differential Diagnosis of Main Forms of Thyrotoxicosis with Low/Suppressed Thyroidal Radioactive Iodine Uptake

	Subacute thyroiditis	Iodine-induced thyrotoxicosis	Thyrotoxicosis factitia
Serum-free thyroid hormones	Increased	Increased	Increased
Serum TSH	Suppressed	Suppressed	Suppressed
Serum thyroglobulin	Increased	Increased	Low/Undetectable
Serum interleukin-6	Increased	Increased	Low
Urinary iodine excretion	Normal	Increased	Normal
Thyroid echogenicity	Markedly reduced and not homogeneous	Slightly reduced	Normal
Thyroid vascularity	Reduced	Reduced	Reduced
Thyroid pain/tenderness	Present	Absent	Absent

## CLINICAL FEATURES

The clinical picture of thyrotoxicosis factitia does not differ from that of classical spontaneous hyperthyroidism. Patients complain of tachycardia, tremors, loss of weight, increased perspiration, heat intolerance, extreme anxiety and nervousness, increased bowel activity, and/or insomnia (Table I). Goiter and ophthalmopathy, as seen in Graves' disease, are absent. Likewise, thyroid pain and tenderness, commonly observed in subacute thyroiditis, are absent (Table I). Cardiovascular complications (e.g., tachyarrhythmias, heart failure, myocardial infarction) as well as bone loss (osteopenia) may occur as a consequence of prolonged excess thyroid hormone ingestion. These complications are more likely to take place in older patients.

Laboratory evaluation reveals the typical increase in free T4 (FT4) and free T3 (FT3) concentrations associated with undetectable serum thyrotropin (TSH) levels. An isolated increase in serum FT3 concentration associated with suppressed serum FT4 levels indicates that the ingested thyroid hormone preparation contained only T3. Circulating autoantibodies to thyroglobulin or thyroperoxidase, as well TSH receptor autoantibodies responsible for Graves' disease, are usually absent. Thyroidal radioactive iodine uptake (RAIU) is characteristically very low or suppressed (Table I). Serum thyroglobulin concentration is typically markedly reduced or undetectable in thyrotoxicosis factitia (Table I). Accordingly, its measurement is a useful tool for differentiating thyrotoxicosis factitia from other thyrotoxic conditions associated with low RAIU values (Table II). Serum interleukin-6, a marker of thyroidal destructive processes, is also undetectable in thyrotoxicosis factitia (Table II). Measurement of thyroid hormones in stools may be useful for identifying the abnormally high fecal excretion of ingested thyroid hormones.

Urinary iodine excretion is normal (Table II). Color flow Doppler sonography of the thyroid shows an absent vascularity and normal–low peak systolic velocity in spite of the thyrotoxic state (Table II).

If the patient denies the surreptitious thyroid hormone intake, it may be necessary to hospitalize him or her to be sure that the deliberate ingestion does not continue. Under strict medical controls, a rapid improvement in the clinical and laboratory features of thyrotoxicosis is usually observed. However, recurrency of thyrotoxicosis is frequently observed after hospital release unless reasons for surreptitious thyroid hormone intake are identified.

## TREATMENT

Treatment of thyrotoxicosis factitia obviously requires withdrawal of thyroid hormones. It may be useful to associate a short-term treatment with  $\beta$ -adrenergic blocking drugs to control tachycardia and tremors promptly. However, for a full recovery of the patient, psychiatric aid or counseling is mandatory in all cases.

## See Also the Following Articles

Anorexia Nervosa • Depression, Thyroid Function and • Thyroglobulin • Thyrotoxicosis: Diagnosis • Thyrotoxicosis, Overview of Causes • Thyrotoxicosis, Systemic Manifestations • Thyrotoxicosis, Treatment

## Further Reading

- Bartalena, L., Bogazzi, F., and Martino, E. (1996). Adverse effects of thyroid hormone preparations and antithyroid drugs. *Drug Safety* **15**, 53–63.
- Bartalena, L., and Vitti, P. (2001). Hyperthyroidism and thyrotoxicosis. In "Endocrinology and Metabolism" (A. Pinchera, ed.), pp. 161–172. McGraw–Hill International, London.
- Bogazzi, F., Bartalena, L., Scarcello, G., Campomori, A., Rossi, G., and Martino, E. (1999). The age of patients with thyrotoxicosis

- factitia in Italy from 1973 to 1996. *J. Endocrinol. Invest.* **22**, 128–133.
- Bogazzi, F., Bartalena, L., Vitti, P., Rago, T., Brogioni, S., and Martino, E. (1996). Color flow Doppler sonography in thyrotoxicosis factitia. *J. Endocrinol. Invest.* **19**, 603–606.
- Bouillon, R., Verresen, L., Staels, F., Bex, M., De Vos, P., and De Roo, M. (1993). The measurement of fecal thyroxine in the diagnosis of thyrotoxicosis factitia. *Thyroid* **3**, 101–103.
- Cohen, J. H., Ingbar, S. H., and Braverman, L. E. (1989). Thyrotoxicosis due to ingestion of excess thyroid hormone. *Endocrine Rev.* **10**, 113–124.
- Hamolsky, M. W. (1982). Truth is stranger than factitious. *N. Engl. J. Med.* **307**, 436–437.
- Mariotti, S., Martino, E., Cupini, C., Lari, R., Giani, C., Baschieri, L., and Pinchera, A. (1982). Low serum thyroglobulin as a clue to the diagnosis of thyrotoxicosis factitia. *N. Engl. J. Med.* **307**, 410–412.
- Rose, E., Sanders, T. P., Webb, W. L., Jr., and Hines, R. C. (1969). Occult factitial thyrotoxicosis. *Ann. Int. Med.* **71**, 309–315.
- Roti, E., Minelli, R., Gardini, E., and Braverman, L. E. (1993). The use and misuse of thyroid hormone. *Endocrine Rev.* **14**, 401–423.

Table I Classification of Known Causes of Thyrotoxicosis, Epidemiology, and Distinctive Diagnostic Features

Group	Disease	Relative frequency (percentage)	Distinctive features	Neck RAIU	Serum TSH	Serum thyroglobulin	
Thyrotoxicosis of thyroidal origin	Associated with hyperthyroidism	Graves' disease	70	Diffuse goiter Ophthalmopathy Positive TRABs	High	Undetectable	High
		Toxic adenoma	5	Single "hot" nodule at thyroid scan	High	Undetectable	High
		Toxic multinodular goiter	20	Multiple "hot" nodules at thyroid scan	High	Undetectable	High
		Iodine-induced thyrotoxicosis	<1	High urinary iodine	Low to high	Undetectable	High
		TSH-secreting adenomas	<1	Inappropriately high TSH level	High	Normal to high	High
		Familial gestational hyperthyroidism	Extremely rare	Pregnancy-associated DNA analysis	Presumably high	Undetectable	High
		Trophoblastic tumors	<1	High chorionic gonadotropin	High	Undetectable	High
		Neonatal thyrotoxicosis	<1	Positive TRABs	High	Undetectable	High
		Non-autoimmune congenital and familial hyperthyroidism	Extremely rare	TSH receptor gene mutations by DNA analysis	High	Undetectable	High
	Associated with thyroid destruction	Subacute thyroiditis	3	Neck pain High erythrocyte sedimentation rate	Low	Undetectable	High
Silent thyroiditis		3	Positive thyroid autoantibodies	Low	Undetectable	High	
Type II amiodarone-induced thyrotoxicosis		<1	High urinary iodine High serum interleukin-6	Low	Undetectable	High	
Thyrotoxicosis of nonthyroidal origin	Factitious thyrotoxicosis	<1	History Low serum thyroglobulin	Low	Undetectable	Low	
	Thyroid hormone intoxication	<1	History Low serum thyroglobulin	Low	Undetectable	Low	
	Dermoid tumors (struma ovarii)	Very rare	Abdominal RAIU	Low	Undetectable	High	
	Metastatic differentiated thyroid cancer	Very rare	Bone RAIU	Low	Undetectable	High	

and overproduction and release of thyroid hormone that lead to hyperthyroidism. In some cases, thyroid nodularities can be felt as a consequence of either long-standing disease or preexisting nodular goiter. Physical examination reveals signs of Graves' ophthalmopathy in 30 to 45% of patients, and when studied with imaging techniques, suggestive findings can be observed in up to 70% of cases. Local myxedema is a peculiar and rare skin manifestation of Graves' disease characterized by edema, inflammation, and lymphocytic infiltration localized mostly to the pretibial dermis. It occurs almost exclusively in patients who also have Graves' ophthalmopathy.

### **Toxic Adenoma**

Toxic adenoma is a quite frequent cause of hyperthyroidism, especially in iodine-deficient countries, where its prevalence has been reported to be as high as 4.5% (10% of all cases of thyrotoxicosis). A lower prevalence (2.7%) has been reported in iodine-sufficient areas.

Toxic adenomas are benign isolated thyroid tumors that function autonomously. The nodular tissue acquires the capability of producing thyroid hormones independently of TSH stimulation. The increased thyroid hormone secretion first suppresses pituitary TSH secretion and eventually leads to overt hyperthyroidism. Because of TSH suppression, the extranodular thyroid tissue becomes functionally quiescent and may undergo some degree of atrophy. Toxic adenomas are more frequent in the aged population and in females. The natural history of toxic adenoma is characterized by slow growth over many years with a progression of its functional properties through various stages. During the early stages, the excess of secretion of thyroid hormones is not sufficient to completely suppress TSH secretion (partial autonomy) and the extranodular tissue. With further growth of the nodule, TSH suppression becomes complete, whereas circulating thyroid hormones are in the upper range of normality (complete autonomy). Eventually, overt thyrotoxicosis ensues with frankly elevated thyroid hormone levels. Recent studies have shown that somatic mutations of the TSH receptor gene are the cause of 20 to 80% of toxic adenomas. In other cases, the Gs- $\alpha$ -subunit of the TSH receptor-coupled adenylcyclase is mutated. Both kinds of mutations cause permanent activation of the TSH receptor intracellular signaling pathway in the absence of the natural ligand (TSH); therefore, they fully explain the autonomous functioning of the nodules.

### **Toxic Multinodular Goiter**

Toxic multinodular goiter is also more frequent in iodine-deficient countries and primarily affects women. Epidemiological studies have clearly shown that toxic multinodular goiter represents the long-term outcome of many long-standing nontoxic goiters; therefore, it is most often found in aged persons. Somatic activating mutations of the TSH receptor have been observed occasionally in toxic multinodular goiter as well. The natural history of toxic multinodular goiter is similar to that of toxic adenoma, with a slow formation of multiple autonomously functioning nodular areas in the setting of an overall nodular goiter. Progression of autonomous function may lead to subclinical hyperthyroidism and overt thyrotoxicosis. The clinical manifestations are scanty due to the advanced age of patients and the slow progression of thyrotoxicosis.

### **TSH-Secreting Adenoma**

Hyperthyroidism caused by excessive TSH secretion by a pituitary adenoma (central hyperthyroidism) is very rare. TSH-secreting adenomas completely or partially lose feedback regulation by thyroid hormone and, therefore, cause sustained stimulation of the thyroid gland, leading to the development of goiter and hyperthyroidism. The severity of thyrotoxicosis is highly variable, ranging from modest elevations to very high levels of thyroid hormones; therefore, the patient may report few or no symptoms or may present with overt thyrotoxic symptoms. TSH-secreting adenomas may cosecrete growth hormone (GH) and prolactin, with clinical presentation of acromegaly and galactorrhea. Most TSH-secreting adenomas are macroadenomas and may lead to hypopituitarism and visual field defects.

### **Human Chorionic Gonadotropin-Dependent Hyperthyroidism**

Human chorionic gonadotropin (hCG) is secreted in large amounts by placental tissue in normal pregnancy and also by trophoblastic tumors such as hydatidiform mola. hCG has a partial homology with TSH and can act as a weak TSH agonist. Therefore, large amounts of hCG in the bloodstream can overstimulate the thyroid gland and cause hyperthyroidism.

### **Trophoblastic Tumors**

In hydatidiform mola, hyperthyroidism is caused by tumor secretion of large quantities of hCG. However, overt thyrotoxicosis is observed in only a minority (10%) of patients with extraordinarily high levels of hCG (>3 million IU/L).

**Hyperemesis Gravidarum**

Hyperemesis gravidarum is characterized by prominent nausea and vomiting, weight loss ketosis, and electrolyte abnormalities. It is a poorly understood early complication of pregnancy associated with inappropriately high levels of circulating hCG that causes clinical or subclinical hyperthyroidism in about one-third of cases.

**Familial Gestational Hyperthyroidism**

This inherited pregnancy-associated form of hyperthyroidism has been described in one family. Hyperthyroidism was due to a mutation in the TSH receptor gene, causing increased sensitivity of the receptor to the effect of hCG. For this reason, hyperthyroidism manifests only during pregnancy and recurs every time an affected woman becomes pregnant.

**Fetal and Neonatal Thyrotoxicosis**

TSABs in the serum of mothers with Graves' disease can cross the placenta and cause fetal and neonatal hyperthyroidism through direct stimulation of the fetal thyroid. The disease can be very severe and is characterized by tachycardia, jaundice, heart failure, and failure to thrive. A goiter is usually present. The thyrotoxicosis is transient and resolves within 3 months after birth because there is no source of TSABs in the neonate.

**Nonautoimmune Congenital and Familial Hyperthyroidism**

Recently, a new form of congenital hyperthyroidism has been reported. The disease is caused by a germ line *de novo* mutation of the TSH receptor gene, causing constitutive permanent activation of the receptor and, therefore, diffuse goiter and overproduction of the thyroid hormone, in turn causing severe hyperthyroidism in the neonate. The clinical presentation is similar to the presentation of neonatal thyrotoxicosis, and the diagnosis should be suspected when no history of Graves' disease is present in the mother.

Familial nonautoimmune hyperthyroidism has been described in two kindreds. In this case, hyperthyroidism is due to dominant inherited activating mutations of the TSH receptor, but because the effect of the mutation is mild, hyperthyroidism and goiter develop only during adult age. The clinical manifestations are mild and variable in patients bearing the same mutation.

**Causes of Destructive (low-RAIU) Thyrotoxicosis****Subacute Thyroiditis**

Transient thyrotoxicosis occurs in approximately 50% of patients with subacute thyroiditis (ST). ST is an inflammatory disorder of the thyroid, probably due to a viral agent. The viral origin is suggested by the frequent association with a history of recent upper respiratory tract infection. Clinically, patients with ST present with pain in anterior neck, variable degrees of thyroid swelling, and systemic symptoms such as malaise and fever. From a histopathological point of view, the thyroid gland indicates a leukocytic and granulomatous infiltrate with follicles disruption. Thyrotoxicosis results from the destruction of thyroid follicles by the inflammatory process with release of preformed thyroid hormones. When present, thyrotoxicosis lasts for 3 to 8 weeks and is sometimes followed by a phase, also transient, of mild hypothyroidism. Complete recovery of thyroid function eventually occurs, but permanent hypothyroidism has also been reported.

**Painless Thyroiditis**

Painless thyroiditis is probably due to an autoimmune thyroid disorder that can generate an inflammation of the gland with chronic lymphocytic infiltration closely resembling that of Hashimoto's thyroiditis. Similar to subacute thyroiditis, the disease is characterized by a transient phase of thyrotoxicosis, but neck pain and general symptoms are usually absent. Circulating thyroid autoantibodies are found in the majority of cases, and the progression to spontaneous permanent hypothyroidism is observed in as many as 20% of cases in long-term follow-up. In some cases, painless thyroiditis is precipitated by radiotherapy in the region of the neck for a variety of neoplasms that induce thyroid damage with discharge of preformed thyroid hormone. Cytokine treatments, particularly with interferon-1 alpha and interleukin-2, have also been shown to cause a clinical syndrome resembling classical painless thyroiditis. These cytokines may also cause classical Graves' disease by acting as generic triggers of thyroid autoimmunity in predisposed individuals. The most likely mechanism in these cases is activation of preexisting thyroid autoreactive T-cell clones.

Findings suggestive of painless thyroiditis are the presence of a goiter, low radioiodine uptake with transient thyrotoxicosis, and occasional progression toward hypothyroidism. These cases may be classified as a variant of Graves' disease with predominant

cytotoxic aspects, quickly leading to a clinical picture of Hashimoto's thyroiditis. Painless thyroiditis may be seldom associated with Graves' ophthalmopathy.

### Postpartum Thyroiditis

Postpartum thyroiditis is a subacute thyroid inflammation that occurs during the early postpartum period in susceptible woman. It represents a variant of painless thyroiditis. Postpartum thyroiditis is a rather common disorder, occurring in 5 to 10% of pregnancies. Most women with postpartum thyroiditis have circulating antithyroglobulin and antithyropoxidase antibodies before or during the onset of the disease. Some human leukocyte-associated antigen (HLA) haplotypes, such as B35, confer a clear-cut predisposition to the disorder, making it distinct from classical Hashimoto's thyroiditis. From a clinical point of view, the disease is similar to painless thyroiditis with absence of nonspecific or local symptoms and with transient thyrotoxicosis that occurs in approximately 50% of cases. Approximately one-third of cases present with a second phase of transient hypothyroidism that may develop up to 10 years later. The risk of recurrent postpartum thyroiditis in subsequent pregnancies is as high as 70% in women who have already had an episode.

### Other Forms of Destructive Thyrotoxicosis

Rarely, destructive thyrotoxicosis can be precipitated by anterior neck injuries. Thyrotoxic crises following thyroid surgery were frequent at the beginning of thyroid surgery era. It has become extremely rare considering the optimal preparation of patients with antithyroid drugs and the refinement of surgical procedures. Finally, thyrotoxicosis may transiently worsen or recur in patients who are treated with radioiodine for Graves' disease, toxic adenomas, and multinodular toxic goiter. This phenomenon can be explained with two mechanisms: ongoing thyroid hyperfunction before radioiodine fully takes effect and radiation-induced thyroid destruction.

### Iodine- and Amiodarone-Induced Thyrotoxicosis

Iodine-induced thyrotoxicosis, secondary to the consumption of large amounts of iodine through the diet or other routes (some medications and diagnostics [e.g., contrast media, disinfectants, drugs] and some foods), has been recently reported with increased frequency. Thyrotoxicosis can precipitate through several mechanisms. When high doses of iodine are

**Table II Thyroid Disorders Predisposing to Iodine-Induced Thyrotoxicosis**

• Autonomous or pretoxic thyroid adenoma
• Nontoxic, autonomous, or pretoxic multinodular goiter
• Euthyroid or "latent" Graves' disease
• Graves' disease in remission after or during antithyroid drug treatment

administered, the normal thyroid responds with inhibition of organification. This is an autoregulatory defense mechanism known as the Wolff–Chaikoff effect. Eventually, an escape phenomenon ensues and thyroid hormone synthesis resumes. The protection mechanism appears to be defective in autonomous thyroid tissue, and excess iodine or even simple dietary supplementation may precipitate hyperthyroidism in patients with a preexisting thyroid disorder through the "jodbasedow" phenomenon. Most of these disorders (Table II) are characterized by thyroid autonomy; therefore, iodine-induced thyrotoxicosis is by far more prevalent in areas of iodine deficiency and in aged patients.

Among drugs, the antiarrhythmic amiodarone deserves a special mention due to the dual mechanism by which it can cause thyrotoxicosis. One tablet of 200 mg of amiodarone contains approximately 75 mg of organic iodide and will release approximately 8 mg of free iodine, a tremendous amount when compared with the daily recommended dose of 200  $\mu$ g. This amount of iodine can precipitate hyperthyroidism in predisposed patients simply through the jodbasedow phenomenon in a manner similar to that of iodine of other sources (type I amiodarone-induced thyrotoxicosis). However, amiodarone has been shown to be directly cytotoxic to thyroid follicular cell *in vitro* and can precipitate a form of thyrotoxicosis similar to that observed in subacute thyroiditis and due to the release of preformed hormones (type II amiodarone-induced thyrotoxicosis). The distinction between the two forms is crucial because the treatments are quite different. A mixed form of amiodarone-induced thyrotoxicosis may occur.

### Thyrotoxicosis of Extrathyroidal Origin

#### Thyrotoxicosis Factitia

Thyrotoxicosis factitia is due to the voluntary surreptitious ingestion of excess thyroid hormone preparations with the purpose of mimicking thyrotoxicosis. It should be distinguished from iatrogenic thyrotoxicosis, which is caused by excessive doses of thyroid

hormones administered by the physician or inadvertently taken by the patient. However, the term has been widely applied to all forms of thyrotoxicosis due to the ingestion of thyroid hormone. True thyrotoxicosis factitia is most often observed in women with psychiatric disturbances. Very often, thyroid hormone is taken to reduce weight or to receive medical attention. Denial of thyroid hormone assumption may be extreme in these patients, and the diagnosis is rarely obtained at history taking. Sometimes, thyroid hormone is inadvertently taken as a component of “herbal” or “alternative” medications, usually for weight-reducing purposes. Finally, accidental grinding of cattle thyroids in hamburger meat has been reported to be the cause of an outbreak of thyrotoxicosis among hamburger consumers in the United States.

### **Struma Ovarii**

Struma ovarii is a rare teratoma of the ovary that may contain functional thyroid follicular tissue, among others. Struma ovarii causes overt thyrotoxicosis only rarely, depending on the amount of follicular tissue present in the neoplasia. Because complete TSH suppression occurs in struma ovarii, the neoplastic thyroid tissue is assumed to be functionally autonomous to cause thyrotoxicosis. A suspicion of the struma ovarii can be confirmed at the time of RAIU by scanning the pelvic area with the probe. The presence of functional thyroid tissue can be demonstrated by the finding of significantly increased uptake of iodine in the ovarian region. Computed tomography (CT) or ultrasound scan will confirm the presence of an ovarian mass.

### **Functional Metastatic Thyroid Cancer**

Differentiated thyroid cancer, even when metastatic and with large tumor burdens, rarely produces physiologically relevant amounts of thyroid hormone. On the other hand, thyroid follicular cancers with extensive bone metastases may cause thyrotoxicosis.

### **See Also the Following Articles**

Graves' Disease • Pituitary Adenomas, TSH-Secreting • Thyroiditis, Postpartum • Thyroiditis, Subacute • Thyrotoxic Storm • Thyrotoxicosis: Diagnosis • Thyrotoxicosis Factitia • Thyrotoxicosis, Systemic Manifestations • Thyrotoxicosis, Treatment • Toxic Adenoma • Toxic Multinodular Goiter

### **Further Reading**

- Bartalena, L., Bogazzi, F., Brogioni, S., *et al.* (1996). Measurement of serum free thyroid hormone concentrations: An essential tool for the diagnosis of thyroid dysfunction. *Horm. Res.* **45**, 142–147.
- Beck-Peccoz, P., Brucker-Davis, F., Persani, L., *et al.* (1996). Thyrotropin-secreting pituitary tumors. *Endocrine Rev.* **17**, 610–638.
- Berglund, J., Christensen, S. B., and Hallengren, B. (1990). Total and age-specific incidence of Graves' thyrotoxicosis, toxic nodular goitre, and solitary toxic adenoma in Malmo 1970–74. *J. Int. Med.* **227**, 137–141.
- Burch, H. B., and Wartofsky, L. (1993). Graves' ophthalmopathy: Current concepts regarding pathogenesis and management. *Endocrine Rev.* **14**, 747–793.
- Chiovato, L., Vitti, P., Bendinelli, G., *et al.* (1994). Detection of antibodies blocking thyrotropin effect using Chinese hamster ovary cells transfected with the cloned human TSH receptor. *J. Endocrinol. Invest.* **17**, 809–816.
- Hamburger, J. I. (1980). Evolution of toxicity in solitary nontoxic autonomously functioning thyroid nodules. *J. Clin. Endocrinol. Metab.* **50**, 1089–1093.
- Marcocci, C., Vitti, P., Cetani, F., *et al.* (1991). Thyroid ultrasonography helps to identify patients with diffuse lymphocytic thyroiditis who are prone to develop hypothyroidism. *J. Clin. Endocrinol. Metab.* **72**, 209–213.
- Mariotti, S., Martino, E., Cupini, C., *et al.* (1982). Low serum thyroglobulin as a clue to the diagnosis of thyrotoxicosis factitia. *N. Engl. J. Med.* **307**, 410–412.
- Martino, E., Bartalena, L., Mariotti, S., *et al.* (1988). Radioactive iodine thyroid uptake in patients with amiodarone-iodine-induced thyroid dysfunction. *Acta Endocrinol. (Copenh)* **119**, 167–173.
- Martino, E., Safran, M., Aghini-Lombardi, F., *et al.* (1984). Environmental iodine intake and thyroid dysfunction during chronic amiodarone therapy. *Ann. Int. Med.* **101**, 28–34.
- Masiukiewicz, U. S., and Burrow, G. N. (1999). Hyperthyroidism in pregnancy: Diagnosis and treatment. *Thyroid* **9**, 647–652.
- Reinwein, D., Benker, G., Konig, M. P., *et al.* (1988). The different types of hyperthyroidism in Europe: Results of a prospective survey of 924 patients. *J. Endocrinol. Invest.* **11**, 193–200.
- Ross, D. S. (1998). Syndromes of thyrotoxicosis with low radioactive iodine uptake. *Endocrinol. Metab. Clinics North America* **27**, 169–185.
- Siegel, R. D., and Lee, S. L. (1998). Toxic nodular goiter: Toxic adenoma and toxic multinodular goiter. *Endocrinol. Metab. Clinics North America* **27**, 151–168.
- Stanbury, J. B., Ermans, A. E., Bourdoux, P., *et al.* (1998). Iodine-induced hyperthyroidism: Occurrence and epidemiology. *Thyroid* **8**, 83–100.
- Tonacchera, M., Vitti, P., Agretti, P., *et al.* (1998). Activating thyrotropin receptor mutations in histologically heterogeneous hyperfunctioning nodules of multinodular goiter. *Thyroid* **8**, 559–564.
- Weetman, A. P., and McGregor, A. M. (1994). Autoimmune thyroid disease: Further developments in our understanding. *Endocrine Rev.* **15**, 788–830.
- Woolf, P. D. (1980). Transient painless thyroiditis with hyperthyroidism: A variant of lymphocytic thyroiditis? *Endocrine Rev.* **1**, 411–420.
- Zimmerman, D. (1999). Fetal and neonatal hyperthyroidism. *Thyroid* **9**, 727–733.

Table I Systemic Effects of Thyrotoxicosis

System	Effects
General	Weight loss, fatigue, heat intolerance, insomnia, nervousness, tremulousness
Skin	Fine, warm, and moist; hyperpigmentation; hyperhidrosis; onycholysis; fine and friable hair; urticaria
Eye	Exophthalmos, lid edema, lid lag, globe lag, chemosis, ophthalmoplegia, optic nerve involvement
Cardiovascular	Tachycardia, overactive heart, widened pulse pressure, bounding pulse; occasionally cardiomegaly, signs of congestive heart failure, paroxysmal tachycardia, atrial fibrillation
Gastrointestinal	Hyperphagia, increased thirst, increased frequency of stools or diarrhea, elevated liver function tests, hepatomegaly
Muscular	Tremulousness, weakness of proximal muscles, quickened and hypermetric reflexes, myopathy, muscle atrophy, periodic paralysis
Nervous	Anxiety, inability to concentrate, irritability, restlessness, lability, depression, psychiatric reactions
Respiratory	Dyspnea
Renal	Mild polyuria, polydipsia
Skeletal	Osteopenia or osteoporosis
Hematopoietic	Anemia (usually normochromic or normocytic), lymphocytosis, splenomegaly, lymphadenopathy, enlarged thymus
Endocrine	Irregular menses or amenorrhea, gynecomastia, decreased fertility
Metabolic	Elevated serum calcium, decreased serum magnesium, increased bone alkaline phosphatase, hypercalciuria

frequent in all forms of thyrotoxicosis and is responsible for the bright-eyed “stare” or “fish eyes” of the patient with thyrotoxicosis (Fig. 3). Lid lag is caused

Table II Clinical Findings in Patients with Graves' Hyperthyroidism and Controls (Percentages)

	Hyperthyroid	Controls
Cases	880	880
Symptoms		
Palpitations	65	13
Increased perspiration	45	7
Heat intolerance	55	8
Weight loss	61	13
Weight gain	12	21
Increased appetite	42	5
Decreased appetite	11	6
Increased number of bowel movements	22	2
Increased appetite with weight loss	24	0
Tiredness	69	41
Irritability	45	18
Nervousness	69	15
Signs		
Fine finger tremor	69	6
Pulse rate $\geq 90$ beats/min	19	78
Atrial fibrillation <sup>a</sup>	3	—
Thyroid size ( $\times$ normal)	$1.9 \pm 0.6$	$1.3 \pm 0.4$

Modified from Nodyke, R. A., Gilbert, F. I., Jr., and Harada, A. S. M. (1988). Graves' disease: Influence of age on clinical Findings. *Arch. Int. Med.* **148**, 626–631.

<sup>a</sup>The presence of atrial fibrillation was not assessed in controls.

by the fact that the upper lid lags behind the globe when the patient is asked to gaze downward; globe lag occurs when the globe lags behind the upper lid when the patient gazes slowly upward. These ocular manifestations appear to be the result of increased adrenergic activity. It is important to differentiate these ocular manifestations from those of infiltrative ophthalmopathy characteristic of Graves' disease.

### Cardiovascular System

The vascular manifestations of thyrotoxicosis constitute some of the most profound and characteristic symptoms and signs of the disorder. Tissue blood flow is increased in response to accelerated metabolism and increased oxygen consumption. Hemodynamic changes in thyrotoxic patients are characterized by an elevated cardiac output and a decreased peripheral vascular resistance. The mechanism responsible for the reduced vascular resistance is unclear. Thyroid hormone itself may be involved directly through its action on smooth muscles of blood vessels. Moreover, the finding in thyrotoxic patients of elevated levels of plasma adrenomedullin and proadrenomedullin-N-terminal 20-peptide, which have a potent vasodilatory activity, raises the possibility that these substances might also be involved in the decrease of vascular resistance in these patients.

Tachycardia is nearly always present, even at rest. The heart rate is also elevated during sleep; this helps to distinguish tachycardia of thyrotoxic origin from that of psychogenic origin. The pulse pressure is widened as a result of both an increase in cardiac



**Figure 1** Dermopathy of Graves' disease. Marked thickening of the skin is noted, usually over the pretibial area. Thickening will occasionally extend downward over the ankle and the dorsal aspect of the foot but almost never above the knee.

output and a decrease in peripheral vascular resistance. Other common cardiovascular symptoms include exercise intolerance and dyspnea on exertion. The latter is usually present with sustained activity but may also arise with an activity as limited as climbing a flight of stairs. Because of the diffuse and forceful nature of the apex beat, the heart may appear to be enlarged on physical examination, but echocardiography is usually normal.

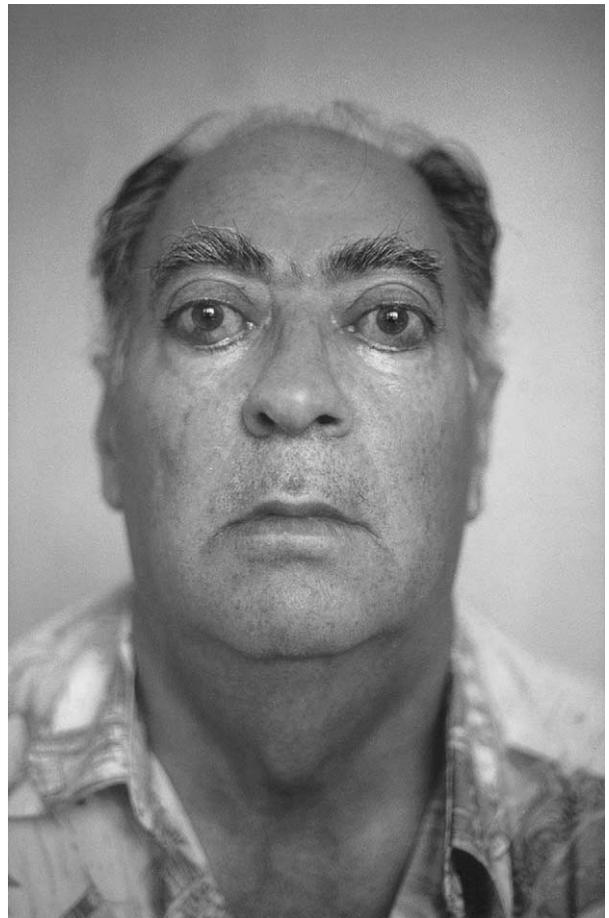
In aged thyrotoxic patients, the cardiovascular manifestations may be limited to resting tachycardia. Other classic thyrotoxic symptoms may be absent, possibly due to the relative paucity of adrenergic activity.

Thyrotoxic patients may have chest pain similar to angina pectoris in nearly all respects, probably caused by either relative myocardial ischemia or coronary artery spasm. However, in older patients, the increased



**Figure 2** Thyroid acropachy. Photograph of the hands in a patient with acropachy shows clubbing and swelling of fingertips.

myocardial oxygen demand due to thyrotoxicosis may unmask coronary artery disease. The plasma level



**Figure 3** Clinical presentation of Graves' ophthalmopathy. Photograph shows proptosis, marked conjunctival injection and chemosis, and retraction of lower eyelids.

of homocysteine, an independent risk factor for cardiovascular disease, in thyrotoxic patients did not differ significantly from that in controls.

Heart sounds are loud and ringing, and a systolic murmur—or even a late diastolic or presystolic murmur—may be present at apex. Auscultation may reveal a systolic ejection murmur and a gallop rhythm caused by rapid flow of blood through the aortic outflow tract. Systolic murmurs may arise from valve prolapse, left ventricular dilatation, or dysfunction of the mitral valve apparatus. Mild edema sometimes occurs in the absence of heart failure. Heart failure rarely occurs in thyrotoxic patients unless an underlying cardiac disease is also present.

Cardiac arrhythmias are common with thyrotoxicosis and are almost invariably supraventricular. Approximately 10% of patients with thyrotoxicosis have atrial fibrillation, and a similar percentage of patients with otherwise unexplained atrial fibrillation are thyrotoxic. This manifestation may be the presenting symptom of thyrotoxicosis, particularly in the elderly, and the risk of developing persistent atrial fibrillation is approximately three times that in normal individuals. Paroxysmal supraventricular tachycardia may be demonstrable or may be suggested by the history. Ventricular premature contractions are rare. Angina pectoris and myocardial infarction may occur rarely in the absence of coronary artery disease.

Nonspecific electrocardiographic changes may occur in thyrotoxicosis. A shortening of the PR interval is common, secondary to the increased rate of conduction through the atrioventricular node.

Thyrotoxicosis alone may determine heart failure in both old patients and (much less often) young patients. In large clinical studies, thyrotoxic patients with heart failure were generally old and, therefore, at risk for underlying heart disease and had chronic thyrotoxicosis. Elderly patients with rhythm disturbances, including atrial fibrillation, are at the greatest risk for heart failure. In the absence of atrial fibrillation, heart failure is rare. In the absence of underlying heart disease or in young patients, the heart failure is thought to be “high output.” High-output heart failure might not be a true heart failure; instead, it might be a circulatory congestion caused by fluid retention. In thyrotoxicosis, cardiac output is potentially near to maximal at rest and cannot increase in response to exercise, stress, surgery, or pregnancy. As a consequence, atrial filling pressures rise, leading to pulmonary and peripheral edema. This situation may be worse if atrial fibrillation is present. Left ventricular function is impaired because the persistent tachyarrhythmia alters this function. Sustained

tachycardia causes abnormal ventricular systolic and diastolic function that resolves when arrhythmia is treated.  $\beta$ -adrenergic receptor blockade-mediated slowing of the heart rate can rapidly reverse even severe degrees of left ventricular dysfunction in thyrotoxic patients.

## Gastrointestinal System

The classical gastrointestinal manifestations of thyrotoxicosis are rapid intestinal transit, increased frequency of semiformal stools, and weight loss from increased caloric requirement or malabsorption. These changes are not necessarily frequent. An increase in appetite, both during and between meals, is a common symptom but is usually not seen in patients with mild disease. In severe disease, the increased intake of food is usually inadequate to meet the increased caloric requirements, and weight loss occurs. Anorexia, rather than hyperphagia, sometimes accompanies severe thyrotoxicosis. It occurs in approximately one-third of elderly patients and contributes to the picture of “apathetic” thyrotoxicosis.

Frequent bowel movements are significantly more common in thyrotoxic patients than in normal controls. Diarrhea is rare. When constipation is present before the development of thyrotoxicosis, bowel function may become normal. More often, stools are less well formed and the frequency of bowel movements is increased. Gastric emptying and intestinal motility are increased, and these changes appear to be responsible for slight malabsorption of fat. Gluten enteropathy and Graves' disease may coexist more frequently than can be accounted for by chance due to their common autoimmune origin.

Hepatic dysfunction occurs in thyrotoxicosis, particularly when the disease is severe; hypoproteinemias and increased serum alkaline phosphatase and transaminase levels may be present. In severe cases, hepatomegaly and jaundice may be found.

## Nervous System

Hyperactivity, emotional lability, distractibility, and anxiety observed in thyrotoxicosis may reflect changes in the nervous system, but the pathogenetic mechanisms remain obscure. The reaction to all sorts of stimuli is distinctly excessive. Examination reveals a fine rhythmic tremor of the hands, tongue, or slightly closed eyelids. Emotional lability causes patients to lose their tempers easily and to have episodes of crying without any apparent reason. Crying may be evoked by merely questioning patients about the

symptom. In rare cases, mental disturbance may be severe. Rarely, patients develop visual or auditory hallucinations or a frank psychosis. It is probable that thyrotoxicosis makes manifest an abnormality already present rather than inducing a psychosis *de novo*. Anxiety is characterized by restlessness, shortness of attention span, and a compulsion to be moving around despite a feeling of fatigue.

Fine hand tremor is a frequent finding and may sometimes mimic that of Parkinsonism, and a preexisting Parkinsonian tremor can be accentuated. Chorea seldom appears as a manifestation of thyrotoxicosis. The neurological manifestation of thyrotoxic crisis rarely includes coma and status epilepticus. Patients with a convulsive disorder may become more difficult to control with the usual medications, and seizures may appear in patients who never manifested such symptoms previously.

The electroencephalogram of most thyrotoxic patients reveals an increased fast wave activity. The basal metabolic rate tends to correlate with the frequency of brain waves, but the correlation is usually poor at the extremes of thyroid abnormality.

## Muscle

Muscle weakness and fatigue are frequent. In most instances, they are not accompanied by objective evidence of local disease of muscle except for the generalized wasting associated with weight loss. Weakness is often most prominent in the proximal muscles of the limbs, causing difficulties in climbing stairs or in maintaining the leg in an extended position. In severe untreated cases, muscle wasting occasionally occurs as a predominant symptom (thyrotoxic myopathy). In extreme forms, the patient may be unable to rise from a sitting or lying position and may be virtually unable to walk.

Muscle manifestations affect men with thyrotoxicosis more commonly than they do women and may overshadow other manifestations of the syndrome. In severe forms, the myopathy involves mainly distal muscles of extremities and the muscles of the trunk and face. The involvement of ocular muscles may mimic myasthenia gravis. Graves' disease occurs in approximately 3 to 5% of patients with myasthenia gravis, and approximately 1% of patients with Graves' disease develop myasthenia gravis. Myasthenia gravis associated with Graves' disease has a mild expression characterized by preferential involvement of the eye muscles. Another myopathy sometimes observed in association with thyrotoxicosis is hypokalemic periodic paralysis. It is characterized by sporadic attacks

(which may last from minutes to many hours), most commonly involving flaccidity and paralysis of legs, arms, and/or trunk, although any muscle can be involved. Episodes can occur spontaneously, after carbohydrate ingestion, or after exercise. Hypokalemic periodic paralysis is most frequent in Asians.

## Respiratory System

Dyspnea is present in the large majority of severe thyrotoxic patients, and several factors may contribute to this condition, including reduction of vital capacity, decreased pulmonary compliance, weakness of the respiratory muscles, and increase in respiratory dead space ventilation. In some cases, it is difficult to separate patients with pure respiratory muscle weakness from patients who have only decreased lung compliance. Manifestations of respiratory muscle dysfunction include rapid shallow respirations, respiratory dyskinesia, hypoventilation, respiratory acidosis, and easy fatigability. Most patients with overt thyrotoxicosis have diminished proximal muscle strength. Pulmonary function returns to normal when the eumetabolic state is restored.

## Renal System

Most of the renal effects in thyrotoxic patients produce no symptoms except mild polyuria. Renal plasma flow and glomerular filtration rate are increased in thyrotoxicosis, probably because of the increase in cardiac output and decrease in peripheral resistance. Intrarenal vasodilation also occurs. The mean 24-h urine creatinine excretion is significantly lower in thyrotoxic patients than in normal individuals. The latter finding has been attributed to loss of muscle mass and occurs despite an increase in urea clearance. These changes are normalized when a normal metabolic state is restored.

Thyrotoxicosis is generally not associated with abnormalities in water metabolism. Serum electrolytes are usually normal. Some thyrotoxic patients have polydipsia, with 24-h urine volumes up to 3 to 4 L. Polyuria in these patients is due to increased thirst, as in primary polydipsia, and could be secondary to an increase of plasma angiotensin II concentration.

## Skeletal System: Calcium and Phosphorus Metabolism

Thyrotoxicosis is associated with an increase of bone turnover and eventually bone loss, especially in

postmenopausal women. Patients with a long-standing history of thyrotoxicosis may have overt osteoporosis and an increased risk of fractures. Bone turnover is increased, but the increase in bone resorption is relatively greater than that of bone formation, so the urinary excretion of calcium, phosphorus, and hydroxyproline is increased. As a consequence of this acceleration in bone resorption, hypercalcemia may occur in a significant proportion of patients with thyrotoxicosis. Total serum calcium may be slightly increased in up to 27% of these patients, and the ionized serum calcium level may be increased in up to 47%. The concentration of alkaline phosphatase and osteocalcin is also frequently increased. Parathyroid hormone and 1,25-dihydroxy-vitamin D3 levels tend to be low as a result of the increased calcium released from bone. Excretion of calcium in the feces is also increased in thyrotoxic patients. The secretions of the gastrointestinal tract are altered in thyrotoxicosis, and the transit time of calcium in the intestine is shortened.

### Hematopoietic System

The red blood cells are usually normal, but the red blood cell mass is increased. The increase in erythropoiesis appears to be due both to a direct effect of thyroid hormones on the erythroid marrow and to an increased production of erythropoietin. A parallel increase in plasma volume also occurs; therefore, hematocrit value is normal.

The most common red blood cell morphological abnormality is microcytosis, which is found in at least 37% of patients. The cause of this change is unclear. Iron deficiency is occasionally reported in thyrotoxic states. Microcytosis usually resolves with the restoration of euthyroidism. Some patients with severe thyrotoxicosis may develop a normocytic anemia. Defective iron use has been shown to occur in thyrotoxic patients and may be responsible for the development of anemia. Approximately 3% of patients with Graves' disease have pernicious anemia, and a further 3% have antibodies to intrinsic factor but normal absorption of vitamin B12. Autoantibodies against gastric parietal cells are present in about one-third of patients with Graves' disease, and the requirements for vitamin B12 and folic acid appear to be increased.

The total white blood cell count is often low because of a decrease in the number of neutrophils. The absolute lymphocyte count is normal or increased,

leading to a relative lymphocytosis. The numbers of monocytes and eosinophils may also be increased.

Blood platelets and the intrinsic clotting mechanism are normal. However, the concentration of factor VIII is often increased and returns to normal when thyrotoxicosis is treated. Furthermore, there is an enhanced sensitivity to coumarin anticoagulants because of an accelerated clearance of vitamin K-dependent clotting factors.

### Endocrine System

Thyrotoxicosis affects the secretion of most pituitary hormones, in particular the secretion of growth hormone (GH), prolactin (PRL), adrenocorticotropin (ACTH), follicle-stimulating hormone (FSH), and luteinizing hormone (LH).

Children with thyrotoxicosis grow more rapidly than do normal children. Growth acceleration in thyrotoxicosis suggests that GH secretion might be greater than normal. However, serum GH concentrations are lower in thyrotoxic patients than in normal individuals. This decrease is probably due to the increased metabolic clearance rate. Serum insulin-like growth factor-1 (IGF-1) concentration is higher in thyrotoxic patients and returns to normal values after restoration of the euthyroid state.

Thyrotoxicosis has several effects on adrenocortical function and adrenocortical hormone metabolism, with an increased clearance of the latter. The half-life of cortisol is shortened, but both the number of bursts of ACTH and the resulting burst of cortisol secretion are increased and maintain serum cortisol levels. The plasma concentration of corticosteroid-binding globulin is normal. The urinary excretion of the free cortisol and 17-hydroxycorticosteroids is normal or slightly increased, whereas the urinary excretion of 17-ketosteroids may be reduced. The rate of turnover of aldosterone is increased, but its plasma concentration is normal. Plasma renin activity is increased, and the sensitivity to angiotensin II is reduced.

$\beta$ -adrenergic receptor blockade ameliorates most of the cardiovascular manifestations of thyrotoxicosis. This suggests that catecholamines play a role in their genesis, but the secretion rate and plasma levels of epinephrine and norepinephrine are normal in thyrotoxic patients. Indeed, the apparent sympathetic hyperactivity appears to be the consequence of a direct effect of thyroid hormones on peripheral tissues.

Thyrotoxicosis during early life may cause delayed sexual maturation, although physical development is normal and skeletal growth may be accelerated.

## Reproductive System

Thyrotoxicosis, after puberty, influences the reproductive function, especially in women. An increase in libido occurs in both genders. The intermenstrual interval may be prolonged or shortened, and menstrual flow initially diminishes and ultimately ceases. Fertility may be reduced. In some women, menstrual cycles are predominantly anovulatory with oligomenorrhea; however, in most women, ovulation occurs. With treatment, menstrual cycles return to their regular pattern. Thyrotoxicosis in prepubertal girls may result in slightly delayed menarche. In premenopausal women with thyrotoxicosis, basal plasma concentrations of LH and FSH are normal but may display an enhanced responsiveness to LH-releasing hormone.

An increase in sex hormone-binding globulin (SHBG) is a prominent feature of thyrotoxicosis and is responsible for many of the alterations in steroid metabolism. Because of the increase in SHBG, the metabolic clearance rates of testosterone and (to some extent) estradiol are decreased. Testosterone levels are elevated because of the increased concentration of SHBG. Free testosterone levels tend to be normal. The metabolic clearance rate of estradiol is normal, suggesting that tissue metabolism of the hormone is increased. Conversion rates of androstenedione to testosterone, estrone, and estradiol, as well as conversion rates of testosterone to dihydrotestosterone, are increased. Extragonadal conversion of androgens to estrogens is increased, and this could be the mechanism responsible for gynecomastia observed in a consistent minority of thyrotoxic men.

## Energy Metabolism: Protein, Carbohydrate, and Lipid Metabolism

One of the most prominent symptoms in the hyperthyroid patient is heat intolerance. The symptom reflects an increase in the basal metabolism of many substrates. The increase in metabolic activity results in increased consumption of adenosine triphosphate (ATP) and oxygen. Despite the increased food intake, a state of chronic caloric inadequacy often ensues, depending on the degree of increased metabolism, and becomes more pronounced with age. In addition to losing fat stores, there is often a loss of muscle mass, making weakness a common complaint. Both synthesis and degradation of proteins are increased, with the latter increased to a greater extent than the

former, so that there is a net decrease in tissue protein content.

The oral glucose tolerance test is often abnormal. The most common abnormality is a faster rise in plasma glucose after glucose ingestion, but some patients have a delayed peak plasma glucose or a peak value that is higher than that in normal individuals. These abnormalities may reflect changes in glucose absorption rather than metabolism given that many patients who have abnormal oral glucose tolerance have normal responses to intravenous glucose administration. Preexisting diabetes mellitus is aggravated by thyrotoxicosis, with one cause being increased degradation of insulin.

Both synthesis and clearance of cholesterol and triglycerides are increased, but the latter effect predominates, so that serum levels are generally low. Plasma phospholipid and low-density lipoprotein cholesterol concentrations fall, whereas high-density lipoprotein cholesterol levels increase. Finally, thyroid hormones may influence cholesterol metabolism by increasing its conversion to bile acid and its clearance through the membrane surface low-density lipoprotein receptors. Although fatty acid synthesis is increased in both adipose tissue and liver, degradation of most lipids appears to be stimulated out of proportion to synthesis; consequently, body lipid deposits become depleted and plasma concentrations of various lipid components fall.

Several studies have investigated the relationship between leptin level and thyroid status. With the exception of a couple of reports suggesting a relative hypoleptinemia, most clinical studies have found no effect of thyrotoxicosis on leptin levels.

## See Also the Following Articles

Thyrotoxicosis: Diagnosis • Thyrotoxicosis Factitia • Thyrotoxicosis, Overview of Causes • Thyrotoxicosis, Treatment • Thyrotoxic Storm

## Further Reading

- Chiovato, L., Mariotti, S., and Pinchera, A. (1997). Thyroid disease in the elderly. *Baillieres Clin. Endocrinol. Metab.* **11**, 251–270.
- Cummings, S. R., Nevitt, M. C., Browner, W., et al. (1995). Risk factors for hip fractures in white women: Study of Osteoporotic Fractures Research Group. *N. Engl. J. Med.* **332**, 767–773.
- Huang, M. J., Li, K. L., Wei, J. S., Wu, S. S., Fan, K. D., and Liaw, Y. F. (1994). Sequential liver and bone biochemical changes in hyperthyroidism: Prospective controlled follow-up study. *Am. J. Gastroenterol.* **89**, 1071–1076.
- Jandresic, D. P. (1990). Psychiatric aspects of hyperthyroidism. *J. Psychosomatic Res.* **34**, 603–615.

- Kendric, A. H., O'Reilly, J. R., and Laszlo, G. (1988). Lung function and exercise performance in hyperthyroidism before and after treatment. *Q. J. Med.* **68**, 615-618.
- Konishi, K., Hayashi, M., and Saruta, T. (1994). Renal tubular acidosis with autoantibody directed to renal-collecting duct cells. *N. Engl. J. Med.* **331**, 1593-1596.
- Rosen, T., and Klemm, W. R. (1992). Cutaneous manifestations of thyroid disease. *J. Am. Acad. Dermatol.* **26**, 885-887.
- Rosner, W. (1990). The functions of corticosteroid-binding globulin and sex hormone-binding globulin: Recent advances. *Endocrine Rev.* **11**, 80-84.
- Thomas, R., and Reid, R. L. (1987). Thyroid disease and reproductive dysfunction: A review. *Obstet. Gynecol.* **70**, 789.
- Tønner, D. R., and Schlecheter, J. A. (1993). Neurologic complications of thyroid and parathyroid disease. *Med. Clinics North America* **77**, 251-263.

Table I Pharmacological Features of Thionamides

	Methimazole	Propylthiouracil
Serum half-life (hours)	4–6	1–2
Serum protein binding (percentage)	—	80–90
Metabolism of drug during illness		
Severe liver disease	Decreased	Normal
Severe kidney disease	Normal	Normal
Levels in breast milk	Higher	Low

Modified from Cooper, D. S. (1996). Treatment of thyrotoxicosis. In "Werner and Ingbar's The Thyroid" (L. E. Braverman and R. D. Utiger, eds.), 7th ed., pp. 713–734. Lippincott-Raven, Philadelphia.

problem, may be observed with both MMI and PTU, but a low dose of MMI is safer than a high dose of either drug. Agranulocytosis may develop suddenly and usually occurs during the first months of therapy. It typically presents with fever and evidence of infection. In addition to prompt discontinuation of the antithyroid drug, treatment of agranulocytosis involves the administration of broad-spectrum antibiotics and growth factors to stimulate recovery of bone marrow. Patients usually recover within 2 to 3 weeks, but a few deaths from this complication have been reported. Hepatitis, vasculitis, and lupus-like syndromes are rare complications but make the discontinuation of antithyroid drugs mandatory. Minor side

Table II Side Effects of Thionamides

<b>Major (&lt; 1%)</b>
Rare
Agranulocytosis
Very rare
Severe vasculitis and Lupus-like syndrome
Cholestatic jaundice
Toxic hepatitis
Aplastic anemia
Thrombocytopenia
<b>Minor (1–5%)</b>
Common
Cutaneous manifestations (e.g., rash, pruritus, urticaria)
Transient granulocytopenia
Arthralgia
Less common
Fever
Gastrointestinal manifestations
Loss of taste sensation

effects are much more frequent, occurring in 1 to 5% of patients. Among them, pruritus, rash, and (less commonly) urticaria are the most prominent manifestations. These side effects may resolve spontaneously despite continued therapy, but they generally call for substitution of one thionamide for the other, although cross-sensitivity to these drugs may occur.

### β-Adrenergic Antagonist Drugs

These drugs are an integral part of the management of hyperthyroidism. Blockade of β-adrenergic receptors rapidly ameliorates some manifestations of thyrotoxicosis such as tremor, palpitation, and anxiety. β-adrenergic antagonists do not affect thyroid hormone synthesis and release; therefore, they should not be used alone except for short periods before and/or after radioiodine therapy in selected patients preparing for thyroid surgery or in patients with self-limited forms of thyrotoxicosis. Since the introduction of propranolol, a number of new agents with a longer duration (e.g., atenolol, metoprolol, nadolol), or with greater cardioselectivity (e.g., atenolol, metoprolol, bisoprolol), have become available. The usual contraindications of β-adrenergic antagonists, such as asthma, should be considered. Despite the initial concern, propranolol or other β-adrenergic antagonists are widely used in thyrotoxic heart disease in view of the notion that tachycardia and tachyarrhythmias are the most prominent factors involved in this condition.

### Iodide and Iodine-Containing Agents

Inorganic iodide given in pharmacological doses (as Lugol's solution or as saturated solution of potassium iodide [SSKI]) decreases its own transport into the thyroid, inhibits iodide organification (the Wolff-Chaikoff effect), and rapidly blocks the release of T4 and T3 from the gland. However, after a few days or weeks, its antithyroid action is lost and thyrotoxicosis recurs or may worsen. The usual dose of Lugol's solution is three to five drops three times daily and that of SSKI is one drop three times daily. Short-term iodide therapy is used to prepare patients for surgery, usually in combination with a thionamide drug. Iodide is also used in the management of severe thyrotoxicosis (thyroid storm) because of its ability to inhibit thyroid hormone release acutely.

Iopanoic acid and sodium ipodate, widely used in the past as oral cholecystographic agents, may be useful in the management of thyrotoxicosis. These compounds have a dual action: they produce a fall

in the serum concentration of thyroid hormones (resulting from the block of thyroid hormone secretion due to the inorganic iodide released from the drug) and inhibit the peripheral conversion of T4 to T3. Some reports have confirmed the efficacy of these compounds as a primary therapy of hyperthyroidism. These agents may be particularly useful when a rapid decrease in T3 levels is desired. Although an early escape from the therapeutic effect has been suggested, other studies have not found this to be the case for up to nearly 2 years of therapy.

### Perchlorate

Perchlorate interferes with accumulation of iodide by the thyroid. In conjunction with thionamides, it has been used successfully in the treatment of amiodarone-induced hyperthyroidism. Gastric irritation and toxic reactions (aplastic anemia) limit the long-term use of perchlorate in the management of hyperthyroidism.

### Glucocorticoids

Glucocorticoids in high doses inhibit the peripheral conversion of T4 to T3. In Graves' hyperthyroidism, glucocorticoids appear to decrease T4 secretion by the thyroid, but the efficiency and duration of this effect are unknown. In severe hyperthyroidism, short-term glucocorticoid administration may be used as a general supportive treatment. The immunosuppressive effect of glucocorticoids in high doses is commonly exploited in the treatment of ophthalmopathy and dermopathy of Graves' disease but not in the management of uncomplicated Graves' hyperthyroidism.

## RADIOIODINE

Among different radioactive isotopes of iodine,  $^{131}\text{I}$  is the agent of choice in the treatment of thyroid hyperfunction. After oral administration, radioiodine is completely absorbed, rapidly concentrated, oxidized, and organified by thyroid follicular cells. The destruction of thyroid cells produced by radioiodine results from the ionizing effects of  $\beta$ -particles that have a path length of 1 to 2 mm. Specifically,  $1\ \mu\text{Ci}$  of  $^{131}\text{I}$  retained per gram of thyroid tissue delivers approximately 70 to 90 rads. Biological effects of radioiodine include necrosis of follicular cells, shorter survival and impaired replication of nondestroyed cells, and vascular occlusion. In the long run, there is atrophy and fibrosis, as well as a chronic inflammatory response, that ultimately may result in thyroid failure.

The goal of radioiodine therapy of hyperthyroidism is to destroy sufficient thyroid tissue to cure hyperthyroidism with one dose of  $^{131}\text{I}$ . The administered dose is calculated on the basis of thyroid size and uptake of  $^{131}\text{I}$  using the following formula:

$$\begin{aligned} \text{Dose (mCi)} &= \text{Estimated Thyroid Weight(g)} \\ &\times \text{Planned Dose } (\mu\text{Ci/g}) / \\ &\quad \text{Fraction 24-h Radioiodine Uptake} \\ &\times 1000. \end{aligned}$$

In other centers, standard fixed doses are given. Because of radiation safety restrictions, in some centers, especially in Germany, repeated 2- to 3-mCi doses of radioiodine are administered. It is a common experience that small glands appear to be destroyed more readily by radioiodine than do larger ones and that toxic adenomas or toxic multinodular goiters are usually more radioresistant than Graves' glands.

The therapeutic effect of radioiodine is delayed, and in some patients up to several months may be required for the complete control of hyperthyroidism. Thus, a course of antithyroid drugs before the administration of radioiodine is frequently used, particularly in cases of severe thyrotoxicosis, to avoid the inconveniences of persistent hyperthyroidism. In selected patients with significant comorbidity, antithyroid drugs may also be administered following radioiodine, but doing so may decrease the efficacy of treatment.

Radioiodine therapy may lead to worsening of Graves' ophthalmopathy when present. This worsening can be prevented by concomitant glucocorticoid therapy, beginning after radioiodide administration.

In the past, the major concerns with radioiodine therapy derived from the possible carcinogenic effects of ionizing radiation and from the risks of congenital malformations in offspring of women treated during their childbearing years. No association was found between radioiodine administration and thyroid cancer in large epidemiological studies. Similarly, there is no evidence that radioiodine therapy for hyperthyroidism increases the risk of leukemia or solid tumors. No association between radioiodine treatment for hyperthyroidism and congenital abnormalities in subsequent offspring has been observed.

## THYROIDECTOMY

Subtotal or near-total thyroidectomy is performed in Graves' disease and toxic multinodular goiter, whereas lobectomy is the procedure of choice in toxic adenoma. Restoration of euthyroidism before surgery is mandatory. The classical approach combines a course of

thionamide treatment to restore and maintain euthyroidism and the preoperative administration of iodide for approximately 10 days to decrease blood flow of the gland. Care must be taken not to discontinue or decrease the dose of antithyroid drugs when iodide is added. We do not favor a preoperative program based only on the use of a  $\beta$ -adrenergic antagonist associated with iodide. Euthyroidism is not restored in these patients because iodide alone normalizes the concentration of thyroid hormones in serum only a few days before the operation but does not produce "tissue" euthyroidism. Furthermore, the risk of thyroid storm is not completely prevented even when high-dose propranolol is administered for several days after surgery.

Possible complications of thyroid surgery include thyroid storm (which is extremely rare nowadays), bleeding, injury to the recurrent laryngeal nerve, and hypoparathyroidism. In particular, the risk of laryngeal nerve injury and hypoparathyroidism cannot be disregarded. Although an incidence of these complications of less than 2% is reported from clinics with wide experience in thyroid surgery, much higher figures of up to 10 to 15% are encountered in some series. These complications are, of course, less frequent when lobectomy for toxic adenoma is performed.

## CHOICE OF THERAPY

The choice of treatment for Graves' hyperthyroidism involves both physician's prejudice and patient's preference. Our policy is to give radioiodine as a first-choice treatment for Graves' hyperthyroidism once euthyroidism has been restored and maintained by a short course (3–6 months) of antithyroid drugs. We do advise this type of treatment for all middle-aged and elderly people, and we offer this option to young adults as well. Our preference for this treatment strategy derives from two considerations. First, the recurrence rate of hyperthyroidism after antithyroid drugs is high. Second, radioiodine therapy for hyperthyroidism is effective, inexpensive, and safe.

If we take into account the possible side effects of long-term thionamide administration, the risks of recurrent thyrotoxicosis, and/or the complications of thyroid surgery, radioiodine appears to be a relatively safe form of treatment for Graves' hyperthyroidism. Acute complications of radioiodine are extremely rare if patients are rendered euthyroid with antithyroid drugs for a period of time sufficient to deplete intrathyroidal stores of hormones.

In patients with Graves' hyperthyroidism and ophthalmopathy, we favor the rapid restoration of

euthyroidism by antithyroid drugs, followed by radical treatment of hyperthyroidism by thyroidectomy or radioiodine, as indicated. In patients with mild or moderate ophthalmopathy, we favor protection from possible exacerbations of ophthalmopathy after radioiodine or surgery with a short-term course of glucocorticoids. In patients with severe ophthalmopathy, we favor radical treatment of hyperthyroidism associated with specific therapy for eye disease consisting of orbital irradiation and high-dose systemic glucocorticoids.

Thionamides are the first-choice treatment in pregnant women with hyperthyroidism. Radioiodine therapy is contraindicated during pregnancy, and surgery is restricted to exceptional cases. Both PTU and MMI cross the placenta and, in excessive doses, may cause hypothyroidism and goiter in the fetus and the neonate. Because PTU is less lipid soluble and more highly protein bound, its placental transfer appears to be lower, so that the preferential use of this drug in pregnancy has been advocated by some, mainly in the United States. However, one study showed no difference between PTU and MMI in suppressing fetal thyroid function. The drug chosen should be given at the lowest possible dose (but not higher than 10 mg MMI or 100 mg PTU daily as maintenance dose) to maintain maternal-free T<sub>4</sub> and free T<sub>3</sub> in the high-normal range.

Graves' disease is the main cause of hyperthyroidism in children and adolescents. In this age group, the peak incidence of the disease is between 11 and 15 years of age, but it may occur in children under 5 years of age. Antithyroid drugs, radioiodine, and surgery have been successfully used in children, but we favor thionamides as the first-choice treatment.

Hyperthyroidism due to toxic adenoma responds well to radioiodine, and the nodule is partially reduced by this form of therapy. In severe hyperthyroidism, pretreatment with thionamides is indicated, but radioiodine should be given before restoration of extra-nodular thyroid function by thyrotropin-stimulating hormone (TSH).

Surgery is a more appropriate choice in large (>5 cm) nodules with associated pressure symptoms (Table III). Young age also favors surgery with respect to radioiodine. Lobectomy is indicated in classic toxic adenoma in otherwise normal thyroid glands. When single hyperfunctioning thyroid nodules occur in a multinodular goiter, subtotal or near-total thyroidectomy is indicated.

Toxic multinodular goiter is more frequent in patients over 50 years of age who may have other non-thyroidal illnesses adding a risk to surgery (Table III).

Table III Factors Favoring Specific Forms of Treatment of Toxic Adenoma and Uninodular and Toxic Multinodular Goiter

<b>Surgery</b>
Young-adult age
Large goiter or adenomas greater than 5 cm
Low radioiodine uptake and large areas of poorly functioning tissue within the goiter
Airway obstruction or compression of other structures
Refusal to take radioiodine
<b>Radioiodine</b>
Small adenomas (< 5 cm)
Old age
Contraindications to surgery

Therefore, many patients with toxic multinodular goiters receive radioiodine. Hyperthyroidism is cured in virtually all cases, although more than one dose of  $^{131}\text{I}$  may be required. However, the reduction in thyroid size is only partial because of the presence of intrathyroidal calcifications, fibrosis, and large areas of nonfunctioning tissue. Nevertheless, in patients with large mediastinal goiters but contraindications to surgery, radioiodine may be used with benefit and partial relief of pressure symptoms.

Surgery is generally regarded as a better option in patients with large goiters, especially if there is evidence of substernal extension and compression of airways and blood vessels. Surgery is also indicated in those instances when radioiodine uptake by the gland is relatively low, as frequently occurs in large multinodular goiters.

## See Also the Following Articles

Thyroidectomy • Thyrotoxic Storm • Thyrotoxicosis: Diagnosis • Thyrotoxicosis Factitia • Thyrotoxicosis, Overview of Causes • Thyrotoxicosis, Systemic Manifestations

## Further Reading

- Allhabadia, A., Daykin, J., Sheppard, M. C., Gough, J. C., and Francklyn, J. A. (2001). Radioiodine treatment of hyperthyroidism: Prognostic factors for outcome. *J. Clin. Endocrinol. Metab.* **86**, 3611–3617.
- Bartalena, L., Pinchera, A., and Marcocci, C. (2000). Management of Graves' ophthalmopathy: Reality and perspectives. *Endocrine Rev.* **21**, 168–199.
- Braga, M., and Cooper, D. S. (2001). Oral cholecystographic agents and the thyroid. *J. Clin. Endocrinol. Metab.* **86**, 1853–1860.
- Fontanilla, J. C., Schneider, A. B., and Sarne, D. H. (2001). The use of oral radiographic contrast agents in the management of hyperthyroidism. *Thyroid* **11**, 561–567.
- Homsanit, M., Sriussadaporn, S., Vannasaeng, S., Peerapatdit, T., Nitiyanant, W., and Vichayanrat, A. (2001). Efficacy of single daily dosage of methimazole vs. propylthiouracil in the induction of euthyroidism. *Clin. Endocrinol.* **54**, 385–390.
- Marcocci, C., Bartalena, L., and Pinchera, A. (1998). Ablative or nonablative therapy for Graves' hyperthyroidism in patients with ophthalmopathy? *J. Endocrinol. Invest.* **21**, 468–471.
- Marcocci, C., Bruno-Bossio, G., Manetti, L., Tanda, M. L., Miccoli, P., Iacconi, P., Bartolomei, M. P., Nardi, M., Pinchera, A., and Bartalena, L. (1999). The course of Graves' ophthalmopathy is not influenced by near-total thyroidectomy: A case-control study. *Clin. Endocrinol.* **51**, 503–508.
- Miccoli, P., Vitti, P., Rago, T., Iacconi, P., Bartalena, L., Bogazzi, F., Fiore, E., Valeriano, R., Chiovato, L., Rocchi, R., and Pinchera, A. (1996). Surgical treatment of Graves' disease: Subtotal or total thyroidectomy? *Surgery* **120**, 1020–1025.
- Palit, T. K., Miller, C. C., and Miltenburg, D. M. (2000). The efficacy of thyroidectomy for Graves' disease: A meta-analysis. *J. Surg. Res.* **90**, 161–165.

**Table 1 Causes of Low Serum TSH Levels in the Absence of Thyrotoxicosis**

Nonthyroidal chronic or acute illness
Starvation and malnutrition, anorexia nervosa
Hypopituitarism
Hypercortisolism
Major depression
Anorexia nervosa
Early pregnancy
Drugs
Dopamine agonists
Somatostatin
Glucocorticoids
Triiodoacetic acid

satisfactory. In most iodine-sufficient countries, a single FT4 measurement is sufficient to confirm or reject the suspicion of thyrotoxicosis. In contrast, in iodine-deficient countries, a significant proportion of hyperthyroid patients (up to 12%) may have normal FT4 levels with elevated FT3 levels (T3 toxicosis). Conversely, FT4 can be falsely elevated in conditions causing reduced peripheral conversion of T4 to T3 such as amiodarone or high-dose propranolol treatment. We prefer to assess both FT4 and FT3 levels together with TSH to obtain a complete baseline panel of the thyroid function status in every patient where the diagnosis of thyrotoxicosis is suspected.

A condition characterized by low or undetectable TSH level and normal free thyroid hormone levels, detected occasionally at routine thyroid function testing or in patients complaining of mild thyrotoxic symptoms, is termed "subclinical thyrotoxicosis."

## DIFFERENTIAL DIAGNOSIS OF THYROTOXICOSIS: LABORATORY AND INSTRUMENTAL INVESTIGATIONS

History and physical examination are usually sufficient to identify the cause of thyrotoxicosis. However, in some cases, a careful differential diagnosis is needed to establish an etiological diagnosis. This is necessary to plan the correct treatment. For many years, radioiodine uptake (RAIU) has represented a mainstay of the differential diagnosis of thyrotoxicosis. RAIU is easily performed by administering a minimal (tracer) dose of radioactive iodine and then measuring the percentage of administered radioactivity accumulated in the thyroid. The upper normal limit of RAIU in iodine-sufficient countries, 24h

after the administration of the tracer, is approximately 25%, whereas it may reach 40% in areas with mild to moderate iodine deficiency. RAIU represents an indirect method of estimating the thyroidal intracellular mechanisms for iodine trapping and organification. Whenever excessive active formation of thyroid hormone takes place in the thyroid gland, RAIU is increased. Therefore, a high RAIU readily identifies true hyperthyroidism (e.g., with thyroid hyperfunction). In contrast, thyrotoxicosis with a low RAIU indicates either thyroidal destruction, with release of preformed hormone, or an extrathyroidal source of thyroid hormone. In thyroid destruction, the damaged follicular cell loses its capability of iodine trapping. Conversely, when exogenous hormones are administered in excess, the suppression of the pituitary secretion of TSH causes a block of the trapping capacity of follicular cells. The only exception to this rule is iodine-induced thyrotoxicosis, in which a low RAIU can be observed because of dilution of the tracer dose in the large body pool of iodine in spite of true hyperthyroidism.

Nowadays, RAIU is not universally performed in the initial assessment of a thyrotoxic patient, and a vast array of laboratory and imaging techniques are available to accurately identify the cause of thyrotoxicosis. However, RAIU may still be used as a starting point to broadly define forms of thyrotoxicosis according to their pathogenesis before proceeding with the use of adjunctive diagnostic tools.

## Graves' Disease

### Anti-TSH Receptor Antibodies

A mainstay in the diagnosis of Graves' disease resides in serum detection of TSH receptor antibodies (TRABs). TRABs can be measured by different methods. TRABs were originally detected with *in vivo* bioassays, but this method has been replaced by *in vitro* systems. The most commonly used assay is based on the displacement of radiolabeled TSH from its receptor by the patients' sera. Antibodies detected with this method have been termed "TSH binding-inhibiting immunoglobulins" (TBIs). This test does not provide any information on the biological activity of the detected antibodies and, therefore, cannot distinguish between antibodies with stimulating activity (TSABs) and those with blocking activity (TSHBAs) that can also be detected in thyroid autoimmune disorders. The stimulating activity of TRABs can be tested by employing cellular systems carrying a functional TSH receptor (e.g., Chinese hamster ovary cells transfected with cloned human TSH receptor) and detecting the

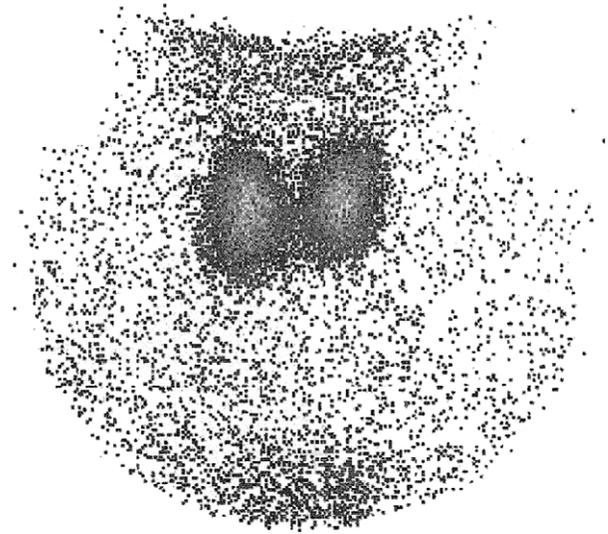
release of cyclic AMP (cAMP) in the culture medium on challenge with serum or purified immunoglobulins (TSAB assay). In a modification of the assay, TSHBABs can be detected as well. TSABs are the cause of hyperthyroidism in Graves' disease. Unfortunately, the assay is quite expensive and requires cell culture capabilities, making it available only to research centers. For clinical purposes, the TBII assay is used most often. By past-generation assays, positive TBII tests are found in 75 to 95% of patients, with a high specificity (99%). Therefore, the TBII test represents an excellent tool in the diagnosis of Graves' disease. In spite of its efficiency, a TBII test is needed only in the minority of cases where the clinical picture is unclear, for example, in the differential diagnosis of hyperemesis gravidarum, in the nodular variants of Graves' disease that must be differentiated from toxic nodular goiter, and in patients with exophthalmos without thyrotoxicosis (euthyroid Graves' disease). Therefore, the TBII test should be considered a second-line test in the diagnosis of Graves' disease. The finding of TBII after long-term treatment with antithyroid drugs may be useful to identify patients prone to recurrence of hyperthyroidism.

#### **Antithyroid Peroxidase and Antithyroglobulin Antibodies**

Antithyroid peroxidase (AbTPO) antibodies can be found by commercial radioimmunoassays in up to 90% of patients with untreated Graves' disease, whereas antithyroglobulin (AbTG) antibodies are less frequently positive (in about 50–80% of cases). However, both antibodies are also present in other forms of thyroid autoimmune disorders, some of which may cause thyrotoxicosis, such as postpartum thyroiditis and silent thyroiditis. A relatively high percentage (up to 25%) of positive tests is also found in normal individuals, especially women. Thus, anti-TPO and anti-TG tests do not establish the diagnosis of Graves' disease as the cause of thyrotoxicosis but may be useful as complementary tests in confirming the presence of thyroid autoimmunity.

#### **Thyroid RAIU and Scan**

A high value of RAIU at the 24th hour is always found as a distinctive feature in untreated hyperthyroid Graves' disease patients. In some cases, the 3rd- or 6th-hour value can be even higher than the 24th-hour value, as an expression of an extremely high iodine turnover. The test is very useful for ruling out transient thyrotoxicosis due to hashitoxicosis or painless or subacute thyroiditis, factitious thyrotoxicosis, and type II amiodarone-induced thyrotoxicosis.



**Figure 1** Thyroid scanning (performed with radioiodine) in Graves' disease.

Thyroid scanning in Graves' disease is useful when coexisting nodules are detected by palpation and their functional status needs to be evaluated. Thyroid imaging can be performed with radioiodine at the time of RAIU or by using  $^{99m}\text{Tc}$ perchnetate (Fig. 1).

#### **Thyroid Ultrasound**

The thyroid gland in Graves' disease hyperthyroidism has a typical ultrasound pattern. Because of both the reduction in the colloid content and the lymphocytic infiltrate, the gland becomes diffusely hypoechoic. A similar pattern is also observed in goitrous autoimmune thyroiditis. Therefore, thyroid ultrasound can be useful for confirming the suspicion of thyroid autoimmunity during the evaluation of thyrotoxicosis. Moreover, thyroid ultrasound scanning allows an accurate measurement of the goiter size. This information is important in choosing the most appropriate treatment. Finally, thyroid ultrasound accurately distinguishes true thyroid nodules from the lobulations that can be felt occasionally at palpation in Graves' disease glands. Therefore, the information provided by thyroid ultrasound is quite useful in the initial evaluation of Graves' disease patients, although it is not strictly needed from a diagnostic standpoint.

The measurement of blood flow to the thyroid gland by color flow Doppler ultrasound has also been used experimentally in Graves' disease patients. In untreated Graves' disease, the color Doppler pattern is characterized by markedly increased signals with a patchy distribution. In the setting of a hypoechoic pattern at ultrasound, the detection of an

increased blood flow allows the distinction from Hashimoto's thyroiditis. Therefore, color Doppler studies of the thyroid gland can be useful, as is RAIU, in distinguishing Graves' disease from other forms of thyrotoxicosis such as amiodarone-induced destructive thyrotoxicosis, subacute thyroiditis, and painless thyroiditis (in which the blood flow is reduced or absent).

### Toxic Adenoma

The presence of a toxic adenoma must always be suspected in a thyrotoxic patient with a single thyroid nodule revealed at neck palpation. In confirming the diagnosis of thyrotoxicosis, it is important to measure both FT4 and FT3 levels because T3 toxicosis is distinctly frequent in toxic adenomas.

At  $^{99m}$ technetium or radioiodine thyroid scanning, the nodule will appear "warm" when only partial autonomy is present, with the extranodular thyroid tissue clearly visible. In this case, parallel thyroid function tests will show a low but detectable TSH and thyroid hormone levels that are normal or at the upper limit of the normal range. When TSH is completely suppressed (e.g., complete autonomy, overt thyrotoxicosis), the scan will show only the autonomous nodule with complete suppression of the extranodular tissue. Ultrasound scanning of the neck provides no direct diagnostic information on the functional property of the nodule, but it is useful in detecting coexisting nodules (which can eventually be cold at scan) and accurately defines the size of the nodule. Preliminary reports have shown a distinctive color Doppler pattern in autonomously functioning thyroid nodules, characterized by an increased blood flow in the nodular tissue, in good correlation with radionuclide scans. This technique is not able to distinguish benign nodules from malignant ones; therefore, it is of limited value. Fine needle aspiration biopsy is not useful in the evaluation of a thyroid hot nodule and provides inconclusive findings that often show undetermined follicular neoplasm. The risk of malignancy in hot nodules is extremely low, although they are occasionally reported. Therefore, in the presence of a low TSH, fine needle aspiration is needed only when coexisting nodules detected by palpation or ultrasound are cold at radionuclide scanning.

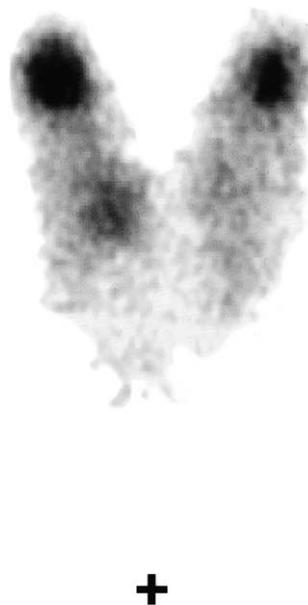
In the past, T3 in suppressive doses was administered for 5 to 7 days to patients with warm nodules to completely suppress TSH secretion, and a scan was performed thereafter (Werner's test). The purpose of the test was to show autonomous radionuclide uptake in the nodular tissue only, confirming the suspicion of

nodular functional autonomy. This test has become obsolete with the availability of ultrasensitive TSH assays.

Further imaging, such as neck X rays, barium swallow, and computed tomography (CT) scans, may be needed in selected patients with large nodules to evaluate the presence of significant tracheal and/or esophageal compression. It is important to remember that CT scan, when done with this purpose, should always be performed without the administration of iodinated contrast media because these may worsen thyrotoxicosis or precipitate it in partially autonomous nodules.

### Toxic Multinodular Goiter

The same alterations of thyroid function tests described in toxic adenomas can be observed in toxic multinodular goiter. Physical findings and history are often sufficient to suspect the presence of a toxic multinodular goiter. Thyroid radionuclide scanning is quite useful for identifying and mapping autonomous nodules and cold nodules that are often coexistent (Fig. 2). Scanning is also useful as an adjunct to TRAB measurement in distinguishing true toxic multinodular goiter from Graves' disease hyperthyroidism overimposed on a preexisting nontoxic multinodular goiter. RAIU is always elevated, unless iodine overload is present, but is not always necessary to establish the diagnosis.



**Figure 2** Thyroid scanning (performed with  $^{99m}$ perchnetate) in toxic multinodular goiter.

Thyroid ultrasound scan is also useful in measuring the size of the goiter. Moreover, relative to radio-nuclide scanning images, it is helpful in identifying cold nodules. Fine needle biopsy should be performed in any palpable dominant nodule that is cold at scan.

### TSH-Secreting Adenomas

Patients with TSH-secreting adenomas present a detectable TSH in the presence of clearly elevated circulating thyroid hormone levels (inappropriate secretion of TSH).

The first step in the evaluation of inappropriate secretion of TSH is making sure that interferences in the measurement of TSH (by heterophilic antibodies) or thyroid hormone levels (by thyroid hormone antibodies) are not the cause of the laboratory findings. Once these interferences are ruled out, further investigations are required to differentiate patients with TSH-secreting pituitary adenomas from those with resistance to thyroid hormone.

TSH-secreting adenomas abnormally secrete the  $\alpha$ -subunit of TSH in molar excess with respect to TSH. A serum  $\alpha$ -subunit/TSH ratio greater than 1 is observed in approximately 90% of patients with TSH-secreting adenomas. High ratios can also be observed in postmenopausal women and even in normal individuals, making this test alone unable to establish the diagnosis. Growth hormone (GH), insulin-like growth factor-1 (IGF-1), and prolactin serum measurements are useful because approximately 30% of TSH-secreting adenomas cosecrete these hormones.

Dynamic tests may be useful for demonstrating the unresponsiveness of TSH in patients with TSH-secreting adenomas. In most TSH-secreting tumors (92%), the TSH level fails to increase in response to a standard TRH stimulation test, whereas a normal or increased response is observed in resistance to thyroid hormone. It can be useful to investigate the response of pituitary TSH to exogenous T3. T3 is administered

orally, and the dose is increased every 3 days (up to 200 mg daily). Before every increase, basal and TRH-stimulated TSH levels are measured together with peripheral markers of thyroid hormone action. Only partial or no suppression of TSH secretion is observed in TSH-secreting adenomas, whereas complete suppression is observed in resistance to thyroid hormone. The test is contraindicated in elderly patients and in patients with arrhythmias and/or coronary artery disease. Available tests in the differential diagnosis of the syndrome of inappropriate secretion of TSH are described in Table II.

Pituitary imaging is very important in confirming the diagnosis. Fully 90% of TSH-secreting adenomas are larger than 1 cm at diagnosis and, therefore, are easily detected at pituitary magnetic resonance imaging (MRI) scanning. As a complement, radiolabeled octreotide pituitary scintigraphy can be used and may be particularly useful for detecting small tumors.

### Human Chorionic Gonadotropin-Dependent Thyrotoxicosis

The diagnosis of thyrotoxicosis during hyperemesis gravidarum can be particularly difficult. Because of weight loss and malnutrition, FT3 levels may be disproportionately low or even normal in comparison with FT4 levels due to a reduced peripheral conversion of T4 to T3. The TSH level is often low during early normal pregnancy, but it is seldom undetectable as it is in true thyrotoxicosis. The only distinctive laboratory feature is an inappropriately high human chorionic gonadotropin (hCG) level, but a large overlap with normal pregnancies exists. Therefore, the diagnosis of thyrotoxicosis in hyperemesis gravidarum relies mainly on the clinical picture and on appropriate exclusion of other more common forms of hyperthyroidism by specific testing. It is important to remember that RAIU, as well as any other *in vivo* radioisotopic procedure, is absolutely contraindicated during pregnancy.

Table II Laboratory Investigations in the Differential Diagnosis of the Syndrome of Inappropriate Secretion of TSH

Test	TSH-secreting adenomas	TRH	Comment
Peripheral markers of thyroid hormone action	High	Normal-high	Nonspecific
$\alpha$ -subunit/TSH molar ratio	>1	1	High during menopause
TSH after T3 suppression test	Unchanged or slightly reduced	Frankly reduced or suppressed	Hazardous in elderly and cardiopathic patients
TSH after TRH	Unchanged	Increased	
Pituitary imaging	Positive	Negative	Confirmatory

The presence of a trophoblastic tumor should be suspected when thyrotoxicosis is diagnosed in an amenorrheic woman, especially when a palpable abdominal mass is present. The diagnosis is readily confirmed by the finding of extremely high circulating hCG levels and a pelvic mass at ultrasonography.

Given the extreme rarity of the disorder, no diagnostic guidelines are available for familial gestational hyperthyroidism. The diagnosis can be suspected only on a history of recurrent pregnancy-associated hyperthyroidism and can be confirmed only by the demonstration of a mutation of the TSH receptor gene, available in very few research laboratories worldwide.

### Fetal and Neonatal Hyperthyroidism

Mothers with a past or current history of Graves' disease should be carefully monitored throughout pregnancy. The persistence of high TRAB levels in the maternal serum by the end of pregnancy, when the transplacental passage is maximal, is a predictor of hyperthyroidism in the neonate. Fetuses of mothers with Graves' disease who have been previously treated with radioiodine or surgery may be at higher risk because they lack the protective effect of antithyroid drugs administered to the mothers. The presence of a fetal heart rate greater than 160 beats per minute, in the absence of other fetal abnormalities, is suggestive of fetal hyperthyroidism. It is very useful to test neonatal cord blood at the time of delivery for thyroid function tests and TRABs. When the mother has been treated with high-dose antithyroid drugs, the neonate should be retested 10 days after birth given that transplacental passage of methimazole or propylthiouracil may initially mask hyperthyroidism. A TRAB-negative neonatal hyperthyroidism, in the absence of a maternal history of Graves' disease, should direct diagnosis to the suspicion of nonautoimmune congenital hyperthyroidism. Nowadays, the diagnosis can be confirmed only by sequencing of the TSH receptor gene.

### Iodine-Induced Thyrotoxicosis

Iodine urinary excretion in patients with thyrotoxicosis due to excessive iodine consumption is always high, and excessive iodine consumption should always be suspected when hyperthyroidism appears abruptly in patients with a history of nodular thyroid disease. A careful history often identifies the source of iodine, and all patients should be asked about recent exposure to any of the compounds listed in Table III. With the

**Table III Common Sources of Iodine Contamination**

Foods	Seaweed and seaweed containing foods (Japanese cuisine)
Food supplements	Kelp and other seaweed derivatives Vitamin supplements
Radiological contrast agents	Intravenous and oral (e.g., Gastrografin, Renografin, iopanoic acid, sodium ipodate)
Antiseptics	Betadine Iodoform gauze
Drugs	Amiodarone Expectorants
Iodine solutions	Lugol's solution, SSKI, KI

exception of type II amiodarone-induced thyrotoxicosis, RAIU is usually low in thyrotoxic patients with heavy iodine contamination, but it is almost never less than 1%, a feature that allows distinction from subacute and painless thyroiditis.

### Amiodarone-Induced Thyrotoxicosis

In the presence of amiodarone-induced thyrotoxicosis, further testing is required to distinguish between the type I (nondestructive) and type II (destructive) forms given that treatments may be radically different. Because of the concomitant underlying presence of thyroid disease, such as Graves' disease or nodular thyroid disease, type I amiodarone-induced thyrotoxicosis differs somewhat from other forms of iodine-induced thyrotoxicosis. It usually can be diagnosed with appropriate tools. Accordingly, RAIU is usually low but by definition never less than 1%. In contrast, in type II amiodarone-induced thyrotoxicosis, RAIU is always less than 1% and often no clear underlying thyroid disorder can be identified. High circulating interleukin-6 levels have been proposed as a useful marker of thyroid tissue destruction, and color flow Doppler ultrasound imaging shows a distinctive absence of vascularization in the gland.

### Subacute, Painless, and Postpartum Thyroiditis

Subacute, painless, and postpartum thyroiditis are classically characterized by a low (<1%) RAIU during the thyrotoxic phase. This test alone, in

the presence of a suggestive clinical presentation, allows the diagnosis in nearly all cases. High-titer anti-TG and anti-TPO antibodies are usually found in postpartum and painless thyroiditis as a marker of prominent thyroid autoimmunity, whereas only weakly and transiently positive tests are occasionally found in subacute thyroiditis. A very high (> 50 mm/h) erythrocyte sedimentation rate is an additional distinctive diagnostic feature in subacute thyroiditis. Other inflammation indexes may be high in subacute thyroiditis, and a mild leukocytosis is often observed. As in other thyroidal destructive processes, interleukin-6 levels are elevated in all three disorders and, therefore, have no differential values. Ultrasound findings are generally characterized by patchy areas of hypoechogenicity in subacute thyroiditis, whereas a more diffuse hypoechoic pattern, closely resembling Hashimoto's thyroiditis, is found in postpartum and painless thyroiditis. The color flow Doppler pattern shows reduced vascularity in all three disorders. Occasionally, and especially when patients are first seen during the recovery or hypothyroid phase, a more subtle picture can emerge from testing, with a low but not nil RAIU and with only mild elevations of the erythrocyte sedimentation rate, making the differential diagnosis more difficult.

### Thyrotoxicosis of Extrathyroidal Origin

An extrathyroidal source of thyroid hormone should always be suspected when more frequent causes of low RAIU thyrotoxicosis have been ruled out. When thyrotoxicosis factitia is suspected, a serum thyroglobulin measurement can be extremely useful in confirming the diagnosis because this disorder represents the only condition (in absence of AbTG) in which thyrotoxicosis is associated with an undetectable thyroglobulin level. Indeed, the presence of AbTG may cause falsely low thyroglobulin levels. Given the high prevalence of thyroid nodules in the general population, especially in iodine-deficient areas, it is also useful to perform ultrasound scanning of the neck because, in the presence of thyroid nodules, thyroglobulin may be elevated in spite of the assumption of exogenous thyroid hormone.

The suspicion of struma ovarii can be confirmed at the time of RAIU by simply scanning the pelvic area with the probe. The presence of functional thyroid tissue is demonstrated by the presence of significantly increased uptake of iodine in the ovarian region. Further imaging (CT or ultrasound scan) will confirm the presence of an ovarian mass.

When the source of thyroid hormone is metastatic thyroid follicular cancer, the presence of the latter is usually evident from the history. Because all patients with differentiated thyroid cancer after thyroidectomy take L-T4 in TSH-suppressive doses, thyroid function tests should be repeated after tapering the medication to rule out iatrogenic thyrotoxicosis. Confirmation is obtained with whole body radioiodine scanning that will show multiple foci of uptake in several skeletal regions.

### See Also the Following Articles

Graves' Disease • Thyroglobulin • Thyrotoxicosis Factitia • Thyrotoxicosis, Overview of Causes • Thyrotoxicosis, Systemic Manifestations • Thyrotoxicosis, Treatment • Thyrotoxic Storm • Toxic Adenoma • Toxic Multinodular Goiter • TSH (Thyroid-Stimulating Hormone; Thyrotropin)

### Further Reading

- Bartalena, L., Bogazzi, F., Brogioni, S., *et al.* (1996). Measurement of serum free thyroid hormone concentrations: An essential tool for the diagnosis of thyroid dysfunction. *Horm. Res.* **45**, 142–147.
- Braverman, L. E., and Utiger, R. D. (eds.) (2000). "The Thyroid: A Fundamental and Clinical Text," 8th ed. J. B. Lippincott, Philadelphia.
- Burch, H. B., and Wartofsky, L. (1993). Graves' ophthalmopathy: Current concepts regarding pathogenesis and management. *Endocrine Rev.* **14**, 747–793.
- Chiovato, L., Vitti, P., Bendinelli, G., *et al.* (1994). Detection of antibodies blocking thyrotropin effect using Chinese hamster ovary cells transfected with the cloned human TSH receptor. *J. Endocrinol. Invest.* **17**, 809–816.
- Faglia, G., Beck-Peccoz, P., Piscitelli, G., *et al.* (1987). Inappropriate secretion of thyrotropin by the pituitary. *Horm. Res.* **26**, 79–99.
- Marcocci, C., Vitti, P., Cetani, F., *et al.* (1991). Thyroid ultrasonography helps to identify patients with diffuse lymphocytic thyroiditis who are prone to develop hypothyroidism. *J. Clin. Endocrinol. Metab.* **72**, 209–213.
- Mariotti, S., Caturegli, P., Piccolo, P., *et al.* (1990). Antithyroid peroxidase autoantibodies in thyroid diseases. *J. Clin. Endocrinol. Metab.* **71**, 661–669.
- Mariotti, S., Martino, E., Cupini, C., *et al.* (1982). Low serum thyroglobulin as a clue to the diagnosis of thyrotoxicosis factitia. *N. Engl. J. Med.* **307**, 410–412.
- Martino, E., Bartalena, L., Mariotti, S., *et al.* (1988). Radioactive iodine thyroid uptake in patients with amiodarone-iodine-induced thyroid dysfunction. *Acta Endocrinol. (Copenh.)* **119**, 167–173.
- Martino, E., Safran, M., Aghini-Lombardi, F., *et al.* (1984). Environmental iodine intake and thyroid dysfunction during chronic amiodarone therapy. *Ann. Int. Med.* **101**, 28–34.
- Masiukiewicz, U. S., and Burrow, G. N. (1999). Hyperthyroidism in pregnancy: Diagnosis and treatment. *Thyroid* **9**, 647–652.
- McDermott, M. T., and Ridgway, E. C. (1998). Central hyperthyroidism. *Endocrinol. Metab. Clinics North America* **27**, 187–203.

- McKenzie, J. M., and Zakarija, M. (1989). Clinical review 3: The clinical use of thyrotropin receptor antibody measurements. *J. Clin. Endocrinol. Metab.* **69**, 1093–1096.
- Ross, D. S. (1998). Syndromes of thyrotoxicosis with low radioactive iodine uptake. *Endocrinol. Metab. Clinics North America* **27**, 169–185.
- Siegel, R. D., and Lee, S. L. (1998). Toxic nodular goiter: Toxic adenoma and toxic multinodular goiter. *Endocrinol. Metab. Clinics North America* **27**, 151–168.
- Tonacchera, M., Vitti, P., Agretti, P., *et al.* (1998). Activating thyrotropin receptor mutations in histologically heterogeneous hyperfunctioning nodules of multinodular goiter. *Thyroid* **8**, 559–564.
- Wolf, P. D. (1980). Transient painless thyroiditis with hyperthyroidism: A variant of lymphocytic thyroiditis? *Endocrine Rev.* **1**, 411–420.
- Zimmerman, D. (1999). Fetal and neonatal hyperthyroidism. *Thyroid* **9**, 727–733.

the technique by blocking axonal transport with colchicine led to the disclosure of TRH-containing neurons in the rat telencephalon and brain stem. The olfactory bulb, hippocampus, periaqueductal gray, medulla oblongata, and dorsal horn of the spinal cord are among the extrahypothalamic brain areas containing TRH-positive cells and fibers, suggesting a physiological role for TRH as a neurotransmitter or neuromodulator in the central nervous system (CNS) in addition to its neuroendocrine role in TSH release. In fact, the majority of total brain TRH is found in extrahypothalamic brain areas. The pituitary gland, gastrointestinal tract, pancreatic islets, and reproductive tract also contain some TRH.

## TRH RECEPTORS AND TRH-DEGRADING ENZYME

The TRH receptor TRH-R1, originally cloned from a pituitary cDNA library, was the only TRH receptor known until recently. Many studies suggest that this receptor belongs to the family of G protein-coupled membrane receptors acting via the inositol phospholipid-calcium-protein kinase C signal transduction pathway. In spite of the wide distribution of TRH in the rat CNS, there is only little TRH-R1 mRNA expression in the brain. In contrast, TRH-binding sites are present in a large number of rat and human brain areas. Part of this discrepancy has been clarified by the demonstration of another TRH receptor, TRH-R2, which was cloned from a rat brain cDNA library. It has approximately 50% amino acid homology with rat TRH-R1 and may be a key component of TRH signaling within the CNS.

The termination of TRH signals is probably mediated by a membrane-bound ectoenzyme designated pyroglutamyl peptidase II, which has a remarkable specificity to hydrolyze only TRH and closely related peptides. Therefore, this enzyme has been termed TRH-degrading ectoenzyme (TRH-DE). The highest activity is present in rat brain. In the pituitary, the enzyme is localized on lactotrophs, where TRH-DE activity and mRNA levels are tightly regulated by estrogen and thyroid hormone. In peripheral tissues, the enzyme is present in retina, lung, and liver. Studies have compared the distribution of TRH-R1, TRH-R2, and TRH-DE in rat brain, showing that TRH-R1 is predominantly expressed in hypothalamic regions and the anterior pituitary, in accordance with the known neuroendocrine functions of TRH. TRH-R2 is widely distributed throughout a number of brain areas, including the thalamus, cerebral and cerebellar

cortex, and habenulae. The complementary expression of both receptor subtypes and the overlapping expression of both receptors with TRH-DE strongly suggest that the two receptors serve highly specific functions in discrete neuroanatomical pathways.

## FUNCTIONS OF TRH

Ever since its discovery, TRH has been known as the major neuropeptide stimulating the anterior pituitary to synthesize and release TSH. It also stimulates release of prolactin from the anterior pituitary. Changes in serum concentrations of thyroid hormone have a strong effect on TRH gene expression in the hypothalamic PVN but not elsewhere, as demonstrated by mRNA *in situ* hybridization studies. TRH immunocytochemical staining intensity and mRNA expression increase in hypothyroidism, with reverse changes in hyperthyroidism. In addition, lesions of the PVN reduce serum TSH concentrations and prevent the increase in serum TSH in primary hypothyroidism. Studies in mice have demonstrated an important role for the TR $\beta$ 2 isoform in this negative feedback effect of thyroid hormone on TRH cells in the PVN. Neuroanatomical tracing studies using transsynaptic viral tracers have identified multisynaptic autonomic connections between neurons in the PVN and the thyroid gland. Because some of these neurons are TRH immunoreactive, TRH cells in the PVN may be involved in thyroid regulation via both classical neuroendocrine and novel neural pathways.

It has been known for many years that serum concentrations of thyroid hormone decrease during food deprivation. The way in which the brain orchestrates this adaptation to starvation has become increasingly clear. In a food-deprived condition, TRH gene expression in the PVN is markedly decreased. The fat-derived hormone leptin plays a pivotal role in this neuroendocrine response to starvation. A decrease in serum leptin signals via the leptin receptor in the hypothalamic arcuate nucleus (Arc) to agouti-related peptide (AgRP) and  $\alpha$ -melanocyte-stimulating hormone (MSH)-containing neurons. These neurons have monosynaptic connections with TRH cells in the PVN that express the melanocortin-4 receptor (MC4R). The MC4R is stimulated by  $\alpha$ -MSH and is inhibited by AgRP. A number of *in vitro* and lesion studies in rats have clearly demonstrated the importance of these pathways for the integrity of the neuroendocrine response to fasting. The effects of starvation on TRH cells in the PVN are less dependent on TR $\beta$ 2 than are the effects of hypo- and

hyperthyroidism, underlining the different roles for TRH cells in the PVN in these two conditions.

In extrahypothalamic brain areas, TRH may be involved in blood pressure regulation, temperature regulation, and carbohydrate metabolism. In TRH knockout mice, hyperglycemia and impaired insulin secretion in response to glucose support the latter. Experimental studies involving the administration of TRH suggest a role for TRH in attention and mood.

## TRH AND PATHOLOGY

Within hours after the onset of a severe physical stressor such as a surgical procedure or trauma, the serum concentration of triiodothyronine (T3) decreases, mainly due to decreased peripheral conversion of thyroxine (T4) to T3. The magnitude of the change in serum T3 roughly correlates with the severity of the illness. Similarly, serum T3 decreases during food deprivation. Serum TSH is typically within the normal range, with a strongly attenuated nocturnal TSH surge. These changes are known as nonthyroidal illness (NTI), as well as the sick euthyroid syndrome and low-T3 syndrome, and they have been interpreted as an attempt of the organism to reduce energy expenditure during illness and starvation. Patients treated in intensive care units for prolonged periods of time have decreased serum T3 and T4 in combination with low-normal serum TSH and dramatically diminished pulsatility. Postmortem studies in human brain have shown decreased gene expression of TRH in the PVN, correlating with the decrease in serum T3 and TSH in serum samples taken just before death, pointing to a role of hypothalamic TRH in the resetting of the thyroid axis during illness.

Van den Berghe and colleagues studied the neuroendocrine component of the decreased activity of the thyroid and somatotrophic axes in critical illness by the intravenous infusion of TRH, growth hormone-releasing peptide-2 (GHRP-2), and growth hormone-releasing hormone. Patients who had been critically ill for several weeks showed a dramatic augmentation of nonpulsatile TSH release by TRH infusion and increased pulsatile TSH secretion after TRH plus GHRP-2 infusion. TRH infusion also increased serum T4 and T3 by 40 to 110%. In subsequent placebo-controlled studies, the same group demonstrated near-normalized serum concentrations of thyroid hormone, increased markers of anabolism, and decreased markers of catabolism after infusion of TRH and GHRP-2 over 5 days. Thus, the hypothesis

of inappropriately decreased neuroendocrine drive in critically ill patients can be tested in randomized clinical trials assessing clinical outcome.

Another clinical condition involving changes in the set point of the hypothalamus–pituitary–thyroid axis is major depression. Some 25% of these patients have serum T4 higher than the reference value, probably resulting from increased T4 production. Serum T3 is often normal but may be decreased in a proportion of patients. Serum TSH is low, but it is mostly within the normal range and shows an attenuated diurnal variation. The latter observation points to changes in hypothalamic function given that the hypothalamic SCN drives the diurnal variation in thyroid hormones. Both increased and unchanged cerebrospinal fluid (CSF) levels of TRH have been reported in depression. Based on the neuroendocrine changes in depression and earlier observations of central effects of TRH, a number of studies have addressed the possibility of a therapeutic role for TRH in depression. The results of these studies have been largely inconsistent, possibly due in part to the pharmacokinetics of orally administered TRH-like peptides. One double-blind crossover study reported strong and positive effects of intrathecal TRH in refractory depressed patients, but the number of patients was rather small.

## CONCLUSION

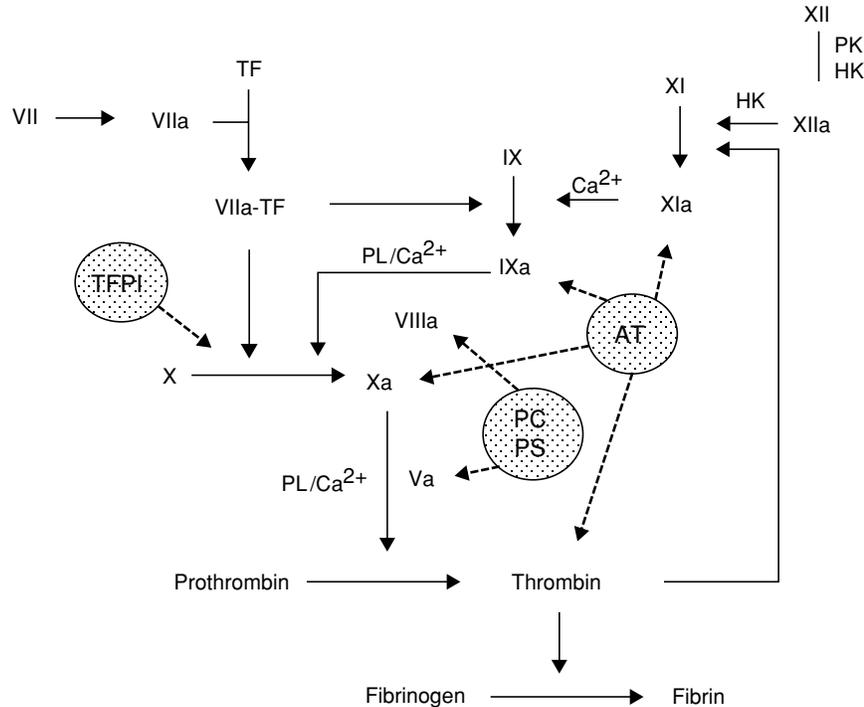
TRH is a classical neuropeptide originally identified as a hypothalamic-releasing factor for TSH. It is important for thyroid physiology and a key element of thyroid set point regulation. Studies in TRH knockout mice and the identification of a TRH receptor mutation in a patient with central hypothyroidism have confirmed these assumptions. Studies during the past decade or so have shown an important role for hypothalamic TRH in the neuroendocrine response to food deprivation. This response includes changes in serum leptin concentrations mediated via the hypothalamic melanocortin system to TRH neurons in the PVN. Studies in critically ill patients have revealed a role for TRH in the neuroendocrine response to disease. Major depression is another example of a pathological condition in which TRH plays a role.

## See Also the Following Articles

Depression, Thyroid Function and • Prolactin (PRL) • TSH (Thyroid-Stimulating Hormone; Thyrotropin)

## Further Reading

- Abel, E. D., Ahima, R. S., Boers, M., Elmquist, J. K., and Wondisford, F. E. (2001). Critical role for thyroid hormone receptor  $\beta 2$  in the regulation of paraventricular thyrotropin-releasing hormone neurons. *J. Clin. Invest.* **107**, 1017–1023.
- Collu, R., Tang, J., Castagné, J., *et al.* (1997). A novel mechanism for isolated central hypothyroidism: Inactivating mutations in the thyrotropin-releasing hormone receptor gene. *J. Clin. Endocrinol. Metab.* **82**, 1361–1365.
- Flier, J. S., Harris, M., and Hollenberg, A. N. (2000). Leptin, nutrition, and the thyroid: The why, the wherefore, and the wiring. *J. Clin. Invest.* **105**, 859–861.
- Fliers, E., Wiersinga, W. M., and Swaab, D. F. (1998). Physiological and pathophysiological aspects of thyrotropin-releasing hormone gene expression in the human hypothalamus. *Thyroid* **8**, 921–928.
- Heuer, H., Schäfer, M. K. H., O'Donnell, D., Walker, P., and Bauer, K. (2000). Expression of thyrotropin-releasing hormone receptor 2 (TRH-R2) in the central nervous system of rats. *J. Comp. Neurol.* **428**, 319–336.
- Jackson, I. M. D. (1998). The thyroid axis and depression. *Thyroid* **8**, 951–956.
- Kalsbeek, A., Fliers, E., Franke, A. N., Wortel, J., and Buijs, R. M. (2000). Functional connections between the suprachiasmatic nucleus and the thyroid gland as revealed by lesioning and viral tracing techniques in the rat. *Endocrinology* **141**, 3832–3841.
- Legradi, G., Emerson, C. H., Ahima, R. S., Flier, J. S., and Lechan, R. M. (1997). Leptin prevents fasting-induced suppression of prothyrotropin-releasing hormone mRNA in neurons of the hypothalamic paraventricular nucleus. *Endocrinology* **138**, 2569–2576.
- Marangell, L. B., George, M. S., Callahan, A. M., *et al.* (1997). Effects of intrathecal thyrotropin-releasing hormone (protirelin) in refractory depressed patients. *Arch. Gen. Psychiatry* **54**, 214–222.
- Rondeel, J. M., Klootwijk, W., Linkels, E., Van Haasteren, G. A., De Greef, W. J., and Visser, T. J. (1995). Regulation of the TRH-like peptide pyroglutamyl–glutamyl–prolineamide in the rat anterior pituitary gland. *J. Endocrinol.* **145**, 43–49.
- Van den Berghe, G., Wouters, P., Weekers, F., *et al.* (1999). Reactivation of pituitary hormone release and metabolic improvement by infusion of growth hormone-releasing peptide and thyrotropin-releasing hormone in patients with protracted illness. *J. Clin. Endocrinol. Metab.* **84**, 1311–1323.
- Yamada, M., Saga, Y., Shibusawa, N., *et al.* (1997). Tertiary hypothyroidism and hyperglycemia in mice with targeted disruption of the thyrotropin-releasing hormone gene. *Proc. Natl. Acad. Sci. USA* **94**, 10862–10867.



**Figure 1** Schematic representation of blood coagulation cascade. Roman numbers refer to clotting factors. Dotted arrows indicate inhibitory activity. AT, antithrombin; HK, high-molecular-weight kininogen; PK, prekallikrein; PL, phospholipids; PC, protein C; PS, protein S; TF, tissue factor; TFPI, tissue factor pathway inhibitor.

glycosylation sites. The molecular structure of t-PA is similar to that of other serine proteases (Fig. 3). It comprises an N-terminal portion, also called the heavy or A-chain, containing several domains typical of serine proteases and involved in fibrin or surface cell receptor binding. The catalytic site of the enzyme is located in the C-terminal portion of the molecule, also termed the light or B-chain. The cleavage of the Arg275-Ile276 peptide bond transforms the single-chain t-PA in a double-chain molecule held together by a disulfide bridge. In contrast to all other serine proteases, both the single-chain and double-chain forms of t-PA are enzymatically active.

The t-PA A-chain is composed of a finger domain, an epidermal growth factor domain, and two kringle domains. The finger domain extends from residues 4 to 50 and is involved in the binding of t-PA to fibrin. The epidermal growth factor domain comprises a sequence between amino acids 50 and 87 and it likely represents the molecular epitope recognized by the hepatic cell receptors involved in the protease or protease-inhibitor complex clearance. The kringle 1 and the kringle 2 domains, which show a high degree of homology with the plasminogen kringles, extend from residues 87 to 176 and from 176 to 262, respectively.

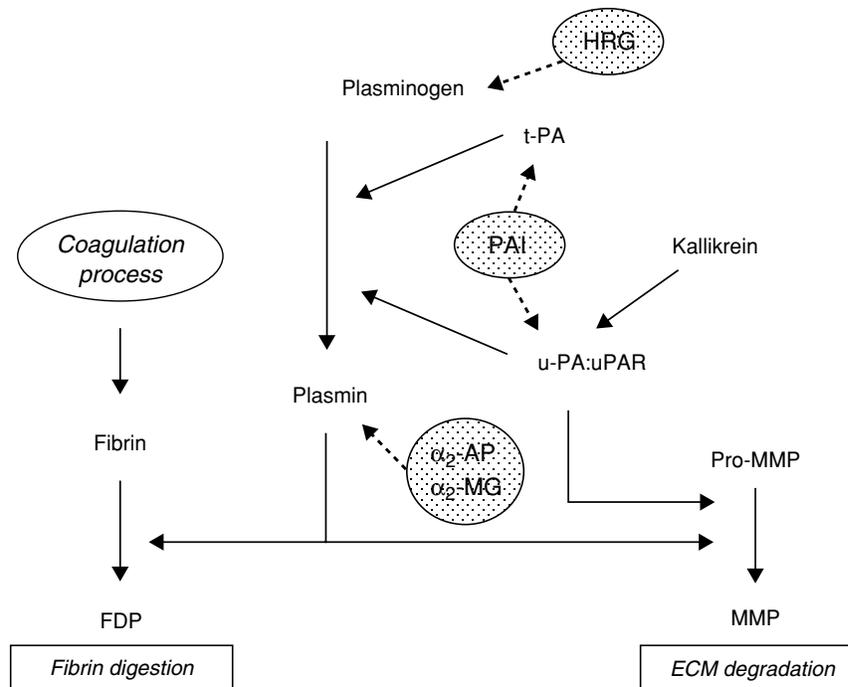
The physiological function of kringle 1 is unknown. In contrast, kringle 2 was found to exhibit an affinity for fibrin, lysine residues, and  $\omega$ -amino acids, such as the  $\epsilon$ -aminocaproic acid.

The catalytic domain in the B-chain of the t-PA molecule is composed of a 230-amino-acid sequence, with the active site comprising the amino acid triad typical of the serine proteases: His-322, Asp-371, and Ser-478.

### T-PA GENE STRUCTURE AND POLYMORPHISMS

The gene for t-PA is located on bands p12-p11 on chromosome 8 and consists of 32.7 kb. It contains 14 exons coding for the protease domains and 13 introns. The complete cDNA is 2530 bp. The proximal promoter sequences contain TATA- and CAAT-boxes and potential recognition sequences for transcription factors. Two 5'-untranslated regions of mRNA, of 209 and 99 bp, respectively, have been found. The 3'-flanking region contains the polyadenylation signal at positions 32,688-32,693.

The first polymorphism found in the t-PA gene consisted of an insertion/deletion polymorphism of



**Figure 2** The fibrinolytic system. Dotted arrows indicate inhibitory activity.  $\alpha_2$ -AP,  $\alpha_2$ -antiplasmin;  $\alpha_2$ -MG,  $\alpha_2$ -macroglobulin; ECM, extracellular matrix; FDP, fibrin(ogen) degradation products; HRG, histidine-rich glycoprotein; MMP, metalloproteinases; PAI, plasminogen activator inhibitor; t-PA, tissue-type plasminogen activator; u-PA, urinary-type plasminogen activator; u-PAR, u-PA receptor.

the 311 bp Alu-repeat sequence in intron 8. Another eight polymorphisms of the t-PA gene have been described and characterized by a single nucleotide substitution. Among these, three were found in linkage disequilibrium with the Alu-repeat polymorphism, namely, a C-7351T substitution in the enhancer region, a T20099C substitution in exon 6, and a T27445A substitution in intron 10. *In vitro* studies failed to demonstrate any influence of the Alu-repeat polymorphism on basal endothelial synthesis of t-PA. No correlation between the Alu-repeat I/D genotype and circulating levels of both t-PA antigen and activity was observed *in vivo*. In contrast, a relationship between this polymorphism and the net local t-PA release rate, as well as an influence on t-PA release after venous occlusion, was seen in healthy subjects. Conflicting results have emerged from studies dealing with the relationship between the Alu-repeat I/D polymorphism of t-PA and the risk of myocardial infarction.

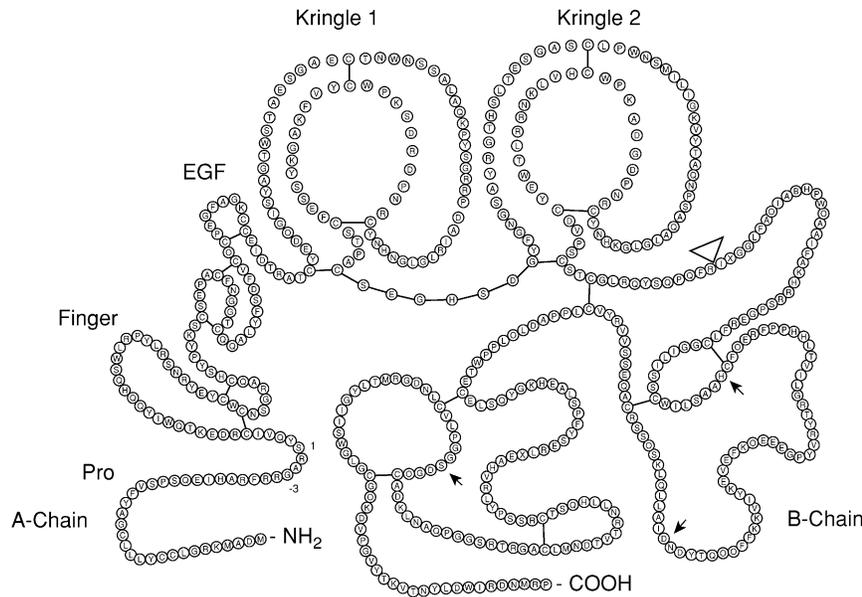
## SYNTHESIS AND SECRETION OF t-PA

t-PA has been isolated from many tissues, such as endothelial cells and smooth muscle cells of vascular

walls, monocytes, mesothelial cells, megakaryocytes, mast cells, cardiac fibroblasts, and neuronal cells. The principal site of t-PA synthesis is the endothelial cell, where the protease is stored in the cytoplasm and released through a constitutive secretory pathway. Moreover, t-PA can be released in the circulation within a few minutes in response to different stimuli through a regulated secretory pathway, suggesting the presence of a cytoplasmic storage pool rather than a *de novo* synthesis of the protein.

Several stimuli have been found to induce *in vivo* acute t-PA release, such as venous occlusion, exercise, 1-desamino-8-D-arginine-vasopressin (DDAVP) infusion, mental stress, and sympathoadrenal activation; in this last case, t-PA may be also coreleased with catecholamines by chromaffin cells. Moreover, *in vitro* stimulation of endothelial cells with thrombin, fibrin, bradykinin, histamine, and acetylcholine is associated with enhanced t-PA secretion; also, an arterial shear stress of 14 to 28 dyn/cm<sup>2</sup> applied to the culture may increase the production of t-PA.

By immuno- and electromicroscopic techniques in human umbilical vein endothelial cells, the site of cytoplasmic storage of t-PA has been identified in small and dense vesicles that are different from the



**Figure 3** Molecular structure of t-PA. The cleavage of the Arg275–Ile276 peptide bond (open arrowhead) transforms the single-chain t-PA in a double-chain molecule. The solid arrows indicate the catalytic triad His-322, Asp-371, and Ser-478. EGF, epidermal growth factor. Adapted from Ny T. *et al.* (1984), with permission.

Weibel-Palade bodies, where t-PA may also be accumulated together with von Willebrand factor. Concomitant release of t-PA and von Willebrand factor *in vivo* is commonly observed after DDAVP infusion, whereas patients with severe von Willebrand's disease are deficient in acute t-PA release.

The t-PA concentration in normal plasma is approximately 5  $\mu\text{g/liter}$ . Since most of the protein circulates in a complex with its inhibitor, PAI-1, the amount of free enzyme activity is usually approximately 0.5 U/ml, corresponding to 1  $\mu\text{g/liter}$ .

### ACTION AND INHIBITION OF $\tau$ -PA

The sole substrate of t-PA is represented by plasminogen, in which t-PA cleaves the Arg561–Val562 peptide bond, thus generating the active protease plasmin. In the absence of fibrin, however, t-PA is a weak activator of plasminogen; conversely, in the presence of fibrin, the affinity of t-PA for plasminogen is greatly increased, leading to an enhanced plasminogen activation of at least 2 orders of magnitude. This finding can be explained by an assembly of both t-PA and plasminogen on the fibrin surface, thus forming a ternary complex; moreover, fibrin-bound plasminogen exhibits an open conformation that favors its activation. The single-chain t-PA has a lower catalytic activity than the double-chain t-PA in the absence of fibrin, but this difference disappears when the molecules bind to fibrin. During lysis of the fibrin clot,

the appearance of additional plasmin-generated carboxyl-terminal lysine residues further promotes the binding of t-PA.

The main physiological inhibitor of t-PA is PAI-1, even though other serpins are known to potentially inactivate the t-PA. An interaction between the positively charged t-PA sequence 298–302 and the negatively charged PAI-1 sequence 350–355 precedes the irreversible reaction between the Arg346–Met347 reactive site of the inhibitor and the serine residue of the t-PA active site. This reaction occurs rapidly in plasma, although fibrin-bound t-PA may also be inhibited by PAI-1. Both free t-PA and t-PA–PAI-1 complexes are rapidly cleared from the circulation; the liver is the main organ involved in this process. In particular, liver endothelial cells and Kupffer cells provide t-PA clearance via a mannose receptor, whereas the hepatocytes may remove both t-PA and t-PA–PAI-1 complexes via the lipoprotein receptor-related protein/ $\alpha$ 2-macroglobulin receptor (LRP). The half-life of free t-PA is approximately 3–4 min, whereas the formation of complex with PAI-1 is associated with an increase in the LRP-mediated clearance of approximately 1 order of magnitude.

### CLINICAL ASPECTS OF $\tau$ -PA

A physiological fluctuation of t-PA activity during a 24 h period, with a marked decrease in the early morning hours, is tightly associated with the circadian

rhythm of PAI-1, which is characterized by a significant increase in the inhibitor's levels early in the morning.

Congenital t-PA deficiency has never been described in humans, suggesting a lethal condition. In contrast, mice with a complete deletion of the t-PA gene revealed normal development and mild glomerulonephritis; after endotoxin injection, t-PA<sup>-</sup>/t-PA<sup>-</sup> mice showed more extended thrombotic lesions than t-PA<sup>+</sup>/t-PA<sup>+</sup> mice. A bleeding diathesis due to an excess of t-PA has been reported in only a few cases. An overexpression of annexin II, a cell receptor for t-PA, on acute promyelocytic leukemia cells seems to be involved in the pathogenesis of the hyperfibrinolytic state typical of this disease.

The endothelial release of t-PA is part of the thromboprotective properties of the vascular surface and represents a counterregulatory mechanism to prevent the progression of a clotting process into occlusive thrombus. A lower level of t-PA production has been observed in the calf veins with respect to the proximal veins, suggesting a possible role in the onset of deep venous thrombosis. An impaired fibrinolytic capacity due to reduced t-PA release or, more often, to PAI-1 excess has been found in up to 40% of patients with deep venous thrombosis; however, the predictive value of both t-PA and PAI-1 levels for the risk of thrombotic events and recurrences has not been confirmed. A predictive value for major cardiovascular events has been ascribed to high t-PA antigen values. This apparently paradoxical finding, however, depends mainly on concomitant increased PAI-1 levels

**Table I Clinical Application of Thrombolytic Therapy with Recombinant t-PA**

Major indications
Acute myocardial infarction
Hemodynamically unstable massive pulmonary embolism
Indications suggested in highly selected cases
Acute peripheral artery occlusion
Massive iliofemoral vein thrombosis
Acute ischemic stroke
Cerebral sinus vein thrombosis
Budd-Chiari syndrome
Central retinal artery and/or vein thrombosis

*Note.* The risk of bleeding must be carefully excluded before treatment.

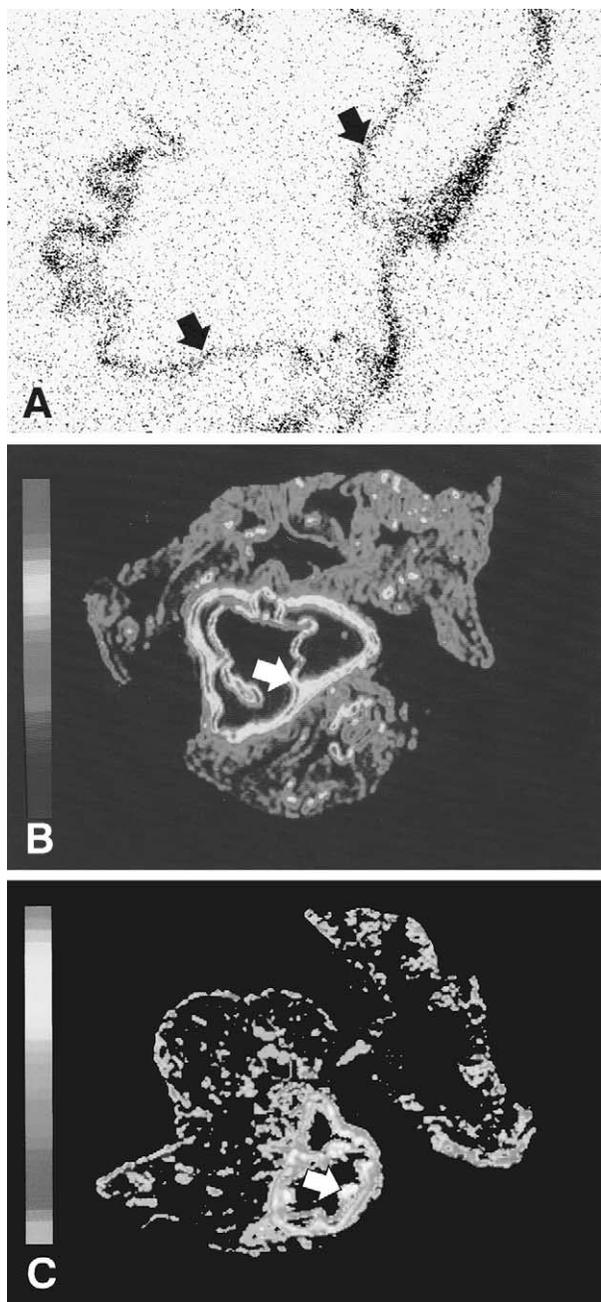
leading to high concentrations of protease-inhibitor circulating complexes but low free active t-PA.

Several drugs may enhance t-PA synthesis and secretion, such as DDAVP, sodium nitroprusside, epinephrine, isoproterenol, sodium butyrate, and triazobenzodiazepines.

Using recombinant DNA technology, the human t-PA molecule (rt-PA) has been produced for therapeutic thrombolysis. The efficacy of this agent in the treatment of acute myocardial infarction and massive pulmonary embolism has been clearly demonstrated. Even though rt-PA has fibrin-selective properties, its use is not completely devoid of the risk of systemic fibrinolysis or of hemorrhagic complications. Therefore, therapeutic thrombolysis is indicated in selected life-threatening cases (Table I). Variants of rt-PA have been constructed with modified pharmacokinetic or functional properties, but clinical trials have failed to demonstrate any superiority to the original molecule.

## Further Reading

- Bachmann, F. (2001). Plasminogen-plasmin enzyme system. In "Hemostasis and Thrombosis: Basic Principles and Clinical Practice" (R. W. Colman, J. Hirsh, V. J. Marder, A. W. Clowes, and J. N. George, eds.), pp. 275-320. Lippincott Williams & Wilkins, Philadelphia, PA.
- Brandt, J. T. (2002). Plasminogen and tissue-type plasminogen activator deficiency as risk factors for thromboembolic disease. *Arch. Pathol. Lab. Med.* **126**, 1376-1381.
- Collen, D. (1999). The plasminogen (fibrinolytic) system. *Thromb. Haemost.* **82**, 259-270.
- Degen, S. J., Rajput, B., and Reich, E. (1986). The human tissue plasminogen activator gene. *J. Biol. Chem.* **261**, 6972-6985.
- Emeis, J. J., van den Eijnden-Schrauwen, Y., van den Hoogen, C. M., de Preister, W., Wesmuckett, A., and Lupu, F. (1997). An endothelial storage granule for tissue-type plasminogen activator. *J. Cell Biol.* **139**, 245-256.
- Kooistra, T., Schrauwen, Y., Arts, J., and Emeis, J. J. (1994). Regulation of endothelial cell t-PA synthesis and release. *Int. J. Hematol.* **59**, 233-255.
- Loscalzo, J., and Schafer, A. (2003). Overview of hemostasis and fibrinolysis. In "New Therapeutic Agents in Thrombosis and Thrombolysis" (A. A. Sasahara and J. Loscalzo, eds.), pp. 1-7. Dekker, New York.
- Marder, V. J. (2001). Thrombolytic therapy. Foundation and clinical results. In "Hemostasis and Thrombosis: Basic Principles and Clinical Practice" (R. W. Colman, J. Hirsh, V. J. Marder, A. W. Clowes, and J. N. George, eds.), pp. 1475-1495. Lippincott Williams & Wilkins, Philadelphia, PA.
- Ny, T., Elgh, F., and Lund, B. (1984). The structure of the human tissue-type plasminogen activator gene: Correlation of intron and exon structure to functional and structure domain. *Proc. Natl. Acad. Sci. USA* **81**, 5355-5359.



**Figure 1** tRAAS and heart valve leaflets (V). By *in situ* hybridization, renin mRNA (A) expression is markedly increased in leaflets as contrasted with surrounding tissue. Shown is the tricuspid valve of the rat heart. Using *in vitro* autoradiography, high-density binding for ACE (B) and Ang II receptors (C) is found in valve leaflets. Shown is the aortic valve of the normal rat heart. Reprinted from Sun and Weber (2003), with permission.

Katwa has shown that cultured leaflet myoFb express (mRNA and protein) requisite components for Ang peptide formation, including angiotensinogen (Ao)

an aspartyl protease (cathepsin D), and ACE. Their ACE and kinase activities include conversion of substrates such as Ang I, bradykinin, and substance P, respectively. Ang II generated by these cells is involved in the autocrine induction of type I collagen expression via up-regulation of TGF- $\beta_1$  expression. This is an angiotensin receptor type I (AT $_1$ ) receptor-mediated event. Contraction of leaflet myoFb grown on a flexible substratum and mediated by  $\alpha$ -smooth muscle actin microfilaments in these myoFb can be induced by Ang II, catecholamines, endothelin-1, or serotonin through the binding of their respective receptors and ligands; papaverine induces relaxation.

### Diseased Hearts

Compared to nonfailing donor heart tissue, the number of ACE transcripts is increased in tissue homogenates prepared from explanted failing human hearts. ACE and TGF- $\beta_1$  expression is likewise increased in myocardium obtained from patients undergoing aortic valve replacement for aortic stenosis and correlate with the extent of fibrosis found in these tissue samples. ACE and Ang II are present in the fibrous tissue that contributes to the sclerosis and the stenosis of diseased human aortic valves. In keeping with the expression of ACE in heart valve leaflets and Ang II in regulating leaflet collagen turnover, Atalar *et al.* have suggested that the ACE-DD genotype is associated with a greater risk of valve deformity following a bout of acute rheumatic fever.

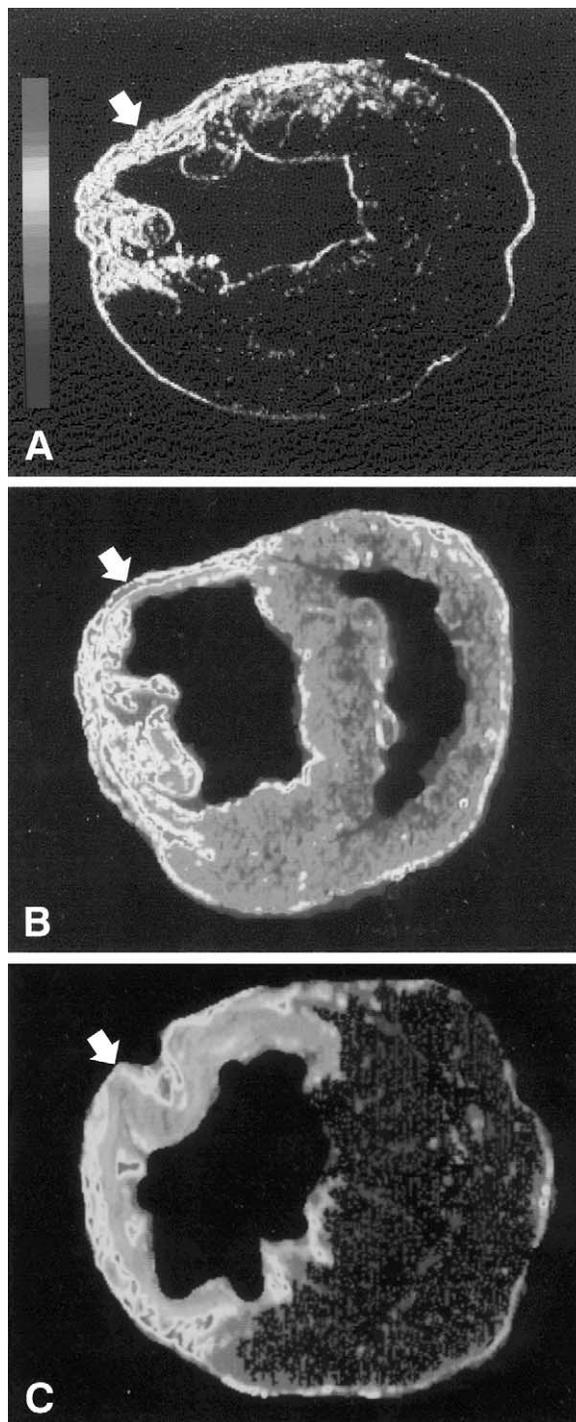
The aorta–coronary sinus concentration gradient for Ang II is decreased in patients with heart failure with increased Ang II appearing in this venous effluent and is accompanied by increased mRNA expression of angiotensinogen and ACE in noncardiomyocyte cells found in explanted failing hearts. ACE-dependent Ang II formation is increased in the infarcted segment of scarred myocardium in autopsied hearts and it has also been reported that expression of angiotensinogen, renin, ACE, and Ang II receptor genes is present in adult cardiomyocytes and up-regulated in response to myocardial ischemia.

Temporal and spatial responses in autoradiographic ACE binding have been assessed by Sun *et al.* in a rat heart model of myocardial infarction (MI). Other forms of injury involving various tissues are present in this model. They include the heart, visceral pericardium, kidneys, and skin. Each serves as a positive control in the analysis of ACE and tissue repair. Low-density ACE binding is present in normal rat myocardium, where renin mRNA expression is not found. High-density ACE binding and renin mRNA

expression appear at the site of anterior MI on day 7 following left coronary artery ligation and are coincident with the appearance of fibrillar collagen. As a fibrillar collagen network forms scar tissue over 8 weeks, the density of ACE binding and the expression of renin (see Figs. 2A and 2B) at this site increase progressively. A large transmural MI in the rat heart is associated with high-density ACE binding and the appearance of fibrosis at sites remote to the infarct, including the noninfarcted left ventricle, interventricular septum, and right ventricle (see Fig. 2B). The appearance of fibrosis at these remote sites is directly related to the extent of infarct injury. When an infarct injury is extensive, the entire myocardium, including infarcted and noninfarcted ventricular tissue, is involved in tissue repair and subsequent structural remodeling by fibrotic tissue.

Noninfarct-related sites of injury serve to further address the relationship between the appearance of ACE and repair. Sham operation includes manual handling of the heart and this leads to inflammation and subsequent fibrosis of the visceral pericardium. Silk ligature placement around the left coronary artery or within skin to close a surgical incision is associated with a foreign-body fibrosis at each site. The appearance of a mural thrombus in the infarcted left ventricle is associated with subsequent endocardial fibrosis and on occasion leads to thromboembolic renal infarction. At each of these sites of injury and repair, high-density autoradiographic ACE binding is temporally and spatially concordant with fibrous tissue formation. Angiotensinogen, renin, ACE, and TGF- $\beta_1$  mRNA levels are increased in the rat right ventricle at 3 and 6 weeks after monocrotaline-induced pulmonary injury. In mice with desmin-deficient cardiomyopathy, areas of cardiac fibrosis are colocalized with up-regulated expression of ACE and TGF- $\beta_1$ .

Potential differences in healing and thereby ACE expression could occur as a result of ischemic versus nonischemic injury. Permanent coronary artery ligation, for example, can impede the delivery of circulating cells and signals to the site of injury and can limit said delivery to that provided by collateral vessels. Nonischemic models of cardiac myocyte necrosis have been examined. They include the following: the endogenous release of catecholamines that accompanies Ang II infusion from implanted minipump or exogenous administration of isoproterenol, and chronic (>3 weeks) administration of aldosterone by minipump in uni-nephrectomized rats on a high-salt diet and which is accompanied by enhanced urinary potassium excretion and subsequent cardiac myocyte



**Figure 2** Renin, ACE, and Ang II receptor expression in the infarcted rat heart. Following MI, high-density renin mRNA expression is observed at the site of MI, as well as in the inflamed pericardium and endocardium at week 1 (A) and remains elevated over the course of 4 weeks (not shown). ACE and Ang II receptor-binding density was also markedly increased at the same sites. (B) ACE and (C) Ang II receptor-binding densities in the infarcted heart at week 4 post-MI. Reprinted from Sun and Weber (2003), with permission.

potassium depletion with cell necrosis. At each site of nonischemic myocyte loss, and irrespective of its etiologic basis, the temporal and spatial appearance of high-density ACE binding is coincident with the deposition of fibrous tissue and resembles the aforementioned responses observed in the infarcted heart following permanent coronary artery ligation. Thus, irrespective of the etiologic basis of injury, the tissue involved in repair, or the presence of ischemic versus nonischemic repair, marked autoradiographic ACE binding is coincident with fibrous tissue formation.

Furthermore, examination of serial heart sections of the infarcted rat heart demonstrates high-density ACE binding to be spatially concordant with marked autoradiographic Ang II (see Fig. 2C) and TGF- $\beta$  receptor-binding densities and mRNA expression of renin TGF- $\beta_1$  and type I collagen by *in situ* hybridization at these sites of fibrosis. Collectively, these findings in various injured tissues indicate that both Ang II and TGF- $\beta_1$  are part of a common signaling pathway involved in promoting repair.

The mRNA expression of ALDO synthase (CYP11B2), integral to the biosynthesis of ALDO, and ALDO production have been demonstrated in rodent heart and vascular tissue. ALDO generation is regulated by Ang II, a low-Na<sup>+</sup> or high-K<sup>+</sup> diet, or adrenocorticotrophic hormone. ALDO is extracted by the heart following MI and the transcardiac ALDO gradient (between aorta and coronary sinus) correlates with a serologic marker of collagen turnover (i.e., procollagen type III amino-terminal peptide) found in coronary venous effluent and associated with left ventricular (LV) dilation and impaired function. ALDO is extracted by the chronically failing human heart of diverse etiologic origins, a response that is blocked by spironolactone. Others have reported that ALDO production is increased in the failing human left ventricle based on coronary sinus levels of ALDO that exceed those found in the aorta. There exists an up-regulated expression of CYP11B2 in the left ventricle of the failing human heart of diverse etiologic origins. Additionally, 11 $\beta$ -hydroxysteroid dehydrogenase, an enzyme that is critical to maintaining the specificity of the mineralocorticoid receptor, given its equal affinity for mineralocorticoids and glucocorticoids, has been found in human cardiac tissue.

## CELLS EXPRESSING tRAAS

### Normal Heart

Cells expressing ACE in normal rat heart valve leaflets have been identified by immunolabeling with a

monoclonal antibody: endothelial cells lining atrial and ventricular leaflet surfaces and fibroblast-like cells residing within the leaflet matrix or myoFb. TGF- $\beta_1$  induces ACE synthesis in cultured rat heart interstitial fibroblasts and is accompanied by their differentiation into myoFb.

### Diseased Hearts

The ACE-positive cells involved in the response to renin expression seen at and remote to the infarct site include macrophages that invade the infarct site and myoFb. Within 24 h of MI, macrophages appear at the interface between viable and necrotic myocardium; by day 3, stromal fibroblasts coaggregate with macrophage clusters bordering on the infarct site. Thereafter, fibroblast differentiation follows, resulting in an  $\alpha$ -smooth muscle actin ( $\alpha$ -SMA)-positive myoFb phenotype; the myoFb then proliferate and migrate into the site of necrosis during the remainder of week 1. A combination of cell growth with spatial control of growth and fibrillar collagen assembly governs the rebuilding of infarcted tissue.

Macrophages and myoFb found at the infarct site also express Ang II receptors, TGF- $\beta_1$ , and TGF- $\beta_1$  receptors. Beyond day 14, only myoFb and endothelial cells are renin- and ACE-positive, coincident with the gradual disappearance of macrophages from the infarct site. Persistent renin expression and high-density ACE and Ang II receptor binding are present at the infarct site in rat hearts for 8 weeks or more post-MI. This is primarily due to  $\alpha$ -SMA-positive myoFb, which remain in infarct scar tissue for prolonged periods of time. In the infarcted human heart, these myoFb persist at the site of MI for years.

Gabbiani *et al.* have shown that myoFb have considerable phenotypic and functional diversity. Immunolabeling with  $\alpha$ -SMA, vimentin, and desmin defines their phenotype at the infarct site. Fibroblast-like cells express vimentin (V). ACE-labeled fibroblasts found in the infarct scar and involved in the expression of fibrillar collagen mRNA are positive for the cytoskeletal proteins  $\alpha$ -SMA and vimentin. These vimentin/actin (VA)-positive myoFb, instrumental to tissue repair including wound contraction, are likewise found in the connective tissue that comprises endocardial fibrosis, pericardial fibrosis, renal infarction, and sites of foreign-body fibrosis. Unlike incised skin, where myoFb contribute to tissue repair and then progressively disappear through programmed cell death (apoptosis) coincident with wound closure and scar tissue formation at week 4, the VA phenotype at the infarct site remains for prolonged periods.

Whether pathologic fibrosis at and remote to MI found in the infarcted heart is related to its persistent myoFb is uncertain.

*In vitro* emulsion autoradiography identifies VA-positive myoFb as expressing Ang II receptors. Together with displacement studies using either an AT<sub>1</sub> receptor antagonist, losartan, or an AT<sub>2</sub> receptor antagonist, PD123177, the great majority of these receptors in the infarcted rat heart are of the AT<sub>1</sub> subtype. MyoFb found at sites of microscopic scarring involving both infarcted and noninfarcted tissue also express mRNA for the fibrogenic cytokine TGF- $\beta$ <sub>1</sub> and TGF- $\beta$  receptors. This has implicated locally produced Ang II at sites of injury in regulating collagen turnover, which has been further suggested by the cardioprotective actions of losartan, an AT<sub>1</sub> receptor antagonist, in attenuating fibrous tissue formation at and remote to the MI. Locally produced Ang II is also involved in regulating *de novo* ALDO production in the infarcted heart, which likewise may contribute to tissue repair. Increased expression of ALDO synthase and ALDO tissue levels, together with increased concentrations of Ang II, have been observed in noninfarcted rat myocardium following coronary artery ligation. Treatment with losartan prevented these responses related to *de novo* ALDO production.

## tRAAS AND TISSUE REPAIR

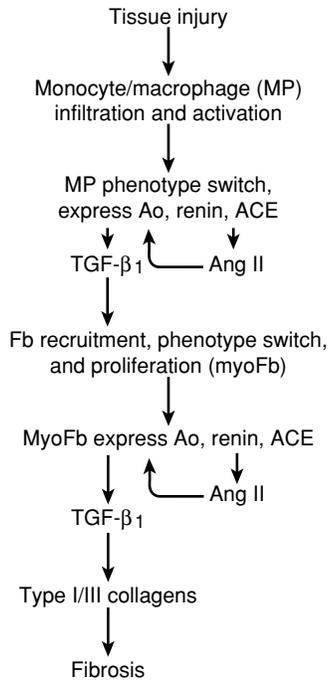
### Normal Heart

ACE imparts connective tissue with metabolic activity. Its substrate utilization involves factors participating in a reciprocal regulation of cell behavior—stimulators or inhibitors of cell growth and functions integral to the formation and degradation of fibrillar collagen. Loose and dense connective tissue formation is a dynamic process during the early growth and development of newborn rats. The contribution of Ang II to this process has been examined in young rats, where treatment of 4-week-old rats with enalapril attenuated the cardiac and vascular accumulation of collagen involving the right and left ventricles, aorta, and systemic arteries compared to untreated, age-matched control rats. No such study has yet been conducted with an Ang II receptor antagonist. In rats with a genetic predisposition to hypertension, treatment with either quinapril or hydralazine during early growth and development prevented the appearance of hypertension in adulthood. However, only quinapril attenuated the expected development of connective tissue seen in age-matched hypertensive controls.

### Diseased Hearts

High-density ACE binding in connective tissue that appears in response to tissue repair indicates that there is marked ACE activity at such sites. ACE activity has been examined in tissue obtained from the failing, infarcted, and noninfarcted human heart tissue. Homogenates of blocks of transmural tissue adjacent to visible scar tissue and obtained at the time of aneurysmectomy were prepared for analysis. Such samples, it was noted, may have contained scar tissue. The ACE activity of this homogenate was compared to that prepared from noninfarcted ventricular tissue obtained at necropsy from persons dying of noncardiac causes. Infarct tissue ACE activity exceeded that of such control tissue severalfold and the extent of activity was related to the severity of tissue damage. In rat heart tissue homogenates prepared from sites remote to a large transmural anterior MI, ACE activity is increased and the extent to which substrate conversion is increased correlates with infarct size. The importance of fibrous versus nonfibrous tissue ACE in determining this heightened ACE activity at infarcted and remote sites could not be determined from these studies.

A paradigm of tissue repair in which ACE and local Ang II are integral to the orderly and sequential nature of repair that eventuates in fibrosis has been proposed by Weber. As shown in Fig. 3, ACE is involved in a two-part *de novo* generation of Ang II within granulation tissue that forms at sites of injury. The first component of local Ang II generation is provided by macrophages. In an autocrine manner, ACE regulates the expression of the fibrogenic cytokine TGF- $\beta$ <sub>1</sub> that determines the phenotype conversion of coaggregating stromal fibroblasts. VA-positive myoFb then generate Ang II, whose autocrine induction of TGF- $\beta$ <sub>1</sub> regulates collagen turnover at sites of fibrous tissue formation, including infarcted and noninfarcted myocardium. It is suggested that Ang II generation at the infarct site is related to the extent of the myoFb response and accordingly the degree of myocyte necrosis and subsequent healing response. An extensive transmural MI and the accompanying inflammatory cell response generate a large amount of Ang II, which reaches distant or remote sites via its diffusion through tissue fluid to promote fibrosis. Accordingly, activation of fibrogenesis is greatest at sites closest to the anterior MI (e.g., interventricular septum) and less so at more remote sites (e.g., right ventricle). Expression of type I and III collagens is greater and persists longer in the septum than in the right ventricle in rat hearts following left coronary artery ligation.



**Figure 3** A schematic representation of the two-part *de novo* generation of Ang II involved in tissue repair following MI. See text for discussion and abbreviations. Adapted from Weber (1997), with permission.

The fibrogenic signals generated at a site of MI are transferred to remote sites through the heart's common interstitial space. These signals (e.g., Ang II) promote postinfarct remodeling at and remote to the site of MI. The persistence of VA-positive, active myoFb perpetuates such remodeling. Circulating and locally generated Ang II contributes to myocardial remodeling post-MI. Elevations in circulating Ang II, due to RAAS activation, represent a signal that gains entry to tissue where it further promotes postinfarct fibrosis by AT<sub>1</sub> receptor binding. Diastolic dysfunction is an outcome of such exuberant, unbridled tissue fibrosis. Ang II, whether derived locally or from the circulation, may further contribute to abnormal tissue stiffness through its induction of myoFb and thereby fibrous tissue contraction.

MyoFb persist in the infarct scar. Furthermore, these cells remain active, expressing ACE, Ang II receptors, TGF- $\beta$  receptors, type I collagen, and TGF- $\beta$ <sub>1</sub>. Prolonged stimulation of collagen production and adverse myocardial remodeling by fibrosis long after the acute phase of healing post-MI are reminiscent of progressive valvular sclerosis that appears years after acute rheumatic valvulitis and where valvular deformity (e.g., leaflet and chordal shortening) is related to contractile myoFb. The persistence of myoFb in

injured kidneys is accompanied by progressive interstitial fibrosis, renal dysfunction, and poor prognosis.

Treatment with either losartan or spironolactone prevents the accompanying accumulation of collagen at sites remote to the MI, suggesting the involvement of Ang II-driven local ALDO production in regulating tissue repair. In circumstances where circulating ALDO is not increased, spironolactone attenuates neointimal thickening following vascular barotrauma, tissue repair at sites of fibrous tissue formation, and vascular injury in stroke-prone rats. These observations further suggest that the autocrine/paracrine properties of locally produced ALDO participate in tissue repair, as proposed by DelCayre and co-workers.

Salutary clinical responses to ACE inhibition are likely to be multifactorial in origin. One important component relates to the prevention of adverse structural remodeling of infarcted and noninfarcted myocardium by fibrous tissue. Evidence supporting a contribution of locally produced Ang II in regulating myoFb collagen synthesis is obtained using pharmacologic probes that interfere with local Ang II generation (i.e., ACE inhibition) or occupancy of its AT<sub>1</sub> receptor prior to circulating RAAS activation. Captopril or enalapril, begun at or close to the onset of MI, reduces infarct size, infarct expansion, and thinning and attenuates the rise in hydroxyproline concentration at the infarct site in dogs with permanent coronary artery occlusion. The potential additional contribution of reduced bradykinin degradation to tissue repair and which would accompany ACE inhibition is under investigation. Bradykinin (BK) is released post-MI and a BK<sub>2</sub> receptor antagonist (Hoe140 or icatibant) accentuates collagen accumulation remote to the MI site.

Losartan, begun on day 1 after coronary artery ligation and at a dose that reduced AT<sub>1</sub> receptor binding by 50%, reduces the infarct scar area. Moreover, the expected rise in tissue Ang II concentration found at the infarct site 3 weeks after coronary artery ligation is markedly attenuated by either delapril or TCV-116, an AT<sub>1</sub> receptor antagonist, introduced on postoperative day 1. These findings raise the prospect that the number of myoFb or their Ang II-generating activity per cell at sites of repair may be influenced by Ang II. Other studies have not found such antagonists to influence fibrosis post-MI. The reason for these divergent findings is unclear.

Fibrous tissue formation at sites remote to MI is also influenced by these pharmacologic interventions. Perindopril, given 1 week after MI, attenuates the endocardial fibrosis that appears in the nonnecrotic

segment of the rat left ventricle. Captopril, commenced at the time of coronary artery ligation, attenuates the expected fibrosis of noninfarcted rat left and right ventricles and proliferation of fibroblasts and endothelial cells that appears at remote sites 1 and 2 weeks following MI. In these circumstances, captopril prevents the rise in LV end diastolic pressure that appears in untreated rats and which does not occur in propranolol-treated rats. Captopril also reduces the inducibility of ventricular arrhythmias in this model. When initiated 3 weeks post-MI, well after the tissue repair process has commenced and progressed, captopril does not prevent fibrosis remote to the infarct site or the rise in ventricular stiffness. Losartan prevents fibrosis at remote sites, but not the cellular proliferation that appears. Others did not find an inhibition of type I and III collagen mRNA expression at remote sites and have suggested posttranslational modification in collagen turnover to explain why fibrosis fails to appear at remote sites.

These favorable tissue protective effects of ACE inhibition or AT<sub>1</sub> receptor antagonism are not confined to the infarcted heart. These interventions prevent the appearance of fibrosis in diverse organs with experimentally induced or naturally occurring tissue injury and where circulating RAAS is not activated. These include the following: pericardial fibrosis post-pericardiectomy; tubulointerstitial fibrosis associated with unilateral ureteral obstruction, toxic nephropathy, cyclosporine, remnant kidney or renal injury following irradiation; cardiovascular sclerosis and glomerulosclerosis that appear in stroke-prone spontaneously hypertensive rats; interstitial pulmonary fibrosis that follows irradiation or monocrotaline administration; and subcutaneous pouch tissue in response to croton oil. The attenuation of fibrous tissue formation by these interventions in diverse organs with various forms of injury supports the importance of local Ang II in promoting fibrosis. A more detailed discussion of Ang II and tissue repair involving systemic organs can be found elsewhere.

A structural remodeling of the cardiovascularity by fibrous tissue accompanies aldosteronism derived from either endogenous or exogenous sources. This fibrogenic phenotype includes intramural arteries of the heart, kidney, pancreas, mesentery, and vasorum of the aorta and pulmonary artery. Cotreatment with a receptor antagonist (e.g., spironolactone, eplerenone), in either nondepressor or depressor dosage, prevents this remodeling, indicating its independence of elevations in blood pressure. In a substudy to the RALES trial, Zannad *et al.* found that a survival benefit was associated with a reduction in

circulating markers of collagen synthesis that presumably reflected an attenuation in ongoing vascular fibrosis. In this regard, urinary excretion of hydroxyproline, a marker of collagen turnover, is increased in adrenalectomized rats treated with ALDO, 1% dietary NaCl, and cortisone. Glucocorticoids, on the other hand, are known to reduce urinary hydroxyproline excretion and their inhibition of collagen formation in bone is associated with osteoporosis.

## SUMMARY

The tRAAS serves to regulate local concentrations of various mediators of inflammation and tissue repair in the heart and other organs. MyoFb-bound ACE is integral to *de novo* generation of Ang II that modulates the expression of TGF- $\beta_1$  and whose autocrine/paracrine properties regulate collagen turnover in heart valve leaflets, in an exteriorized portion of the normal extracellular matrix, and at sites of fibrous tissue formation that appear in response to various forms of injury involving diverse tissues. Persistent myoFb and their ACE at the infarct site or diseased heart valve leaflets contribute to a sustained metabolic activity that can account for a progressive fibrosis at these sites. It is such adverse structural remodeling by fibrous tissue that eventuates in ischemic cardiomyopathy, a major etiologic factor in the appearance and progressive nature of chronic cardiac failure, or in heart valve deformity, a major cause of chronic circulatory failure.

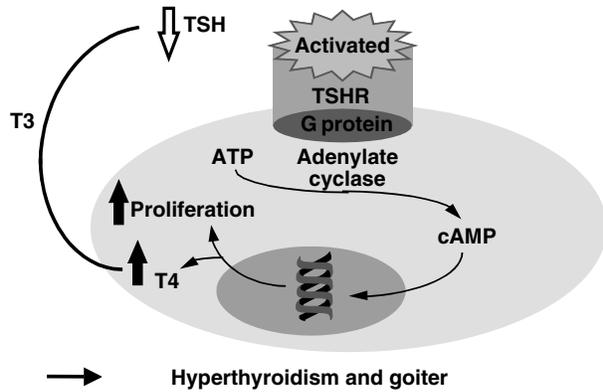
## See Also the Following Articles

Aldosterone in Congestive Heart Failure • Aldosterone Receptors • Baroreceptor Responses • Captopril • Carbenoxolone • Mineralocorticoids and Mineralocorticoid Excess Syndromes • Primary Aldosteronism (PAL) • Renal Vein Renin • Renin

## Further Reading

- Atalar, E., Tokgozoglul, S. L., Alikasifoglu, M., Ovunc, K., Aksoyek, S., Kes, S., and Tuncbilek, E. (2003). Angiotensin-converting enzyme genotype predicts valve damage in acute rheumatic fever. *J. Heart Valve Dis.* **12**, 7–10.
- Delcayre, C., and Swynghedauw, B. (2002). Molecular mechanisms of myocardial remodeling: The role of aldosterone. *J. Mol. Cell. Cardiol.* **34**, 1577–1584.
- Katwa, L. C., Tyagi, S. C., Campbell, S. E., Lee, S. J., Cicila, G. T., and Weber, K. T. (1996). Valvular interstitial cells express angiotensinogen, cathepsin D, and generate angiotensin peptides. *Int. J. Biochem. Cell Biol.* **28**, 807–821.
- Sappino, A. P., Schürch, W., and Gabbiani, G. (1990). Differentiation repertoire of fibroblastic cells: Expression of cytoskeletal

- proteins as marker of phenotypic modulations. *Lab. Invest.* **63**, 144–161.
- The SOLVD Investigators (1991). Effect of enalapril on survival in patients with reduced left ventricular ejection fractions and congestive heart failure. *N. Engl. J. Med.* **325**, 293–302.
- The SOLVD Investigators (1992). Effect of enalapril on mortality and the development of heart failure in asymptomatic patients with reduced left ventricular ejection fractions. *N. Engl. J. Med.* **327**, 685–691.
- Sun, Y., and Weber, K. T. (2003). RAS and connective tissue in the heart. *Int. J. Biochem. Cell Biol.* **35**, 919–931.
- Weber, K. T. (1997). Extracellular matrix remodeling in heart failure: A role for *de novo* angiotensin II generation. *Circulation* **96**, 4065–4082.
- Weber, K. T. (1999). Angiotensin II and connective tissue: Homeostasis and reciprocal regulation. *Regul. Pept.* **82**, 1–17.
- Zannad, F., Alla, F., Dousset, B., Perez, A., and Pitt, B. (2000). Limitation of excessive extracellular matrix turnover may contribute to survival benefit of spironolactone therapy in patients with congestive heart failure: Insights from the randomized aldactone evaluation study (RALES). Rales Investigators. *Circulation* **102**, 2700–2706.



**Figure 1** Signal transduction in an autonomously functioning thyroid cell. The cAMP cascade is stimulated by the constitutively activated TSH receptor (or less frequently by the  $G_{s\alpha}$  protein). cAMP stimulates thyroid function and growth, ultimately resulting in a phenotype of hyperthyroidism and goiter.

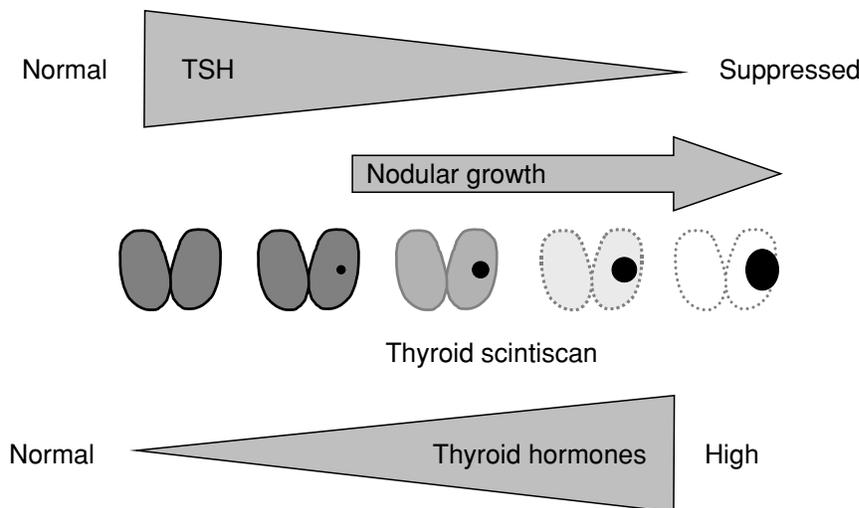
At this stage, serum levels of thyroid hormones are usually still normal but the TSH level is already reduced and the function in normal thyroid tissue is suppressed (subclinical hyperthyroidism). If this stage is passed by further autonomous growth and/or function, which causes an excess in the serum thyroid hormone level, thyrotoxicosis becomes clinically manifest (Fig. 2).

Over the past years, several European studies suggested that somatic activating TSH receptor mutations constitute the most frequent molecular cause

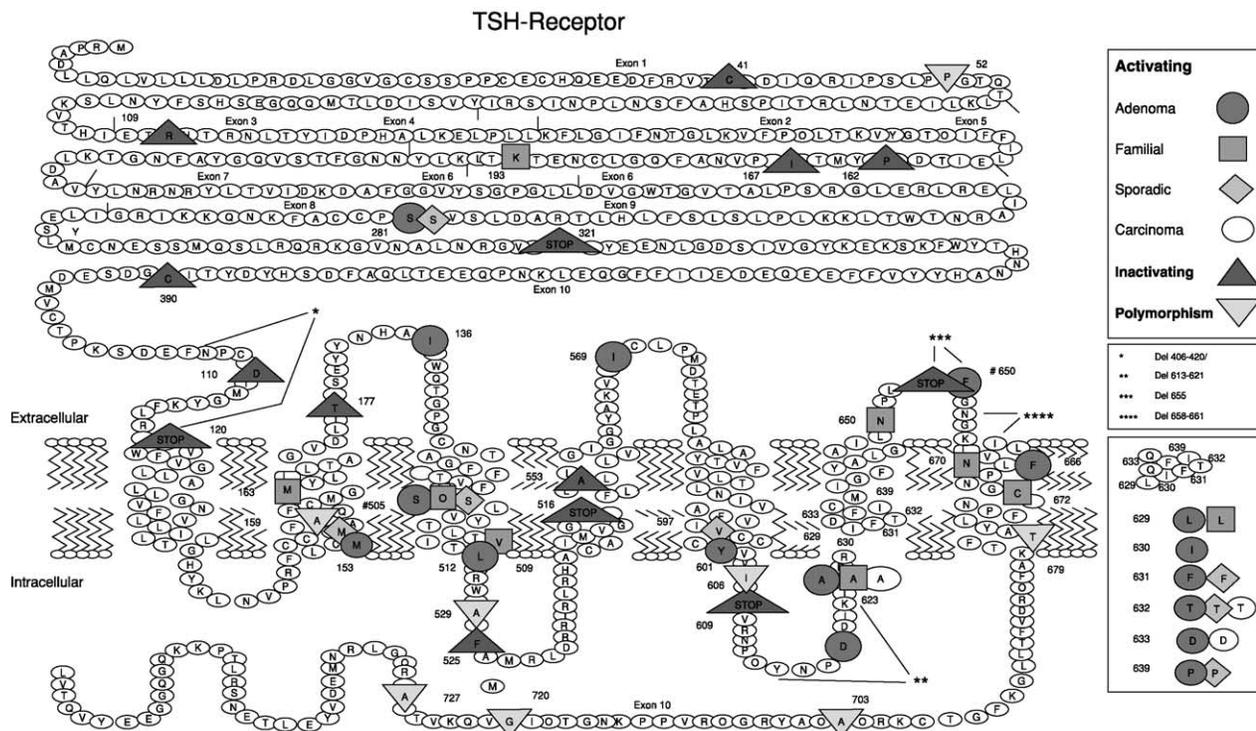
of toxic adenoma. An overall prevalence of ~50% was reported for somatic TSH receptor mutations, whereas mutations in the  $G_{s\alpha}$  protein were found to occur in less than 10% of TA. Figure 3 shows a compilation of activating TSH receptor mutations identified in toxic adenoma. It should be noted that ~20% of TA are monoclonal but do not harbor a TSH receptor or a  $G_{s\alpha}$  protein mutation, which suggests the presence of further, as yet unknown mutations in other genes.

### CLINICAL FEATURES—DIAGNOSIS

Clinical features and presenting complaints in a patient with toxic adenoma can be attributed to (1) the nodule or, if present, a goiter and (2) symptoms of hyperthyroidism. Thus, a patient may present to the doctor because of a lump in the neck, intolerance of tight necklaces, or an increase in collar size. Moreover, mechanical symptoms such as difficulty in swallowing and, very rarely, difficulty in breathing due to local esophageal or tracheal compression by the nodule may be apparent. However, more often, the patient presents with features suggestive of hyperthyroidism, such as nervousness, hyperactivity, weight loss despite increased appetite, palpitations, tremor, and heat intolerance (Table I). Of note, these classical symptoms of hyperthyroidism are frequently absent in elderly patients with TA or TMG.



**Figure 2** Development of thyroid autonomy. A thyroid cell with a somatic gain-of-function TSH receptor mutation has a growth and functional advantage over unaffected thyrocytes. The autonomous nodule increasingly contributes to the production of thyroid hormone by the thyroid independent of the TSH level, which in turn starts to decrease. If the nodule reaches a critical mass, thyrotoxicosis becomes clinically manifest, TSH levels become low, and surrounding tissue is suppressed. On scintiscan, the toxic adenoma appears as a “hot” nodule with suppression of surrounding thyroid tissue.



**Figure 3** Full spectrum of TSH receptor mutations identified in toxic adenoma, toxic multinodular goiters, and patients with autosomal dominant nonautoimmune hyperthyroidism (information can be accessed from the TSH receptor database at <http://www.uni-leipzig.de/~innere/TSH>).

**Table I Clinical Manifestations of Thyrotoxicosis**

**Complaints**

- Nervousness
- Fatigue
- Increased perspiration
- Heat intolerance
- Tremor
- Hyperactivity
- Palpitation
- Appetite change (usually increase)
- Weight change (usually loss)
- Menstrual disturbances

**Clinical features**

- Hyperactivity
- Tachycardia/arrhythmia
- Systolic hypertension
- Warm, moist smooth skin
- Tremor
- Hyperreflexia
- Muscle weakness

*Note.* Classic features of hyperthyroidism are often absent in the elderly.

Diagnosis of a toxic adenoma is based on three characteristics: (1) Abnormal results are obtained on thyroid function tests. Typically, there is overt hyperthyroidism {suppressed TSH, elevated thyroid hormones [triiodothyronine (T3) and (T4)]}. However, dependent on the autonomous cell mass, TA can also be associated with subclinical hyperthyroidism (TSH suppressed, T3 and T4 normal) or may even be present in a still euthyroid patient (TSH normal, T3 and T4 normal). (2) A palpable and/or sonographically localized thyroid nodule is present. (3) There is increased radioiodine or <sup>99m</sup>Tc-pertechnetate uptake in the nodule concomitant with a decreased uptake in the surrounding extranodular thyroid tissue. This typical scintiscan finding of a TA has led to the notion of a “hot” nodule (Fig. 4). If a TA is suspected in a patient with euthyroid function, it is sometimes necessary to induce an artificial state of TSH suppression by administration of thyroid hormones. In this state, a “suppression” scan can be performed, in which all normal tissue, but not the autonomous tissue, will show decreased uptake of the



administration of excessive amounts of iodine, e.g., in form of contrast medium widely used for angiography or computer tomography scanning. Iodine-induced hyperthyroidism is a common problem in regions with iodine deficiency. Thereby, “iodine contamination,” which can be easily demonstrated by the elevated urinary iodine excretion, may even trigger the onset of a life-threatening thyroid storm.

Due to the somatic molecular defect, there is practically no spontaneous cure in toxic adenoma, which indicates that once the TA becomes clinically manifest, definite treatment is indicated. Importantly, ~97% of TA are benign thyroid lesions, so that patients can be confidently reassured about the “good nature” of the toxic nodule.

## TREATMENT

The treatment of a toxic adenoma involves (1) symptomatic treatment of hyperthyroidism and (2) ablation of the autonomously functioning nodule.

The first step is to render a thyrotoxic patient euthyroid. This is achieved by the administration of so-called thyrostatic drugs, e.g., carbimazole, that inhibit thyroperoxidase, an enzyme essential for the production of thyroid hormones. Due to the long half-life of the predominantly produced thyroid hormone T<sub>4</sub> (7 days) and the intrathyroidal storage of large amounts of thyroid hormone, normalization of thyroid hormone levels usually takes several weeks. Frequent side effects of thyrostatic treatment are skin rashes (1–5%), whereas effects on the bone marrow (leukocytopenia or thrombocytopenia) constitute a rare (<0.5%) but potentially life-threatening complication (in the case of agranulocytosis), which requires immediate cessation of the thyrostatic drug.

Classical symptoms of hyperthyroidism, such as nervousness, sweating, tremor, and palpitations, are relieved faster and controlled by the additional administration of beta-blockers, preferably propranolol. Beta-blockers antagonize the autonomous sympathetic nervous system (and inhibit the peripheral conversion of T<sub>4</sub> to T<sub>3</sub>), which is greatly stimulated by the excess of thyroid hormones.

Once euthyroidism has been achieved, ablation of the toxic adenoma can be achieved either by thyroid surgery, usually in form of a hemithyroidectomy, if it is a solitary nodule, or by <sup>131</sup>I radioiodine therapy. The major advantage of thyroid surgery is that it immediately “cures” thyrotoxicosis; the obvious drawback is that it involves an invasive procedure with

potential side effects from anesthesia and/or thyroid surgery-specific complications, such as hypoparathyroidism and laryngeal nerve palsy.

In contrast, radioiodine treatment takes advantage of the characteristic feature of the “hot” nodule, i.e., increased iodine uptake. Thus, <sup>131</sup>I is preferentially accumulated in the autonomous tissue, which is progressively destroyed by the <sup>131</sup>I beta radiation. As a consequence, abrogation of hyperthyroidism by radioiodine therapy is a gradual process and may take several months, during which the patient requires further symptomatic treatment of hyperthyroidism.

As an alternative procedure, percutaneous ethanol injection of the toxic nodule has been employed and more than 200 patients with toxic nodules have been studied. The treatment usually involves three to eight ethanol injections into the nodule under ultrasound control. Studies from specialist centers have demonstrated that it is a cost-effective, reliable, and safe strategy if performed in experienced hands. Since there is only limited follow-up and evaluation in comparison with standard therapy of toxic adenoma, percutaneous ethanol injection is widely regarded as an alternative treatment for patients with contraindications to surgery or radioiodine therapy (e.g., old age, severe nonthyroidial illness, hemodialysis, pregnancy).

Thus, the clinician’s decision regarding the discussed treatment options for toxic adenoma will be based on several criteria, e.g., the patient’s age, the severity of hyperthyroidism, the size and structure of the thyroid gland, concomitant other medical conditions, and the patient’s personal preference.

## See Also the Following Articles

Medullary Thyroid Carcinoma • Pituitary Adenomas, TSH-Secreting • Thyroid, Aging and • Thyroid Carcinoma • Thyroid Imaging • Thyrotoxicosis: Diagnosis • Thyrotoxicosis, Overview of Causes • Toxic Multinodular Goiter • TSH (Thyroid-Stimulating Hormone; Thyrotropin)

## Further Reading

- Hicks, D. G., LiVolsi, V. A., Neidich, J. A., Puck, J. M., and Kant, J. A. (1990). Clonal analysis of solitary follicular nodules in the thyroid. *Am. J. Pathol.* **137**, 553–562.
- Krohn, K., and Paschke, R. (2001). Progress in understanding the etiology of thyroid autonomy. *J. Clin. Endocrinol. Metab.* **86**, 3336–3345.
- Laurberg, P., Pedersen, K. M., Vestergaard, H., and Sigurdsson, G. (1991). High incidence of multinodular toxic goitre in the elderly population in a low iodine intake area vs. high incidence of

- Graves' disease in the young in a high iodine intake area: Comparative surveys of thyrotoxicosis epidemiology in East-Jutland Denmark and Iceland. *J. Intern. Med.* **229**, 415–420.
- Parma, J., Duprez, L., Van Sande, J., Cochaux, P., Gervy, C., Mockel, J., Dumont, J., and Vassart, G. (1993). Somatic mutations in the thyrotropin receptor gene cause hyperfunctioning thyroid adenomas. *Nature* **365**, 649–651.
- Paschke, R., and Ludgate, M. (1997). The thyrotropin receptor in thyroid diseases. *N. Engl. J. Med.* **337**, 1675–1681.
- Reinwein, D., Benker, G., König, M. P., Pinchera, A., Schatz, H., and Schleusener, A. (1988). The different types of hyperthyroidism in Europe: Results of a prospective survey of 924 patients. *J. Endocrinol. Invest.* **11**, 193–200.
- Studer, H., Peter, H. J., and Gerber, H. (1989). Natural heterogeneity of thyroid cells: The basis for understanding thyroid function and nodular goiter growth. *Endocr. Rev.* **10**, 125–135.
- Trülsch, B., Krohn, K., Wonerow, P., Chey, S., Holzapfel, H. P., Ackermann, F., Führer, D., and Paschke, R. (2001). Detection of TSH receptor and G<sub>sα</sub> mutations in 75 toxic thyroid nodules by denaturing gradient gel electrophoresis. *J. Mol. Med.* **78**, 684–691.



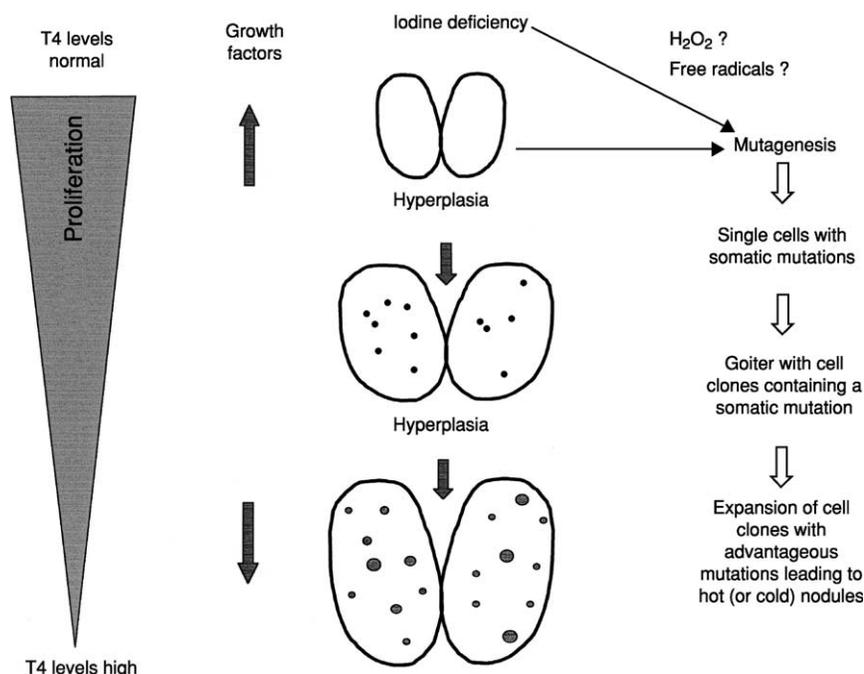
**Figure 1** Scintiscan of a patient with toxic multinodular goiter and subclinical hyperthyroidism. Note areas of increased uptake (arrows) in the right thyroid lobe and the isthmus region and decreased uptake in the surrounding thyroid tissue.

Diffuse goiters often undergo nodular transformation later in life and evolve into euthyroid multinodular goiter. If autonomously functioning nodules within these nontoxic multinodular goiters produce thyroid hormones in excess, the serum hormone level increases and causes thyrotoxicosis. Similar to solitary toxic thyroid adenomas, somatic mutations in the TSH receptor have also been detected in hot nodules of toxic multinodular goiters (see Fig. 4). Moreover, these mutations have also been identified in microscopic autonomously functioning areas of euthyroid goiters. Growth advantage as the result of an advantageous somatic mutation (e.g., in the TSH receptor or  $G_{s\alpha}$  protein) appears in a single thyrocyte and induces monoclonal growth. Thus, a monoclonal origin has also been demonstrated in nodules within a toxic multinodular goiter. However, other nodules within

the same goiter appear to be polyclonal. As yet, their precise etiology is unknown but epigenetic, e.g., environmental, effects that cause a growth advantage in a number of cells simultaneously have been proposed to explain polyclonal thyroid nodules.

### TOXIC MULTINODULAR GOITER VERSUS GRAVES' DISEASE

In iodine-deficient areas, the distinction between toxic autoimmune and nonautoimmune forms of hyperthyroidism (e.g., Graves' disease and toxic multinodular goiter) is often complicated by findings that do not correspond to typical diagnostic criteria. In this regard, several possibilities may be encountered. (1) The disorder may be incorrectly classified as



**Figure 2** Hypothesis for the evolution of clonal thyroid nodules in multinodular goiters. Hyperplasia due to iodine shortage increases the likelihood for somatic mutations that may confer a growth and functional advantage. Further expansion of cell clones ultimately results in “hot” and “cold” nodules that characteristically prevail together in multinodular goiters. Modified with permission from Krohn, K., and Paschke, R., *J. Clin. Endocrinol. Metab.* **86**(7), 3336–3345, 2001, The Endocrine Society.

nonautoimmune hyperthyroidism due to the failure to detect TSH receptor antibodies (TRAB) with the routinely used *in vitro* assays (TBII assay). This is illustrated by the detection of TRAB with more sensitive cell culture assays in sera of hyperthyroid patients who show a scintiscan compatible with toxic multinodular goiters (e.g., Graves’ disease superimposed on a preexisting nodular thyroid gland, which is prevalent in iodine-deficient regions) or in patients who show a scintiscan with diffuse uptake and have no other clinical signs of Graves disease (termed “diffuse” autonomy). (2) Nonautoimmune hyperthyroidism occurring as the result of TSH receptor germline mutations without detectable TSH receptor antibodies might clinically masquerade as Graves’ disease.

#### AUTOSOMAL-DOMINANT NONAUTOIMMUNE HYPERTHYROIDISM AS A RARE FORM OF TOXIC MULTINODULAR GOITER

Unusually, in some patients there may be a family history of thyroid autonomy and a characteristic course of frequent relapses of hyperthyroidism following thyrostatic therapy or partial thyroidectomy. Depending on the age of onset (neonatal period

to adulthood), these patients may present with a diffusely enlarged goiter or a TMG. The underlying cause of this condition is an activating germline mutation in the TSH receptor gene, which can be confirmed through molecular analysis from leukocyte DNA extracted from a routine blood sample. Patients with an activating TSH receptor germline mutation require ablative thyroid treatment in the form of a total thyroidectomy and sometimes in addition an ablative dose of radioiodine to prevent further relapses. Genetic counseling is also mandatory as the condition is autosomal dominantly inherited.

#### CLINICAL FEATURES AND DIAGNOSIS

The diagnosis of TMG may be suggested by clinical inspection, thyroid palpation, or ultrasound examination. Generally, toxic multinodular goiter is seen more often in the elderly. In this patient group, the clinical features of hyperthyroidism are often not apparent to the clinician; i.e., often the main concern for consultation is that of weight loss and worry about cancer, and tremor, sweating, and hyperkinesia are absent. The presence of atrial fibrillation and congestive heart failure in the presence of a nodular goiter should alert the clinician to the diagnosis, especially

when there is accompanying weight loss. Biochemical diagnosis of hyperthyroidism is achieved with standard thyroid function tests.

Examination of the goiter and neck structures reveals the degree of enlargement and multinodularity of the gland as does ultrasonography. Imaging by scintiscan indicates which nodules are functional and helps to ascertain the extent of possible retrosternal extension of the goiter. Fine-needle aspiration cytology should be performed on prominent cold nodules within a TMG.

## TREATMENT

The management of patients with TMG will to some extent depend on the age of the patient, the severity of hyperthyroidism, the size of the thyroid gland, the presence of additional nonfunctioning thyroid nodules, and the coexistence of other medical illness. The first requirement must involve measures to render the patient euthyroid as well as arrange for treatment of a multinodular goiter that may be large and compressive. There is controversy as to whether treatment for the subclinical hyperthyroidism in such patients is indicated. There is evidence, however, that bone mineral density is reduced in postmenopausal women with TMG and there is also an increased risk of atrial fibrillation. In these situations, anti-thyroid treatment

is beneficial although the effect on bone density has yet to be established.

Therapy for the toxic nodular goiter patient is either by radioiodine ( $^{131}\text{I}$ ) administration or by surgical resection of the gland (see Fig. 3). If the patient is hyperthyroid, appropriate anti-thyroid drug therapy must be given so that euthyroidism is present prior to operation. If heart failure is present, this can be managed with standard therapy, including anticoagulants if atrial fibrillation is present.

Surgical treatment is indicated for large goiters with or without compressive features especially in the younger age group. A hemithyroidectomy, subtotal thyroidectomy, or near-total thyroidectomy may be performed depending on the extent of disease as assessed preoperatively and at surgery. Relative contraindications to surgery include advanced age and general ill health. Following surgery, recurrent hyperthyroidism will occur in less than 10–20% of patients but hypothyroidism will be seen in up to 70% of patients, depending on the extent of surgery. Potential long-term complications of surgery include hypocalcemia due to hypoparathyroidism or hoarseness due to laryngeal nerve lesion, both of which are rare ( $\leq 1\%$ ) if the operation is performed by an experienced thyroid surgeon.

Radioiodine therapy is effective in terms of both eradicating hyperthyroidism and reducing thyroid

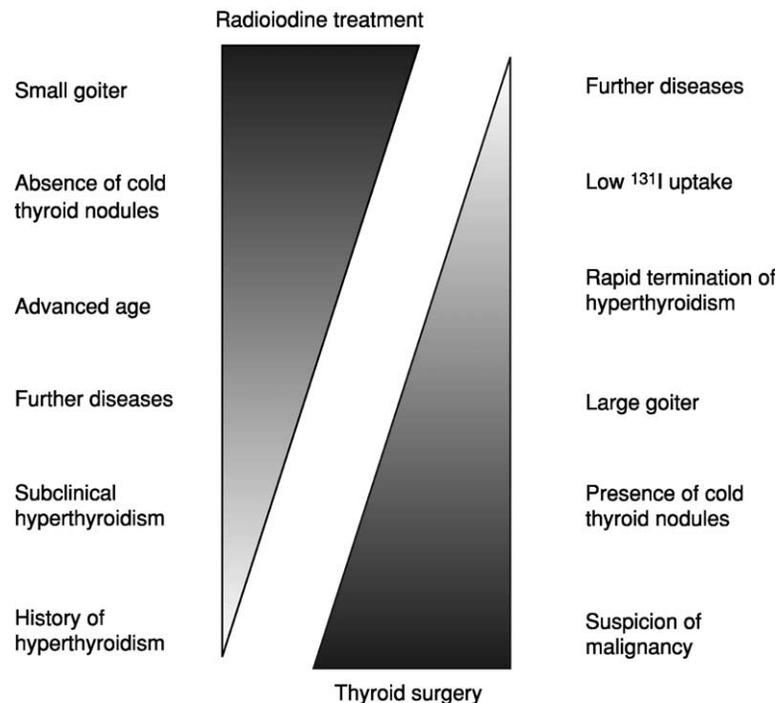


Figure 3 Clinical features influencing the treatment strategy of toxic adenomas and toxic multinodular goiters.

gland volume. Usually, anti-thyroid drugs must be given before and after  $^{131}\text{I}$  therapy (generally for weeks up to few months) to ensure euthyroidism. Postradioiodine hypothyroidism occurs in approximately 31% of patients after an 8-year follow-up period and must be checked for on a regular basis because it evolves gradually.

A review of 253 patients with toxic multinodular goiter treated at the Mayo Clinic over a period of 18 years showed that both surgery and  $^{131}\text{I}$  treatment are effective for this condition. Surgery was effective sooner, thereby minimizing cardiac effects of coexisting tachycardia. The surgical approach also rarely required re-treatment and did remove any coexisting malignancy.  $^{131}\text{I}$  was associated with a 20% requirement for a second treatment. Posttreatment hypothyroidism was higher in the surgical group but this can be readily treated by oral L-thyroxine administration.

The long-term management of patients with toxic multinodular goiter is directed at the detection of thyroid dysfunction (hypothyroidism or a relapse of the hyperthyroid state), relapse of goiter, and treatment of surgical complications, e.g., calcium administration in the case of hypoparathyroidism. After surgery, thyroxine is often administered routinely, independent of thyroid function, although there is no evidence to suggest that this is beneficial unless the patient is hypothyroid. Follow-up of patients with toxic multinodular goiter (as well as other forms of thyroid disease) should be performed on a routine basis by experienced doctors or thyroid specialists.

### See Also the Following Articles

Goitrogens, Environmental • Nontoxic Goiter • Thyroid, Aging and • Thyroid Disease, Genetic Factors in • Thyroid

Imaging • Thyroid Nodule • Thyrotoxicosis: Diagnosis • Thyrotoxicosis, Overview of Causes • Toxic Adenoma

### Further Reading

- Baltisberger, B. L., Minder, C. E., and Burgi, H. (1995). Decrease of incidence of toxic nodular goitre in a region of Switzerland after full correction of mild iodine deficiency. *Eur. J. Endocrinol.* **132**, 546–549.
- Corvilain, B., Dumont, J. E., and Vassart, G. (2000). Toxic adenoma and toxic multinodular goiter. In “Werner and Ingbar’s The Thyroid” (L. E. Braverman and R. D. Utiger Lippincott, eds.), 8th ed., pp. 564–572. Raven Press, Philadelphia, PA.
- Duprez, L., Parma, J., Van Sande, J., Allgeier, A., Leclere, J., Schwartz, C., Delisle, M. J., Decoux, M., Orgiazzi, J., and Dumont, J. (1994). Germline mutations in the thyrotropin receptor gene cause non-autoimmune autosomal dominant hyperthyroidism. *Nat. Genet.* **7**, 396–401.
- Erickson, D., Gharib, H., Hongzhe, L. I., and van Heerden, J. A. (1998). Treatment of patients with toxic multinodular goiter. *Thyroid* **8**, 277–282.
- Führer, D., Warner, J., Sequeira, M., Paschke, R., Gregory, J., and Ludgate, M. (2000). Novel TSHR germline mutation (met463-val) masquerading as Graves’ disease in a large Welsh kindred with hyperthyroidism. *Thyroid* **10**, 1035–1041.
- Kopp, P., Kimura, E. T., Aeschmann, S., Oestreicher, M., Tobler, A., Fey, M. F., and Studer, H. (1994). Polyclonal and monoclonal thyroid nodules coexist within human multinodular goiters. *J. Clin. Endocrinol. Metab.* **79**, 134–139.
- Krohn, K., Wohlgemuth, S., Gerber, H., and Paschke, R. (2000). Hot microscopic areas of iodine deficient euthyroid goiters contain constitutively activating TSH receptor mutations. *J. Pathol.* **192**, 37–42.
- Laurberg, P., Pedersen, K. M., Vestergaard, H., and Sigurdsson, G. (1991). High incidence of multinodular toxic goitre in the elderly population in a low iodine intake area vs. high incidence of Graves’ disease in the young in a high iodine intake area: Comparative surveys of thyrotoxicosis epidemiology in East-Jutland Denmark and Iceland. *J. Intern. Med.* **229**, 415–420.
- Studer, H., Peter, H. J., and Gerber, H. (1989). Natural heterogeneity of thyroid cells: the basis for understanding thyroid function and nodular goiter growth. *Endocr. Rev.* **10**, 125–135.

composed of exons 1 to 6, variant I is composed of exons 1 to 5 and 6B, and variant II is composed of exons 1 to 5,  $\alpha$ , and 6B. The main transcript (NCBI accession no. NM\_003236) is observed at approximately 4.5 to 4.8 kb. Poorly characterized shorter transcripts of 1.6 and 0.35 kb have been reported in some cultured cell lines, but their importance is unclear. The half-life of TGF- $\alpha$  mRNA varies widely and can be decreased or increased, for example, by the von Hippel-Lindau tumor suppressor in the 786-0 renal cell carcinoma line or by autoinduction in keratinocytes, respectively. In adult tissues, TGF- $\alpha$  mRNA is present in numerous cell types of endodermal, ectodermal, and mesodermal origin. Expression is highly variable and inducible, for example, in the nervous system following damage. Similarly, expression is induced during inflammation and repair in activated fibroblasts, macrophages, monocytes and neutrophils, and eosinophils. The role of TGF- $\alpha$  in various tissues is discussed in more detail later. TGF- $\alpha$  is also expressed during embryonic development in numerous tissues, including lung, liver, digestive tract, testis, ovary, mammary gland, and several areas of the central nervous system (CNS). Finally, TGF- $\alpha$  expression is up-regulated in several disease states, including multiple types of carcinoma, psoriasis, cystic fibrosis, idiopathic pulmonary fibrosis, palmar fibromatosis, and Menetrier's disease of the gastric mucosa.

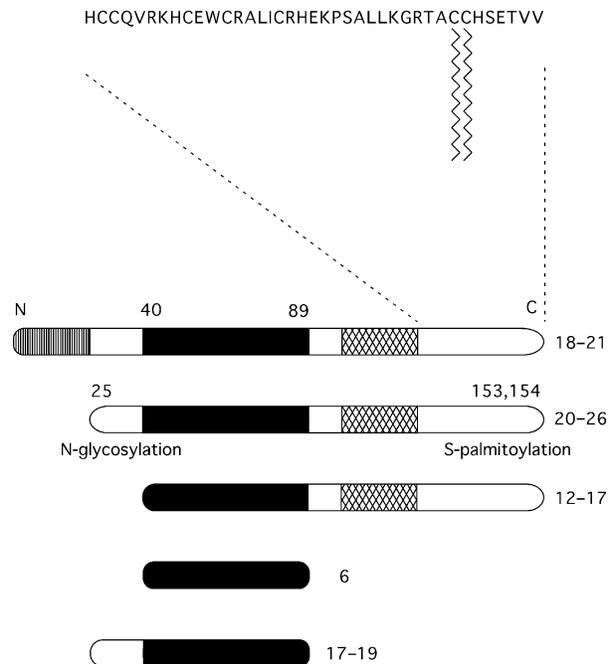
### Maturation

#### Precursor

Although soluble human TGF- $\alpha$  is a 50-amino acid polypeptide, it is synthesized as a 160-amino acid, 18-kDa, type 1 membrane protein termed pro-TGF- $\alpha$  (Fig. 1). The N-terminal 99-amino acid region of the precursor undergoes heterogeneous N- and O-linked glycosylation, raising the apparent size to 20 to 22 kDa. N-linked glycosylation is required for maturation in some, but not all, cell types. This structural heterogeneity is observed in the soluble TGF- $\alpha$ , which varies in apparent size from 6 to 20 kDa. In addition to glycosylation and disulfide bridge formation, pro-TGF- $\alpha$  is covalently modified by palmitoylation (C16:0) of two cysteine residues (153 and 154) located near the C terminus.

#### Secretory Pathway

The C terminus of wild-type pro-TGF- $\alpha$  contains a class I PDZ-binding motif (TVV), which is required for export from the endoplasmic reticulum. Therefore, different export signals may operate for the C-terminal splice variants, both of which are also expressed on the cell surface. Early in the export



**Figure 1** Human TGF- $\alpha$  polypeptides. The cytoplasmic domain of pro-TGF- $\alpha$  (residues 125–160) contains a putative basolateral sorting motif (dileucine 146 and 147), the palmitoylation sites (dicysteine 153 and 154), and a C-terminal type I PDZ-binding motif (residues 158–160) that are required for export from the endoplasmic reticulum. The schematics of human TGF- $\alpha$ , based on the annotation in the SwissProt database entry TGFA\_HUMAN, illustrate the various forms arising during maturation, including the N-terminal signal peptide (residues 1–23, vertical lines), propeptide (residues 24–39), EGF-like domain (residues 40–89, solid), and transmembrane region (residues 99–124, hatched lines). One consensus N-linked glycosylation site exists at Asn(25) in the propeptide. Sizes to the right of each form are apparent values in kilodaltons reported for various cell types using gel electrophoresis.

pathway, the C terminus of wild-type pro-TGF- $\alpha$  binds a protein termed TACIP18 (pro-TGF- $\alpha$  cytoplasmic domain-interacting protein 18). This protein is also known by several other names, including mda-9 (melanoma differentiation-associated gene-9) and SNTA1 (syntenin 1) or SDCBP (syndecan-binding protein) because it binds syndecans among several other proteins as well as the lipid phosphatidylinositol(4,5)bisphosphate. The interaction with pro-TGF- $\alpha$  appears to require both PDZ domains in TACIP18, although the role of this multifunctional protein in pro-TGF- $\alpha$  trafficking remains to be elucidated. A cytoplasmic protein with two PDZ domains, termed p59, also binds the C-terminal residues of pro-TGF- $\alpha$ , although the N-terminal PDZ domain of p59 alone is sufficient for binding. This protein is the human orthologue of the murine GRASP55 protein, which has been implicated in the assembly and stacking of

intermediate Golgi cisternae. As with TACIP18/syntaxin 1, it remains to be shown whether or not p59/GRASP55 is involved in the trafficking of pro-TGF- $\alpha$  through the early Golgi compartment and what its function is in this process. On leaving the Golgi compartment, pro-TGF- $\alpha$ , like its receptor, is sorted to the basolateral surface in polarized epithelial cells. Targeting appears to employ more than one intrinsic sorting signal in the cytoplasmic region, but the C-terminal PDZ-binding motif is not required.

#### **Proteolysis–Ectodomain Shedding**

At the cell surface, mature TGF- $\alpha$  is released from pro-TGF- $\alpha$  by proteolysis (Fig. 1). Proteolytic release of extracellular domains is a feature of many different cell surface proteins and is generically termed ectodomain shedding. The soluble TGF- $\alpha$  released from the cell surface can stimulate either the host cell (autocrine) or proximal cells (paracrine). Although the relative rates of proteolysis at the N- and C-terminal sites can vary, current genetic and biochemical data indicate that the tumor necrosis factor- $\alpha$ -converting enzyme (TACE), also known as ADAM17 (a disintegrin and metalloprotease 17), is responsible for both events. The rate of proteolysis is increased via intracellular protein kinase C and mitogen-activated protein (MAP) kinase signaling pathways induced by a variety of factors, including phorbol esters, growth factors, serum, and retroviral transformation. In contrast, overexpression of the tetraspanin protein CD9, which binds the ectodomain of pro-TGF- $\alpha$ , inhibits the increase in shedding in response to phorbol esters or serum. However, the recent identification of a complex containing CD9, pro-heparin-binding (HB)-EGF, and the HB-EGF shedding protease ADAM10 suggests the possibility that CD9 overexpression may artifactually disrupt pro-TGF- $\alpha$  proteolysis. In gastric juice, the seven C-terminal residues of TGF- $\alpha$  are removed by pepsin to produce TGF- $\alpha$ (1–43), thereby reducing its biological activity by more than 50%.

#### **Structure**

The three-dimensional structure of the mature TGF- $\alpha$  polypeptide has been determined in a solution using nuclear magnetic resonance spectroscopy. The crystal structure of TGF- $\alpha$  bound to its receptor is described later. TGF- $\alpha$  belongs to the type I subset (InterPro IPR006210) of the large family of EGF-like domains (InterPro IPR006209). The solution structure possesses the set of three disulfide bonds, which link cysteine residues 8–21, 16–32, and 34–43, and two pairs of antiparallel  $\beta$ -sheets (composed of residues 19–33 and 38–46) that characterize EGF-like domains.

## **Receptor**

At the cell surface, the EGF receptor (also termed HER or ErbB) family binds TGF- $\alpha$ . The EGF receptor (HER1 or ErbB1), as either a homo- or heterodimer with other members of the family, binds TGF- $\alpha$  with the highest affinity (low nanomolar range), similar to its affinity for six other structurally similar ligands: EGF, HB-EGF, amphiregulin, betacellulin, epiregulin, and epigen. TGF- $\alpha$  appears to bind certain other members of the EGF receptor family, but the affinity is several orders of magnitude lower. High-affinity binding by the EGF receptor can be inhibited in trans, termed transmodulation (e.g., via agonists that stimulate protein kinase C activity).

#### **Expression**

The EGF receptor is ubiquitously expressed on adherent cell types, but with the exception of certain early progenitors, it is absent from most hemopoietic cells. The levels of receptor expression vary widely among cell types, during development, and in several disease states such as psoriasis and multiple types of solid tumor.

#### **Ligand Binding**

##### **Structure**

The crystal structure of TGF- $\alpha$  bound to a soluble fragment of the EGF receptor comprising the N-terminal 501 residues has allowed resolution to 2.5 Å. More than one-third of the accessible surface area of TGF- $\alpha$  is buried on binding to the receptor, and nearly two-thirds of the amino acid residues of TGF- $\alpha$  make contact with the receptor. In the back-to-back, 2:2 (TGF- $\alpha$ :EGF receptor) complex, the TGF- $\alpha$  molecules are bound on opposite sides and separated by more than 70 Å. Binding appears to stabilize the structure of TGF- $\alpha$  relative to the form in free solution; a short third  $\beta$ -strand is generated (residues 4–6) as part of the principal  $\beta$ -sheet formed by residues 19 to 33 and the C terminus becomes more ordered. Ligand binding stabilizes an active conformation of the receptor, which favors dimerization and in which allosteric inhibition of the intracellular protein tyrosine kinase domain by the extracellular region is relieved and the kinase activity is enhanced.

#### **Pro-TGF- $\alpha$ as an EGF Receptor Ligand**

Early studies gave rise to the conclusion that soluble mature TGF- $\alpha$  or pro-TGF- $\alpha$  can directly activate EGF receptors. Accordingly, the soluble form was

suggested to mediate autocrine and paracrine signaling, whereas the membrane precursor was suggested to stimulate EGF receptors located on neighboring cells, a mechanism termed juxtacrine signaling. In some studies, including inhibition of shedding by CD9, stimulation by pro-TGF- $\alpha$  was reported to be even more effective than that by soluble TGF- $\alpha$ . However recent studies have questioned this idea and indicated instead that although pro-TGF- $\alpha$  can bind the EGF receptor, this interaction does not lead to receptor signaling in the absence of shedding. Thus, although it is clear that pro-TGF- $\alpha$  binds to the EGF receptor, it remains unclear whether juxtacrine signaling by pro-TGF- $\alpha$  can occur in some cell-cell interactions or whether it only appears to exist and localized shedding at cell-cell contacts actually leads to signaling.

#### **Intracellular Signaling**

It has been suggested that pro-TGF- $\alpha$  itself may be able to act as a receptor and stimulate intracellular signals at the cell surface in response to a juxtacrine interaction with the EGF receptor. Although protein serine, threonine and tyrosine kinase activity, and proteins of 86 and 106 kDa associate with pro-TGF- $\alpha$  at the surface of Chinese hamster ovary (CHO) cells overexpressing pro-TGF- $\alpha$ , the identity of these proteins and whether or not they transduce signals from a pro-TGF- $\alpha$ -EGF receptor complex are unclear. A review of the multiple intracellular signaling pathways stimulated by the EGF receptor in various cell types is beyond the scope of this article, although a recent review is cited in the Further Reading section.

#### **Internalization and Degradation**

Although in many cellular assays, including receptor autophosphorylation, TGF- $\alpha$  and EGF produce very similar results, in some cases TGF- $\alpha$  and EGF have been found to possess differential potency. One suggestion as to why this should occur arises from the observation that TGF- $\alpha$  and EGF cause the EGF receptor to follow different internalization pathways. Whereas EGF causes signaling to be terminated by degradation of the ligand and receptor within the lysosomal compartment, TGF- $\alpha$  dissociates from the receptor at less acidic conditions, and this in turn causes a smaller amount of ligand and receptor degradation and allows a fraction of the internalized ligand and receptor to recycle back to the cell surface. However, in some cell types, TGF- $\alpha$  is degraded very efficiently by insulin-degrading enzyme (IDE) within the mildly acidic endosomal compartment.

## **PHYSIOLOGICAL FUNCTIONS IN HEALTH AND DISEASE**

Much is derived from animal, particularly rodent, models and *in vitro* studies with cell lines. These are not always apparently consistent with *in vivo* results, and it is clear that growth and differentiation can differ among cell types and be fundamentally affected by the context in which they are analyzed. Furthermore, in transgenic mutant lines, compensatory activity by other EGF receptor ligands may affect phenotypic presentation.

#### **Tissue Distribution**

TGF- $\alpha$  and the EGF receptor are expressed in many developing and adult tissues, where TGF- $\alpha$  has been implicated in cell migration, growth, and differentiation. The physiological importance of TGF- $\alpha$  is equally pleiotropic, encompassing the organization, homeostasis, protection, and repair of several types of tissues. Notable among the targets of TGF- $\alpha$  are various epithelial cells, and often also associated mesenchymal cells, in the gastrointestinal tract, lung, liver, kidney, mammary gland, dermis, gonads, skeletal muscle, and nerve cells within the central and peripheral systems. Although TGF- $\alpha$  commonly acts via autocrine or paracrine signaling in solid tissues, it can also mediate paracrine signaling by activated macrophages, monocytes, neutrophils, and eosinophils. Many studies have taken advantage of murine strains containing either decreased or increased levels of TGF- $\alpha$ . A TGF- $\alpha$ -deficient line termed waved-1 (*wa-1*) carries a naturally occurring mutation that causes a 5- to 30-fold decrease in TGF- $\alpha$  expression. TGF- $\alpha$  knockout lines have also been engineered. In addition, various lines have been engineered to overexpress TGF- $\alpha$  under the control of a variety of promoters.

#### **Cancer**

TGF- $\alpha$  is a potent mitogen for many different types of epithelial and mesenchymal cells, whose transgenic overexpression under a variety of different nonspecific or tissue-specific promoters causes different types of hyperproliferative and preneoplastic lesions, including benign papillomas, adenomas, cysts, and fibroses. However, in most tissues, tumorigenesis is absent without secondary oncogenic factors such as treatment with chemical carcinogens, altered expression of *bona fide* cellular and viral oncogenes, and viral

infection (particularly hepatitis B and hepatitis C). Nonetheless, a few types of epithelial tumors have been reported as a consequence of transgenic TGF- $\alpha$  overexpression, notably hepatocellular carcinoma, mammary adenocarcinoma, pancreatic carcinoma, and coagulation gland carcinoma *in situ*. Increased expression of TGF- $\alpha$  has been reported in numerous primary epithelial tumors, including carcinomas of the liver, pancreas, breast, kidney, lung, epidermis, prostate, larynx, gallbladder, gastrointestinal tract, and ovary. In several studies, the importance of autocrine growth stimulation was suggested by increased EGF receptor expression, for example, in gastric and ovarian carcinomas. Furthermore, blocking autocrine stimulation with an anti-TGF- $\alpha$  antibody inhibits growth of primary ovarian cancer cells and the GEO colon carcinoma cell line. Finally, increasing or decreasing TGF- $\alpha$  expression by expression of sense or antisense cDNA increases or decreases the tumorigenicity of GEO cells, respectively. The level of TGF- $\alpha$  expression alone is not prognostic for most carcinomas, including breast, larynx, gallbladder, lung, and colon, but it is apparently indicative of longer, progression-free survival of ovarian carcinoma patients. Other factors must be considered before productive TGF- $\alpha$  signaling is assumed. For example, studies have shown that TACE, which is required for shedding a variety of growth factors such as TGF- $\alpha$ , is also required for tumorigenesis in a rodent model. Finally, TGF- $\alpha$  is chemotactic for vascular endothelial cells and promotes angiogenesis in experimental models. However, there is no direct evidence of an angiogenic role in primary tumors.

## Role in Specific Tissues

### **Hemopoietic System**

TGF- $\alpha$  plays a role in wound repair and is differentially secreted by granulocytes during inflammatory responses in many different tissues. Expression is induced in activated macrophages that are recovered, for example, from wound sites or that are stimulated in culture from which soluble TGF- $\alpha$  is released. Monocytes and neutrophils store TGF- $\alpha$  in poorly characterized cytoplasmic granules. Eosinophils constitutively produce TGF- $\alpha$ , and transcription is increased by treatment with the cytokines interleukin-3 (IL-3) and granulocyte-macrophage colony-stimulating factor (GM-CSF). However, secretion of soluble TGF- $\alpha$  is increased by GM-CSF but not by IL-3. Eosinophils are a major source of TGF- $\alpha$  at malignant sites in the oral epithelium. Finally, TGF- $\alpha$  is

expressed in bone marrow by progenitor cells in the myeloid lineage, consistent with the observation that progenitor cells within this lineage are the only hemopoietic cells to express the EGF receptor.

### **Gastrointestinal System**

Many parts of the gastrointestinal tract synthesize TGF- $\alpha$  to enhance the epithelial cell migration and proliferation required to compensate for continual cell loss and to protect the integrity of the epithelium from lesions caused, for example, by low pH or oxidation. Consequently, transgenic overexpression of TGF- $\alpha$  causes foveolar hyperplasia, and TGF- $\alpha$ -knockout mice and overexpressing mice have an increased and decreased susceptibility, respectively, to experimentally induced colitis. Hyperplasia similar to that seen in transgenic mice and TGF- $\alpha$  overexpression by the gastric epithelium occur in Menetrier's disease, although the etiology of this condition is unknown. Although TGF- $\alpha$  stimulates cell division, migration, and prostaglandin estradiol (PGE<sub>2</sub>) synthesis in gastric epithelial cells in culture, the protective effect against mucosal injury in rodent models appears to be mediated indirectly by stimulation of calcitonin gene-related peptide (CGRP) release from capsaicin-sensitive sensory neurons. TGF- $\alpha$  also inhibits stimulated secretion of gastric acid (e.g., by gastric mucosa treated with histamine) and of chloride ions (e.g., by intestinal epithelial cells treated with carbachol). The latter has been shown to occur by transactivation of the EGF receptor via increased shedding of TGF- $\alpha$  and subsequent autocrine signaling.

### **Lung**

In the lung, TGF- $\alpha$  is expressed in different epithelial and mesenchymal cell types throughout fetal development and is likely to regulate growth and maturation. In the adult lung, TGF- $\alpha$ , arising also from alveolar macrophages, reduces acute damage caused, for example, by inhalation of noxious substances at least in part by attenuating inflammation. Transgenic overexpression of TGF- $\alpha$  by murine lung epithelial cells causes pulmonary fibrosis and remodeling, including enlarged airspaces, whereas TGF- $\alpha$  knockout mice exhibit reduced pulmonary fibrosis following chemical insult.

### **Liver**

Most studies on the role of TGF- $\alpha$  in the liver have focused on neoplasia and regeneration. The level of TGF- $\alpha$  mRNA increases after partial hepatectomy, and the level of TGF- $\alpha$  in the serum correlates with the degree of regeneration. In animal models,

autocrine mitogenic stimulation arises from expression induced in hepatocytes. In fetal liver, TGF- $\alpha$  is expressed by bile ducts and hepatocytes, whereas in adults, the expression of TGF- $\alpha$ , but not the EGF receptor, is lost in hepatocytes. The mitogenic action of bile acids toward bile duct cholangiocytes arises from a stimulation of TGF- $\alpha$  ectodomain shedding. Constitutive expression of TGF- $\alpha$  in transgenic mice leads to increased hepatocyte proliferation, liver enlargement, and neoplasia. In addition to its mitogenic activity, TGF- $\alpha$  has been shown to possess antiapoptotic activity by virtue of its ability to inhibit Fas-mediated hepatic apoptosis.

### **Kidney**

Expression of TGF- $\alpha$  in the fetal and adult kidney is detectable but low. Expression in rodents has been reported to be complicated by antibody cross-reactivity and to be localized to the proximal tubules of the adult renal cortex. Transgenic overexpression of TGF- $\alpha$  in the kidney causes renal enlargement, glomerular mesangial expansion, and cyst formation. Furthermore, early-stage lesions in adult polycystic kidney disease are positive for a proximal tubule marker protein and display increased levels of TGF- $\alpha$ , although end-stage cysts show markers for different tubule regions.

### **Mammary Tissue**

TGF- $\alpha$  is expressed in mammary epithelial cells and plays a role in mammary development and lactation. Overexpression of TGF- $\alpha$  in mice causes abnormal duct morphology and hyperplasia and can lead to failure of lactation and mammary adenocarcinoma formation. In regard to lactation itself, the production of whey acidic protein (WAP) by rodent mammary epithelial cells is inhibited by TGF- $\alpha$  produced when the alveolar extracellular matrix is disrupted. TGF- $\alpha$  is present in milk, but at a much lower concentration than is EGF. Transgenic overexpression of TGF- $\alpha$  in mammary epithelia under the WAP promoter restricts high-level expression to pregnancy and lactation and increases the incidence of preneoplastic and neoplastic lesions with strain-dependent penetrance.

### **Reproductive Organs**

In human oocytes the level of TGF- $\alpha$  is greatest in primordial follicles, but expression by theca and granulosa cells increases with development, reaching a maximum at the lutein cell stage of the mid-luteal phase. Expression by rodent theca cells is indirectly enhanced by follicle-stimulating hormone (FSH) via increased estradiol synthesis by granulosa cells.

TGF- $\alpha$  overexpression in a murine transgenic line attenuates the action of gonadotropins on ovarian steroidogenesis, although in the brain an opposing mechanism, possibly acting on hypothalamic release of luteinizing hormone-releasing hormone (LHRH), enhances female sexual maturation. In the latter example, the action of TGF- $\alpha$  on hypothalamic astrocytes is proposed to stimulate PGE<sub>2</sub> release by glial cells, followed in turn by LHRH release by neurons in response to PGE<sub>2</sub> stimulation. In the testes, expression is strongest in Leydig cells, and weak expression has been reported in other cell types, although the testes appear normal in TGF- $\alpha$  null mice.

### **Skin**

TGF- $\alpha$  is produced by embryonic epidermal layers and in the adult stratified epidermis. It stimulates keratinocyte migration *in vitro*, and an autocrine pathway induces gene expression of TGF- $\alpha$  and other EGF receptor ligands, although the increased level of TGF- $\alpha$  mRNA is more stable and persists far longer than do the amphiregulin and HB-EGF transcripts. The level of TGF- $\alpha$  in cutaneous exudates increases after damage caused by abrasion or ultraviolet irradiation, consistent with an autocrine role in wound repair. The phenotypes of mice overexpressing TGF- $\alpha$  under the control of a keratin K14 or K1 promoter suggest a developmental role for TGF- $\alpha$  in controlling epidermal growth, with overexpression causing increased epidermal thickness in some areas of skin or wrinkling, respectively. TGF- $\alpha$ -deficient mice develop abnormal external and middle ear epidermal morphology, deranged hair follicles causing curly whiskers and wavy fur, and prematurely opened eyes at birth. A similar phenotype occurs in waved-2 mice, which carry a mutant EGF receptor gene, and is exacerbated by TACE deficiency. In studies using primary tissue from psoriasis patients, the level of TGF- $\alpha$  was increased in all layers of the epidermis relative to nonpsoriatic skin, although mRNA was increased only in the subcorneal layers. Finally, although several previous studies found no evidence of linkage, when considered as a single group, different mutations located in conserved regions of the 3' UTR of the human TGF- $\alpha$  gene show a weak link with nonsyndromic cleft lip with or without cleft palate and nonsyndromic cleft palate only.

### **Muscle**

Smooth muscle cells in several different types of embryonic and adult tissue have been shown to express TGF- $\alpha$ . Early reports suggested that TGF- $\alpha$  affects contractility in addition to its mitogenic action on

myocytes. Transgenic overexpression of TGF- $\alpha$  leads to a decrease in skeletal muscle mass, and a role in the early differentiation of myoblasts and in pyloric muscle hypertrophy also has been proposed.

### Nervous System

TGF- $\alpha$  is widely expressed in numerous different neuronal and glial cells of the CNS as well as in the peripheral nervous system, where it has roles in growth, development, and repair. For example, TGF- $\alpha$  regulates the proliferation of neural stem cells in the forebrain subependyma and is also a tropic factor for cultured mesencephalic dopaminergic neurons, which are severely depleted from the substantia nigra in TGF- $\alpha$  null mice. TGF- $\alpha$  regulates the differentiation of cultured retinal precursor cells. A role in neuronal damage and repair is indicated by the ability of TGF- $\alpha$  to indirectly reduce damage caused by neuronal excitotoxicity via glial cell stimulation and by the rapid and transient up-regulation of TGF- $\alpha$  expression by motor neurons following nerve crush in rodents. Mutation of the TGF- $\alpha$  open reading frame has been excluded from Alström syndrome and autosomal dominant Parkinson's disease, although the latter is linked to chromosome 2p13.

Several roles for TGF- $\alpha$  in controlling behavior have been reported using rodent models. Transgenic male mice overexpressing TGF- $\alpha$  are more aggressive and less able to cope with stress, although transgenic females display the opposite phenotype relative to nontransgenic individuals. An endocrine link is suggested by the observation that the aggressive behavior of males is lost following castration. Intriguingly, TGF- $\alpha$  exhibits rhythmic expression in the supra-chiasmatic nucleus (SCN) of the hypothalamus, where it appears to inhibit daily cycles of locomotor activity by paracrine stimulation of proximal neurons in the subparaventricular zone.

### Endocrine Glands

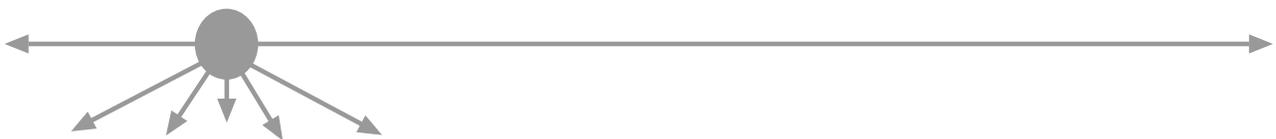
In addition to the endocrine tissues already mentioned, TGF- $\alpha$  is synthesized by the anterior pituitary, thyroid, parathyroid, adrenal cortex and medulla, and pancreas. In the pituitary gland, TGF- $\alpha$  synthesized locally mediates the mitogenic activity of estrogen *visa paracrine* signaling, but much remains to be understood regarding the physiological role of TGF- $\alpha$  in the other endocrine glands.

### See Also the Following Articles

EGF and Related Growth Factors • Pituitary Tumors, Molecular Pathogenesis • Receptor Tyrosine Kinase

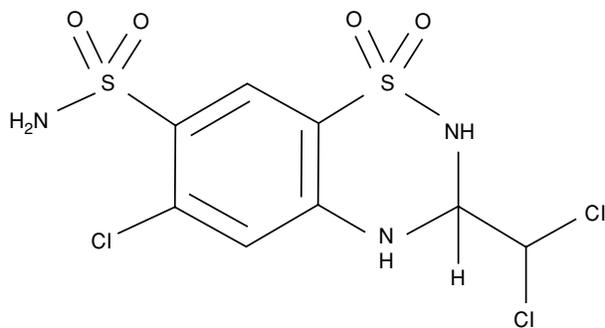
### Further Reading

- Borrell-Pages, M., Rojo, F., Albanell, J., Baselga, J., and Arribas, J. (2003). TACE is required for the activation of the EGFR by TGF-alpha in tumors. *EMBO J.* **22**, 1114–1124.
- Coffey, R. J., Romano, M., and Goldenring, J. (1995). Roles for transforming growth factor-alpha in the stomach. *J. Clin. Gastroenterol.* **21**, S36–S39.
- Garrett, T. P., McKern, N. M., Lou, M., Elleman, T. C., Adams, T. E., Lovrecz, G. O., Zhu, H. J., Walker, F., Frenkel, M. J., Hoyne, P. A., Jorissen, R. N., Nice, E. C., Burgess, A. W., and Ward, C. W. (2002). Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor alpha. *Cell* **110**, 763–773.
- Jorissen, R. N., Walker, F., Pouliot, N., Garrett, T. P., Ward, C. W., and Burgess, A. W. (2003). Epidermal growth factor receptor: Mechanisms of activation and signalling. *Exp. Cell Res.* **284**, 31–53.
- Kumar, V., Bustin, S. A., and McKay, I. A. (1995). Transforming growth factor alpha. *Cell Biol. Intl.* **19**, 373–388.
- Lee, D. C., Fenton, S. E., Berkowitz, E. A., and Hissong, M. A. (1995). Transforming growth factor alpha: Expression, regulation, and biological activities. *Pharmacol. Rev.* **47**, 51–85.
- Schlessinger, J. (2002). Ligand-induced, receptor-mediated dimerization and activation of EGF receptor. *Cell* **110**, 669–672.
- Xian, C. J., and Zhou, X. F. (1999). Roles of transforming growth factor-alpha and related molecules in the nervous system. *Mol. Neurobiol.* **20**, 157–183.



TRH

see Thyrotropin-Releasing Hormone



**Figure 1** Molecular structure of trichlormethiazide. Note the common thiazide backbone of 1,2,4-benzothiadiazine-1,1-dioxide, with additional radicals, including a  $\text{CHCl}_2$  group at position R3 and a Cl at R6.

This increases the stability and duration of action to approximately 24 h, with a relative oral natriuretic maximal response in human of approximately 2.1-fold above chlorothiazide at a dose of 1–4 mg/day and an equieffective intravenous chloruretic response of 0.01 mg/kg in the dog.

### BIOAVAILABILITY, PHARMACODYNAMICS, AND DOSAGE

Whereas chlorothiazide is poorly absorbed in the gastrointestinal tract, the uptake of other thiazides is more efficient, with a diuretic response occurring within 1 h of administration. The plasma half-life of thiazides ranges from 6 to 44 h, with TCMZ displaying a duration of action of approximately 24 h. Uptake of thiazides into erythrocytes, where they likely bind to carbonic anhydrase, may slow plasma clearance. The recommended dosage of TCMZ is 1 to 4 mg/day, with initial doses given twice daily. The onset of diuretic action is usually seen at 2 h, with the maximum effect at 4 h. Antihypertensive effects are long-lasting, requiring as with most other thiazides only one dose per day. Such a response is probably related to natriuresis and not to diuretic potency, since equivalent doses of loop diuretics have weaker antihypertensive activity. The reason may be traced to the fact that thiazide-induced natriuresis is persistent for several hours, thus resulting in efficient control of  $\text{Na}^+$  balance in hypertensive subjects. In contrast, loop diuretics with a shorter duration of action promote greater natriuresis for 6 to 8 h (up to 20–25% of the filtered load of  $\text{Na}^+$ ); thereafter, they paradoxically enhance  $\text{Na}^+$  reabsorption by a mechanism involving volume depletion and stimulation of the renin–angiotensin–aldosterone system through the remainder of the day. This often

results in a net positive  $\text{Na}^+$  balance over time, if repeated doses are not given and the patient is not placed on a  $\text{Na}^+$ -restricted diet.

Another interesting feature of TCMZ and other thiazides is that since their site of action is distal to that of furosemide and other loop diuretics, chronic treatment does not induce the compensatory adjustments that tend to offset their efficacy by inducing hypertrophy of distal tubular  $\text{Na}^+$  transporters. This translates into distal reuptake of  $\text{Na}^+$  and  $\text{Cl}^-$  ions that have escaped reabsorption in the thick ascending loop of Henle. This is indeed the rationale for combination therapy employing a loop diuretic in conjunction with a thiazide or  $\text{K}^+$ -sparing agent to potentiate the diuretic response in massively edematous patients.

Dose reduction may be appropriate for TCMZ in chronic renal failure, although a substantial lack of activity is common for thiazides in such a setting.

### SIDE EFFECTS AND TOXICITY

TCMZ ( $\text{LD}_{50}$  orally in rats: >20,000 mg/kg) shares with other thiazides a number of minor side effects, seldom requiring withdrawal of therapy. These side effects include hypokalemia with metabolic alkalosis (linked to volume depletion), hyperuricemia (also resulting from volume contraction with enhanced urate reabsorption), and hypotension (particularly if the patient is aggressively diuresed through combination treatment with loop agents or is concurrently treated with other antihypertensives). Rarely, volume depletion is such that prerenal azotemia may occur, which is usually reversible on rehydration and/or volume expansion. Most of these effects are seen within the first few weeks of treatment, so that prolonged surveillance of serum  $\text{K}^+$  or urate levels is usually not needed if the patient has not experienced major side effects from the outset. Thiazides may also promote hypercalcemia, hypophosphatemia, hyponatremia, and hypomagnesemia, with the last side effect resulting in dizziness, lethargy, confusion, and muscle weakness. TCMZ may impair glycemic control in diabetics, through a host of mechanisms involving impaired glyconeogenesis, reduced glycogenolysis, and possibly diminished insulin responsiveness to an oral glucose load. Thiazides also increase cholesterol and triglyceride levels in plasma, possibly enhancing atherosclerotic risk.

### DRUG INTERACTIONS

Hypotension may occur when TCMZ is associated with angiotensin-converting enzyme (ACE) inhibitors. Beta-blockers may increase hyperglycemia in type II

diabetes mellitus. Bile acid-binding resins are known to impair gastrointestinal absorption of thiazides. Nonsteroidal anti-inflammatory drugs antagonize diuretic potency and the antihypertensive effects of thiazides, likely by blocking the natriuretic activity of endogenous prostanoids. As a matter of fact, this interaction is not unique to thiazides, but rather applies to most antihypertensive agents. Lithium toxicity may result from impaired lithium excretion during concurrent therapy. Cyclosporin toxicity and the risk of gout are increased by thiazides, so that concurrent therapy should possibly be avoided.

## CLINICAL USE

TCMZ [Achletin (Toyama); Anatron (Tobishi); Anistadin (Maruko); Aponorin (Kodama); Carvacron (Taiyo); Diurese (American Urologicals); Esmarin (Merck); Fluitran (Shionogi); Flutra (Schering); Intromene (Teikoku); Kubacron (Kayaku); Metahydrin (Merrell Dow); Naqua (Schering); Salurin (Syntex); Tachionin (Sana); Tolcasone (Rhone-Poulenc Rorer); Triflumen (Serono)] has been extensively used for the treatment of hypertension and edematous states. Its efficacy has been validated by a number of clinical trials, in which it was compared with other diuretics, calcium channel blockers, beta-blockers, ACE inhibitors, and other drugs. All studies confirm a mild but significant antihypertensive efficacy, although combination treatment is often necessary, with the exception of some subsets of hypertensives, such as African Americans or volume-expanded individuals, who show brisk responses to thiazide therapy. The latter concept has been highlighted by several trials that emphasized the cost-to-benefit ratio of thiazides as first-line agents in antihypertensive therapy compared to the much more expensive drugs that have been introduced. However, a major limitation to the use of TCMZ and other thiazides is renal insufficiency, in which reduced filtration and tubular delivery impair activity in NaCl transport, markedly diminishing their clinical effectiveness.

Another area of interest in the use of TCMZ and related compounds is the prevention of recurrent nephrolithiasis, with particular emphasis on calcium oxalate stones. The agent has been shown to markedly reduce calciuria in patients with primary or secondary hypercalciuria, thus effectively reducing the recurrence of stones. The mechanism of action involves the enhanced uptake of  $\text{Ca}^{2+}$  by distal tubular cells once they become hyperpolarized by intracellular  $\text{Cl}^-$  depletion.

An uncommon but useful indication of thiazide diuretics is symptomatic treatment of polyuria in

nephrogenic diabetes insipidus, either the X-linked or the autosomal-recessive form. Here, TCMZ and other thiazides have been shown to reduce the free water clearance by diminishing the volume delivered to the distal tubule. Natriuresis appears once again to be responsible, by reducing plasma volume and enhancing proximal tubular reabsorption of salt and water, thus decreasing the flow through the distal nephron.

## See Also the Following Articles

Bartter's Syndrome • Hypercalciuria

## Further Reading

- Ellison, D. H. (2000). Divalent cation transport by the distal nephron: Insights from Bartter's and Gitelman's syndromes. *Am. J. Physiol. Renal Physiol.* **279**, F616–F625.
- Gamba, G., Saltzberg, S. N., Lombardi, M., Miyanosita, A., Lytton, J., Hediger, M. A., Brenner, B. M., and Hebert, S. C. (1993). Primary structure and functional expression of a cDNA encoding the thiazide-sensitive, electroneutral sodium–chloride cotransporter. *Proc. Natl. Acad. Sci. USA* **90**, 2749–2753.
- Hatch, M., and Vaziri, N. D. (1994). Do thiazides reduce intestinal oxalate absorption? A study *in vitro* using rabbit colon. *Clin. Sci.* **86**, 353–357.
- Hebert, S. C., Mount, D. B., and Gamba, G. (2003). Molecular physiology of cation-coupled  $\text{Cl}^-$  cotransport: The SLC12 family. *Pflugers Arch.* May 9 (electronic publication).
- Honda, T., Hamada, M., Shigematsu, Y., Matsumoto, Y., Matsuoka, H., and Hiwada, K. (1999). Effect of antihypertensive therapy on aortic distensibility in patients with essential hypertension: Comparison with trichlormethiazide, nicardipine and alacepril. *Cardiovasc. Drugs Ther.* **13**, 339–346.
- Huonker, M., Sorichter, S., Schmidt-Trucksass, A., Irmer, M., Staiger, J., and Keul, J. (1999). Effectiveness of digitoxin versus trichlormethiazide/amiloride in congestive heart failure NYHA class II/III and sinus rhythm. *Cardiovasc. Drugs Ther.* **13**, 233–241.
- Insogna, K. L., Ellison, A. S., Burtis, W. J., Sartori, L., Lang, R. L., and Broadus, A. E. (1989). Trichlormethiazide and oral phosphate therapy in patients with absorptive hypercalciuria. *J. Urol.* **141**, 269–274.
- Jackson, E. K. (2001). Diuretics. In “Goodman and Gilman's The Pharmacological Basis of Therapeutics” (J. G. Hardman, L. E. Limbird, Goodman, and A. Gilman, eds.), 10th ed., pp. 757–788. McGraw-Hill, New York.
- Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. (1997). The Sixth Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. *Arch. Int. Med.* **157**, 2413–2446.
- Monroy, A., Plata, C., Hebert, S. C., and Gamba, G. (2000). Characterization of the thiazide-sensitive  $\text{Na}^+-\text{Cl}^-$  cotransporter: A new model for ions and diuretics interaction. *Am. J. Physiol. Renal Physiol.* **279**, F161–F169.
- National Intervention Cooperative Study in Elderly Hypertensives Study Group. (1999). Randomized double-blind comparison of a calcium antagonist and a diuretic in elderly hypertensives. *Hypertension* **34**, 1129–1133.
- Rose, B. D. (1991). Diuretics. *Kidney Int.* **39**, 336–352.

## THYROID FUNCTION IN THE FETUS AND INFANT

### Fetal Thyroid Physiology

During the first 8 to 16 weeks of gestation, the human fetal thyroid synthesizes only minute amounts of thyroxine ( $T_4$ ) and triiodothyronine ( $T_3$ ). There is limited transfer of maternal  $T_4$  and  $T_3$  to fetal blood, and serum total  $T_4$  and  $T_3$  levels in the early gestational fetus are low. Although the relative placental barrier to thyroid hormone transfer is maintained throughout pregnancy, the transplacental passage of some thyroid hormone may preserve fetal central nervous system development and function in the infant with thyroid agenesis. After mid-gestation, fetal hypothalamic expression of TSH-releasing hormone (TRH) and the production of TSH and  $T_4$  rise progressively to peak during the month prior to term. Total  $T_3$  levels in fetal blood remain low until 26 to 30 weeks of gestation, reflecting the immaturity of type I deiodinase activity (which converts  $T_4$  to  $T_3$ ) and the high levels of type III deiodinase activity (which converts  $T_4$  to the inactive reverse  $T_3$  and converts  $T_3$  to diiodothyronine) in fetal and placental extrathyroidal tissues. Type II deiodinase mediates the local production of  $T_3$  from  $T_4$  in selected fetal tissues such as the brain.

### Neonatal Thyroid Physiology

At delivery, the sudden cool temperature stimulates a dramatic rise in plasma TSH during the first hour of life. The levels of TSH rise to a peak of more than  $50 \mu\text{U/ml}$  at 30 min of age, stimulating a rise in serum total and free  $T_4$  ( $FT_4$ ) and  $T_3$  ( $FT_3$ ). In full-term infants with normal thyroid function, TSH levels are generally below  $20 \mu\text{U/ml}$  at 24 h of age and are 1.3 to  $16.0 \mu\text{U/ml}$  by 4 days of age. By 6 weeks of age, TSH concentrations approximate those of adults. In full-term infants, serum  $T_4$  and  $FT_4$  concentrations fall gradually during the subsequent 4 to 6 weeks, and serum  $T_3$  levels rise gradually after the neonatal period to achieve mature levels by 2 to 12 weeks of age. By 6 months of age, the levels of  $T_4$  and  $FT_4$  remain slightly higher than those of older children and adults, and the circadian pattern of TSH secretion begins to appear. These observations suggest that hypothalamic-pituitary regulation of TSH secretion matures gradually until 6 weeks to 6 months of postnatal age.

### Thyroid Function after Premature Birth

In preterm infants, the neonatal rise of TSH and the rise in  $T_4$ ,  $FT_4$ , and  $FT_3$  are blunted. By 4 days of age, TSH levels have declined to 0.8 to  $6.9 \mu\text{U/ml}$ .  $T_4$  and  $FT_4$  levels in preterm infants exceeding 30 to 32 weeks gestation increase progressively over the ensuing 4 to 8 weeks to values comparable to those of term infants. In contrast, levels of  $T_4$  and  $FT_4$  in very low-birthweight infants (less than 30 weeks or less than 1200–1500 g) decline progressively after day 1, reaching a nadir at 1 to 2 weeks of age. TSH is low at 5 to 7 days of age ( $0.8 \mu\text{U/ml}$ ) and increases to  $6.1 \mu\text{U/ml}$  at 5 weeks of age, paralleled by an increase in  $FT_4$  and  $FT_3$ . This relative hypothyroxinemia is most profound in those infants born most prematurely. In the majority of cases, serum TSH concentrations are not elevated.

## TSH STRUCTURE AND GLYCOSYLATION

TRH has been shown to be necessary for TSH synthesis, posttranslational glycosylation, and secretion of a fully bioactive TSH molecule from the pituitary. TSH is composed of an  $\alpha$ -subunit (identical to that of luteinizing hormone [LH], follicle-stimulating hormone [FSH], and human chorionic gonadotropin [hCG]) and a unique  $\beta$ -subunit. Both of these subunits undergo posttranslational glycosylation in response to TRH. The glycosylation patterns influence the bioactivity of the TSH molecule. There are two oligosaccharide chains on the  $\alpha$ -subunit and one on the  $\beta$ -subunit. Highly sialated chains (as in recombinant TSH) have decreased bioactivity, decreased hepatic clearance, and longer half-life. Altered TSH glycosylation resulting in altered bioactivity is seen in the following:

1. Mixed hypothyroidism (central hypothyroidism with elevated TSH): oligosaccharide chains having increased mannose with decreased TSH bioactivity
2. Primary hypothyroidism: increased sialation, decreased TSH bioactivity, and elevated TSH levels
3. Resistance to thyroid hormone: increased galactose N-acetylglucosamine and increased TSH bioactivity
4. TSH-secreting pituitary adenomas: variable biological activity

5. Nonthyroidal illness: transient central hypothyroidism due to decreased TRH secretion, decreased TSH surge, and decreased TSH bioactivity

The TSH  $\beta$ -subunit gene was cloned in 1988, allowing research leading to production of recombinant TSH. Key domains of the  $\beta$ -subunit allowing ligand receptor interaction were identified, potentially permitting engineering of TSH molecules with increased bioactivity. Glycosylation patterns affect bioactivity; the endogenous TSH  $\beta$ -subunit has oligosaccharides with sialic acid and galactose, or sulfate and N-acetylgalactosamine, whereas recombinant TSH has sialated oligosaccharides only. For TSH to act, both the  $\alpha$ - and  $\beta$ -subunits interact with the G protein-coupled receptor.

## REGULATION OF TSH SYNTHESIS AND RELEASE

### TSH-Releasing Hormone

Like GH, TSH is synthesized in anterior pituitary cells in response to a releasing hormone. TRH is secreted from the hypothalamus and transported to the pituitary via the venous portal system. Dopamine and somatostatin (both of which inhibit TSH release) are also transported from the hypothalamus. The TRH (G protein-coupled) receptor has been identified on pituitary cells and also in the central and peripheral nervous systems.

### Circadian Variation in TSH Concentration

In persons of all ages after infancy (6 months), TSH secretion normally occurs in a circadian pattern with lower concentrations in the afternoon, a nocturnal surge beginning after 1900 h, and higher concentrations between 2200 and 0400 h. Peak TSH levels between 2200 and 0400 h are followed by a gradual decline between 0400 and 1000 h to nadir TSH values. Small frequent pulses occur as an overlay on this overall "sin wave"-shaped circadian pattern. Thus, at least one-third of the trophic influence of TSH on the thyroid gland occurs during the hours of sleep.

Despite the clear description of the daily pattern of TSH variation in the literature, the neuroendocrine influences that result in this circadian pattern have not yet been fully characterized. Factors regulating the timing and amplitude of the TSH surge may include an endogenous surge in hypothalamic TRH secretion at night, onset of sleep, and/or altered pituitary sensitivity to endogenous TRH. Substances that appear

to influence pituitary sensitivity to TRH include  $\alpha$ -adrenergic stimuli, somatostatin, endogenous opiates, dopamine, cortisol, leptin, and GH concentrations.

## TSH FUNCTION AT THE RECEPTOR

### TSH Receptor Binding

TSH is secreted from the pituitary into the blood, stimulating synthesis and release of  $T_4$  and  $T_3$  from the thyroid gland as well as influencing uptake of iodine. Both subunits of TSH interact with the G protein-coupled cell membrane receptor to generate cyclic AMP (cAMP) within the cell. The TSH receptor has an extracellular glycosylated portion and an intracellular portion with a cytoplasmic "tail" involved in G protein coupling and signal transduction. Both portions undergo posttranslational processing, influencing receptor-binding affinity of TSH. Mature complex glycosylation of the extracellular portion is required for TSH binding; if there are only mannose-type sugars, TSH binds with only low affinity.

### Results of TSH Receptor Binding

TSH receptor binding leads to increased synthesis of iodothyronines and to their storage in thyroid colloid in association with thyroglobulin. TSH receptor binding also leads to increased release of fully formed  $T_4$  and small amounts of  $T_3$  from the colloid. Plasma  $T_4$  and  $T_3$  circulate in the serum bound to thyroid-binding globulin and albumin, with only small amounts that are "free" or unbound.  $FT_4$  undergoes intracellular deiodination to  $FT_3$ , which interacts with DNA to influence cellular mRNA and protein synthesis. Both  $T_4$  and  $T_3$  provide negative feedback at the hypothalamus and pituitary to modulate TRH and TSH secretion.

### TSH Receptor Mutations

A TSH receptor mutation database has been established to allow easy access to the many identified TSH receptor mutations and their clinical consequences. At least 12 loss-of-function mutations have been found in patients with partial resistance to TSH or in rare cases of congenital hypothyroidism. A truncated TSH receptor with complete inactivation has been identified in infants with congenital hypothyroidism from an inbred kindred. Their thyroid glands were hypoplastic but not ectopic. Activating mutations of the TSH receptor in focal tissue may result in hyperfunctioning thyroid nodules or multinodular goiter. Expression of

the mutation in all thyroid tissue may result in thyrotoxicosis presenting during the neonatal period (and persisting beyond 6 months of age), late infancy, childhood, or adulthood.

## ELEVATED TSH SECRETION

### Primary Hypothyroidism

#### **Congenital Hypothyroidism**

Primary hypothyroidism (reduced function of the thyroid gland itself with low FT<sub>4</sub> and high TSH) is the most common form of hypothyroidism both in the general population and in cancer survivors. There is now universal screening for congenital primary hypothyroidism in the United States. American newborn screening programs measure either TSH or total T<sub>4</sub> by radioimmunoassay in blood spot samples obtained after 24 h of age. Hypothyroxinemia is defined as a total T<sub>4</sub> value below 90% of the samples screened on that day. Measurement of TSH in hypothyroxinemic infants identifies those with primary hypothyroidism. TSH values often exceed 50 to 100 μU/ml in infants with thyroid agenesis and other forms of permanent primary hypothyroidism. TSH levels exceeding 16 to 20 μU/ml between 24 and 96 h of age, or 5.0 to 7.7 μU/ml between 1 and 12 months of age, may reflect primary thyroid dysfunction.

#### **Hypothyroidism in Preterm Infants**

It should be recognized that very low-birthweight infants are at an eightfold increased risk for the development of transient primary hypothyroidism with low T<sub>4</sub> levels and marked elevations in TSH. Transient primary hypothyroidism in preterm infants may, in some cases, be induced by transdermal absorption of iodine antiseptics. On the other hand, in iodine-deficient geographic regions, transient hypothyroidism is more likely related to insufficient iodine intake. Transient primary hypothyroidism may be accompanied by a delayed rise in TSH that is not detected until several weeks after birth; thus, a number of investigators now recommend rescreening very low-birthweight infants (<1500 g or <30 weeks gestation) at 2 to 6 weeks of age. The prevalence of permanent congenital primary hypothyroidism in preterm infants is comparable to that in term infants.

#### **Acquired Primary Hypothyroidism**

Primary hypothyroidism may be acquired after infancy and may represent a partially functioning ectopic or hypoplastic thyroid gland or be associated with thyroid peroxidase antibodies (as in Hashimoto's

thyroiditis). In acquired primary hypothyroidism, TSH may be as elevated as 1800 μU/ml. The frequency of primary hypothyroidism is increased in chromosome abnormality syndromes such as Turner, Klinefelter, and Down syndromes. Most patients with mildly elevated TSH concentrations have been considered to have mild or "compensated" hypothyroidism. However, mild primary hypothyroidism has potential clinical importance given that children with mild TSH elevation and T<sub>4</sub> levels within normal limits typically grow less well than do other children. In addition to slow growth, mild hypothyroidism causes fatigue, dry skin, constipation, increased sleep requirement, and cold intolerance. Thyroid replacement therapy in mild hypothyroidism improves growth velocity, lipid profile, and energy level.

#### **Primary Hypothyroidism after Cancer**

Primary hypothyroidism (with TSH elevation) is most likely to occur in patients who have received mantle irradiation for Hodgkin's disease, craniospinal radiation for medulloblastoma, or total body irradiation prior to bone marrow transplantation. Primary hypothyroidism is rarely isolated in cancer survivors, who may also have GH deficiency, adrenocorticotrophic hormone (ACTH) deficiency, and/or pubertal disorders.

### Thyroid Hormone Resistance and TSH-Secreting Adenoma

In thyroid hormone resistance and TSH-secreting adenoma, TSH is measurable (inappropriately) in the presence of elevated T<sub>4</sub> and elevated FT<sub>4</sub>. In resistance, this biochemical picture is associated with few or no symptoms of hyperthyroidism. Patients may be euthyroid or even hypothyroid. In contrast, in TSH-secreting adenomas, patients are clinically hyperthyroid. Pituitary TSH-secreting adenomas are uncommon in general and are particularly rare in the pediatric age group.

## DEPRESSED TSH SECRETION

### Hyperthyroidism

Low serum TSH concentration is an appropriate response to excessive thyroid hormone levels resulting from either endogenous or exogenous source. Graves' hyperthyroidism is a result of antibodies to the TSH receptor binding in a manner that activates the intracellular G protein signaling pathway as if TSH were binding at the receptor. Hyperthyroidism in

McCune–Albright syndrome is an example of a chronically activated G protein pathway.

## Central Hypothyroidism

### **Reduced TSH Surge**

A blunted or absent nocturnal TSH surge is a characteristic of central hypothyroidism, suggesting a loss of the normal circadian variation in TRH (and thus a reduction in total TSH release) and loss of one-third to one-half of the daily trophic stimulus to the thyroid gland. In central hypothyroidism, FT<sub>4</sub> is low or in the lowest third of the normal range, with a normal daytime TSH. Measurement of the nocturnal TSH surge has improved sensitivity for detection of central hypothyroidism compared with the TRH test.

### **Low-Normal T<sub>4</sub> and FT<sub>4</sub>**

Many patients with central hypothyroidism maintain normal daytime TSH concentrations and iodothyronine concentrations that are just below the normal range or in the lowest third of the normal range. These patients remain clinically puzzling and, therefore, are often not treated with thyroid replacement medication. Failure to recognize and treat central hypothyroidism can result in poor growth in children and in a less than optimal state of health in children and adults. Subtle hypothyroidism may be associated with depression and mild hyperlipidemia.

### **A Cause of Short Stature**

In children referred to endocrinology for evaluation of apparent idiopathic short stature, approximately 13% had isolated central hypothyroidism (a blunted TSH surge associated with low or low-normal FT<sub>4</sub> in the absence of any other pituitary hormone disturbance). The incidence was 33% of children with height shorter than  $-2$  SD who had an FT<sub>4</sub> in the lowest third of the normal range. Growth velocity significantly improved in children during treatment with levothyroxine compared with that in children who had a normal TSH surge and who were otherwise clinically similar. Children with known hypopituitarism also show blunting of the nocturnal TSH surge. In these conditions, delayed TSH response to TRH is not always observed.

### **Thyroid Function after Head Injury**

Disturbances in thyroid function are commonly observed in head-injured patients, including low T<sub>3</sub> and low T<sub>4</sub>, usually with a normal level of TSH. Reverse T<sub>3</sub> is often elevated. Thyroid axis injury may still be

quite difficult to identify in newly injured patients because of the prolonged half-life of T<sub>4</sub> (7 days).

### **Central Hypothyroidism after Cancer**

TRH testing in children after radiotherapy suggests that hypopituitarism is more common than was suspected previously. Central hypothyroidism is recognized in as many as 65% of patients after brain tumor or nasopharyngeal tumor, in more than 35% after bone marrow transplant, and in as many as 15% after leukemia, suggesting that central hypothyroidism (hypothalamic–pituitary–thyroid dysregulation) may be quite common in cancer survivors. In contrast, primary hypothyroidism occurs most commonly after total body radiation or radiation to the nasopharynx, neck, or spine. Central and mixed hypothyroidism both occur in patients after radiation to the head and include brain tumors, nasopharyngeal tumors, and total body radiation. Chemotherapy alone can cause primary or central hypothyroidism, but the frequency of this occurrence is much less than that after radiotherapy. If TSH secretion is not tested until GH deficiency becomes apparent, the diagnosis of hypothyroidism may be delayed in approximately one-third of patients.

### **Central Hypothyroidism in Adults**

Central hypothyroidism can be quite difficult to recognize in adults where slowed growth rate is not available as a sign. Symptoms of central hypothyroidism (e.g., asthenia, edema, drowsiness, adynamia, skin dryness) may be of gradual onset and be unrecognized until therapy is begun and the patient feels better.

## Mixed Hypothyroidism

Mixed hypothyroidism consists of central hypothyroidism with TSH elevation. It has been described in survivors of childhood cancer and in women with Sheehan syndrome (postpartum pituitary necrosis). With reduced TRH from the hypothalamus, TSH may be abnormally glycosylated and of lower biological activity. Thus, mild elevation of serum TSH concentrations (5–14 mU/L) may be seen in central hypothyroidism; the differentiation from primary hypothyroidism can be made by documentation of a blunted or absent TSH surge. In Sheehan syndrome patients, elevated 24-h TSH was observed with blunting of the TSH surge, suggesting dysregulated release of biologically inactive TSH.

Mixed hypothyroidism may reflect either separate injuries to the thyroid gland and hypothalamus (e.g.,

radiation injury of both structures) or central hypothyroidism in which the biological activity of the secreted TSH is reduced. Using baseline FT<sub>4</sub> and TSH alone, 92% of those with central hypothyroidism and 27% of those with mixed hypothyroidism may remain undiagnosed.

## DIAGNOSIS OF HYPOTHYROIDISM

### Individual Set Point for TSH and FT<sub>4</sub>

FT<sub>4</sub> is currently the best measure of thyroid status along with serum TSH. FT<sub>4</sub> remains fairly stable in an individual over the years at an optimal “set point” for thyroid function. If FT<sub>4</sub> production from an injured thyroid gland declines (as in mild primary hypothyroidism), the intact pituitary secretes more TSH. Thus, in primary hypothyroidism, a serum sample for TSH at 0800 h is the best test. In contrast, in central hypothyroidism, hypothalamic–pituitary release of TSH is itself impaired, and as a result, FT<sub>4</sub> declines; the patient is just as hypothyroid, but the TSH concentration is not usually elevated.

### Free Thyroxine

FT<sub>4</sub> below the normal range without TSH elevation is clearly suggestive of central hypothyroidism. However, FT<sub>4</sub> concentration may, alternatively, be in the lowest third of the normal range. FT<sub>4</sub> and T<sub>4</sub> values are similar to those seen in mild primary hypothyroidism. The “red flag” of an elevated TSH is absent because the ability to increase TSH secretion is impaired; however, daytime serum TSH is not usually low. A blunted or absent nocturnal TSH surge in central hypothyroidism suggests loss of the normal circadian variation in TRH release. Central hypothyroidism can be difficult to diagnose because of this subtle clinical and laboratory presentation. Decline in FT<sub>4</sub> over time (years) after injury or radiation therapy can be the most important observation; however, baseline (prior to radiation therapy) measurements of FT<sub>4</sub> are not always available for comparison. It is important to maintain a high index of suspicion and to recognize that FT<sub>4</sub> levels in the lowest third of the normal range are compatible with thyroid dysfunction.

### Monitoring

Yearly measurements of TSH and FT<sub>4</sub> and growth surveillance are recommended in children at risk for hypothyroidism such as girls with Turner syndrome, children with Down syndrome, and childhood cancer

survivors. Earlier diagnosis of mild hypothyroidism will allow earlier intervention to improve growth velocity and quality of life. In a healthy person with slow growth or other symptoms of hypothyroidism, criteria for starting thyroid hormone therapy without further diagnostic testing include (1) TSH above 4 mU/L at 0800 or 0900 h or above 3 mU/L between 1000 and 2000 h (regardless of the FT<sub>4</sub> value) and (2) FT<sub>4</sub> at or below the lower limits of the normal range for the assay (regardless of the TSH value). If FT<sub>4</sub> is in the upper two-thirds of the normal range and TSH does not meet the preceding criteria, no thyroid therapy is needed, and thyroid status and growth (in children) should be reviewed in 1 year in patients at risk. If FT<sub>4</sub> is in the lowest one-third of the normal range without TSH elevation, both the TRH test and the TSH surge should be performed. The combination of history of head injury or cranial or total body radiation, slow growth rate, normal weight gain, no intercurrent illness, delayed bone maturation, and declining FT<sub>4</sub> or FT<sub>4</sub> below the lower limits of normal or TSH above the upper limits of normal should be diagnostic of hypothyroidism.

## THERAPY FOR HYPOTHYROIDISM

### Starting Thyroid Hormone Therapy

Thyroid hormone doses at initiation of therapy are much higher in infants (10–12 μg/kg BW/day) and toddlers (5–8 μg/kg BW/day) than in children (3 μg/kg BW/day) and adolescents or adults (1.25–1.50 μg/kg BW/day). These doses approximate 100 μg per meter squared per day.

Typical initial thyroid replacement dose in healthy young people with TSH of less than 30 mU/L and without risk of cardiac decompensation is 3 μg levothyroxine per kilogram body weight given daily in the morning. If TSH is more elevated or if there are concerns about medical stability, starting with one-quarter of this dose and increasing by a quarter dose each month can permit more gradual physiological and psychological adjustment to the new metabolic state. In general, because levothyroxine has a long half-life of 5 to 6 days, it is useful to measure thyroid levels only after 4 weeks.

### Thyroid Dose Adjustment

Thyroid dose should then be adjusted as follows. In primary hypothyroidism (initial TSH elevation above normal with no evidence of central hypothyroidism), TSH is the most useful test to monitor during

therapy. The target should be a TSH of 0.5 to 1.5 mU/L (or  $\mu$ U/ml), without symptoms of overreplacement. In central or mixed hypothyroidism, therapy with levothyroxine rapidly suppresses TSH before resolution of clinical symptoms, so TSH is not useful to measure unless the initial TSH was elevated. The recommendation is to monitor FT<sub>4</sub> during therapy, adjusting the dose to achieve FT<sub>4</sub> just above the middle of the normal range and without symptoms of hypo- or hyperthyroidism.

### Thyroid Replacement and Adrenal Function

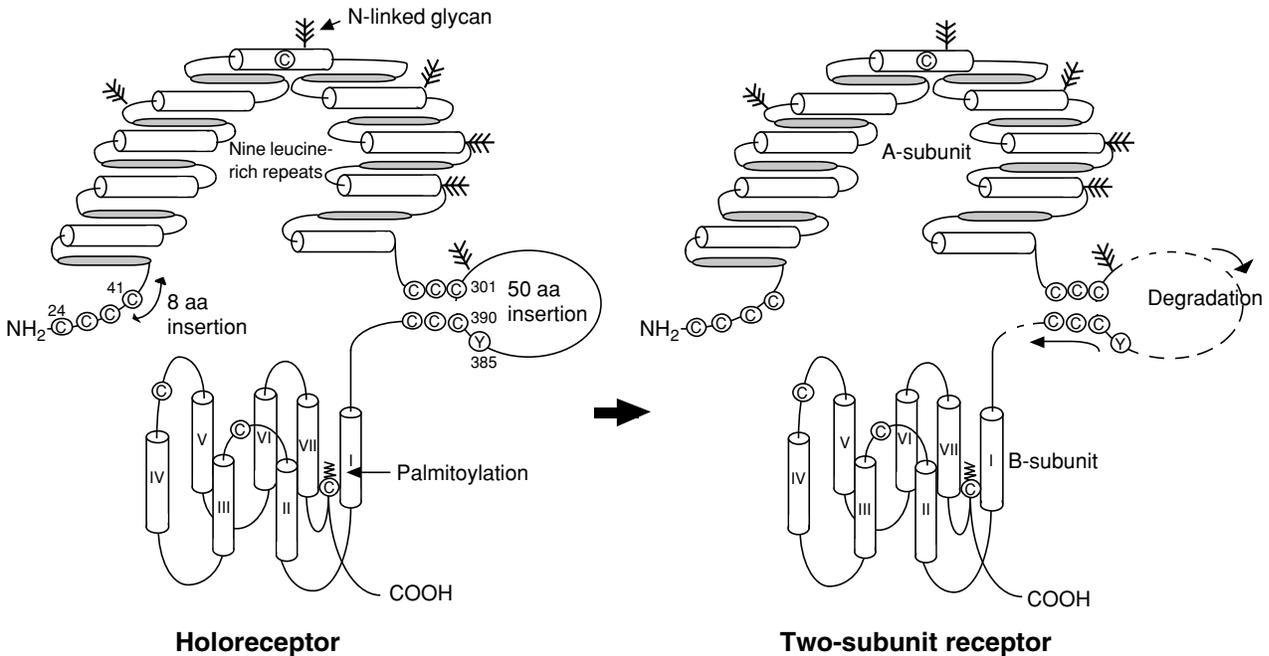
Because thyroid therapy can result in improved metabolic clearance of many substances such as cortisol, thyroid hormone replacement can result in clinical decompensation of patients with unrecognized adrenal insufficiency. In patients at risk, it is necessary to evaluate for primary adrenal insufficiency and/or hypothalamic ACTH deficiency and to provide such patients with hydrocortisone prior to initiating thyroid therapy.

### See Also the Following Articles

Hypothalamus–Pituitary–Thyroid Axis • Pituitary Adenomas, TSH-Secreting • Pituitary Gland Anatomy and Embryology • Thyrotropin-Releasing Hormone (TRH) • TSH Receptor (Thyrotropin Receptor)

### Further Reading

- Cohen, L. E. (2000). Genetic regulation of the embryology of the pituitary gland and somatotrophs. *Endocrine* **12**, 99–106.
- Haddow, J. E., Palomaki, G. E., Allan, W. C., *et al.* (1999). Maternal thyroid deficiency during pregnancy and subsequent neuropsychological development of the child. *N. Engl. J. Med.* **341**, 549–555.
- Parks, J. S., Brown, M. R., Hurley, D. L., *et al.* (1999). Heritable disorders of pituitary development. *J. Clin. Endocrinol. Metab.* **84**, 4362–4370.
- Persani, L. (1998). Hypothalamic thyrotropin-releasing hormone and thyrotropin biologic activity. *Thyroid* **8**, 941–946.
- Pitukcheewanont, P., and Rose, S. R. (1997). Nocturnal TSH surge: A sensitive diagnostic test for central hypothyroidism in children. *The Endocrinologist* **7**, 226–232.
- Rapaport, R., Rose, S. R., and Freemark, M. (2001). Hypothyroxinemic in the preterm infant: The benefits and risks of thyroxine treatment. *J. Pediatr.* **139**, 182–188.
- Rose, S. R. (2000). Disorders of thyrotropin synthesis, secretion, and function. *Curr. Opin. Pediatr.* **12**, 375–381.
- Rose, S. R. (2001). Cranial irradiation and central hypothyroidism. *Trends Endocrinol. Metab.* **12**, 97–104.
- Scully, K. M., and Rosenfeld, M. G. (2002). Pituitary development: Regulatory codes in mammalian organogenesis. *Science* **295**, 2231–2235.
- Sklar, C. A. (1997). Growth and neuroendocrine dysfunction following therapy for childhood cancer. *Pediatr. Clinics North America* **44**, 489–503.
- Trulzsch, B., Nebel, T., and Paschke, R. (1999). The thyrotropin receptor mutation database. *Thyroid* **9**, 521–522.
- Weintraub, B. D., and Szkudlinski, M. W. (1999). Development and *in vitro* characterization of human recombinant thyrotropin. *Thyroid* **9**, 447–450.
- Yeung, S.-C. J., Chiu, A. C., Vassilopoulou-Sellin, R., and Gagel, R. F. (1998). The endocrine effects of non-hormonal antineoplastic therapy. *Endocrine Rev.* **19**, 144–172.



**Figure 1** Schematic representation of the TSHR. The holoreceptor (left) and the two-subunit receptor (right) are shown.

of 8 and 50 aa unique to TSHR as compared with LHR and FSHR are included in this domain (aa 38–45 and 317–367). The second domain is the transmembrane/cytoplasmic region, a characteristic of the G protein-coupled receptor superfamily.

TSHR undergoes a variety of posttranslational modifications, including glycosylation, intramolecular cleavage, disulfide formation, tyrosine sulfation, and palmitoylation.

There are six potential N-linked glycosylation sites (N<sub>x</sub>S/T) on TSHR, all of which appear to be actually glycosylated. The TSH holoreceptor can be detected by Western blotting as a doublet of ~100 and 120 kDa. The ~100 kDa protein is a precursor with mannose-type oligosaccharides located in the endoplasmic reticulum, which gradually is transformed into the ~120 kDa mature protein with complex oligosaccharides in the Golgi apparatus and then is expressed on the cell surface. Thus, TSHR is heavily glycosylated and contains an ≈25 kDa glycan, which corresponds to 40% of the molecular mass of the A subunit (see below).

TSHR undergoes intramolecular cleavage into two subunits: the TSH-binding A-subunit and the transmembrane B-subunit. This cleavage apparently occurs at multiple sites on the cell surface of the mature receptor, presumably by a membrane-associated matrix metalloproteinase, resulting in the removal of a peptide corresponding to an approximately 50 aa insertion

(C peptide). These subunits are linked by disulfide bridges that are reduced by a cell surface enzyme, possibly protein disulfide isomerase, to release the A-subunit, a phenomenon called “receptor shedding.” Shedding may also involve the removal of the cysteine residues at the N terminus of the B-subunit by cleavage. Cleavage does not affect the receptor function in terms of high TSH-binding and TSH-mediated signal transduction. However, it is of interest that TSAb appear to preferentially recognize the shed A-subunit.

There are 11 cysteines in the ectodomain, clustering in three distinct regions: the N and C termini of the A-subunit and the N terminus of the B-subunit. Two additional cysteines are also present in the extracellular loops. Although it has not been confirmed which cysteines form disulfide bridges with each other, it is likely that cysteines in the second and third clusters link the A- and B-subunits.

Other posttranslational modifications include the following: (1) sulfation of the tyrosine at aa 385 in the ectodomain, which is required for high-affinity binding for TSH, but not TSAb; (2) palmitoylation of the cysteine at aa 699 in the C-terminal cytoplasmic tail of the receptor, which positively controls the rate of intracellular trafficking of the receptor; and (3) the formation of oligomers, which rapidly dissociate into monomers on TSH binding.

TSH and presumably also autoantibodies bind to the TSHR ectodomain with high affinity. The binding site(s) appears to be highly conformational and consists of multiple, discontinuous amino acid sequences, which span the entire ectodomain. The binding sites for TSH and the autoantibodies likely overlap but may not be identical. For example, the midregion of the ectodomain (particularly aa 201–211 and 222–230) is a TSH-binding site, whereas the N and C termini of the ectodomain seem to be critical for TSAb and TBAb binding, respectively. As mentioned above, the TSHR A-subunit rather than the holoreceptor may be the primary autoantigen in Graves' disease.

Because of its highly complex tertiary structure and because of a variety of posttranslational modifications, the functional receptor can be expressed only in mammalian cells. In addition to the full-length receptor, the entire ectodomain linked to the single transmembrane region of CD8 or a glycosylphosphatidylinositol anchor and the soluble, truncated receptor, approximately corresponding to the A-subunit, but not the whole ectodomain itself, have also been functionally expressed.

Neither the three-dimensional structure of TSHR nor the mechanisms for transmembrane signaling are known. However, the LRR region of the TSHR ectodomain is modeled based on the known structure of LRRs in ribonuclease inhibitor (RI). This region appears to have a nonglobular, horseshoe shape whose concave surface likely interacts with TSH.

## TSHR GENE

The TSHR gene consists of 10 exons and 9 introns and spans > 60 kb on human chromosome 14q. Exons 1–9 encode most of the ectodomain, with each exon corresponding to a LRR, and exon 10 encodes part of the ectodomain and the transmembrane/cytoplasmic region. These findings indicate that the TSHR gene is derived from the integration of multiple genes, each encoding a LRR into a prototypic intronless G protein-coupled receptor gene. The full-length TSHR cDNA is approximately 4 kb long and contains a single open reading frame of 2292 bp.

Multiple TSHR mRNAs are detected in the thyroid tissue; there are major transcripts of ~4 kb and several other minor transcripts. The former correspond to the full-length receptors with distinct 3'-untranslated regions and the latter correspond to alternatively spliced, truncated receptors lacking the transmembrane/cytoplasmic region.

The promoter region of the TSHR has characteristics typical of those of housekeeping genes with multiple transcription start sites. TSH induces a transient increase and subsequent down-regulation of TSHR expression under *in vitro* culture conditions, whereas TSHR mRNA levels are relatively stable *in vivo*. The minimal sequence required for thyroid-specific expression and negative regulation by TSH is located at –195 to –39 bp of the 5'-flanking region.

TSHR is primarily expressed in the basolateral membrane of thyroid follicular epithelial cells, but extrathyroidal expression including fat tissues and retro-orbital fibroblasts is noted, which may be relevant for extrathyroidal manifestations of Graves' disease.

## AUTOIMMUNITY

As mentioned above, the TSHR autoantibodies act as agonists in most patients with Graves' disease (TSAb) or as antagonists in some patients with autoimmune hypothyroidism (TBAb). To detect these antibodies, two assays have become available. One is a bioassay for TSAb that measures agonist-mediated cAMP production using thyroid cells of various species or mammalian cells stably expressing recombinant TSHR. The other is competition of antibodies for radiolabeled TSH binding to TSHR [TSH-binding inhibiting antibody (TBIAb)]. Although the TBIAb assay has long used porcine thyroid membrane as an antigen, a highly sensitive TBIAb assay with recombinant human TSHR has become commercially available. However, this assay cannot discriminate between stimulating and blocking antibodies.

New methods to detect direct binding of TSHR autoantibodies to TSHR have also been developed with recombinant receptors. These assays detect any type of antibody, not only TSAb and TBAb, but also neutral antibodies with no biological activities. The results obtained with these direct binding assays are closely correlated with TBIAb values.

TSHR autoantibodies can be experimentally elicited by traditional immunization approaches using the soluble TSHR protein with classical adjuvants, which, however, do not usually have TSAb activities. In contrast, immunization of mice with syngeneic cells coexpressing TSHR and major histocompatibility complex class II antigen or genetic immunization with plasmid- or adenovirus-encoded TSHR can effectively induce TSAb and Graves'-like hyperthyroidism. Since Graves' disease is an antibody-mediated disease, it has been widely believed that the T helper 2 cell (Th2)-based immune response

is predominant in Graves' disease. However, studies with these animal models challenge this concept, suggesting that the Th1 immune response may be more critical than previously anticipated or that the Th1/Th2 paradigm may be too simplistic to explain the pathogenesis of Graves' disease.

### NATURALLY OCCURRING MUTATIONS

Numerous mutations in the TSHR gene causing loss- or gain-of-function have been identified. Constitutively activating mutations were found in hyperfunctioning thyroid adenomas and autosomal-dominant congenital nonautoimmune hyperthyroidism. Most of the mutations are point mutations and are localized in transmembrane/cytoplasmic regions. TSHR is "noisy," that is, it displays significant constitutive activity even in the absence of agonist. Unliganded ectodomain likely constrains the receptor activity. In contrast, loss-of-function mutations in congenital hypothyroidism associated with resistance to TSH can be found in any region of the receptor. Furthermore,

a unique mutation (K183R), which increases the sensitivity to human choriogonadotropin and causes a type of familial gestational hyperthyroidism, has also been identified.

### See Also the Following Articles

Antithyroid Drugs • G Protein-Coupled Receptors • Graves' Disease • Pituitary Tumors, Molecular Pathogenesis • Thyroid Autoimmunity • TSH Function and Secretion

### Further Reading

- Nagayama, Y. (2002). The thyroid stimulating hormone receptor. *In* "Thyroid Eye Disease: Diagnosis and Treatment" (J. J. Dutton and B. G. Haik, eds.), pp. 19–27. Decker, New York.
- Rapoport, B., Chazenbalk, G. D., Jaume, J. C., and McLachlan, S. M. (1998). The thyrotropin receptor: Interaction with thyrotropin and autoantibodies. *Endocr. Rev.* **19**, 673–716.
- Rapoport, B., and McLachlan, S. M. (2001). Thyroid autoimmunity. *J. Clin. Invest.* **108**, 1253–1259.
- Szkudlinski, M. W., Fremont, V., Ronin, C., and Weintraub, B. D. (2002). Thyroid-stimulating hormone and thyroid-stimulating hormone receptor structure–function relationship. *Physiol. Rev.* **82**, 473–502.

peptides that include TRH and flanking and intervening sequences. These peptides themselves have important intracellular or extracellular actions, notably prepro-TRH, which can stimulate TSH gene expression. In fact, the potentiating effects of prepro-TRH on TRH-induced TSH secretion may be mediated by folliculostellate cells in the anterior pituitary.

The hypothalamus exerts a dominant stimulatory role on TSH secretion, and this effect is mediated by TRH. Evidence for this comes from a number of observations. TRH exerts a dose-dependent increase on TSH secretion *in vivo* and *in vitro*, and TSH levels decrease following lesions to the hypothalamus or after hypothalamic-pituitary dissociation. The presence of TRH in hypophyseal portal blood in physiologically relevant concentrations and the occurrence of hypothyroidism in animals treated with a neutralizing antibody to TRH add further substance to the importance of TRH in the regulation of TSH secretion and the maintenance of normal thyroid function. In humans, TSH levels increase within 2 to 5 min of intravenous TRH administration, peak at 20 to 30 min, and gradually return to normal by 2 to 3 h. This increase is paralleled later by a rise in serum T<sub>3</sub> and T<sub>4</sub> concentrations, which peak at 3 and 8 h, respectively. However, TRH does not exert its effects simply by stimulating TSH release; rather, it also stimulates TSH synthesis through the induction of TSH subunit gene transcription and translation. This action involves calcium influx and activation of protein kinase C and is modulated by both cyclic AMP (cAMP) and the pituitary-specific transcription factor Pit-1, which also regulates the expression of the prolactin (PRL) and growth hormone (GH) genes and mediates the action of TRH on PRL gene expression in cultured GH<sub>3</sub> cells. As a result, mutations of this factor result in congenital TSH deficiency as well as deficient lactotroph and somatotroph function.

TRH regulates TSH by acting at a posttranslational level as well as at a transcriptional level and by modifying the processing of the oligosaccharide groups of TSH. TSH needs to be fully glycosylated for maximal biological activity, and this observation explains the clinical finding that some patients with central hypothyroidism and slightly elevated basal TSH concentrations secrete TSH with reduced biological activity that increases after TRH administration.

### Interactions between TRH and Thyroid Hormones

The stimulatory effect of TRH on the synthesis and secretion of TSH is counteracted by the direct

pituitary inhibition of TSH synthesis and release by thyroid hormones, a process involving the intrapituitary conversion of T<sub>4</sub> to T<sub>3</sub>. However, thyroid hormones can also modulate the expression of the TRH and TRH receptor genes. For example, the number of TRH receptors on thyrotrophs increases in hypothyroidism and can be reduced by thyroid hormone replacement. In contrast, in pituitary tumor GH<sub>4</sub>C<sub>1</sub> cells, TRH reduces T<sub>3</sub> receptor number and T<sub>3</sub> responsiveness, and this may represent a further site of feedback interaction between T<sub>3</sub> and TRH at the level of the pituitary. The levels of TRH mRNA in hypothalamic paraventricular nuclei also increase in hypothyroidism and are reduced by thyroid hormone treatment. Furthermore, rats that have lesions of their hypothalamic paraventricular nuclei do not have the normal rise of TSH and TSH subunit mRNA after induction of primary hypothyroidism. These observations indicate that the paraventricular nuclei are another target for the action of thyroid hormones in the regulation of TRH gene expression and secretion, providing a further mechanism for the thyroidal control of TSH secretion.

### Role of Somatostatin

Somatostatin was initially isolated and characterized on the basis of its ability to inhibit GH release from the anterior pituitary gland. It has since been shown to inhibit the secretion of TSH in both animals and humans. Somatostatin inhibits both basal and TRH-stimulated TSH secretion from cultured rat anterior pituitary cells, and this effect is greater in the presence of hypothyroidism. These observations led to the hypothesis that TSH release was regulated by the hypothalamus through a dual-control system involving stimulation by TRH, on the one hand, and inhibition by somatostatin, on the other. The biological relevance of this proposal was confirmed in studies using antiserum directed against somatostatin. For example, administration of antisomatostatin antiserum to rats increases both basal serum TSH concentrations and the serum TSH response to cold stress and TRH. In humans, somatostatin reduces the elevated TSH concentrations in patients with primary hypothyroidism, reduces the serum TSH response to TRH, and prevents the release of TSH after administration of dopamine antagonist drugs. Despite these acute effects, chronic treatment with somatostatin or its analogues does not cause hypothyroidism, presumably because the sensitivity of thyrotrophs to any reduction in thyroid hormone concentration overrides the inhibitory effect of somatostatin in the long term.

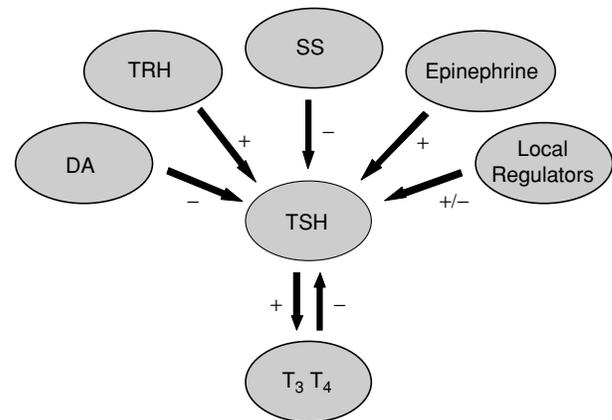
## Role of Neurotransmitters

An extensive network of neurotransmitter neurons, projecting from the midbrain and other regions, terminates on the cell bodies of the hypophysiotropic neurons and within the interstitial spaces of the median eminence where they regulate neuropeptide release into hypophyseal portal blood. However, these interacting neural networks are complex, and it is not surprising that attempts to dissect the relative contributions of different neurotransmitter systems on the regulation of TSH release have proved to be difficult. Nevertheless, consensus views have developed about the roles of several different neurotransmitter pathways, notably the catecholaminergic system.

Studies using central neurotransmitter agonist and antagonist drugs have identified the existence of stimulatory  $\alpha$ -adrenergic and inhibitory dopaminergic pathways in the control of TSH secretion in rats, and although the precise central effects of these pathways are uncertain, it is clear that dopamine and epinephrine exert opposing actions on TSH release directly at the pituitary level. Furthermore, both of these molecules are present in hypophyseal blood in physiologically relevant concentrations and in higher concentrations than in peripheral blood. Dopamine causes a dose-dependent inhibition of TSH release from cultured rat and bovine anterior pituitary cells, mediated by the DA<sub>2</sub> class of dopamine receptors that are negatively coupled to adenylate cyclase, and also inhibits  $\alpha$ -subunit and TSH- $\beta$ -subunit gene transcription. In contrast to dopamine, adrenergic activation causes dose-dependent stimulation of TSH release in cultured anterior pituitary cells, an effect that is mediated by high-affinity  $\alpha$ -1 adrenoreceptors. Quantitatively, the adrenergic release of TSH is nearly equivalent to that induced by TRH, and together they have additive effects on TSH secretion, suggesting activation of separate intracellular pathways.

In humans, it is well recognized that dopamine exerts a physiological inhibitory effect on TSH release, and some data suggest a stimulatory  $\alpha$ -adrenergic pathway. However, the catecholaminergic control of TSH secretion in humans appears to be a fine-tuning mechanism rather than a matter of primary importance, and as with somatostatin analogues, chronic administration of catecholaminergic agents does not result in alterations in thyroid status, reflecting the presence of compensatory mechanisms to maintain TSH secretion and euthyroidism.

The effects of a number of other substances on the release of TSH have been studied, and actions have



**Figure 1** Schematic representation of the principal primary and secondary modulators of thyrotroph function. DA, dopamine; SS, somatostatin. Local regulators include opioid peptides, neurotensin, neuromedin B, vasopressin, vasoactive intestinal polypeptide, tumor necrosis factor, and interleukin-1. Plus sign (+) indicates stimulation; minus sign (-) indicates inhibition.

been described for serotonergic pathways; opioid peptides; neuropeptides such as neurotensin, neuromedin B, vasopressin, and vasoactive intestinal polypeptide; and cytokines such as tumor necrosis factor and interleukin-1. However, in the majority of cases, adequate physiological studies have not been undertaken, and the precise roles of these substances in the regulation of TSH release are unclear. The major factors regulating the release of TSH are illustrated in [Fig. 1](#).

## PHYSIOLOGICAL ALTERATIONS IN TSH RELEASE

Alterations in TSH secretion may be apparent in the pattern or degree of change in basal TSH concentration or in the pattern and degree of TSH responses to TRH or dopamine receptor blockade.

Serum TSH concentrations show clear circadian and ultradian variation. In humans, serum TSH concentrations begin to rise before the onset of sleep, reaching a peak between 2300 and 0400 h and declining thereafter to reach a nadir at approximately 1100 h. Furthermore, a seasonal variation in serum TSH concentrations has been reported in patients with primary hypothyroidism receiving thyroxine replacement, and some of these patients have higher TSH concentrations during winter than during summer. However, the central mechanisms underlying TSH pulsatility and rhythmicity are unknown, and although pulsatile TRH release can influence

TSH pulse amplitude, the major input may be from the suprachiasmatic nuclei in the hypothalamus.

Serum TSH concentrations can also vary according to changes in body temperature, age, and caloric restriction. Exposure to cold in rats induces an increase in hypothalamic TRH secretion and a subsequent rise in serum TSH concentrations. A similar response is apparent in human neonates, although this increase is unusual in adults and is usually small when present. Aging causes a slight reduction in TSH secretion due to a readjustment of the threshold of TSH inhibition by thyroid hormones secondary to increased pituitary conversion of  $T_4$  to  $T_3$ , increased uptake of  $T_4$  by thyrotrophs, or reduced  $T_4$  and  $T_3$  clearance. These observations highlight the need for caution in using TSH assays alone when assessing thyroid function in the elderly. Caloric restriction induces a small reduction in basal and TRH-stimulated serum TSH concentrations despite a decrease in serum  $T_3$  concentrations, which in the extreme condition of anorexia nervosa may be secondary to increased serum cortisol concentrations. Recent evidence also indicates a critical role for leptin in mediating hypothalamic responses to caloric withdrawal.

### PATHOLOGICAL ALTERATIONS IN TSH RELEASE

Apart from thyroid and hypothalamic-pituitary disease, stress, nonthyroidal illness, and neuropsychiatric disease all can modify TSH secretion. In humans, surgical stress causes both a transient acute lowering of TSH release and chronic abolition of the nocturnal increase in serum TSH concentration. This occurs despite a fall in serum-free  $T_3$  concentrations, whereas serum-free  $T_4$  concentrations do not change. As with the effects of caloric restriction, these stress phenomena are similar to the altered neuroregulation of TSH that can occur in nonthyroidal illness and in some neuropsychiatric disorders such as anorexia nervosa and depression. Although basal serum TSH concentrations are usually normal in patients with both acute and chronic nonthyroidal illness, they can be either low or slightly raised. However, the mechanisms involved in the central suppression of thyrotroph function apparent in severe nonthyroidal illness are unclear, although alterations in glucocorticoid, cytokine, leptin, opiodergic, dopaminergic, and somatostatinergic activity all have been implicated.

The development of highly sensitive assays for the measurement of TSH has rendered TRH testing

obsolete and enabled thyrotoxicosis and euthyroidism to be reliably distinguished. With the exception of thyrotoxicosis secondary to inappropriate TSH secretion (from a TSH-secreting pituitary tumor or thyroid hormone resistance), serum TSH concentrations are low or undetectable in patients with thyrotoxicosis, irrespective of the cause, as a result of increased negative feedback of thyroid hormones on the thyrotrophs. However, following treatment, serum TSH concentrations remain low or undetectable for 4 to 6 weeks after  $T_4$  and  $T_3$  levels have returned to normal, reflecting a delay in synthesis or secretion of TSH by the previously suppressed thyrotrophs. In primary hypothyroidism, serum TSH concentrations are usually in excess of 20 mU/L in patients who are symptomatic, whereas concentrations between 5 and 20 mU/L are usually associated with serum thyroid hormone concentrations in the lower part of their normal ranges. After initiation of  $T_4$  replacement, serum TSH concentrations fall slowly (4–8 weeks) in comparison with the increase in  $T_4$  concentrations (days). This differential response allows the identification of poorly compliant patients where raised  $T_4$  and TSH values may be observed following  $T_4$  ingestion only for the few days before the clinic visit. Low serum TSH in the presence of normal thyroid hormone concentrations, termed “subclinical hyperthyroidism,” is common in patients with multinodular goiter. Such patients were not usually treated in the past, although the observation that osteopenia and atrial fibrillation are associated with subclinical hyperthyroidism now favors early treatment. In contrast, the treatment of subclinical hypothyroidism, where TSH levels are elevated in the presence of normal  $T_4$  and  $T_3$  concentrations, is more contentious.

In the majority of patients with pituitary disease, basal serum TSH concentrations and serum TSH responses to TRH or dopamine receptor blockade are normal or reduced; an absent TSH response to TRH is suggestive of a pituitary lesion. Occasional patients with a pituitary tumor causing stalk compression and mild hyperprolactinemia have a delayed TSH response to TRH. In patients with hypothalamic disease, TSH responses to TRH can be normal, suppressed, or delayed, with peak TSH responses not occurring until 60 min after TRH administration. In patients with prolactin-secreting pituitary tumors, thyroid status depends on the size of the tumor. Patients with macroadenomas may have impaired TSH responses to TRH and dopamine antagonism or, less commonly, to overt hypothyroidism. However, patients with microprolactinomas are almost invariably euthyroid, although the serum TSH response

to dopamine receptor blockade is characteristically exaggerated. Clinically important disruption of the hypothalamic–pituitary–thyroid axis is uncommon in patients with acromegaly or Cushing’s disease. In acromegaly, central hypothyroidism is rare and usually confined to large tumors. In hypercortisolism, a slight lowering of free  $T_4$  and TSH levels, secondary to inhibition by cortisol, is not uncommon but recovers quickly with treatment of the pituitary disease.

### See Also the Following Articles

Hypothalamus–Pituitary–Thyroid Axis • Iodine Deficiency • Neurotransmitters, Overview • Pituitary Gland Anatomy and Embryology • Resistance to Thyroid Hormone (RTH) • Somatostatin Analogs • Thyroid Disease, Epidemiology of • Thyroid Hormone Action • Thyrotoxicosis: Diagnosis • Thyrotoxicosis, Overview of Causes • Thyrotropin-Releasing Hormone (TRH) • Toxic Adenoma

### Further Reading

- Hofland, L. J., and Lamberts, S. W. (1996). Somatostatin receptors and disease: Role of receptor subtypes. *Baillieres Clin. Endocrinol. Metab.* **10**, 103.
- Jackson, I. M. D. (1982). Thyrotropin-releasing hormone. *N. Engl. J. Med.* **306**, 145–155.
- McIver, B., and Gorman, C. A. (1997). Euthyroid sick syndrome: An overview. *Thyroid* **7**, 125.
- Nillni, E. A., and Sevarino, K. A. (1999). The biology of prothyrotropin releasing hormone-derived peptides. *Endocrine Rev.* **20**, 599–648.
- Peters, J. R., Foord, S. M., Dieguez, C., and Scanlon, M. F. (1983). TSH neuro-regulation and alteration in disease states. *Clin. Endocrinol. Metab.* **12**, 669–694.
- Scanlon, M. F. (2001). Thyrotropin-releasing hormone and thyroid-stimulating hormone. In “Endocrinology” (L. J. DeGroot and J. L. Jameson, eds.), 4th ed. W. B. Saunders, Philadelphia.
- Scanlon, M. F., and Toft, A. D. (1996). Regulation of thyrotropin secretion. In “Werner and Ingbar’s The Thyroid” (L. F. Braverman and R. B. Utiger, eds.), 7th ed. Lippincott–Raven, Philadelphia.

thyroid ablation by surgery or radioiodine. This aggressive transformation of the pituitary tumor resembles that occurring in Nelson's syndrome after adrenalectomy for Cushing's disease.

By light microscopy, tumoral thyrotropes generally appear to be chromophobic, with atypical nuclei and mitoses, and so they are often mistakenly recognized as pituitary malignancy or metastasis from distant carcinomas. By electron microscopy, these tumors are monomorphous and characterized by a poorly developed Golgi apparatus and a low number of small secretory granules aligned mainly under the plasma membrane.

Approximately 75% of TSH-omas secrete TSH alone, which is often accompanied by an unbalanced hypersecretion of the  $\alpha$ -subunit of glycoprotein hormones. Hypersecretion of growth hormone (GH) and/or prolactin (PRL), resulting in acromegaly and/or amenorrhea/galactorrhea syndrome, is present in approximately 25% of tumors, whereas the occurrence of mixed TSH/gonadotropin adenomas is rare and no association with adrenocorticotropic hormone (ACTH) hypersecretion has been documented. This may be due to the fact that GH and PRL share with TSH the common transcription factor Pit-1.

As for most pituitary tumors, the pathogenesis of TSH-omas remains largely unknown. Screening studies for genetic abnormalities that may be responsible for tumor formation were generally negative. Anecdotal reports showed overexpression of the Pit-1 gene, absence of thyroid hormone receptor isoforms, and (recently) mutations in the beta isoform of the receptor leading to an altered feedback mechanism that is possibly responsible for the tumor formation. However, available data concern only a low number of tumors and are too preliminary to draw definite conclusions on transcriptional and/or expression abnormalities in TSH-omas.

## CLINICAL MANIFESTATIONS

Patients with TSH-omas present with signs and symptoms of hyperthyroidism that are sometimes milder than expected on the basis of circulating thyroid hormone levels. Signs and symptoms of hyperthyroidism are frequently associated with those related to the mass effects of the pituitary adenoma. The latter frequently prevail over those of thyroid hyperfunction. Visual field defects are present in approximately 41%, headache in 23%, and partial or total hypopituitarism in 25% of patients (Table I).

Most patients have a long history of thyroid dysfunction, often misdiagnosed as Graves' disease, and

**Table I** Clinical Features of Patients with TSH-omas

Feature	Patients with TSH-omas
Goiter	182/201 (90.5)
Previous thyroidectomy	99/315 (31.4)
Severe hyperthyroidism	50/196 (25.5)
Thyroid nodule(s)	50/71 (70.4)
Macroadenomas	235/279 (84.2)
Visual field defects	64/156 (41.0)
Headache	28/124 (22.5)
Amenorrhea <sup>a</sup>	29/86 (33.7)
Galactorrhea <sup>a</sup>	14/41 (34.1)
Acromegaly	49/312 (15.7)

*Source.* Updated from data published through June 2003.

*Note.* Figures are number of patients/total for whom data have been reported in the literature. Percentages are in parentheses.

<sup>a</sup>Data include women with or without hyperprolactinemia.

approximately one-third had inappropriate thyroidectomy or radioiodine thyroid ablation. In some acromegalic patients, signs and symptoms of hyperthyroidism may be clinically missed because those of acromegaly overshadow them. The presence of goiter, frequently uni- or multinodular (~90% of reported cases), is the rule. Also, in patients with previous thyroidectomy, thyroid residue may regrow as a consequence of TSH hyperstimulation. Progression toward functional autonomy or differentiated carcinomas seems to be infrequent. Bilateral exophthalmos occurred in a few patients who subsequently developed autoimmune thyroid disorders, whereas unilateral exophthalmos due to orbital invasion by pituitary tumor was reported in three patients with TSH-omas.

## DIAGNOSIS

### Serum Thyroid Hormone and TSH Levels

Serum total and free thyroid hormone levels are definitely high in patients with TSH-omas, whereas TSH levels may be elevated or in the normal range (Table II). In some patients with TSH-omas, the findings of normal TSH in the presence of high levels of free thyroxine (FT4) and free triiodothyronine (FT3) were demonstrated to be due to an increased biological activity of secreted TSH molecules.

Particular clinical situations and possible laboratory artifacts causing a biochemical profile similar to that characterizing central hyperthyroidism should be considered. Measuring free, instead of total, thyroid hormones may recognize most of the confusing conditions, including genetic alterations or drugs that may cause quantitative/qualitative alterations of

Table II Differential Diagnosis between TSH-omas and RTH

Parameter	TSH-omas	RTH	P
Female/Male ratio	1.29	1.41	NS
Familial cases (percentages)	0 <sup>a</sup>	86	<0.01
TSH (mU/L)	2.9 ± 0.5	2.4 ± 0.4	NS
FT4 (pmol/L)	39.8 ± 4.1	31.2 ± 2.3	NS
FT3 (pmol/L)	13.9 ± 1.2	12.3 ± 1.0	NS
SHBG (nmol/L)	121 ± 19	60 ± 5	<0.02
Lesions at CT or MRI (percentages)	98	7	<0.01
High $\alpha$ -subunit levels (percentages)	71	2	<0.01
High $\alpha$ -subunit/TSH m.r. (percentages)	82	2	<0.01
Absent or impaired TSH response to TRH test (percentages)	93	3	<0.01
Abnormal TSH response to T3 suppression test <sup>b</sup> (percentages)	100	100	NS

Note. Only patients with intact thyroid were taken into account. Data are obtained from patients followed at our institute and are expressed as means  $\pm$  SE.

<sup>a</sup>Excluding the three families with Men-1 reported in the literature.

<sup>b</sup>Werner's test (80–100  $\mu$ g T3 for 8–10 days). Quantitatively normal responses to T3 (i.e., complete inhibition of both basal and TRH-stimulated TSH levels) have never been recorded in either group of patients. Although abnormal in quantitative terms, TSH response to T3 suppression test was qualitatively normal in RTH patients.

T4-binding globulin, albumin, or transthyretin leading to increases in thyroid hormone levels, particularly T4. Laboratory artifacts, such as those caused by circulating anti-T4 and/or anti-T3 autoantibodies that interfere in most immunometric assays, may cause falsely high levels of free thyroid hormones, whereas heterophilic antibodies directed against or cross-reacting with mouse immunoglobulin G (IgG), as well as anti-TSH autoantibodies in patients previously treated with bovine TSH or contaminated pituitary extracts, may lead to incorrect evaluation of the actual TSH concentrations.

### Pituitary Glycoprotein Hormone $\alpha$ -Subunit

A helpful diagnostic tool for the diagnosis of TSH-omas is the determination of serum concentrations of the  $\alpha$ -subunit, the subunit common to all pituitary glycoprotein hormones. Secretion of the  $\alpha$ -subunit in these tumors is in excess not only of the TSH- $\beta$  subunit but also of the intact TSH molecule, resulting in an  $\alpha$ -subunit/TSH molar ratio generally higher than that recorded in controls matched for age and sex and with similar circulating levels of TSH, luteinizing hormone (LH), and follicle-stimulating hormone (FSH) (Table II).

### Parameters of Peripheral Thyroid Hormone Action

Because patients with central hyperthyroidism may present with mild signs and symptoms of thyroid

hormone overproduction, the measurements of several parameters of peripheral thyroid hormone action have been proposed to quantify the degree of hyperthyroidism. Some of these parameters, and in particular sex hormone-binding globulin (SHBG) and cross-linked carboxyterminal telopeptide of type I collagen (ICTP), have been used to differentiate hyperthyroid patients with TSH-omas from those with RTH (Table II); patients with TSH-omas have levels of these indexes into the hyperthyroid range, whereas they are into the normal range in RTH patients.

### Dynamic Testing

Several stimulatory and inhibitory tests are useful for the diagnosis of TSH-omas. Classically, the T3 suppression test has been used to assess the presence of a TSH-oma because complete inhibition of TSH secretion after T3 (80–100  $\mu$ g/day for 8–10 days) has never been recorded in patients with TSH-omas (Table II). It is worth noting that in patients with previous thyroid ablation, T3 suppression seems to be the most sensitive and specific test in assessing the presence of a TSH-oma. In approximately 93% of patients, TSH levels do not increase after TRH injection. The lack of TSH response to TRH may also be useful in unusual situations where TSH-omas coexist with primary hypothyroidism.

The majority of patients with TSH-omas maintain the sensitivity to native somatostatin or its analogues, and these tests may be predictive of the efficacy of long-term treatment in most patients.

## Imaging Studies

When considering the diagnosis of a TSH-oma, full imaging studies, particularly nuclear magnetic resonance imaging (MRI) or high-resolution computed tomography (CT), are mandatory. Various degrees of suprasellar extension or sphenoidal sinus invasion are present in two-thirds of cases. The positivity of Octreoscan (scintigraphy with  $^{111}\text{Indium-pentetreotide}$ ) may also be helpful in the diagnostic workup of particular cases.

## Differential Diagnosis

In a patient with signs and symptoms of hyperthyroidism, the presence of elevated thyroid hormone and detectable TSH levels rules out primary hyperthyroidism. When the existence of central hyperthyroidism is confirmed and the presence of methodological interferences is excluded, several diagnostic steps have to be carried out to differentiate a TSH-oma from RTH (Table II). Indeed, the possible presence of neurological signs and symptoms (e.g., visual defects, headache) or clinical features of concomitant hypersecretion of other pituitary hormones points to the presence of a TSH-oma. Furthermore, the presence of alterations of pituitary content at MRI or CT scan strongly supports the diagnosis of a TSH-oma. Nevertheless, the differential diagnosis may be difficult when the pituitary adenoma is undetectable by MRO or CT scan or in the case of confusing lesions such as empty sella or pituitary incidentalomas. In these cases, elevated  $\alpha$ -subunit concentrations or high  $\alpha$ -subunit/TSH molar ratios and TSH unresponsiveness to TRH stimulation and/or to T3 suppression tests favors the presence of a TSH-oma. Moreover, in contrast to RTH patients, familial cases of TSH-omas have never been documented. Finally, an apparent association between TSH-oma and RTH was recently reported in a young Japanese woman, although genetic investigations of possible mutations in T3 receptor  $\beta 1$  were not carried out. Nonetheless, the occurrence of TSH-omas in RTH patients is theoretically possible and, therefore, should be carefully considered.

## TREATMENT AND FOLLOW-UP

Surgical resection is the recommended therapy for TSH-producing pituitary tumors. However, a radical removal of large adenomas, which still represent the majority of TSH-omas, is particularly difficult due to the local invasion of the tumor. Particular attention has to be paid to presurgical preparation of the

patient. Antithyroid drugs or octreotide along with propranolol should be used to restore euthyroidism. If surgery is contraindicated or declined, as well as in the case of surgical failure, pituitary radiotherapy may be undertaken with the recommended dose of no less than 45 Gy fractionated at 2 Gy per day or 10 to 25 Gy in a single dose if a stereotactic gamma unit is available. The criteria of cure of patients operated and/or irradiated for TSH-omas have not been clearly established due to the rarity of the disease and the great heterogeneity of parameters used. In particular, clinical remission of hyperthyroidism with normalization of thyroid hormones, TSH, and  $\alpha$ -subunit or  $\alpha$ -subunit/TSH molar ratio, and with the disappearance of neurological symptoms, has been considered for the evaluation of the efficacy of surgery or radiotherapy in patients with TSH-omas. In our experience, undetectable TSH levels 1 week after surgery are likely to indicate complete adenectomy provided that presurgical treatments were stopped at least 10 days before surgery. The most sensitive and specific test to document the complete removal of the adenoma remains the T3 suppression test. In fact, only patients in whom T3 administration completely inhibits basal and TRH-stimulated TSH secretion appear to be truly cured. No data on the recurrence rates of TSH-omas in patients judged to be cured after surgery or radiotherapy have been reported. Although earlier diagnosis has improved the surgical cure rate of TSH-omas, several patients require medical therapy to control the hyperthyroidism. Dopamine agonists, and particularly bromocriptine, have been employed in some TSH-omas with variable results, with positive effects being observed mainly in some patients with mixed PRL/TSH adenomas. Today, the medical treatment of TSH-omas rests on long-acting somatostatin analogues such as octreotide LAR and lanreotide SR. Treatment with these analogues leads to a reduction of TSH and  $\alpha$ -subunit secretion in nearly all cases, with restoration of the euthyroid state in the majority of them. During octreotide therapy, tumor shrinkage occurs in approximately half of patients and vision improvement occurs in 75%. Resistance to octreotide treatment has been documented in only 4% of cases. Whether somatostatin analogue treatment may be an alternative to surgery and irradiation in patients with TSH-omas remains to be established.

## See Also the Following Articles

Gonadotropin-Secreting Tumors • Hypothalamus–Pituitary–Thyroid Axis • Pituitary Tumors, Clonality • Pituitary

Tumors, Molecular Pathogenesis • Resistance to Thyroid Hormone (RTH) • Thyroid Hormone Action • Toxic Adenoma • TSH Function and Secretion

### Further Reading

- Ando, S., Sarlis, N. J., Krishnan, J., Feng, X., Refetoff, S., Zhang, M. Q., Oldfield, E. H., and Yen, P. M. (2001). Aberrant alternative splicing of thyroid hormone receptor in a TSH-secreting pituitary tumor is a mechanism for hormone resistance. *Mol. Endocrinol.* **15**, 1529–1538.
- Beck-Peccoz, P., Brucker-Davis, F., Persani, L., Smallridge, R. C., and Weintraub, B. D. (1996). Thyrotropin-secreting pituitary tumors. *Endocrine Rev.* **17**, 610–638.
- Brucker-Davis, F., Oldfield, E. H., Skarulis, M. C., Doppman, J. L., and Weintraub, B. D. (1999). Thyrotropin-secreting pituitary tumors: Diagnostic criteria, thyroid hormone sensitivity, and treatment outcome in 25 patients followed at the National Institutes of Health. *J. Clin. Endocrinol. Metab.* **84**, 476–486.
- Caron, P., Arlot, S., Bauters, C., Chanson, P., Kuhn, J. M., Pugeat, M., Marechaud, R., Teutsch, C., Vidal, E., and Sassano, P. (2001). Efficacy of the long-acting octreotide formulation (octreotide-LAR) in patients with thyrotropin-secreting pituitary adenomas. *J. Clin. Endocrinol. Metab.* **86**, 2849–2853.
- Chanson, P., Weintraub, B. D., and Harris, A. G. (1993). Octreotide therapy for thyroid stimulating-secreting pituitary adenomas: A follow-up of 52 patients. *Ann. Intern. Med.* **119**, 236–240.
- Losa, M., Giovanelli, M., Persani, L., Mortini, P., Faglia, G., and Beck-Peccoz, P. (1996). Criteria of cure and follow-up of central hyperthyroidism due to thyrotropin-secreting pituitary adenomas. *J. Clin. Endocrinol. Metab.* **81**, 3086–3090.
- Refetoff, S., Weiss, R. E., and Usala, S. J. (1993). The syndromes of resistance to thyroid hormone. *Endocrine Rev.* **14**, 348–399.
- Sanno, N., Teramoto, A., and Osamura, R. Y. (2001). Thyrotropin-secreting pituitary adenomas: Clinical and biological heterogeneity and current treatment. *J. Neuro-Oncol.* **54**, 179–186.
- Terzolo, M., Orlandi, F., Bassetti, M., Paccotti, P., Cortelazzi, D., Angeli, A., and Beck-Peccoz, P. (1991). Hyperthyroidism due to a pituitary adenoma composed of two different cell types, one secreting alpha-subunit alone and another cosecreting alpha-subunit and thyrotropin. *J. Clin. Endocrinol. Metab.* **72**, 415–421.

Table I TNF Superfamily

Superfamily no.	Ligand name	Other names	Phenotypes associated with mutations	Functional observations
TNFSF1	LT- $\alpha$	TNF- $\beta$ , lymphotoxin- $\alpha$	Absence of lymphonodi	
TNFSF2	TNF	Tumor necrosis factor, TNF- $\alpha$ , cachectin	Defective germinal center formation	
TNFSF3	LT- $\beta$	Lymphotoxin- $\beta$ , TNF- $\gamma$	Absence of peripheral lymphonodi	
TNFSF4	OX40L		Defective T-cell responses	
TNFSF5	CD40L	CD154	Defective T-cell and IgG responses	
TNFSF6	FasL	CD95L	Autoimmunity	
TNFSF7	CD27L	CD70		
TNFSF8	CD30L			Malignant lymphocyte disorders
TNFSF9	4-1BBL		Defective T-cell responses	
TNFSF10	TRAIL	Apo-2L		
TNFSF11	RANKL	TRANCE, OPGL, ODF	Osteopetrosis, abnormal B-cell and T-cell development	Mammary gland development
TNFSF12	TWEAK	DR3L, APO3L		Monocyte and natural killer cell cytotoxicity
TNFSF13	APRIL			B-cell responses
TNFSF13B	BLYS	BAFF, THANK, TALL1		B-cell responses
TNFSF14	LIGHT			
TNFSF15	VEGI	TL1		Vascular endothelial cell growth inhibitor
TNFSF18	AITRL	TL6		Inhibits T-cell receptor-dependent apoptosis

immune function. CD95 ligand (CD95L, FasL) plays an important role in several types of physiological apoptosis. TNF-related apoptosis-inducing ligand (TRAIL, Apo-2L) is a potent endogenous activator of the cell death pathway. TNF-related activation-induced cytokine (TRANCE, RANKL) mediates survival of dendritic cells and is required for osteoclast differentiation and activation in the skeleton.

The receptors for these proteins also constitute a TNF receptor (TNFR)-related gene superfamily (Table II). Molecular studies and databases permitted the identification of a group of factors with structural similarities and diverse functions, and the term TNF-TNFR superfamily evolved. The TNF-TNFR superfamily is evolutionally highly conserved.

Two types of TNFRs have been described: TNF receptor I (TNFRSF1A, p55-R in mice; p60-R or TNFR60 in humans; CD120a) and TNFR II (TNFRSF1B, p75-R, p80-R, TNFR80, CD120b). TNF receptor I contains a death domain (DD) and TNF receptor II has a TNFR-associated factor (TRAF) domain. The extracellular ligand-binding region of all TNF superfamily receptors is characterized by a variable number of cysteine-rich domains,

permitting the formation of trimeric structures. TNFRs can be cleaved proteolytically and exist as soluble receptors. Most DD-containing genes derive from a common ancestor, and a subsequent diversification occurred in the TRAF domain.

The cytoplasmic domains of TNFS function as docking sites for signaling molecules. Signaling occurs through two cytoplasmic adaptor proteins, TRAFs and DD molecules. The binding of TNF to TNF receptor I causes the association of adaptors such as TNFR-associated DD protein (TRADD) and Fas-associated death domain (FADD) that ultimately causes caspase activation and cell death. Like other TNF family members, CD95L (FasL) is a homotrimeric molecule. Because DD molecules associate with one another, CD95L binding leads to clustering of the DDs of CD95 (Fas). FADD then binds through its own DD to the clustered receptor DDs.

TRAFs are critical mediators of the cell-activation and antiapoptotic functions of the TNFR superfamily. In particular, TRAF2 is important for protection from apoptosis, as is clear from the observation that cells from TRAF2-deficient mice are overly sensitive to TNF-induced apoptosis.

Table II TNF Receptor Superfamily

Superfamily no.	Receptor name	Other names	Phenotypes associated with mutations	Functional observations
TNFRSF1A	TNFR1	CD120a, p55-R, TNFR60, TNF-R-I	Impaired clearance of bacterial pathogens	
TNFRSF1B	TNFR2	CD120b, p75-R, TNFR80, TNF-R-II	Increased sensitivity to bacterial pathogens	
TNFRSF3	LT- $\beta$ R	TNFR2-RP, TNF-R-III	Absence of lymphonodi	
TNFRSF4	OX40	CD134, ACT35, TXGP1L	Defective T-cell responses	
TNFRSF5	CD40		Defective Ig class switching	
TNFRSF6	FAS	CD95, APO-1, APT1	Autoimmunity	
TNFRSF6B	DeR3			Secreted decoy receptor
TNFRSF7	CD27	Tp55, S152	Defective T-cell responses	
TNFRSF8	CD30	Ki-1		
TNFRSF9	4-1BB	CD137, ILA		Probable role in T-cell responses
TNFRSF11A	RANK	TRANCE-R	Osteopetrosis, abnormal B-cell development	Mammary gland development
TNFRSF12	DR3	TRAMP, WSL-1, APO-3		
TNFRSF14	HveA	HVEM, ATAR, TR2		T-cell proliferation
TNFRSF16	NGFR		Defective sensory neuron	
TNFRSF17	BCMA	BCM		B-cell responses
TNFRSF18	AITR			
TNFRSF19	Troy	Taj		

## BIOLOGICAL ACTIVITIES OF TNF

Cells of the macrophage/monocyte lineage are the predominant source of TNF *in vivo*, but a number of other cell types are capable of producing TNF (Table III). The most potent stimulus for TNF production is bacterial lipopolysaccharide (LPS). The macrophage is sensitive to LPS so that maximal rates of TNF production are observed when LPS is present in barely detectable concentrations. TNF is also produced by exposure to a wide variety of exogenous and endogenous factors (Table III). The exogenous factors include molecules associated with infectious pathogens, such as enterotoxins, fungal cell walls, and viral envelopes. Chemicals such as phorbol esters and calcium ionophores are also active in TNF induction. The endogenous factors that induce TNF include cytokines such as TNF, interleukin (IL)-1, IL-2, granulocyte macrophage colony-stimulating factor, and MIP-1 $\alpha$ . TNF production occurs as part of the conserved response to a wide range of cell stressors and threatening stimuli.

The biological effects of TNF are summarized in Table IV. TNF is a key mediator of inflammation and host defense against microbes, parasites, or neoplasia.

Endothelial cells significantly regulate the coagulation and fibrinolytic pathways via the expression of

tissue factor and thrombomodulin. TNF causes endothelial cells to have procoagulant activity by enhancing the expression of tissue factor and suppressing cofactor activity for the anticoagulant protein C. The effects of TNF eventually result in increased thrombin formation and decreased fibrinolysis and thus promote the intravascular deposition of fibrin. TNF also induces the enhanced expression of HLA antigens and adhesion molecules that participate in the adherence of leukocytes and platelets to the endothelial surface. Endothelial leukocyte adhesion molecule-1 is synthesized rapidly by endothelial cells in response to TNF and significantly mediates neutrophil adhesion. Vascular adhesion molecule-1 induced by TNF interacts specifically with lymphocytes but not neutrophils. Importantly, TNF up-regulates the expression of several adhesion molecules on endothelial cells that interact with lymphocytes, macrophages, and neutrophils and therefore promotes the recruitment of these cells to inflammatory sites.

The activation of T cells requires the presentation of antigen in conjunction with HLA class I or class II antigens. TNF up-regulates MHC class I expression on a number of cells as well as macrophages. Treatment with TNF alone causes a minor induction of MHC class II expression on macrophages, but

**Table III Cell Sources and Factors That Stimulate TNF Production**

Cell sources
Astrocytes
B cells
Basophils
Bone marrow cells
Eosinophils
Fibroblasts
Keratinocytes
Kupffer cells
Langerhans cells
Macrophages/monocytes
Mast cells
Microglial cells
Neutrophils
Natural Killer cells
T cells
Tumor cells
Endogenous stimuli
Antibodies to surface antigens
Antigen-antibody complex
C5a
Dead cells
Fc receptor cross-linking
Granulocyte macrophage colony-stimulating factor
Hypoxia
Interleukin (IL)-1, -2, -7
Interferon- $\alpha$ , - $\beta$ , - $\gamma$
Leukotrienes
MIP-1 $\alpha$
Nitrites
Oxidized lipids
Oxygen radicals
Tumor necrosis factor
Exogenous stimuli
Bacterial cell walls
Calcium ionophores
Endotoxin
Enterotoxin
Flavonoids
Fungal cell walls
Heat-killed bacteria
Irradiation
Lipopolysaccharides
Parasites
Phorbol esters
Synthetic lipid A
Ultraviolet
Viral envelopes

TNF can augment synergistically the effects of interferon- $\gamma$  on MHC class II expression. TNF causes increased phagocytosis and the release of oxygen radicals from macrophages and neutrophils.

TNF causes fever and anorexia because it crosses into the region of the hypothalamic center that regulates body temperature and appetite. The inflammatory effects of TNF are important in the development of cerebral inflammation and edema during meningitis. TNF also participates in the development of the characteristic plaques in multiple sclerosis.

TNF-induced activation of pulmonary endothelium and margination of neutrophils with degranulation lead to the development of adult respiratory distress syndrome, characterized by pulmonary edema, hypoxia, and high mortality.

TNF-treated adipocytes are depleted of lipids and exogenous lipid uptake is blocked. TNF up-regulates the expression of acute-phase proteins in serum and suppresses albumin biosynthesis. TNF stimulates the biosynthesis of circulating lipids, and excessive lipogenesis may contribute to the hypertriglyceridemia observed in the acute-phase response.

TNF has proved to be sufficient to elicit profound biological effects as an autocrine, paracrine, and endocrine mediator. One of the principal biological effects of TNF is that it triggers the release of other cytokines that enhance the biological activities of TNF. TNF operates as a part of the cytokine network system.

## TNF AND DISEASES

TNF displays diverse biological effects, including the ability to modify immunological, metabolic, and physiological responses. The local effects of TNF improve host defense systems by increasing immune function and stimulating inflammation. However, its overproduction may cause immunological and inflammatory diseases. TNF has been closely linked to the pathophysiological processes of many acute and chronic human diseases (Table V). TNF is rapidly released in large amounts in response to a wide variety of endogenous and exogenous stimuli. TNF may be the most important mediator of metabolic changes in the host compared to other cytokines. TNF triggers a cascade of subsequent mediators that enhance cellular and humoral changes. Thus, many consequences of TNF release, which may be beneficial when restricted to a local site, can be harmful to the host.

Table IV Biological Effects of TNF on Various Cells

Cell	Effects	Cell	Effects
Adipocytes	Lipoprotein lipase ↓		Phagocytosis ↑
Bone	Osteoclast activation		Platelet-activation factor ↑
Cardiovascular	Hypotension		Prostaglandins ↑
	Myocardial suppression	Metabolic	ACTH ↑
Endothelial	Adhesion molecules (ELAM-1, VCAM-1) ↑		Catecholamines ↑
	Cytokines ↑		Cortisol ↑
	HLA antigens ↑		Glucagon ↑
	Thrombomodulin ↓		Hyperaminoacidemia
	Tissue factor ↑		Hyperglycemia/hypoglycemia
Fibroblasts	Cytokines ↑		Insulin resistance ↑
	Epidermal growth factor receptor ↑		Lactic acidosis ↑
	Growth ↑	Musculoskeletal	Amino acid release ↑
Hematological	Hematopoiesis ↓		Myalgia
	Iron ↓	Nervous	Anorexia
	Neutrophilia/neutropenia		Fever ↑
	Procoagulant activity ↑		Hypothalamic-pituitary response ↑
Liver	Acute-phase proteins ↑	Neutrophils	Adherence ↑
	Albumin biosynthesis ↓		Chemotaxis ↑
	Hepatic necrosis ↑		Cytotoxicity ↑
Lymphocytes	Antibody ↑		Degranulation ↑
	Cytokines ↑		Oxygen radicals ↑
	Mitogenesis ↑		Phagocytosis ↑
Macrophages	Cytokines ↑	Pulmonary	Leukocyte margination ↑
	Cytotoxicity ↑		Respiratory arrest
	HLA antigens ↑	Tumor	Apoptosis ↑
	Leukotriens ↑		Differentiation ↑
	Oxygen radicals ↑		

TNF is produced in patients with septic shock, and in some cases serum levels of TNF predict mortality. TNF is the main mediator of tissue injury and septic shock. Administration of inhibitors of TNF prevents the development of tissue injury and septic shock. The protection conferred by TNF inhibitors is believed to result from the combined effect of inhibiting TNF toxicity and interrupting the cytokine cascade.

Cachexia is frequently observed in cancer patients and infectious diseases. Cachexia may be caused by chronic exposure to TNF. Cachexia is characterized by catabolism of proteins and lipid stores that leads to a loss of body weight. Increased serum triglyceride levels have been reported in severe malaria. Since serum TNF levels in malaria increase in relation to illness severity, and serum triglyceride levels increase when TNF is administered to the host, TNF seems to be involved in malarial hypertriglyceridemia. Severe anemia is a major component of malarial pathology. TNF makes an important contribution to the anemia of malaria. Dyserythropoiesis and

erythrophagocytosis seen in mice infected with malaria can be reproduced by administration of TNF. Erythropoiesis in mice infected with malaria can be increased by the injection of anti-TNF antibody.

Endogenous TNF released in response to infection triggers the characteristic manifestations of bacterial meningitis since mortality from cerebral meningitis correlates with the levels of TNF detected in the serum and cerebrospinal fluid.

There is convincing evidence that TNF plays an important role in the pathogenesis of autoimmunity. Genetically associated differences in TNF production and polymorphisms in the promoters of TNF have been associated with predisposition to or severity of various autoimmune diseases. TNF has been implicated in the cytotoxicity to islet cells. TNF injection induces a state of insulin resistance that is similar to the metabolic effects of diabetes.

TNF production in the inflamed joint has been implicated in the mediation of rheumatoid arthritis. Clinical investigations in which the activity of TNF in

**Table V Diseases Associated with TNF**

Adrenal medullary necrosis	Leprosy
Adult respiratory distress syndrome	Leukemia
AIDS	Malaria
Anemia	Meningitis
Asthma	Multiple sclerosis
Autoimmune disease (SLE)	Myocardial ischemia
Cancer cachexia	Obesity
Carcinogenesis	Osteoporosis
Colitis	Pollinosis
Contact dermatitis	Renal failure
Delayed allergy	Rheumatoid arthritis
Diabetes	Septic shock
Fever	Stroke
Hepatitis	Transplantation (GVH, HVG)
Immediate allergy	Trauma
Immune suppression	Tuberculosis

SLE, systemic lupus erythematos; GVH, graft versus host; HVG, host versus graft.

rheumatoid arthritis patients was blocked with anti-TNF monoclonal antibody have provided evidence that TNF regulates the production of other cytokines, angiogenesis, and recruitment of inflammatory cells

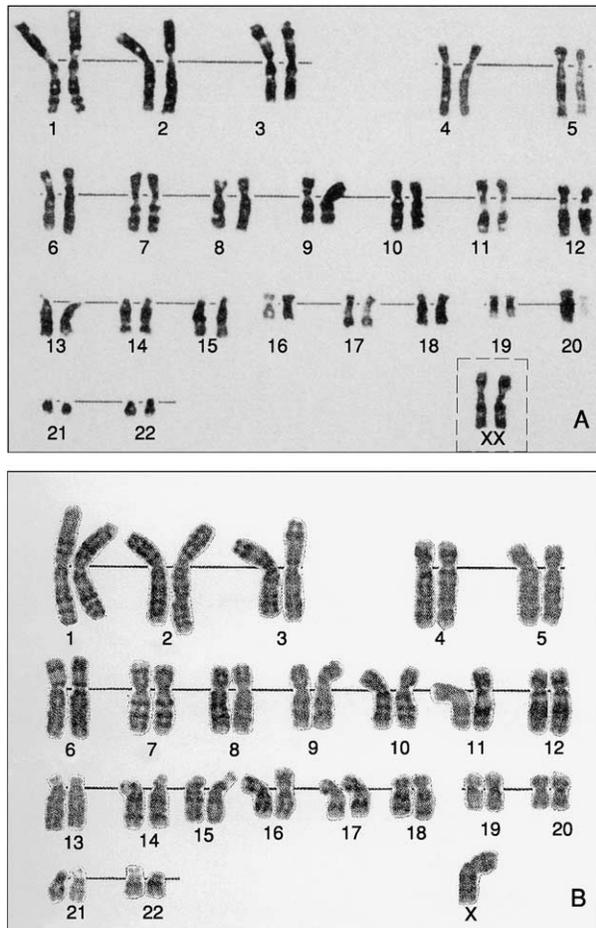
into joints. The important causal role of TNF in rheumatoid arthritis is also suggested by the prolonged period of remission following anti-TNF antibody therapy.

### See Also the Following Articles

Adipocytokines • Cytokines, Constitutive Secretion • Cytokines, Evolutionary Aspects and Functions • Cytokines, Extracellular Transport and Processing • Lipid Second Messengers and Receptors • Nuclear Factor- $\kappa$ B and Glucocorticoid Receptors • Thyrotoxic Bone Disease

### Further Reading

- Beutler, B. (1992). "Tumor Necrosis Factors." Raven Press, New York.
- Bona, C. A., and Revillard, J.-P. (2000). "Cytokines and Cytokine Receptors." Harwood, Amsterdam.
- Locksley, R. M., Killeen, N., and Lenardo, M. J. (2001). The TNF and TNF receptor superfamilies: Integrating mammalian biology. *Cell* **104**, 487–501.
- Remick, D. G., and Friedland, J. S. (1997). "Cytokines in Health and Disease." Dekker, New York.
- Thomson, A. (1991). "The Cytokine Handbook." Academic Press, New York.
- Thomson, A. (1994). "The Cytokine Handbook," 2nd ed. Academic Press, New York.



**Figure 1** (A) Normal karyotype and (B) karyotype of a girl with classic Turner syndrome. Reprinted from Ranke (1997), with permission.

division, instead of longitudinal cell division (leading to an isochromosome), major and minor deletions of either the long arm or the short arm of the X chromosome, or by the formation of a ring X chromosome.

Features of TS may also be seen in cases where, in addition to one of the cell lines mentioned, cells with material from the Y chromosome are also present (e.g., karyotype 45,X/46,XY). The presence of Y chromosome material may cause the development of gonadoblastoma. Therefore, probing for Y chromosome material should be performed in any TS patient with evidence of virilization or when a marker chromosome (a sex chromosomal fragment of unknown origin, i.e., X versus Y) is found. This can be achieved by DNA hybridization or fluorescence *in situ* hybridization using a Y centromeric or short arm probe and may require probing of multiple tissues.

## Karyotype/Phenotype Correlation

The wide range of developmental abnormalities in TS suggested that a number of different X-localized loci are responsible for the complete Turner phenotype. It is widely accepted that dosage (haploinsufficiency) of specific X-Y homologous genes that escape X inactivation causes the characteristic Turner somatic features and associated cognitive defects.

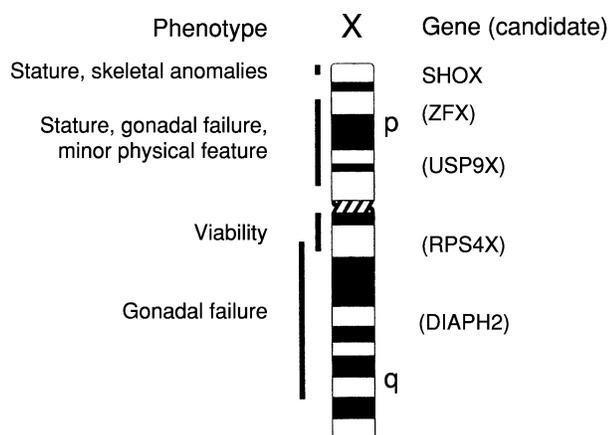
The important studies of the relationship between karyotype and phenotype in TS by Ferguson-Smith, published in 1965, first suggested that genes responsible for the TS phenotype could be assigned to specific regions on the X chromosome. Subsequent studies using chromosome banding techniques have shown: (1) a locus near the tip of the X chromosome involved in short stature; (2) proximal loci on Xp (short arm) involved in other TS anatomic features; (3) the association of the proximal long arm (Xq) with greater viability; and (4) a “critical region” for normal ovarian function from Xq13–q26, excluding part of Xq22. A rough phenotype map is shown in Fig. 2. With the further development of molecular techniques, it became possible to refine the short stature critical region to the extreme tip of the short arm of the X and Y chromosomes, the pseudo-autosomal region (PAR1). Subsequently, the PAR1 SHOX (short stature homeobox) gene was identified. It has been suggested that haploinsufficiency of this gene

**Table 1** Some Typical Chromosome Changes and Their Relative Frequency in Patients with Turner Syndrome ( $n = 649$ )

Karyotype	Number of patients	Percentage (%)
45,X	396	60.9
46,X,Xp-	6	0.9
46,X,Xq-	5	0.8
46,X,i(Xq)	38	5.9
45,X = 46,XX	86	13.3
45,X = 46,X,i(Xq)	41	6.3
45,X = 46,X,r(X)	30	4.6
Complete mosaics and rare deviations (e.g., 46,X,r(X); 45,X/47,XXX; 45,X/46,XX/47,XXX; 45,X/46,X,Xp-)	42	6.5

Based on collective statistics from Ranke, Book, Lenko, Prely, and Lippe. Reprinted from Ranke (1997), with permission.

*Note.* 46,X,Xp-, short arm missing; 46,X,Xq-, long arm missing; 46,X,i(Xq), isochromosome of two long arms; 46,X,r(X), ring chromosome.



**Figure 2** Turner syndrome critical regions and (candidate) genes. SHOX, short stature homeobox gene; ZFX, zinc-finger protein gene, the X-linked counterpart of ZFY, a closely related homologue identified during the search for the testis-determining gene on the Y chromosome; USP9X, the human homologue of a fruit fly gene [*Drosophila fat facets related X (DFFRX)*], involved in oogenesis and eye development; RPS4X, a gene on Xq that encodes an isoform of ribosomal protein S4 and might be associated with edema-related features; DIAPH2, human homologue of the *Drosophila* diaphanous gene; DIAPH2 might be required for normal ovarian function. Reprinted from Zinn and Ross (2000), with permission.

leads to the growth failure seen in TS. A number of lines of research have suggested other candidate genes for some TS phenotypes.

## CLINICAL FEATURES

### General

In addition to short stature and gonadal dysgenesis, there are a multiplicity of findings in patients with TS, occurring with varying frequencies (Table II, Fig. 3). The degree of the severity of the abnormal appearance seems to be partially dependent on the underlying chromosome disorder and is usually most marked when there is a complete loss of an X chromosome (karyotype 45,X). A number of features listed in Table II are considered a part of a skeletal growth disturbance, whereas other features can be explained by lymphatic obstruction or germ cell chromosomal defects. The growth failure and ovarian dysgenesis are the features occurring with the highest frequency in TS.

### Growth Failure

Growth is reduced in virtually all of the patients. Newborns with TS delivered at term are smaller than average girls; body length is reduced by approximately 3 cm and body weight by approximately 500 g. Postnatal growth rate appears to be in the normal

range during the first 2 to 3 years of life. Thereafter, height velocity shows a gradual decrease compared with healthy girls. In most girls, there is a lack of pubertal growth spurt because the ovaries are non-functional. Due to the delayed epiphyseal fusion, most untreated girls continue to grow until their late teens or beyond. The mean adult height of women with TS of white European origin ranges from approximately 142 to 147 cm in northwestern Europe; this is approximately 20 cm smaller than the mean of the normal female population. Figure 4 shows the Dutch–Swedish–Danish reference values for height in TS, which are based on data from a large multinational study. Convincing evidence that patients with a 45,X karyotype differ from those with another chromosomal pattern has not been documented. Although the absence of one of the X chromosomes results in the typical features of TS, such as short stature, the mechanisms leading to the stunted growth in girls with TS are still unknown. Growth failure in TS is not due to growth hormone (GH) deficiency but resistance to GH or growth factors might play a role.

### Ovarian Dysgenesis

In TS, the ovaries develop apparently normally during the first 3 months *in utero*. Thereafter, oocytes are rapidly lost and connective tissue transformation takes place. However, there is wide variation in the loss of germ cells in girls with TS, such that 5–10% retain sufficient ovarian function for puberty to commence spontaneously, though in most girls incompletely. Only in a few cases does spontaneous menstrual bleeding occur and in most of these cases it persists for only a short period of time. Occasionally, “spontaneous” pregnancies have been reported.

## DIAGNOSIS

### Prenatal

Most prenatally detected cases of TS are discovered incidentally during chorionic villous sampling or amniocentesis performed for unrelated reasons, the most common being advanced maternal age, which itself is not associated with an increased incidence of TS. Certain ultrasound findings (such as increased nuchal translucency, coarctation of the aorta, renal abnormalities) or an abnormal maternal multiple-marker screening ( $\alpha$ -fetoprotein, hCG, inhibin A, and unconjugated estriol) are frequently seen in TS, but should not be considered diagnostic of the disorder, however. Even when the prenatal diagnosis has been made by karyotype, chromosomes should be reevaluated postnatally.

Table II Physical Features and Their Frequency in Turner Syndrome

Feature	Frequency <sup>a</sup>	Feature	Frequency <sup>a</sup>
<b>Eyes</b>	20–39%	Start of hair growth in nape of neck pointing upward	
Ptosis		Pterygium colli	
Epicanthus (“Mongol-fold”)		<b>Chest</b>	60–79%
Myopia		Scutiform thorax	
Strabismus		Nipples apparently wide apart (not real, appearance results from scutiform thorax)	
Nystagmus		Inverted nipples	
<b>Ears</b>	40–59%	<b>Skeleton</b>	40–59%
Deformed auricles		Wide angle of arm (cubitus valgus)	
Otitis media		Short metacarpal bones (e.g., fourth metacarpal)	
Impaired hearing		Spongiose bone structure	
<b>Mouth, jaw</b>	60–79%	Scoliosis	
High arched palate (palatus arcuatus)		<b>Heart vessels</b>	40–50%
Small lower jaw (micrognathia)		Stenosis of aortic isthmus	
Defective dental development		Bicuspid aortic valve	
<b>Skin, skin appendages</b>	60–79%	Aortic dilation/aneurysm	
Lymphedema of hands and feet		<b>Kidneys</b>	40–59%
Increased number of pigmented naevi		Renal malformation (e.g., horseshoe kidney)	
Increased body hair growth		Renal aplasia	
Fingernail and toenail dysplasia		Changes in the renal pelvis and ureters	
Increased skin ridge patterns (dermatoglyphics)		Vessel abnormalities	
Alopecia		<b>Ovaries</b>	80–100%
Vitiligo		Gonadal dysgenesis	
<b>Neck</b>	60–79%	<b>Growth</b>	80–100%
Short, thick neck		Small for gestational age	
Low nape of neck/hairline		Growth retardation after birth	

Source. Reprinted from Ranke (1997), with permission.

<sup>a</sup>Frequencies reported in the literature show some variation. Ranges are therefore given for frequency, rather than exact figures.

## Postnatal

TS is most frequently discovered during childhood. The age at which the diagnosis is confirmed depends on the severity of the symptoms. The physical features are not always immediately obvious in every girl with TS or might even be absent in some cases (Table II). Clinicians should always consider the diagnosis in any female patient with unexplained growth failure or pubertal delay.

## MANAGEMENT IN CHILDHOOD, ADOLESCENCE, AND ADULTHOOD

### Short Stature

Although girls with TS are not GH deficient, treatment with biosynthetic recombinant human GH accelerates height velocity. To determine the effect of GH on adult height, the attained adult height is

compared to the individually predicted adult height at the start of GH treatment based on references for untreated girls with TS. The method of choice for predicting adult height in patients with TS is the Projected Adult Height method. This method assumes that the standard deviation (SD) score of the adult height is equal to the SD score of height for chronological age in girls with TS at any moment before final height is reached. In many countries, TS is an accepted indication for GH treatment although controversy exists regarding the effects of GH treatment on adult height since the mean increase in adult height varies between 0.2 and 16.0 cm in different reports. For most girls, a minimal gain in adult height of approximately 10 cm can be achieved when GH is started before the age of approximately 8 or 9 years with a daily GH dose of approximately 0.045 mg/kg/day (1.4 mg/m<sup>2</sup>/day ~ 4 IU/m<sup>2</sup>); the standard daily GH dose for children with GH deficiency is 0.023 to 0.032 mg/kg/day. Higher GH doses (0.06 or

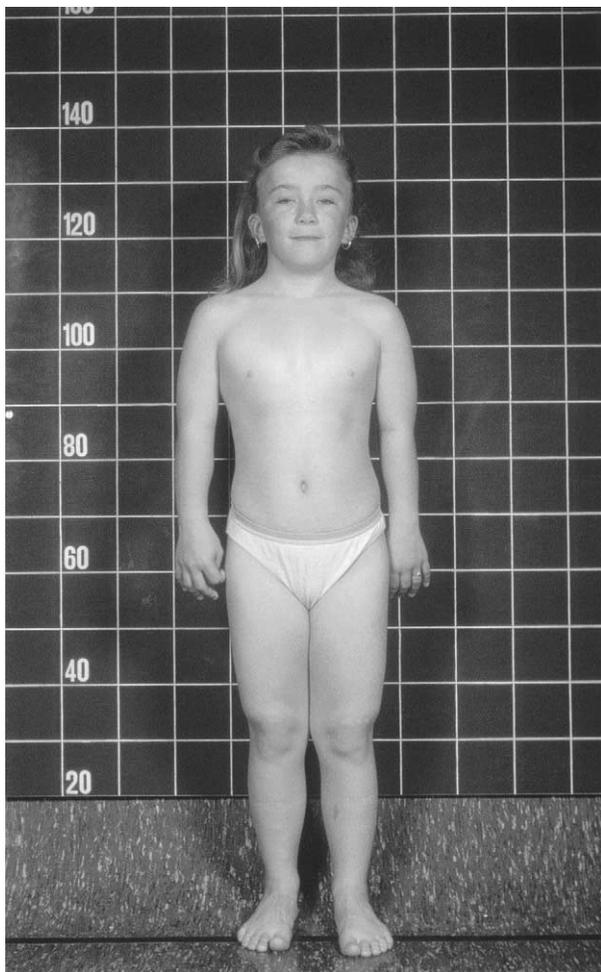


Figure 3 Girl with Turner syndrome.

0.09 mg/kg/day) or the addition of the synthetic anabolic steroid oxandrolone to GH may be more effective, but the efficacy on adult height has yet to be proven.

Although supraphysiological GH dosages are given for several years, no negative side effects of clinical importance of long-term GH treatment have been found, other than a very slight increase in the incidence of edema in the first weeks of the treatment and in the incidence of kyphoscoliosis, diabetes mellitus type 2, and benign intracranial hypertension. During GH treatment, a disproportionately greater increase in foot size than in height is often found in girls with TS. Since it has been shown that untreated adult women with TS have relatively large feet as well, this disproportionate growth of feet must be considered mainly a part of the natural development in TS. GH had no adverse effects on glucose levels, but did induce higher levels of insulin that decreased after

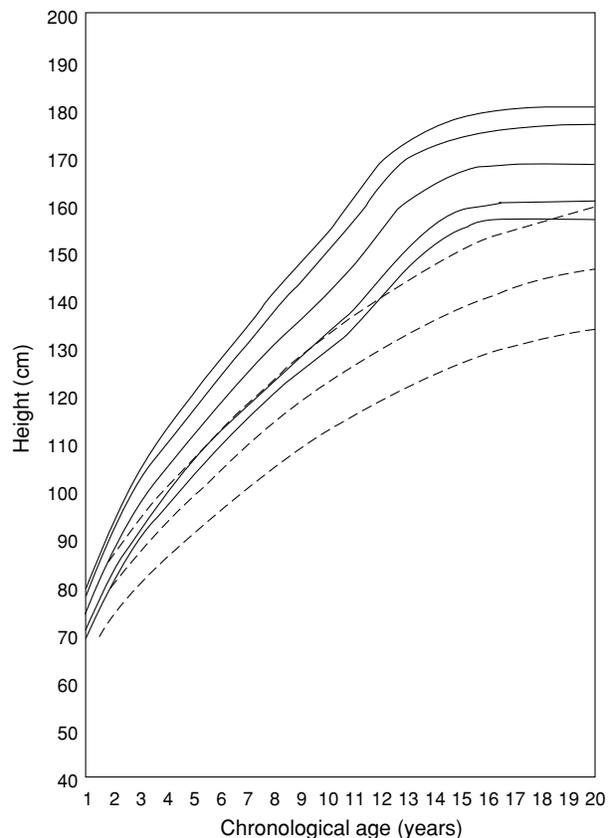


Figure 4 Reference curve for healthy Dutch girls (3rd, 10th, 50th, 90th, and 97th percentiles, indicated by solid lines) and for untreated girls with Turner syndrome (northern European references: 3rd, 50th, and 97th percentiles, indicated by dashed lines).

discontinuation of GH. Long-term GH treatment seems to be safe, but continued observation into adulthood is required.

## Gonadal Dysgenesis

### Induction of Puberty

To induce puberty, continuous treatment with estrogens must be given to girls with TS. Although 5–10% retain sufficient ovarian function for puberty to start spontaneously, most girls show a progressive ovarian failure and need estrogen therapy for complete breast development and withdrawal bleeding. The authors prefer the use of the natural estrogen  $17\beta$ -estradiol over synthetic estrogens. The optimal age to start estrogen therapy is a point of discussion. Although it has been shown that lower doses may stimulate growth, it has been suggested that estrogen therapy be postponed to delay closure of the epiphyseal growth plates and, consequently, to prolong the growth phase. However, delay of pubertal development may have

serious psychosocial consequences. Therefore, it is important to maximize the growth before higher levels of estrogens, required for complete pubertal development, will result in epiphyseal fusion. In the authors' view, when GH treatment has been optimized, that is, starting early (before 8–9 years of age) and with an adequate dose (0.045 mg/kg/day), it is not necessary to delay the induction of puberty: treatment can start at approximately 12 years of age with low-dose estrogens [5 to 10  $\mu$ g 17 $\beta$ -estradiol/kg/day or 50 to 100 ng ethinyl-estradiol/kg/day or Premarin (a conjugated estrogen used particularly in the United States) at a dose of 0.3 to 0.625 mg]. The dose must be adjusted to the response of the individual patient. However, the progression of puberty (from Tanner breast stage 2 to stage 4–5 with regular withdrawal bleeding) should take a similar length of time as in spontaneous puberty, that is, approximately 2–3 years. Cyclic progestin, such as medroxyprogesterone or dydrogesterone, at 5–10 mg daily for 12–14 days each month, should be added in order to induce withdrawal bleeding and consequently prevent endometrial hyperplasia. It should be started in accordance with normal pubertal development, that is, before Tanner breast stage 4 is reached, which is usually after 2 years on estrogen.

#### **Hormonal Replacement Therapy in Adulthood**

To treat the short- and long-term consequences of estrogen deficiency in women with TS, continuation of hormonal replacement therapy (HRT) is highly recommended. The available studies indicate that HRT has beneficial effects on bone density, risk of cardiovascular disease, and hyperinsulinism. In addition, HRT might have positive effects on general well-being, cognition, and psychosocial functioning. Natural estrogens are recommended either orally, at least (the equivalent of) 2 mg 17 $\beta$ -estradiol, or transdermally, 25–50  $\mu$ g/24 h, combined with cyclic progestin. Since overweight and hypertension can be exacerbated by exogenous estrogen, particularly when administered by the oral rather than the transdermal route, blood pressure should be checked at regular intervals.

#### **Fertility**

Ninety-five percent of women with TS are infertile. Abnormal germ cell development combined with accelerated oocyte loss results in a percentage of only 5–10% of Turner girls who will have oocytes remaining in their ovaries at the age of puberty. Those with follicles remaining may develop ovulatory cycles and will have a short period of fertility prior to entering into a premature menopause. In general, outcomes of

spontaneous pregnancy in Turner women are poor, with miscarriage rates of 30% and a high incidence of structural and chromosomal abnormalities that may affect 25% of live births.

The availability of oocyte donation combined with *in vitro* fertilization offers these women a realistic change of conceiving and delivering a healthy child. Although some case series indicate that conception rates rival those obtained in non-Turner women, the pregnancies are more frequently complicated by miscarriage and the need for caesarean section.

Pregnancy is associated with maternal morbidity. Those with cardiovascular abnormalities may be at risk of aortic dissection, and prepregnancy screening for preexisting dilation of the aortic root is therefore recommended. In addition to possibly increasing the rate of miscarriage, the hypoplastic uterus frequently observed in adult TS may be prone to rupture in late pregnancy. Furthermore, an increased risk of cephalopelvic disproportion at delivery exists. The mechanical and physiological stresses of pregnancy may be reduced by avoiding multiple pregnancy.

The future may offer additional fertility options for Turner women. The ability to cryopreserve, thaw, and mature oocytes *in vitro* makes it feasible to preserve for later fertilization those few oocytes remaining at prepuberty and technique is being extensively investigated. Although it is still only a research tool, this technique may provide the possibility of pregnancy with the patient's own oocytes. Though it is an attractive proposition, it would not overcome the high risk of fetal abnormalities observed in spontaneous pregnancies in Turner women.

### **Other Medical Problems**

#### **Cardiovascular Disease**

The most frequently observed major congenital cardiovascular abnormality in TS is the coarctation of the aorta, which can be corrected completely by surgical intervention. Morphological abnormalities of the aortic valves, which are often found, are mostly not of clinical significance in childhood. The incidence of hypertension is higher in girls and adults with TS than in healthy peers. The mechanisms leading to hypertension in these individuals are unknown and not clearly related to the cardiac or renal congenital abnormalities. Obesity, common in adolescent and adult individuals with TS, contributes to high blood pressure. Rupture of an aneurysm or dissection of the aorta in childhood has been rarely described and is associated with hypertension. However, even in the

absence of predisposing factors such as aortic valve abnormalities or hypertension, this life-threatening event occurs more frequently than in the normal population and can even occur at a younger age. Therefore, evaluation and follow-up of the cardiovascular dimensions and blood pressure are required, with adequate treatment when indicated.

The chronic estrogen deficiency known to affect many adult women with TS is likely to be associated with increased cardiovascular morbidity, since it is becoming clear that estrogens not only confer cardioprotectivity by lowering harmful circulating lipids, but also through direct antioxidant effects, through a change in the vascular reactivity, and through its interaction with smooth vascular muscles.

### **Ocular and Hearing Problems**

The eyes often show slight changes in the position and form of the palpebral fissure or ptosis can be observed; this does not affect function. It is important that visual disorders and squinting should be diagnosed and treated as early as possible to prevent further deterioration or permanent damage.

Conductive hearing loss and sensorineural hearing loss are common in girls and women with TS. The outer ear, middle ear, and inner ear are all affected and hearing problems and ear malformations correlate with the karyotype. Frequent periods of otitis media are found in 50–80% of patients. With audiometric measurements, a sensorineural dip in the midfrequencies is

often found. In a young Turner girl, this seldom leads to hearing impairment, but the dip broadens and the depth progresses over time, leading to social hearing problems later. A high-frequency hearing loss (presbycusis) is also added to the dip at >35 years of age, leading to rapid hearing loss and social hearing problems at a younger age than in the normal population.

### **Skin and Neck**

Swelling of the backs of the hands and dorsum of the feet due to lymphedema is not exclusive to TS but is a striking feature, particularly in the newborn. In most girls, the lymphedema shows a tendency to regress after birth. Some patients have an increased number of pigmented naevi, which are usually benign but often show a tendency to grow during puberty. Some clinical evidence has shown that when recombinant human growth hormone is administered for short stature, the growth of melanocytic naevi is boosted without signs of malignant transformation. The patient's neck often appears short and thick-set. In addition, a pterygium colli (webbed neck) can be found (Fig. 5) and can be corrected by surgery. However, keloid scars develop more often than usual after surgical interventions.

### **Skeleton and Bone Mineral Density**

In addition to the "classical" sign of cubitus valgus, occasionally Madelung's deformity of the wrist (the ulna projects beyond the level of the hand because



**Figure 5** Turner girl with pterygium colli (webbed neck).

the wrist is displaced toward the palm of the hand) or some other minor deformities of the hands and feet are found. In addition, girls and women with TS have relatively large hands and feet relative to their height. Bone mineral content (BMC) is decreased in adults with TS, whereas BMC is normal in prepubertal girls with TS. It is well known that treatment with estrogens is pivotal in order to avoid rapid bone mineral loss in adolescents and young adults. Furthermore, studies have shown that GH-treated Turner girls receiving low-dose estrogens do not have osteopenia. There have been suggestions that the reduced BMC and bone mineral density (BMD) are a function of the syndrome per se in combination with the estrogen deficiency. Both clinical and epidemiological data show that the reduced BMD does indeed lead to increased fractures and osteoporosis. Gravholt and co-workers observed that the increased risk of fractures in TS was already present in childhood and persisted throughout all age groups, favoring the explanation that chromosomal aberrations contribute at least to some extent to the decreased BMD and increased risk of fractures. Furthermore, there are indications that other multiple discrete hormone insufficiencies in women with TS contribute to the low BMD and high risk of fractures.

#### **Renal Disorders**

Renal disease occurs up to nine times more frequently in girls and women with TS. Up to one-third of individuals with TS have structural renal malformation. Rotational abnormalities and double-collecting systems are found most frequently. Although many of these abnormalities are without significant major renal complications, some may result in an increased risk of hypertension, urinary tract infections, or hydro-nephrosis. Usually a renal ultrasound was carried out at the time of diagnosis.

#### **Thyroid Disease**

Ample clinical and epidemiological evidence shows that the risk of endocrine disease and especially thyroid disease is increased in TS. Up to 50% of women with TS have anti-thyroid antibodies, and autoimmune thyroiditis is common: almost one-quarter of adult patients have hypothyroidism compared with 1% in the general population. Thyreotoxicosis does not occur more frequently in TS.

#### **Glucose Metabolism and Insulin Sensitivity**

Increased frequency of abnormal glucose tolerance, reduced insulin sensitivity, hyperinsulinemia, and inappropriately low insulin secretion are seen in TS. In

a register study covering the entire population in Denmark, an increased relative risk of 4.38 was found. A new and unexpected association was that not only type 2, but also type 1 diabetes is more prevalent in TS. Adiposity, which is seen in a large proportion of the adult Turner population, and sometimes decreased physical fitness are important factors in the development of type 2 diabetes.

#### **Hepatic Function**

Higher serum levels of hepatic enzymes, such as alanine aminotransferase,  $\gamma$ -glutamyl transferase, and alkaline phosphatase, have been found in adult Turner women than in controls. These elevated levels of hepatic enzymes and proteins do not seem to be associated with overt hepatic disease. Epidemiological evidence, however, suggests that cirrhosis of the liver is more frequent in TS.

#### **Cancer**

The relative risk of most cancers is low in TS and comparable to that of the general population, with the exception of the colon and rectum. In two separate registers in Denmark, it was found that the relative risk of these cancers was five to seven times that of the background population. Since sex hormone substitution may lead to an increased risk of cancer of the breast and reproductive organs, concerns have been raised about a possible increased risk in Turner women on HRT. However, no increased risk in the two Danish registers have been found. Gonadoblastoma might develop in the ovaries of girls and women with TS with Y chromosome material. The risk of this occurring has previously been estimated as 25–40%, although newer data suggest a lower risk of only 7–10%. Nevertheless, gonadectomy is still recommended to exclude malignancy with absolute certainty.

## **COGNITIVE AND PSYCHOSOCIAL ABILITIES**

No increase in the prevalence of mental retardation is associated with TS, except for a few patients with a small ring X chromosome that fails to undergo X inactivation. Numerous studies have documented that girls with TS have a neurocognitive profile consisting of visuospatial information processing deficits in the context of relatively preserved verbal skills. In particular, persons with TS are impaired in the coding and transforming of visuospatial information. Relative to typically developing controls, girls with TS also are described as exhibiting difficulty with tasks of left/

right discrimination, road map skills, mental rotation, line orientation, and integration of motor and visual/perceptual skills.

It has been reported that girls with TS are vulnerable to specific behavioral problems, particularly hyperactivity and inattention, as well as learning disabilities.

In conjunction with behavioral problems, girls with TS often exhibit difficulties with social development and functioning. They have been described as showing delayed emotional maturity, difficulty understanding social cues, needing more structure to socialize, poor peer relations, shyness, and poor body image. Impairments in self-concept and self-esteem increase during adolescence but interestingly improve with initiation of estrogen replacement therapy. Abnormalities in cognitive and psychosocial abilities in persons with TS likely reflect underlying atypical brain development and function.

Girls with TS have a typically female pattern of development, with unambiguous female gender identification. Dating and initiation of sexual activities, however, may be somewhat delayed or infrequent. It is not clear whether this reflects some underlying genetic or hormonal influence on behavior or simply discomfort, because of the issues of short stature, physical anomalies, and infertility with which women with TS must cope.

### See Also the Following Articles

Constitutional Delay of Growth and Puberty (CDGP) • Infertility, Overview • Noonan Syndrome • Short Stature and Chromosomal Abnormalities • *SHOX* Disorders

### Further Reading

- Ferguson-Smith, M. A. (1993). Genotype-phenotype correlations in the Turner syndrome. In "Basic and Clinical Approaches to Turner Syndrome" (I. Hibi and K. Takano, eds.), pp. 17-26. Elsevier, Amsterdam, The Netherlands.
- Gravholt, C. H., Juul, S., Naeraa, R. W., and Hansen, J. (1998). Morbidity in Turner syndrome. *J. Clin. Epidemiol.* **51**, 147-158.
- Karlberg, J., Albertsson-Wikland, K., Naeraa, R. W., Rongen-Westerlaken, C., and Wit, J. M. (1993). Reference values for spontaneous growth in Turner girls and its use in estimating treatment effects. In "Basic and Clinical Approach to Turner Syndrome" (I. Hibi and K. Takano, eds.), pp. 83-92. Elsevier, Amsterdam, The Netherlands.
- Ranke, M. B. (1997). "Turner Syndrome: A Clinical Overview." Oxford Clinical Communications, Oxford, UK.
- Rao, E., Weis, B., Fukami, M., et al. (1997). Pseudoautosomal deletions encompassing a novel homeobox gene cause growth failure in idiopathic short stature and Turner syndrome. *Nat. Genet.* **16**, 3-4.
- Rosenfeld, R. G., Attie, K. M., Frane, J., et al. (1998). Growth hormone therapy of Turner's syndrome: Beneficial effect on adult height. *J. Pediatr.* **132**, 319-324.
- Rovet, J. (1990). The cognitive and neuropsychological characteristics of females with Turner syndrome. In "Sex Chromosome Abnormalities and Behaviour: Psychological Studies" (D. B. Berch and B. G. Bender, eds.), pp. 1-77. Westview Press for the American Association for the Advancement of Science, Boulder, CO.
- Saenger, P., Albertsson-Wikland, K., Conway, G. S., et al. (2001). Recommendations for the diagnosis and management of Turner syndrome. *J. Clin. Endocrinol. Metab.* **86**, 3061-3069.
- Sas, T., de Muinck Keizer-Schrama, S., Stijnen, T., et al. (1999). Normalization of height in girls with Turner's syndrome after long-term growth hormone treatment: Results of a randomized dose-response trial. *J. Clin. Endocrinol. Metab.* **84**, 4607-4612.
- Sas, T. C. J., de Muinck Keizer-Schrama, S. M. P. F., Stijnen, T., Aanstoot, H. J., and Drop, S. L. S. (2000). Carbohydrate metabolism during long-term growth hormone treatment and after discontinuation of growth hormone treatment in girls with Turner syndrome participating in a randomized dose-response study. *J. Clin. Endocrinol. Metab.* **85**, 769-775.
- Sas, T. C. J., Gerver, W. J., de Bruin, R., et al. (1999). Body proportions during long-term GH treatment in girls with Turner syndrome participating in a randomized dose-response trial. *J. Clin. Endocrinol. Metab.* **84**, 4622-4628.
- Turner, H. H. (1938). Syndrome of infantilism, congenital webbed neck, and cubitus valgus. *Endocrinology* **23**, 566-574.
- Zinn, A., and Ross, J. (2000). Critical regions for Turner syndrome phenotypes on the X chromosome. In "5th International Turner Symposium—Optimizing Health Care for Turner Patients in the 21st Century" (P. Saenger and A. Pasquino, eds.), pp. 19-28. Elsevier, Amsterdam, The Netherlands.

to need any further attention. For clinical studies, a more exact location of the testes than “scrotal,” “subinguinal,” or “inguinal” should be used, preferably by measuring the distance in millimeters between the testis and a defined point (e.g., the pubic tubercle or the external inguinal orifice).

The term cryptorchidism is often used to describe the condition of undescended testes. This term implies that the testes are “hidden,” that they can neither be seen nor palpated. Since in clinical practice even inguinal or subinguinal palpable testes should be considered undescended, the term cryptorchidism is not always correct. “Nondescended” and “retained” (Latin: *retentio testis*) are synonyms of undescended testes that are sometimes used.

The term ectopic testis has been coined to describe a position of the testis that is outside the normal path of the testis during descent. This diagnosis generally cannot be made before surgery. Only after surgery can the true differentiation between a low inguinal and an ectopic testis be made.

It is difficult to distinguish bilaterally retained testes from testicular atrophy. The differentiation between the two may be aided by a human chorionic gonadotropin (hCG) stimulation test. Significantly increased testosterone levels indicate the presence of testes. During the first 2 years of life, measurement of inhibin B in serum is valuable since in boys this hormone is only produced in Sertoli cells. It is more common to find uni- rather than bilaterally undescended testes, and unilaterally retained testes have a better prognosis for future fertility.

## INCIDENCE

The classical epidemiological studies by Scorer in the late 1950s in England are often cited. Using his quite strict definitions, 2.7% of boys with a birth weight of > 2.5 kg had undescended testes at birth, but testicular descent proceeded during the first postnatal months so that at age 3 months only 0.9% were considered to have incompletely descended testes. In premature infants, the incidence is much higher (21% if < 2.5 kg), reflecting the late normal descent *in utero*. When Scorer’s study was repeated 20 years later using the same methodology, the incidence at birth and at 3 months was 3.4 and 1.4%, respectively, suggesting a negative trend in fetal testicular development during this period.

In a recent report, the incidence of undescended testes was found to be twice as high in Denmark as in Finland using the same criteria for diagnosis. This

remarkable geographical difference is paralleled by a similar difference in the incidence of testicular cancer during the past few decades, suggesting a possible common environmental cause of testicular maldevelopment.

## MECHANISMS OF TESTICULAR DESCENT

The translocation of the fetal testicles from the lower abdominal wall to their final destination in the scrotum is lead through the inguinal canal by the gubernaculum, one end of which is anchored in the epididymal tail and the other in what is to become the bottom of the scrotum. Initially, the gubernaculum fills the canal, and through successive shortening and swelling of its intra- and extracanalicular parts it is believed to aid the testis through the canal. The precise hormonal regulation of this series of events is incompletely known. In addition to nonhormonal homeobox genes, some endocrine factors have been suggested to have a crucial role in one or several of the events leading to testicular descent.

Undescended testicles are often found in boys with hypogonadotropic hypogonadism or androgen insensitivity. This indicates a role of androgenic hormones in the final passage through the inguinal canal. Patients with persistent Müllerian duct syndrome due to a loss of function of anti-Müllerian hormone or its receptor also have retained testicles. Therefore, this hormone has been suggested to be a key factor for testicular descent. However, since in this syndrome the testes are closely linked to the uterus and fallopian tubes, this may in itself hinder descent. It has been found that in mice the newly described *Ins13* gene is important for gubernacular development and testicular descent. No human disease has yet been ascribed to this factor.

## ABNORMAL FUNCTIONS OF UNDESCENDED TESTICLES

### Spermatogenesis

Numerous animal experiments have proven that spermatogenesis is blocked at a premeiotic stage if the testes are retained in or moved into the abdomen at any time after birth. This spermatogenic arrest is reversed if the testes are transferred to the scrotum or even if they are cooled in an abdominal position with a cooling device. These experiments prove that a temperature less than 37°C is needed to complete

spermatogenesis. In man, the difference between abdominal and scrotal temperature is 2 or 3°C. Germ cells, including spermatogonia and spermatocytes, seem to be the most temperature-sensitive cells, but impaired Sertoli cell function at 37°C may also be of importance.

On the other hand, not all undescended testicles will attain normal spermatogenesis if they are transferred to the scrotum by surgery or by hormonal treatment. As many as 40% of successfully treated men with primary undescended testicles are reported to have grossly subnormal sperm quality. Although the treatment may be harmful, this suggests that a number of undescended testicles have other deficiencies in spermatogenesis unrelated to their abdominal or inguinal position. However, it is not possible to distinguish the testes with primary defects in spermatogenesis from those that will normalize when put in a scrotal position.

### Steroidogenesis

Animal experiments have demonstrated that even if the testes are translocated to the abdomen, normal blood levels of testosterone can be maintained. However, there is a subtle difference: The Leydig cells are enlarged and have a lower sensitivity to stimulation by Lutropin (synonyms: luteinizing hormone, LH, interstitial cell-stimulating hormone) or hCG. It is not known whether this is due to a primary effect of temperature on the Leydig cells or if it is secondary to damage to the seminiferous tubules, causing an abnormal paracrine hormonal milieu.

In man, normal testosterone levels in blood are generally seen even when the testicles are undescended. Thus, a stimulation test with hCG can be used to differentiate between bilaterally nonpalpable retained testicles and the complete absence of testes (testicular atrophy). It is important to remember that even a small testis that has a poor prognosis with regard to spermatogenesis should not be removed if the other testis is absent. This small testis may mean the difference between lifelong testosterone replacement and no medication.

### Malignancies

It is frequently cited that undescended testicles often develop malignancies and therefore should either be placed into the scrotum or removed. The basis for this assumption lies in the finding of an overrepresentation of undescended testicles (with or without successful

treatment) in cohorts of patients with testicular cancer. Five to 10% of men with testicular tumors have a history of undescended testes. However, the true incidence of testicular cancer in a group of men with undescended testes has not been well documented. In a Swedish retrospective study, all 3000 boys with undescended testes from a defined geographical area were operated on. These patients, as well as 30,000 boys operated on for inguinal hernia, were searched for in the Swedish Cancer Registry when they were adults. In the cryptorchidism cohort, 4 cases of testicular cancer occurred versus 0.54 expected, yielding a relative risk of 7.4 (95% confidence interval, 2.0–19.0). The hernia controls had no increased risk. The most prevalent testicular cancer was seminoma, which has a good prognosis following modern treatment. This figure contrasts with the observation by the Copenhagen group of scientists that as many as 1 or 2% of previously undescended testicles show microscopic carcinoma *in-situ* (CIS) cells. Such testes are known to often develop cancer at a later stage. Small dysgenetic testicles are more prone to develop CIS cells than normal-sized testes. Thus, it is recommended that an undescended testis should be brought into a palpable position so that tumor formation can be detected at an early stage, even in cases in which there is little hope of spermatogenesis. Since testosterone production might be maintained even in a small and soft testicle, it should not be removed unless the contralateral testis is normal.

### TREATMENT

There is much controversy regarding the methods of treatment of undescended testicles, despite the numerous articles that have been published. The major controversies relate to the method (hormonal or surgical) and the timing of treatment.

#### Hormonal Treatment

Probably based on the common finding of undescended testicles in boys with hypogonadotropic hypogonadism or androgen insensitivity, stimulation of the testes through administration of hCG or gonadotropin-releasing hormone (GnRH) has been advocated as the logical treatment. Direct administration of systemic testosterone has the drawback of treating the whole body as much as the testis and its surroundings and in a few trials has not been shown to be very effective.

### **Human Chorionic Gonadotropins**

In many countries, injection of hCG for 3–5 weeks is the recommended treatment of undescended testicles. The following is a typical treatment regimen: injection of 500 IU hCG intramuscularly twice weekly for 5 weeks for boys younger than 6 years old and doubling the dose for older boys. In uncontrolled studies, this treatment is reported to result in testicular descent in approximately 50% of patients; the nonresponders are referred for surgery. However, a meta-analysis of randomized controlled studies indicates that the success rate is only 19%, compared to 5% for placebo groups.

The short-term side effects of hCG include pain of injections, increased frequency of erections, and sometimes development of pubic hair. Animal experiments and some clinical observations have raised concern regarding possible long-term adverse effects of hCG on the testis. Thus, impaired spermatogenesis has been noted in dogs treated with hCG at the time of puberty. In rats, injection of hCG or LH causes an acute inflammatory response in the testicular interstitium, with increased intratesticular pressure, extravasation of leukocytes, and signs of edema. A similar microscopical picture has been reported in testicular biopsies taken within a few days of completed unsuccessful hCG treatment. Furthermore, it has been shown that the number of apoptotic spermatogonia is markedly increased following a course of hCG treatment compared to that in biopsies taken from boys operated on without prior hCG treatment. In one study, a positive correlation was found between the number of apoptotic spermatogonia within 1 month after treatment in childhood and the follitropin, follicle-stimulating hormone, levels as adults, approximately 20 years later. Testicular volume in adulthood was negatively correlated with the initial spermatogonial apoptosis, suggesting a quantitative difference in spermatogenesis secondary to the post-hCG apoptosis.

### **Gonadotropic-Releasing Hormone**

When this hormone became available as a therapeutic modality, several studies demonstrated good results on undescended testicles. However, later controlled studies were unable to repeat these results, and it has been concluded that GnRH treatment is most successful in boys with retractile or high scrotal testes. If results of the professional examination during the neonatal period are available, such information is of great help for predicting whether hormonal treatment (hCG or GnRH) will be effective: Only boys who had palpable testes at birth will respond to hormonal

treatment. Also, success is rare if the testicle is not palpable at the time of later examination. The mechanism of action seems to be increased release of pituitary gonadotropins, similar to the hCG treatment but with a more moderate stimulation of Leydig cell testosterone production. Thus, it can be considered similar to a mild hCG treatment. No side effects of GnRH treatment of undescended testes have been reported.

### **Surgery (Orchidopexy)**

Open surgery is the most direct way of transferring an abdominal or inguinal testis into the scrotum. This method has been used for a very long time. The major complications are surgical: In addition to the small risk of general anesthesia, the blood supply to the testis might be damaged during surgery. A high incidence of testicular atrophy after surgical orchidopexy has been reported. The latter complications may occur in departments in which “scrotal and inguinal surgery” (including orchidopexy and hernia repair) is considered to be an activity for surgeons in training. In skilled hands, this complication should be rare if the testis is not situated very high in the abdomen. Surgery has the added advantage that it provides opportunities for direct observation of the testis and, in selected cases, biopsy for microscopical investigations.

### **Comparison between Hormonal and Surgical Treatment**

The main advantages of hormonal treatment are that it is inexpensive and hospitalization and the trauma of anesthesia and surgery are avoided (when successful). This method also provides an opportunity to test the capacity of the testicles to produce testosterone if blood samples are taken at the end of hormone administration.

The major reason for treating undescended testicles is to save future fertility. Unfortunately, there are no published studies in which patients have been randomized to surgical or hormonal treatment and followed up into adulthood. Therefore, a scientifically based conclusion regarding the advantages of the two approaches in terms of future fertility cannot be made. The possible adverse effects of hCG on future spermatogenesis and the poor efficacy of hCG or GnRH argue against this as the first-line therapy; instead, a primary surgical approach is favored, provided that a skilled surgeon is available.

## AGE AT WHICH UNDESCENDED TESTICLES SHOULD BE TREATED

Again, the scientific foundations for recommendations are weak. The single classical report by the Swiss pathologist Hedinger is generally referred to as an argument for early treatment. He compared microscopical sections from undescended and normally descended testicles of boys who had died suddenly. The number of spermatogonia per seminiferous tubular cross section was determined. He noted that in normally descended testicles, there was a steady decline in the number of spermatogonia from birth until 1 or 2 years of age, followed by a slow increase that continued until puberty. In the undescended testes, the initial decline was similar, but the normal increase in spermatogonia after age 2 years did not occur.

The Hedinger study seems to show that undescended testes should be brought into the scrotum before the age of 2 years. It is assumed that spermatogenesis will be improved after successful treatment. Repeat biopsies have been studied in only two publications: One showed improved spermatogenesis, one did not. It is not known whether the catch-up of spermatogenesis is more complete after early rather than after later treatment. Nevertheless, treatment before age 2 has become the standard of treatment in many countries. If this approach is followed, initial surgery is to be preferred rather than hormonal treatment since the latter seems to be even less effective at ages younger than 4 years.

Final recommendations about the most suitable age for treatment of undescended testes must await results of controlled, randomized studies of treatment at various ages. The final outcome to be evaluated in such studies should be spermatogenesis and fertility in adulthood. This requires very long-term prospective studies. Careful measurements of testicular size during the first years after treatment may to some extent predict spermatogenesis and should be performed in future studies.

## CONCLUSIONS

The condition of undescended testicles is the most common inborn developmental abnormality in boys,

affecting approximately 1% of the entire male population. The cause may be hormonal deficiencies or, in most cases, unknown. Untreated, undescended testicles always have severely impaired spermatogenesis, but testosterone production is generally acceptable. The risk of malignancy in undescended testicles is increased. If the testes are moved into the scrotum, future spermatogenesis will be improved but not to the full extent. Both the choice of treatment (hormonal or surgical) and the most appropriate age of treatment (1–4 years) are controversial. However, based on current knowledge, skilled surgery at an early age, rather than hormonal treatment, seems to be advisable.

## See Also the Following Articles

Agonadism, Male and Female • Androgen Biosynthesis and Gene Defects • Beckwith-Wiedemann Syndrome (BWS) • Delayed Puberty and Hypogonadism, Male • Endocrine Disruptors and Male Sexual Differentiation • Hypergonadotropic Hypogonadism • Hypothalamic Hypogonadism • Klinefelter's Syndrome • Newborn Ambiguous Genitalia Management • Prader-Willi Syndrome • Testes, Embryology of

## Further Reading

- Docimo, S. G. (1995). The results of surgical therapy for cryptorchidism: A literature review and analysis. *J. Urol.* **154**(3), 1148–1152.
- Dunkel, L., Taskinen, S., Hovatta, O., Tilly, J. L., and Wikstrom, S. (1997). Germ cell apoptosis after treatment of cryptorchidism with human chorionic gonadotropin is associated with impaired reproductive function in the adult. *J. Clin. Invest.* **100**(9), 2341–2346.
- Hutson, J. M., Hasthorpe, S., and Heyns, C. F. (1997). Anatomical and functional aspects of testicular descent and cryptorchidism. *Endocr. Rev.* **18**(2), 259–280.
- Pyorala, S., Huttunen, N. P., and Uhari, M. (1995). A review and meta-analysis of hormonal treatment of cryptorchidism. *J. Clin. Endocrinol. Metab.* **80**(9), 2795–2799.
- Thonneau, P. F., Candia, P., and Mieusset, R. (2003). Cryptorchidism: Incidence, risk factors, and potential role of environment; An update. *J. Androl.* **24**(2), 155–162.
- Toppiari, J., and Kaleva, M. (1999). Maldescendus testis. *Horm. Res.* **51**(6), 261–269.

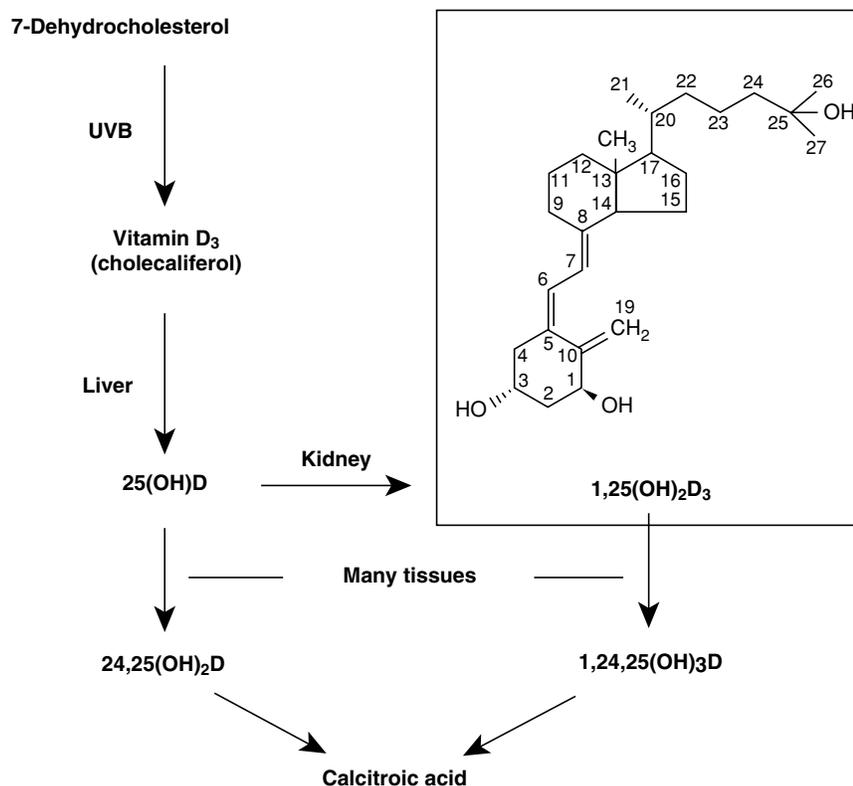


Figure 1 Structure of 1,25(OH)<sub>2</sub>D and the principal pathways for its production and metabolism.

D in some foods are widely different from those claimed on the packaging.

### Transport

Following its synthesis in the skin or absorption from the gastrointestinal tract, vitamin D circulates bound to vitamin D-binding protein. Vitamin D-binding protein is encoded by a gene located on chromosome 4 and is structurally related to  $\alpha$ -feto-protein and albumin. It circulates at a concentration of 4–8  $\mu$ M and there is one binding site for a vitamin D metabolite on each molecule. Thus, only approximately 5% of the binding sites are usually occupied. The affinities of the vitamin D metabolites for vitamin D-binding protein are in the following order: 25(OH)D<sub>3</sub> > 24,25(OH)<sub>2</sub>D<sub>3</sub> > vitamin D<sub>3</sub>  $\approx$  vitamin D<sub>2</sub> > 1,25(OH)<sub>2</sub>D<sub>3</sub> > 1,24,25(OH)<sub>3</sub>D<sub>3</sub>.

Most circulating vitamin D-binding protein is synthesized in the liver. Its circulating concentrations are increased by pregnancy and estrogen therapy and are decreased under hypoproteinemic conditions (e.g., liver disease, malnutrition, nephrotic syndrome). The role of vitamin D-binding protein is probably to

prevent rapid fluctuations in the levels of the vitamin D metabolites. Vitamin D-binding protein deficiency in humans has not been identified but a knockout mouse does exist. This animal is more susceptible to hypocalcemia when fed a vitamin D-deficient diet than is a wild-type animal. Despite its size (58 kDa), vitamin D-binding protein appears to be filtered at the glomerulus and is then reabsorbed in the proximal tubule via the receptor protein, megalin, a member of the low-density lipoprotein receptor family. Mice deficient in megalin who survive long enough to develop a bone phenotype have severe rickets. Approximately 10% of vitamin D metabolites circulate bound to albumin. Once vitamin D metabolites have entered their target cell, their transport to the appropriate intracellular organelle may be facilitated by a different group of intracellular vitamin D-binding proteins, related to the heat shock proteins.

### Activation

The first step in the activation of vitamin D is its 25-hydroxylation, to form 25-hydroxyvitamin D [25(OH)D]. This takes place in the parenchymal

cells of the liver and is mediated by the enzyme from the cytochrome P450 family, coded for by the gene CYP27. This enzyme was originally characterized as providing 27-hydroxylation of cholesterol in the synthesis of bile acids. It is likely that other enzymes are also involved in 25-hydroxylation. There appears to be very little regulation of the 25-hydroxylase and only in very advanced liver failure does its activity become limiting to normal vitamin D metabolism.

25(OH)D formed in the liver returns to the circulation bound to vitamin D-binding protein. If it is to be further hydroxylated, this takes place in the proximal tubular cells of the kidney. The entry of 25(OH)D into these cells is dependent on the action of megalin. Further activation of 25(OH)D is accomplished by 25-hydroxyvitamin D 1 $\alpha$ -hydroxylase, which is also a mitochondrial P450 enzyme encoded by the CYP1 $\alpha$  gene. It shows significant homology to the 25-hydroxylase. The proximal tubular cells that contain the 1 $\alpha$ -hydroxylase have parathyroid hormone (PTH) receptors, which stimulate enzyme activity. This accounts for the increased levels of 1,25-dihydroxyvitamin D [1,25(OH)<sub>2</sub>D] in states of hyperparathyroidism. The 1 $\alpha$ -hydroxylase is also expressed in distal tubular cells and, again, megalin is present in these cells. In contrast to the proximal tubule, the PTH receptor is not present, but the receptor for calcitonin is and appears to have a stimulatory effect on 1 $\alpha$ -hydroxylase activity. 1,25(OH)<sub>2</sub>D directly inhibits the expression of the CYP1 $\alpha$  gene at both sites in the renal tubule.

The kidney is clearly the most important site of 1,25(OH)<sub>2</sub>D synthesis since anephric individuals have dramatically reduced concentrations of the hormone. However, 1,25(OH)<sub>2</sub>D is still measurable in such patients, suggesting that there is extrarenal synthesis of the hormone. This 1 $\alpha$ -hydroxylase gene is expressed in skin but even patients who are deficient in the gene still have measurable levels of 1,25(OH)<sub>2</sub>D, suggesting that other enzymes might mediate 1-hydroxylation of 25(OH)D. Extrarenal synthesis of 1,25(OH)<sub>2</sub>D clearly occurs in granulomatous diseases and in some lymphomas, where unregulated production of the hormone results in hypercalcemia. During pregnancy, the placenta produces 1,25(OH)<sub>2</sub>D.

An alternative fate for 25(OH)D is to be hydroxylated on C-24, to produce 24,25-dihydroxyvitamin D. This conversion is mediated by vitamin D 24-hydroxylase, encoded by the gene CYP24. This enzyme can also 24-hydroxylate 1,25(OH)<sub>2</sub>D and is regarded as the principal pathway for deactivating vitamin D metabolites. As a consequence, it is widely expressed throughout the body, particularly in vitamin

D target tissues, including the proximal and distal tubules of the kidney. Vitamin D metabolites then undergo further hydroxylations and side chain cleavage, resulting in the production of water-soluble calcitroic acid, which has no biological activity.

## REGULATION OF VITAMIN D METABOLISM

The key points of regulation of vitamin D metabolism are the 1 $\alpha$ -hydroxylase and the 24-hydroxylase (Table I). As noted above, the 1 $\alpha$ -hydroxylase gene is directly regulated by PTH, calcitonin, and 1,25(OH)<sub>2</sub>D, the latter two hormones having the predominant effect under normocalcemic conditions, whereas the effect of PTH predominates over that of calcitonin under hypocalcemic conditions. Phosphate concentration is also an important regulator of 1 $\alpha$ -hydroxylase activity, which is suppressed by high concentrations of phosphate and stimulated by low concentrations. Calcium concentration impacts on 1 $\alpha$ -hydroxylase activity, mainly through its effect on circulating levels of PTH, though there is some evidence that it can directly influence enzyme activity as well. There is also evidence for effects of other growth-related hormones, such as estrogen, prolactin, and growth hormone, but these appear to be relatively minor.

1,25(OH)<sub>2</sub>D is the principal regulator of 24-hydroxylase activity. This is a genomic effect mediated by vitamin D-response elements in the promoter for the 24-hydroxylase gene. In the renal proximal tubule, PTH inhibits 24-hydroxylase activity, though it may have an opposite effect in the distal tubule. The induction of 24-hydroxylase by 1,25(OH)<sub>2</sub>D also leads to accelerated catabolism of 25(OH)D. This explains the low levels of this metabolite seen in patients treated with calcitriol and in those suffering from primary hyperparathyroidism.

Table I Physiological Regulators of the Principal Enzymes in Vitamin D Metabolism

1 $\alpha$ -Hydroxylase	24-Hydroxylase
PTH $\uparrow$	1,25(OH) <sub>2</sub> D $\uparrow$
Phosphate $\downarrow$	PTH $\downarrow$
1,25(OH) <sub>2</sub> D $\downarrow$	
Calcitonin $\uparrow$	

Note. Arrows indicate the effect of the regulatory factors on enzyme activity.

## ACTIONS OF VITAMIN D AND ITS METABOLITES

### Mechanisms of Action

The classic mechanism of action of vitamin D metabolites on their target tissues is similar to that of other steroid hormones and is mediated by the vitamin D receptor (VDR), a member of the steroid hormone receptor superfamily. Both  $1,25(\text{OH})_2\text{D}$  and  $25(\text{OH})\text{D}$  bind to this receptor, though the affinity of  $1,25(\text{OH})_2\text{D}$  for the receptor is 1000-fold higher. However, it should be remembered that  $25(\text{OH})\text{D}$  circulates in concentrations 1000-fold higher than those of  $1,25(\text{OH})_2\text{D}$ , so both may contribute to receptor activation. The vitamin D metabolites bind to the VDR, which then forms a heterodimeric complex with the retinoic acid X receptor. This complex then interacts with vitamin D-response elements associated with a variety of genes (see Fig. 2). The VDR gene contains nine exons and encodes a protein of 427 amino acids. In humans, the gene is located on chromosome 12q. The DNA-binding domain is located in the amino-terminal region and consists of a double zinc-finger structure. The ligand-binding domain is in the C-terminal region of the molecule. Mutations of the VDR have been identified in humans and result in vitamin D resistance, producing the clinical picture of rickets. Polymorphisms of the VDR gene have also been observed. These variations in nucleic acid sequence do not result in differences in the amino acid sequence of the expressed protein, but may impact on the transcription or stability of the mRNA for the VDR. Despite extensive clinical studies, it remains unclear whether these polymorphisms are consistently associated with changes in calcium metabolism or bone mass.

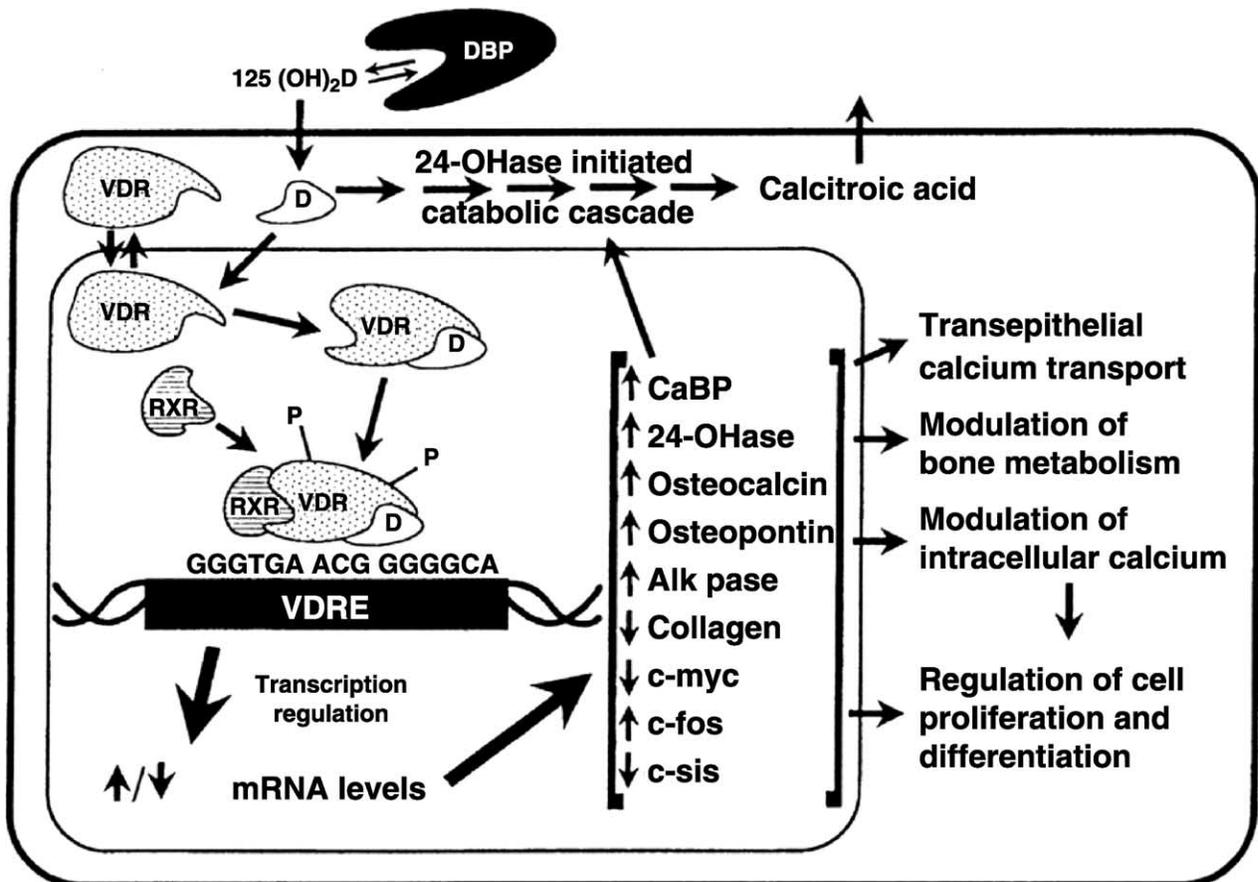
In addition to this classic mechanism of action of vitamin D metabolites, rapid responses, which are thought to occur too rapidly to be accounted for by a genomic mechanism of action, have been observed. These rapid responses include changes in intracellular calcium and transcaltachia, which is a rapid transport of calcium from the intestinal lumen to the circulation. The structure-activity relationships of vitamin D metabolites in stimulating these rapid effects are different from that for the classic genomic effects of vitamin D and some of these effects have been demonstrated in cells lacking the VDR. These observations have raised the possibility of a different vitamin D receptor being present on the cell membrane.

### Calcium Metabolism

$1,25(\text{OH})_2\text{D}$  is one of the classic calcitropic hormones. In humans and animals lacking the vitamin D receptor, the only abnormality observed at birth is in hair follicle development. Subsequently, however, hypocalcemia and the development of rickets are observed. The abnormalities in bone development can be prevented by restoration of normal serum calcium concentration, suggesting that maintenance of normocalcemia and maintenance of normophosphatemia are the principal biological roles of the vitamin D metabolites. It is likely that this is achieved primarily by increasing the absorption of calcium in the small intestine. Vitamin D receptors are found throughout the small intestine but are most abundant in the duodenum. Administration of calcitriol to animals or humans leads to increases in intestinal calcium absorption, though the precise mechanism by which this occurs is not fully understood.  $1,25(\text{OH})_2\text{D}$  regulates the production of a calcium-binding protein, calbindin D8k, within the intestinal epithelial cell. This is mediated by a vitamin D-response element in the gene for calbindin D9k. It is thought that this calcium-binding protein facilitates the transport of calcium through the intestinal epithelial cells.  $1,25(\text{OH})_2\text{D}$  may also facilitate the entry of calcium from the intestinal lumen into the epithelial cell, possibly through a channel.  $1,25(\text{OH})_2\text{D}$  also stimulates the activity of an ATP-dependent calcium pump at the basolateral membrane.  $1,25(\text{OH})_2\text{D}$  also stimulates intestinal phosphate absorption, though this mostly takes place in the distal small bowel.

The direct effects of  $1,25(\text{OH})_2\text{D}$  on bone work in concert with those already described in the intestine, to maintain or increase serum calcium concentrations.  $1,25(\text{OH})_2\text{D}$  acts on osteoblasts and their precursors, causing the production of RANK-L, which binds to preosteoclasts to stimulate their development into osteoclasts. This leads to an increase in osteoclastic bone resorption.  $1,25(\text{OH})_2\text{D}$  also directly stimulates alkaline phosphatase activity and the production of osteopontin and osteocalcin by osteoblasts. Although it has been suggested that  $1,25(\text{OH})_2\text{D}$  may directly influence skeletal mineralization, the majority of evidence suggests that this occurs indirectly, as a result of vitamin D effects on serum calcium and phosphate concentrations.

In the kidney, it is clearly established that  $1,25(\text{OH})_2\text{D}$  induces its own catabolism by the 24-hydroxylase and that it inhibits its own production.



**Figure 2** Mechanism of action of  $1,25(\text{OH})_2\text{D}$  in target cells. Free  $1,25(\text{OH})_2\text{D}$  enters cells and interacts with the vitamin D receptor (VDR), which becomes phosphorylated. A complex is formed with the retinoic acid X receptor (RXR), which then interacts with the vitamin D-response elements (VDRE) of a number of genes regulated by this hormone. Reprinted from Holick, M. F. (1999). *Vitamin D: Photobiology, metabolism, mechanism of action and clinical applications*. In "Primer on the Metabolic Bone Diseases and Disorders of Mineral Metabolism" (M. J. Favus, ed.), 4th ed., pp. 92–98. American Society of Bone and Mineral Research, Philadelphia, with permission.

It has been suggested that it has effects on calcium and phosphate reabsorption in the renal tubule but this remains uncertain.

$1,25(\text{OH})_2\text{D}$  also acts on the parathyroid glands, where its binding to the vitamin D receptor directly decreases expression of the PTH gene. This genomic effect is reinforced by the action of  $1,25(\text{OH})_2\text{D}$  to increase serum calcium and thus diminish PTH secretion. The direct effects of vitamin D metabolites on PTH secretion are particularly important in the management of the secondary hyperparathyroidism of chronic renal failure. They may also be important in the prevention and treatment of primary hyperparathyroidism, since this condition may be exacerbated by the low levels of  $25(\text{OH})\text{D}$  with which it is frequently associated.

### Other Tissues

$1,25(\text{OH})_2\text{D}$  receptors are expressed in a wide variety of tissues other than those classically involved in mineral and bone homeostasis. In keratinocytes and in some white blood cells or their precursors,  $1,25(\text{OH})_2\text{D}$  has anti-proliferative and pro-differentiation effects. Thus, in psoriasis, it appears that the hyperproliferative state of skin cells can be controlled with the use of vitamin D analogues. There is also experimental evidence that some leukemic cell lines show similar responses to vitamin D metabolites. As a result, there has been an interest in using  $1,25(\text{OH})_2\text{D}$  for treating leukemias and other tumors, such as those of the breast, colon, and prostate. Despite the existence of promising preliminary results

in these areas for a number of years, the vitamin D metabolites are not yet established as having a therapeutic role in any of these malignant conditions. Vitamin D metabolites may also have direct actions on muscle cells, on adipocytes, in immune regulation, and on endocrine tissues, such as the pancreatic beta cell.

## ASSESSMENT OF VITAMIN D STATUS

25(OH)D is the principal circulating vitamin D metabolite and it is the entity that should be assessed when determining an individual's vitamin D status. Because ingested or endogenously produced calciferols are converted to 25(OH)D with very little regulation, serum levels of this metabolite accurately reflect both excess and deficiency states. As noted above, 25(OH)D circulates bound to vitamin D-binding protein, so any condition associated with hypoproteinemia may produce falsely depressed levels of 25(OH)D. Caution is required in interpreting measurements of vitamin D metabolites in this situation, because it is uncommon to measure vitamin D-binding protein in routine clinical practice.

When assessing vitamin D status, it is important to consider whether the "normal" range—which varies with latitude—can be regarded as optimal. This has been addressed by Malabanan *et al.*, who demonstrated that vitamin D supplementation suppressed parathyroid hormone levels only in subjects whose baseline serum 25-hydroxyvitamin D was less than 50 nmol/liter. This suggests that 50 nmol/liter is a more appropriate target concentration for serum 25-hydroxyvitamin D, though some cross-sectional studies suggest that it may be as high as 100 nmol/liter.

Measurements of serum 1,25(OH)<sub>2</sub>D are also widely available. The value of measurement of 1,25(OH)<sub>2</sub>D in clinical medicine is relatively small. It is sometimes helpful in elucidating the cause of hypercalcemia, being high when hypercalcemia is a direct consequence of overproduction of 1,25(OH)<sub>2</sub>D, as in some granulomatous conditions and lymphomas. The widespread impression that 1,25(OH)<sub>2</sub>D is the only biologically active vitamin D metabolite frequently leads to the inappropriate use of vitamin D assays. This belief results in the expectation that levels of 1,25(OH)<sub>2</sub>D are all that will influence calcium metabolism, whereas this is clearly not the case. As vitamin D deficiency develops, serum 25(OH)D declines. In response to this decline, secondary hyperparathyroidism develops and with it there are increases in serum 1,25(OH)<sub>2</sub>D. Thus, the paradoxical situation can develop, wherein patients who have clinical and

histological evidence of osteomalacia have a serum 1,25(OH)<sub>2</sub>D level that is either normal or supranormal.

In summary, measurement of serum 25(OH)D suffices for the detection of vitamin D deficiency or vitamin D intoxication and measurement of 1,25(OH)<sub>2</sub>D is usually indicated only in the diagnosis of difficult cases of hypercalcemia.

## THERAPEUTIC USE OF VITAMIN D AND ITS METABOLITES

### Vitamin D Deficiency and Osteoporosis

It is becoming increasingly recognized that vitamin D deficiency is common in the elderly, particularly those who are no longer fully independent and therefore less exposed to sunlight. The problem is often greater at higher latitudes, though it can also occur in very hot climates where the sun is often avoided because of the heat. Vitamin D deficiency leads to secondary hyperparathyroidism and a resulting increase in bone loss.

Physiological supplements of calciferol (e.g., 400 IU/day) reduce parathyroid hormone concentrations in elderly subjects and lead to increases in bone density, particularly at the femoral neck. Similar changes in biochemical end-points can be achieved with regular sunlight exposure for 15–30 min daily. Two large studies have assessed the effect on fracture rates of calciferol supplementation alone. Lips *et al.* showed no difference in fracture incidence in 2578 Dutch men and women over the age of 70 years randomized to calciferol 400 IU/day or placebo, whereas Heikinheimo *et al.* showed that 150,000 IU of vitamin D annually reduced symptomatic fracture rates by 25% in a cohort of 800 elderly subjects in Finland. Two other major studies have been reported in which calcium was coadministered with calciferol to elderly subjects. Chapuy *et al.* demonstrated a reduction of more than 25% in nonvertebral and hip fracture rates in a cohort of 3000 elderly women studied over a period of 3 years. Dawson-Hughes *et al.* demonstrated a reduction of nonvertebral fracture rates by more than 50% in 400 older men and women randomized to calcium 500 mg/day plus 700 IU vitamin D or to placebo. It is not possible to determine whether the calcium, the vitamin D, or the combination of the two was the component essential to the success of these two studies, but the findings do point to the possibility of a major reduction in morbidity in elderly patients using a safe and inexpensive intervention. Vitamin D supplementation seems to produce no benefit in early postmenopausal women

or those who are already vitamin D replete, as defined above.

The dose–response relationship between calciferol intake and serum 25(OH)D levels is relatively flat up to intakes of several thousand international units per day. Thus, this intervention is safe and doses of 400–1200 IU/day are routinely used. Calciferol is stored in adipose tissue and has a half-life of many weeks. Therefore, it can be administered in larger doses less frequently. Doses of at least 500,000 IU can be given at once to correct deficiency and the authors routinely maintain normal vitamin D status with monthly oral supplements of 50,000 IU. There are anecdotal reports of wide variability in the bioavailability of different preparations of calciferol and dietary vitamin D intakes also vary widely from country to country. Therefore, each region needs to determine what regimen of replacement is safe and effective in that environment.

The use of vitamin D as a physiological supplement is fundamentally different from the use of high doses of calciferol or 1 $\alpha$ -hydroxylated vitamin D metabolites (e.g., alfacalcidol, calcitriol) to pharmacologically manipulate intestinal calcium absorption. Both of these strategies bypass the normal homeostatic controls of vitamin D metabolism and therefore carry a significant risk of hypercalcemia and hypercalciuria. The use of pharmacological doses of calciferol has not been demonstrated to confer any beneficial effects on bone density. Trials of the use of alfacalcidol or calcitriol in the prevention and treatment of osteoporosis have shown mixed results, including both significant increases and significant decreases in bone density and fracture rates. Adverse outcomes have been most common in studies of osteoporosis prevention in women and in studies in males. This variability in results may be attributable to a different balance of the potentially beneficial effects on intestinal calcium absorption versus the potentially damaging effects on osteoclast recruitment in the different populations studied. As a result, vitamin D metabolites are not regarded as a first-line therapy for osteoporosis.

In conclusion, suboptimal vitamin D status is very common in the elderly, mainly because of reduced sunlight exposure. The provision of a daily intake of 400–800 IU is a straightforward, safe, and inexpensive means of prevention and appears to produce substantial reductions in fracture rates.

### Other Conditions

Failure of the 1 $\alpha$ -hydroxylation of 25(OH)D is probably the single most important contributor to the

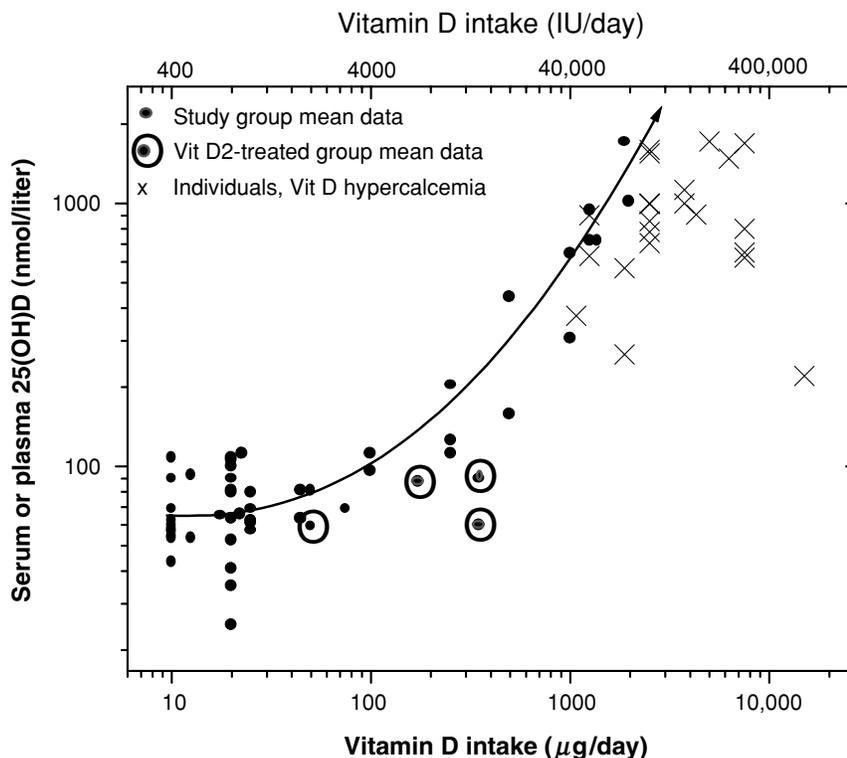
hypocalcemia of renal failure and the resulting development of renal bone disease. Thus, the availability of 1,25(OH)<sub>2</sub>D (calcitriol) and 1 $\alpha$ -hydroxyvitamin D (alfacalcidol) as pharmaceuticals has revolutionized the management of calcium metabolism in renal failure. The development of hypercalcemia can be dose-limiting when using vitamin D metabolites to reverse secondary hyperparathyroidism in renal failure. This has led to the development of synthetic vitamin D analogues that suppress parathyroid hormone secretion but have less effect on serum calcium.

The 1 $\alpha$ -hydroxyvitamin D metabolites have also revolutionized the management of other conditions associated with either hypocalcemia (e.g., hypoparathyroidism) or hypophosphatemia (e.g., X-linked-hypophosphatemia, oncogenic osteomalacia). Previously, these conditions were managed with very large doses of calciferol. Because of its long half-life, any overdosage resulted in sustained hypercalcemia, often leading to renal failure.

As noted above, the anti-proliferative actions of the vitamin D metabolites are being used in the management of psoriasis and experimental work is continuing regarding their use in some cancers. There is also epidemiological evidence suggesting that high vitamin D levels are associated with a lower incidence of cardiovascular disease. This work is potentially subject to a number of biases (e.g., people who are more physically active and therefore at lower risk of cardiovascular disease spend more time outdoors and therefore have better vitamin D status). Intervention studies will be necessary before the significance of these findings can be determined and vitamin D supplementation is not a standard part of cardiovascular disease prevention at the present time.

### SAFETY OF VITAMIN D AND ITS METABOLITES

The use of replacement doses of calciferol is a very safe intervention, as would be expected since the intention is to restore circulating levels of 25(OH)D to the levels that are present in the ambulant population. Thus, in the study by Chapuy in which more than 1600 women were treated with calciferol, the only individual who developed hypercalcemia was subsequently found to have primary hyperparathyroidism. A similar zero-incidence of significant hypercalcemia has been reported by other investigators using low-dose regimens of calciferol administration.



**Figure 3** Dose–response relationship between daily vitamin D intake and mean 25(OH)D concentration. The solid points show mean results for groups of adults consuming the indicated doses of vitamin D, usually vitamin D<sub>3</sub>. Results for groups of adults that were definitely consuming vitamin D<sub>2</sub> are shown by the circled points. The results represented by × are for individuals who had hypercalcemia as a result of vitamin D consumption. The toxic intakes were of vitamin D<sub>2</sub>, not D<sub>3</sub>. Reprinted from Vieth, R. (2001). In “Nutritional Aspects of Osteoporosis” (P. Burckhardt, R. Heaney, and B. Dawson-Hughes, eds.), pp. 173–195. Academic Press, San Diego, CA, with permission.

In contrast, with the use of pharmacological doses of calciferol, there are case reports of severe hypercalcemia, often of long duration and sometimes associated with renal failure. Vieth reviewed this issue and concluded that the relationship of vitamin D dose to serum 25(OH)D concentration is relatively flat up to a daily calciferol intake of 10,000 IU (Fig. 3). It is likely that doses of up to 10,000 IU/day are safe in individuals who have no conditions that predispose them to hypercalcemia (e.g., primary hyperparathyroidism, sarcoidosis) and fully documented cases of toxicity have occurred only with intakes of 40,000 IU/day or more. Calcitriol and alfacalcidol frequently cause hypercalcemia and hypercalciuria, but their short half-lives result in this being readily corrected by dose adjustment. However, close monitoring is necessary when these drugs are being introduced or their doses increased. Changes in calcium intake can also affect serum calcium concentrations in patients using these compounds.

The hypercalcemia of vitamin D intoxication has usually been attributed to increased intestinal calcium absorption, but it has been demonstrated that bisphosphonates are effective in treating some cases, indicating that increased bone resorption may also contribute. Hyperresorption of bone has certainly been demonstrated in animal models of hypervitaminosis D. The long duration of hypercalcemia associated with calciferol intoxication means that this compound is substantially less safe at high doses than its more active and shorter half-life metabolites, such as alfacalcidol and calcitriol.

### See Also the Following Articles

Bone Remodeling, Dynamics of • Hypercalcemia and Hypercalcemia Treatment • Hypercalciuria • Hyperparathyroidism, Primary • Hyperphosphatemia • Hypocalcemia, Therapy • Osteoporosis in Older Men • Osteoporosis in Older Women • Osteoporosis, Overview • Parathyroid

Hormone (PTH) • Renal Osteodystrophy • Vitamin D Deficiency, Rickets, and Osteomalacia

### Further Reading

- Adams, J. S., and Hollis, B. W. (2002). Vitamin D: Synthesis, metabolism and clinical measurement. *In* "Disorders of Bone and Mineral Metabolism" (F. L. Coe and M. J. Favus, eds.), 2nd ed., pp. 157–174. Lippincott, Philadelphia, PA.
- Feldman, D., Malloy, P. J., and Gross, C. (2001). Vitamin D. *In* "Osteoporosis" (R. Marcus, D. Feldman, and J. Kelsey, eds.), 2nd ed., pp. 257–303. Academic Press, San Diego, CA.
- Lips, P. (2001). Vitamin D deficiency and secondary hyperparathyroidism in the elderly: Consequences for bone loss and fractures and therapeutic implications. *Endocr. Rev.* **22**, 477–501.
- Reid, I. R. (2001). Vitamin D and its metabolites in the management of osteoporosis. *In* "Osteoporosis" (R. Marcus, D. Feldman, and J. Kelsey, eds.), 2nd ed., pp. 257–303. Academic Press, San Diego, CA.

failure or delay of mineralization in the growth plates and joints as well as in other skeletal sites.

## VITAMIN D DEFICIENCY

Calcium is absorbed in the upper small intestine under the control of the active vitamin D hormone, 1,25-dihydroxyvitamin D<sub>3</sub> [1,25(OH)<sub>2</sub>D<sub>3</sub>], also known as calcitriol. In this article, we use these terms interchangeably. Vitamin D is synthesized in the skin with the aid of sunlight or it is ingested in the diet. The distribution of vitamin D in food is quite restricted, being substantial mainly in oily fish. Because dietary vitamin D is so limited, some countries have fortified various foods (milk, cereals, orange juice, etc.) in order to avoid widespread vitamin D deficiency. Even in these countries, nutritional sources of vitamin D may be limited and exposure to sunlight is critical for adequate endogenous production of vitamin D.

Vitamin D deficiency is characterized by undermineralized bone and excess osteoid. Normally, most osteoid is rapidly mineralized and a small rim of unmineralized matrix can be found on bone trabeculae. However, in rickets or osteomalacia, the level of unmineralized osteoid is increased. Initially, osteoid surface and volume are increased, but osteoid thickness and lag times (delay in mineralizing osteoid) are normal. In more severe cases, osteoid thickness exceeds 15 μm and mineralization lag time exceeds 100 days. In the most florid cases, no mineralization may be detected. The strength of unmineralized osteoid is greatly diminished compared to that of normally mineralized bone, leading to metabolic bone disease and increased susceptibility to fracture. In the case of childhood rickets, a number of additional skeletal abnormalities are found, including bowing of the long bones, joint pain and swelling, bone pain, and abnormalities of the teeth. In severe vitamin D insufficiency, infants may show delayed development and even respiratory failure because breathing is impaired due to involvement of the thoracic cage with rickets.

Any step in the pathway of calcitriol synthesis or action that is defective can result in rickets/osteomalacia. The usual causes of rickets/osteomalacia are listed in Table I and the pathway involved is shown in Fig. 1. The most common cause of rickets/osteomalacia worldwide is vitamin D deficiency due to inadequate vitamin D ingestion combined with insufficient synthesis in the skin. This can be caused by inadequate amounts of vitamin D in the diet and/or by malabsorption of vitamin D in the gastrointestinal tract in association with inadequate sunlight

**Table I Etiology of Rickets or Osteomalacia**

Nutritional: vitamin D deficiency
Deficient synthesis due to inadequate sunlight exposure
Dietary deficiency of vitamin D and/or calcium
Renal insufficiency
Inadequate 1α-hydroxylase activity
Gastrointestinal disorders
Malabsorption syndrome
Hepatobiliary disease
Pancreatic disease
Tumor-induced osteomalacia
Excess production of phosphatonin
Hereditary causes
X-linked hypophosphatemic rickets
1α-Hydroxylase deficiency (vitamin D-dependent rickets type I)
Hereditary vitamin D-resistant rickets (vitamin D-dependent rickets type II)
Autosomal dominant hypophosphatemic rickets
Miscellaneous causes
Acidosis
Phosphate depletion
Renal tubular disorders

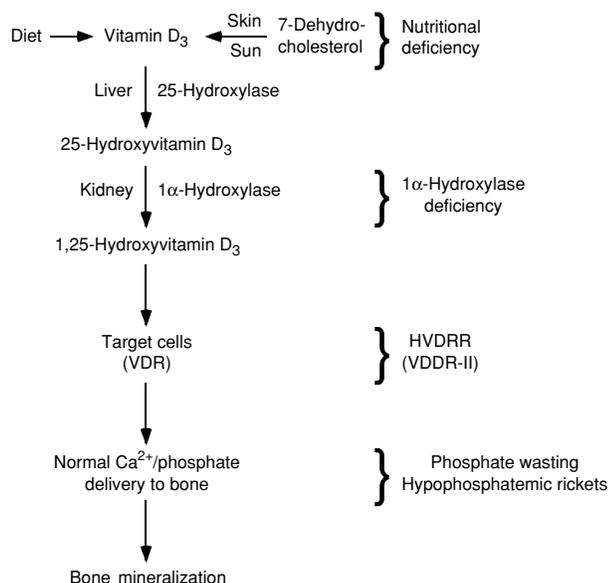
exposure. Since vitamin D supplementation of milk or other dietary products reduces nutritional causes of rickets/osteomalacia in the United States and some other developed countries, other etiologies are becoming more important in the developed world. However, avoidance of dairy products and the use of sunscreens or the wearing of traditional clothing that prevents sunlight exposure contribute to vitamin D deficiency even in developed countries. Furthermore, the elderly or others confined to an indoors existence in nursing homes has led to an upsurge in vitamin D deficiency.

Other causes of rickets/osteomalacia include the inability to synthesize calcitriol due to renal failure or a genetic defect in 1α-hydroxylase, the key enzyme involved in calcitriol synthesis. Abnormalities that cause excess loss of phosphate include genetic defects or tumors that produce phosphaturia. Finally, mutations in the vitamin D receptor (VDR), the protein that mediates calcitriol actions in the intestine and other target tissues, can cause rickets in children. Each of these entities is described later.

## MECHANISM OF VITAMIN D ACTION

### Vitamin D Receptor

Although called a vitamin, 1,25-dihydroxyvitamin D (calcitriol) is actually a member of the steroid



**Figure 1** Pathways of calcitriol synthesis and action. The synthetic pathway of calcitriol (1,25-dihydroxyvitamin D<sub>3</sub>) is shown. Defects in the pathway that interfere with normal mineralization of bone and cause osteomalacia or rickets are indicated. From Feldman, D., and Malloy, P. J. (2003). "Encyclopedia of Hormones," Vol. 3, pp. 620–628. Reprinted with permission from Elsevier.

hormone family. Together with PTH, it regulates calcium metabolism and bone mineralization. Calcitriol also promotes other physiologic activities, including exerting prodifferentiation, antiproliferative, and immunosuppressive actions. Calcitriol acts through the VDR, a member of the steroid–thyroid–retinoid receptor gene superfamily, although some data suggest a rapid nongenomic action as well. The VDR acts as a ligand-activated transcription factor and regulates gene transcription. The VDR has a modular structure comprising an N-terminal DNA-binding domain (DBD) and a C-terminal ligand-binding domain (LBD). The DBD enables the VDR to interact with vitamin D response elements (VDREs), specific nucleotide sequences located in promoter regions of vitamin D-responsive genes. The VDR LBD specifically binds calcitriol and triggers a series of molecular events leading to the activation of target genes. Gene transactivation by the VDR requires its binding as a heterodimer with the retinoid X receptor (RXR). RXR heterodimerizes with a number of other receptors in the steroid–thyroid–retinoid receptor gene superfamily, including thyroid hormone receptor, retinoic acid receptor, and the peroxisome proliferator activating receptor. Coactivator proteins are also recruited to the VDR following ligand binding. These proteins act as bridging factors linking the

VDR–RXR heterodimer to the preinitiation complexes and RNA polymerase II. Coactivators include a family of closely related p160 proteins, such as SRC1/NcoA1, TIF2/GRIP1/NcoA2/SRC2, and pCIP/RAC3/ACTR/AIB1/SRC3. Other coactivators include a group of proteins collectively called vitamin D receptor-interacting proteins (DRIPs). These coactivator proteins associate with nuclear receptors in a ligand-dependent manner to enhance transactivation of target genes.

### Pathway of Calcitriol Action

How do cells respond to calcitriol? After the hormone is produced and secreted by the kidney, it is transported in the blood either bound to vitamin D-binding protein (DBP) or in the free state. The free hormone is lipid soluble and readily gains entry into cells by permeating the lipid bilayer of the cell membrane, although evidence of an active step in cell entry is currently being developed. Cells that respond to calcitriol have VDRs located in the cell nuclear compartment, where they are loosely associated with RXR. Once inside the cell, calcitriol encounters the VDR, where it becomes bound in the ligand-binding pocket within the LBD. The binding causes the VDR to bind more tightly to RXR and to interact with VDREs on target genes. When calcitriol occupies the ligand-binding pocket, helix H12 in the VDR LBD swings into a new position, locking the hormone inside the receptor. The repositioning of helix H12 forms a binding cleft for interaction with coactivators. The coactivators that are recruited by the VDR, such as SRC-1, are then able to remodel the chromatin so that assembly of the transcriptional apparatus can occur. The VDR can then recruit the DRIP complex and begin activating gene transcription.

## ACQUIRED CAUSES OF RICKETS AND OSTEOMALACIA

### Nutritional Deficiency

Vitamin D deficiency causes a sequence of metabolic abnormalities that combine to cause osteomalacia or rickets. The insufficiency of vitamin D, if severe, can result in diminished intestinal calcium absorption and hypocalcemia. The calcium-sensing receptor present in the parathyroid gland detects the low serum calcium level. This leads to an increase in PTH synthesis and results in secondary hyperparathyroidism. The elevated levels of PTH attempt to correct the hypocalcemia by three actions

on bone and kidney. One action aimed at correcting the calcitriol deficiency is the stimulation of renal  $1\alpha$ -hydroxylase activity to cause increased renal production of calcitriol. However, this step requires adequate vitamin D and 25-hydroxyvitamin D substrate, which is absent in vitamin D deficiency. Therefore, this compensation fails to correct the deficiency. A second PTH action on bone is to increase bone resorption to raise calcium and phosphate flux from bone to serum. In other words, bone is sacrificed in an attempt to maintain normal serum calcium, thus contributing to the metabolic bone disease caused by calcitriol deficiency. A third PTH action is to increase renal calcium reabsorption and decrease calciuria. However, while decreasing calciuria, PTH increases phosphaturia, thus reducing the availability of this mineral component for bone mineralization. The overall result of vitamin D deficiency is hypocalcemia, secondary hyperparathyroidism, and defective mineralization of osteoid leading to the development of osteomalacia.

It is clear that adequate calcium and vitamin D are required for bone health. The vitamin D requirement is estimated to be a minimum of 400 IU per day, although elderly subjects may require 600 IU or more. Most infants and children in the United States have adequate vitamin D intake because of vitamin D supplementation of milk. During adolescence, when consumption of milk and dairy products diminishes, decreased dietary vitamin D may adversely affect calcium absorption and thereby lead to impaired skeletal health. Vitamin D insufficiency is greatly increased in the elderly population, which is at risk for both osteoporosis and osteomalacia. This problem is worse in countries that do not supplement milk or other foods with vitamin D and where sunlight exposure may be inadequate.

Vitamin D is not present in substantial levels in many foodstuffs except for oily fish. In the United States, milk is ostensibly fortified with 400 IU of vitamin D per quart, although actual amounts vary. Ultraviolet (UV) radiation from sunlight is necessary for endogenous production of vitamin D in the skin. Since dietary vitamin D is so limited, sunlight is essential for most individuals to achieve normal vitamin D status. The solar irradiation causes photolysis of the precursor 7-dehydrocholesterol in the skin and allows for its conversion to vitamin D. However, many individuals have reduced intake of milk or dairy products and also do not have adequate sun exposure, leading to vitamin D insufficiency. This problem is exacerbated by several modern trends, including avoidance of dairy products to control weight or elevated

cholesterol levels and avoidance of sunlight to reduce the risk of skin cancer. Also, many elderly individuals have limited mobility and do not have the opportunity for adequate sunlight exposure. These circumstances have led to a substantial increase in the frequency of osteomalacic bone disease and fractures in the elderly, especially among the nursing home population. Some populations wear traditional dress covering most of the body and preventing UV rays from reaching the skin. During winter, reduced sunlight further diminishes vitamin D synthesis, especially in the northern latitudes, and worsens the tendency toward vitamin D insufficiency. Races with dark skin have additional difficulty obtaining adequate vitamin D synthesis because the melanin in the skin is a natural sunscreen that reduces the penetration of UV rays to the layers of the dermis where vitamin D is formed.

Deficient dietary calcium intake may also cause osteomalacia or exacerbate the problem of vitamin D insufficiency. There exists a spectrum of etiologies ranging from pure vitamin D deficiency with normal calcium intake to adequate vitamin D sufficiency with inadequate calcium intake. Many cases worldwide are due to vitamin D insufficiency combined with relative calcium insufficiency. Certain dietary habits, especially common in selected populations, may exacerbate vitamin D insufficiency by inhibiting vitamin D absorption or increasing the metabolic clearance of 25-hydroxyvitamin D, including the consumption of chapattis, an East Asian bread made from wheat flour with high phytate levels that impairs calcium and vitamin D absorption. Increased risk of osteomalacia is associated with a number of factors, including inadequate sunlight exposure, living at high altitude, vegetarianism, calcium-deficient diets, and high-fiber diets. A combination of inadequate dietary vitamin D and lack of UV light exposure and exacerbated by other factors determines the risk and severity of the osteomalacic bone disease.

### Renal Osteodystrophy

Renal failure is associated with complex abnormalities of vitamin D, secondary hyperparathyroidism, and calcium metabolism. Phosphate retention due to inadequate kidney function causes hypocalcemia by complexing calcium and inhibiting renal  $1\alpha$ -hydroxylase activity and therefore causes diminished calcitriol synthesis. Also, as kidneys shrink and renal functional tissue declines,  $1\alpha$ -hydroxylase activity is further diminished, leading to deficiency of calcitriol production. These changes combine to

cause secondary hyperparathyroidism, compounding the bone abnormality. The constellation of hypocalcemia, secondary hyperparathyroidism, and calcitriol deficiency causes osteomalacia and renal osteodystrophy. Recognition of this sequence of events has led to important changes in the medical management of renal failure in an attempt to prevent the development of renal osteodystrophy. Two major therapeutic measures are (i) the use of phosphate binders to minimize and/or prevent the development of phosphate retention with elevated phosphate concentration and its consequent hypocalcemia and (ii) supplementation with calcitriol to avoid deficiency of active vitamin D.

### Gastrointestinal Problems and Malabsorption of Vitamin D

In areas of the world that do not fortify foodstuffs with vitamin D, diet is estimated to contribute only one-fourth to one-third of the daily requirement of vitamin D. Therefore, gastrointestinal malabsorption may worsen vitamin D insufficiency but often is not the sole cause of osteomalacia. Some gastrointestinal problems are also associated with increased metabolic clearance of vitamin D metabolites, compounding the malabsorption problem. Gastrointestinal diseases associated with osteomalacia include celiac disease (gluten enteropathy), cirrhosis, biliary obstruction, pancreatic insufficiency, inflammatory bowel disease, and postgastroectomy or jejunioileal bypass surgery. Patients receiving total parenteral nutrition, usually because of chronic bowel disease, develop osteomalacia due to inadequate mineral or vitamin D supplementation. Anticonvulsant therapy increases the metabolic clearance of vitamin D metabolites, requiring vitamin D supplementation.

### Tumor-Induced Osteomalacia

Some small mesenchymal tumors (hemangiopericytomas, fibromas, angiosarcomas, etc.) can cause phosphaturia and hypophosphatemia, leading to osteomalacia. The syndrome is known as tumor-induced osteomalacia (TIO) or oncogenic osteomalacia. The mechanism involves the synthesis and secretion of excessive amounts of a phosphaturic factor called phosphatonin. The major candidates for this role are FGF23 and frizzled-related protein 4 (FRP-4). How FGF23 and FRP-4 causes phosphaturia is not completely understood. However, findings from the study of X-linked hypophosphatemia (XLH) and autosomal dominant hypophosphatemic rickets (ADHR)

shed light on TIO and the mechanism for all three phosphaturic entities is discussed below. The tumors causing TIO originally were thought to be benign and of mesenchymal origin, but malignant tumors have also been reported to cause the syndrome. Often, the tumors are small and difficult to locate. The levels of calcitriol are inappropriately low (hypophosphatemia should stimulate calcitriol production) and so the tumor product is also thought to interfere with renal  $1\alpha$ -hydroxylation. The osteomalacia responds to treatment with large phosphate supplements to restore phosphate levels to normal. The syndrome can be cured by successfully removing the tumor.

## GENETIC CAUSES OF RICKETS

### Hypophosphatemic Rickets

#### **X-Linked Hypophosphatemic Rickets**

XLH is an X-linked dominant disorder caused by renal phosphate wasting and results in severe skeletal abnormalities and growth retardation. The primary mechanism is defective phosphate reabsorption in the renal proximal tubule that results in excessive phosphaturia. The clinical presentation is usually not apparent until 6–12 months of age and ranges from mild abnormalities of the bones to severe rickets and osteomalacia. Children exhibit rachitic bone deformities, including enlargement of the wrists and knees and bowing of the lower extremities. Defects in tooth development and premature cranial synostoses may also be present. Low or inappropriately normal circulating levels of calcitriol are found, despite the hypophosphatemia (Table II). The low serum phosphate normally should cause an increase in  $1\alpha$ -hydroxylase activity and enhanced calcitriol production, suggesting that XLH may also result in abnormal regulation of  $1\alpha$ -hydroxylase.

The gene causing XLH has been cloned and named *PHEX*, which is a phosphate-regulating gene with homologies to endopeptidases located on the X-chromosome. The *PHEX* gene is homologous to a family of endopeptidases that includes endothelin-converting enzyme-1 and neutral endopeptidase. The *PHEX* gene encodes a 749-amino acid membrane-bound protein that is expressed in bone, adult ovary, lung, and fetal liver. A number of genetic defects in the *PHEX* gene have been found in patients with XLH. Since many of the mutations are inactivating mutations, the X-linked dominant expression of the disorder is likely caused by a haploinsufficiency rather than the result of a dominant negative effect. The hypothesis for action is that the *PHEX* enzyme

Table II Comparison of Genetic Causes of Rickets<sup>a</sup>

	VDDR-I	HVDRR	XLH	ADHR
Gene	<i>CYP27B1</i>	VDR	<i>PHEX</i>	<i>FGF23</i>
1,25(OH) <sub>2</sub> D <sub>3</sub>	Low	High	(Normal) <sup>b</sup>	(Normal) <sup>b</sup>
PTH	High	High	Normal	Normal
Calcium	Low	Low	Normal	Normal
Phosphate	Low	Low	Low	Low
Alopecia	No	Yes	No	No

<sup>a</sup>Abbreviations used: VDDR-I, vitamin D-dependent rickets type I; HVDRR, hereditary vitamin D-resistant rickets; XLH, X-linked hypophosphatemic rickets; ADHR, autosomal dominant hypophosphatemic rickets.

<sup>b</sup>Inappropriately normal relative to decreased serum phosphate concentration.

inactivates a phosphaturic factor (phosphatonin) important for reabsorbing phosphate from the glomerular filtrate. Inactivating mutations lead to excessive phosphaturia due to inadequate destruction of phosphatonin and therefore excess urinary phosphate loss.

#### Autosomal Dominant Hypophosphatemic Rickets

ADHR is an autosomal dominant disorder caused by urinary phosphate wasting that has findings commonly seen in other phosphate-wasting disorders. Patients have short stature, bone pain, rickets, and osteomalacia. Patients also have inappropriately normal serum calcitriol concentrations (Table II).

The gene causing ADHR has been cloned and named *FGF23*. It is one of the phosphatonins. The *FGF23* gene is a member of the fibroblast growth factor family. The *FGF23* gene encodes a 251-amino acid protein that is expressed in bone, heart, lymph node, thymus, and liver. Three unique mutations have been found in the *FGF23* gene in patients with ADHR. These mutations occur in two arginine residues located in a consensus proteolytic cleavage site. The mutations provide a gain-of-function activity that results in the autosomal dominant negative transmission. Excessive *FGF23* activity causes phosphaturia and hypophosphatemic rickets.

#### Mechanism of Phosphate Loss in XLH, ADHR, and TIO

How are *PHEX* and *FGF23* involved in the pathophysiology of XLH, ADHR, and TIO? One hypothesis is that under conditions of normal phosphate regulation the *PHEX* enzyme regulates the bioavailability of *FGF23*, *FRP-4*, and possibly other putative phosphatonins. As phosphatonins are secreted from cells, some of

them are degraded to inactive metabolites by the membrane-bound *PHEX* endopeptidase. The remaining active phosphatonin enters the circulation and interacts with a receptor protein on the renal tubule cells. Binding of *FGF23* transmits a signal to down-regulate the activity of the sodium-dependent phosphate cotransporter (NPT2) in the kidney, thus decreasing phosphate reabsorption. In XLH patients, the mutant *PHEX* protein, or lack thereof, is unable to degrade phosphatonin. This leads to excess amounts of *FGF23* in the circulation. As a result, the signal to down-regulate NPT2 activity is magnified, leading to renal phosphate wasting. In ADHR, on the other hand, mutations in *FGF23* prevent the proteolytic processing step by *PHEX* and therefore there is an overabundance of active *FGF23* that then leads to down-regulation of NPT2 activity, which causes phosphate wasting. In TIO, tumors overexpress *FRP-4* and *FGF23* and elevated secretion by the tumors also leads to phosphate wasting in this condition. In all three diseases, phosphate wasting results in hypophosphatemia and undermineralization of bone, causing rickets.

#### Vitamin D-Dependent Rickets I

Vitamin D-dependent rickets type I is also known as pseudo-vitamin D deficiency type I and pseudo-vitamin D deficiency rickets. We refer to the entity as 1 $\alpha$ -hydroxylase deficiency since it has been shown to be caused by mutations in the cytochrome P450 enzyme, 25-hydroxyvitamin D-1 $\alpha$ -hydroxylase (1 $\alpha$ -hydroxylase). The human 1 $\alpha$ -hydroxylase gene (*CYP27B1*) is located on chromosome 12 near the gene encoding the VDR. A number of mutations that disrupt 1 $\alpha$ -hydroxylase activity are found scattered throughout the entire region of the *CYP27B1* gene. 1 $\alpha$ -Hydroxylase deficiency is a rare autosomal recessive disease that is manifested at an early age. Patients with 1 $\alpha$ -hydroxylase deficiency exhibit hypocalcemia, elevated PTH levels, increased alkaline phosphatase, and low urine calcium (Table II). Affected children present with hypotonia, muscle weakness, growth failure, and rickets. Tetany and convulsions may occur with severe hypocalcemia. Patients have normal serum concentrations of 25-hydroxyvitamin D<sub>3</sub> but low levels of 1,25(OH)<sub>2</sub>D<sub>3</sub> due to the defective synthesis of 1,25(OH)<sub>2</sub>D<sub>3</sub>. PTH infusion does not increase circulating 1,25(OH)<sub>2</sub>D<sub>3</sub> levels, consistent with a defect in 1 $\alpha$ -hydroxylase activity. Very large doses of vitamin D or 25-hydroxyvitamin D<sub>3</sub> are required for adequate treatment of 1 $\alpha$ -hydroxylase deficiency; often, 20,000 to more than 100,000 IU of vitamin D daily is needed.

On the other hand, modest doses of the active hormone  $1,25(\text{OH})_2\text{D}_3$  ( $0.25\text{--}2\ \mu\text{g}/\text{day}$ ) tend to be sufficient to restore calcium to normal and heal the rickets. Since mutations in  $1\alpha$ -hydroxylase block  $1,25(\text{OH})_2\text{D}_3$  synthesis, treatment with calcitriol bypasses the defect and reverses the abnormalities caused by the disease.

### Hereditary $1,25$ -Dihydroxyvitamin D-Resistant Rickets

Hereditary  $1,25$ -dihydroxyvitamin D-resistant rickets (HVDRR) is also known as vitamin D-dependent rickets type II or pseudo-vitamin D deficiency type II. HVDRR is an autosomal recessive disease characterized by early onset rickets, hypocalcemia, secondary hyperparathyroidism, and normal or elevated serum  $1,25(\text{OH})_2\text{D}$  levels. Some patients also have total body alopecia (Table II).

HVDRR is caused by mutations in the VDR. A number of mutations have been found throughout the VDR. Missense mutations in the DNA-binding domain usually occur in highly conserved amino acids and prevent the VDR from binding to DNA. Nonsense mutations have been found that result in truncation of the VDR protein and loss of hormone responsiveness. Mutations that cause exon skipping or affect RNA splicing also result in truncation of the VDR. A number of missense mutations have been identified in the ligand-binding domain. Some of these mutations occur in amino acids that are contact points for  $1,25(\text{OH})_2\text{D}_3$ . In one patient, Arg<sup>274</sup>, the contact point for the 1-hydroxyl group of  $1,25(\text{OH})_2\text{D}_3$ , was mutated to leucine (Arg<sup>274</sup>Leu), whereas in a second patient histidine<sup>305</sup>, the contact point for the 25-hydroxyl group of  $1,25(\text{OH})_2\text{D}_3$ , was mutated to glutamine (His<sup>305</sup>Gln). Other mutations, such as Phe<sup>251</sup>Cys, Gln<sup>259</sup>Pro, and Arg<sup>391</sup>Cys, have been identified that prevent RXR heterodimerization with VDR. A mutation in helix H12 (Glu<sup>420</sup>Lys) was found to prevent coactivator binding. All these mutations cause some degree of vitamin D resistance.

Interestingly, one patient exhibited all the signs of HVDRR including alopecia but no mutations in the VDR. Vitamin D resistance in this patient was apparently due to the abnormal expression of a hormone response element-binding protein. This case highlights the fact that proteins other than VDR are involved in  $1,25(\text{OH})_2\text{D}_3$  signaling and that defects in these proteins may also cause the HVDRR syndrome.

Most children with HVDRR do not respond to supraphysiological doses of calcitriol. However,

these children can be successfully treated for their hypocalcemia by normalizing their serum calcium using chronic intravenous infusions of calcium. The intravenous calcium infusions are given nightly over a period of many months. This treatment bypasses the intestinal defect in calcium absorption caused by the mutant VDR. The treatment eventually results in normalization of serum calcium levels, correction of secondary hyperparathyroidism, and normal mineralization of bone and healing of rickets on X-ray. The clinical improvement can be sustained if adequate serum calcium and phosphorus concentrations are maintained. Some patients can sometimes be maintained thereafter by oral calcium. However, patients do not show improvement in their alopecia with treatment.

### SUMMARY

Proper supply of calcium and phosphate to the skeleton is critical for the synthesis of normal bone and the prevention of rickets or osteomalacia. Calcitriol and PTH regulate mineral metabolism and adequate delivery of calcium and phosphate to the bone-forming sites. The most common cause of rickets/osteomalacia worldwide is vitamin D deficiency. Fortification of dietary foodstuffs or vitamin D supplements given to individuals at risk would greatly reduce the impact of this devastating disease of the skeleton. Also, defects at any stage in the pathway of hormone synthesis or action can cause rickets or osteomalacia. Other important causes are calcium deficiency, chronic renal failure with inadequate synthesis of calcitriol and gastrointestinal diseases that result in malabsorption of calcium or vitamin D. Many forms of rickets are due to genetic errors in the synthesis or action of calcitriol or in conditions that cause renal phosphate loss. Improvements in our understanding of the molecular mechanisms by which these defects cause rickets or osteomalacia have led to improved diagnostic and treatment strategies.

### See Also the Following Articles

Osteoporosis, Overview • Parathyroid Hormone (PTH) • Renal Osteodystrophy • Vitamin D

### Further Reading

- Bishop, N. (1999). Rickets today—Children still need milk and sunshine. *N. Engl. J. Med.* **341**, 602–604.
- Feldman, D., Glorieux, F. H., and Pike, J. W. (eds.) (1997). "Vitamin D," Chaps. 31–56. Academic Press, San Diego.

- Feldman, D., Malloy, P. J., and Gross, C. (2001). Vitamin D: Biology, action and clinical implications. In "Osteoporosis" (R. Marcus, D. Feldman, and J. Kelsey, Eds.), pp. 257-304. Academic Press, San Diego.
- Holick, M. F. (1994). McCollum Award Lecture, 1994: Vitamin D—New horizons for the 21st century. *Am. J. Clin. Nutr.* **60**, 619-630.
- Kumar, R. (2002). New insights into phosphate homeostasis: Fibroblast growth factor 23 and frizzled-related protein-4 are phosphaturic factors derived from tumors associated with osteomalacia. *Curr. Opin. Nephrol. Hypertens.* **11**, 547-553.
- Lips, P. (2001). Vitamin D deficiency and secondary hyperparathyroidism in the elderly: Consequences for bone loss and fractures and therapeutic implications. *Endocr. Rev.* **22**, 477-501.
- Malloy, P. J., Pike, J. W., and Feldman, D. (1999). The vitamin D receptor and the syndrome of hereditary 1,25-dihydroxyvitamin D-resistant rickets. *Endocr. Rev.* **20**, 156-188.
- Miller, W. L., and Portale, A. A. (1999). Genetic causes of rickets. *Curr. Opin. Pediatr.* **11**, 333-339.
- Takeyama, K., Kitanaka, S., Sato, T., Kobori, M., Yanagisawa, J., and Kato, S. (1997). 25-Hydroxyvitamin D<sub>3</sub> 1alpha-hydroxylase and vitamin D synthesis. *Science* **277**, 1827-1830.
- White, K. E., Jonsson, K. B., Carn, G., Hampson, G., Spector, T. D., Mannstadt, M., Lorenz-Depiereux, B., Miyauchi, A., Yang, I. M., Ljunggren, O., Meitinger, T., Strom, T. M., Juppner, H., and Econs, M. J. (2001). The autosomal dominant hypophosphatemic rickets (ADHR) gene is a secreted polypeptide overexpressed by tumors that cause phosphate wasting. *J. Clin. Endocrinol. Metab.* **86**, 497-500.

by neurologic complications of brain hemangioblastomas or by metastatic renal cell cancer.

Significant advances have been made in the understanding of the VHL gene and its pathway. These advances have led to improvements in genetic diagnostics, the availability of presymptomatic screening, and new treatment methods, including organ-sparing surgeries, with the intent of increasing the quality of life and life-span of affected individuals.

## History

More than 100 years ago, incomplete clinical descriptions of the disorder of what has come to be known as VHL were first reported. Illustrations of angiomas of the eye and similar vascular lesions of the cerebellum began appearing in reports as early as the 1860s. By 1904, Eugen von Hippel (1867–1938), a German ophthalmologist, published descriptions of retinal angiomas in several generations of family members in a small number of kindreds. In 1926, Arvid Lindau (1892–1958), a Swedish pathologist in Lund, published his thesis recognizing that retinal angiomas and cerebellar hemangioblastomas as well as cysts in the kidney, pancreas, and epididymis were part of a familial syndrome. Among the subsequent reports that refined the clinical understanding of VHL, Melmon and Rosen's landmark summary in 1964 established the first diagnostic criteria and included renal cancer. Linkage of the VHL gene to the short arm of chromosome 3 was reported by Seizinger and colleagues in 1988. Finally, in 1993, Latif and colleagues identified the VHL tumor suppressor gene by positional cloning strategies.

## GENETICS

### Molecular Genetics

The VHL gene is located at chromosome 3p25–p26 and is a tumor suppressor gene. This gene is also conserved in many species, including insects through mammals, and thus its functions are considered to be fundamental to life. Mutations causing loss or inactivation of the wild-type allele from the unaffected parent are thought to transform a cell into the clonal progenitor of a tumor.

Mutations causing VHL have been detected in all three exons and some intronic regions of the VHL gene. Mutation types identified include missense mutations, nonsense mutations, and partial and complete deletions of the gene. In 1996, Zbar and international collaborators reported more than 137 different *VHL*

mutations identified from patients in North America, Europe, and Japan. It is possible that exon 2 of the VHL gene may have a specific role in RCC development, as 45% of patients with mutations in exon 2 have RCC.

### VHL Complex

The VHL gene product (pVHL) binds with elongin C, which binds elongin B, Cul2, and RBX1 to form a multimeric complex. The VHL complex targets the hypoxia-inducible factors (HIF) HIF1- $\alpha$  and HIF2- $\alpha$  for ubiquitin-mediated degradation. Under normoxic conditions, the complex degrades HIF. Under hypoxic conditions, HIF is not degraded and it overaccumulates. Increased levels of HIF are associated with increased transcription of a number of downstream genes, including vascular endothelial growth factor (VEGF), erythropoietin, platelet-derived growth factor, Glut1, and transforming growth factor- $\alpha$ . Mutations of the VHL gene in the  $\alpha$ -domain (elongin C binding) or the  $\beta$ -domain (HIF targeting) have been shown to be associated with increased HIF levels. VEGF is a known promoter of tumor angiogenesis, which may account for the vascular nature of VHL tumors.

### Genetic Counseling, Family Screening, and Molecular Diagnosis

VHL is transmitted in an autosomal-dominant pattern of inheritance with reproductive consequences. Each offspring of an affected parent has a 50% chance of inheriting the mutated copy of the VHL gene, putting them at risk for VHL tumors and for transmitting the trait. Offspring who inherit the affected parent's wild-type allele are not at risk for VHL and cannot transmit the trait for VHL.

Detection of the *VHL* mutation was possible in 93 of 93 VHL families, as reported by Stolle and colleagues in 1998. Laboratories utilize multiple molecular techniques, including DNA sequencing, for the detection of *VHL* point mutations. Gene deletions may be detected by quantitative Southern blotting and, in some instances, fluorescence *in situ* hybridization is available for *VHL* mutation analysis in a growing number of Clinical Laboratory Improvement Amendment (CLIA)-certified clinical laboratories.

Genetic testing should occur within the context of genetic counseling as recommended by the American Society of Human Genetics, the American Society of Clinical Oncologists, and other medical societies. Before being tested, patients make the decision to be

**Table I Approach to Diagnosis of von Hippel-Lindau Syndrome Assisted by DNA Testing**

Obtain a geneticist, genetic counselor, and/or physician to assist:

- (1) Clinically document von Hippel-Lindau syndrome (VHL) in a family member.
- (2) Analyze the member's peripheral blood lymphocyte DNA for a mutation in the VHL gene.
- (3) If a mutation is identified,<sup>a</sup> the DNA-tested family member's first-degree relatives who wish are counseled, sign informed consent, and have their blood cells' DNA tested for the same mutation. All those found positive for the VHL gene mutation, then have DNA testing offered to their remaining first-degree relatives.
- (4) Clinical screening is offered to all those found to carry the VHL gene mutation.

*Note.* VHL penetrance has been estimated at > 90%.

<sup>a</sup>*De novo* VHL in sporadic cases may have clinical manifestations, but patients may not have a mutation identified in their peripheral blood leukocyte DNA. The offspring of patients are considered to be at risk unless they are shown to have a normal VHL gene.

tested, after discussing the medical and psychosocial implications for themselves and their families. Post-test genetic counseling continues through disclosure of genetic results.

Discovery of a germ-line genetic mutation is an indication for a lifetime of periodic screening for VHL tumors and cysts. Early detection and management of VHL-related tumors may result in decreased morbidity and mortality and improved quality of life. At-risk biological relatives need genetic counseling, clinical screening, and/or genetic testing to clarify VHL risk. Presymptomatic detection of this highly treatable disease is aided by genetic diagnostic testing (see Table I). The choice to be tested must be entirely voluntary after informed consent is given and a consent document is then signed.

*De novo* mutations present a special challenge in genetic counseling and genetic diagnostic testing. The frequency of new germ-line mutations in *VHL* is not known. Sgambati and co-workers in 2000 reported that in a registry of 181 kindreds, 42 (23%) included a first in family diagnosis. When 2 (4.8%) of the 42 were further analyzed by additional molecular investigations, VHL mosaicism was demonstrated. New mutations may arise as any one of three types of sporadic VHL mosaicism, depending on which tissues have cells with a *VHL* mutation. Mosaicism is present in varying degrees and may include VHL gene mutations in somatic tissue only, in somatic plus germ tissue (spermatocyte, oocyte), or in germ tissue only.

## CLINICAL MANIFESTATIONS

Common manifestations of VHL are renal cell carcinoma and cysts, pheochromocytoma, pancreatic neuroendocrine tumors, benign cysts of the pancreas, hemangioblastomas of cerebellum or spinal cord and medulla, retinal angiomas, endolymphatic sac tumor, and papillary cystadenomas of the epididymis and broad ligament.

The VHL clinical phenotypic expression is variable among families, which is thought to be mainly attributable to genotypic differences. Variability within families is also seen with regard to age of onset and severity of disease. In the latter case, other influencing factors, including modifier genes and lifestyle choices, such as cigarette smoking, are being studied. Classification of VHL into phenotypic subtypes has been correlated with genotypes to the extent possible (Table II). Type I, type IIA, and type IIB were initially described, and type IIC has been described in a number of reports and was the focus of the report from the laboratory of Kaelin and colleagues in 2001.

Age of onset of VHL is variable, with a mean age of onset of 25, 30, and 37 years for patients with tumors of the eye, CNS, and kidney, respectively. However, there are reports of children with retinal angioma at 1 year of age, with CNS hemangioblastoma at 11 years of age, and clear cell RCC at 16 years of age. Pheochromocytomas can occur in children as young as 6 years of age. In contrast, VHL expression can be very delayed; the diagnosis can occasionally be delayed until the eighth decade, supporting the need for lifelong monitoring of individuals at risk.

**Table II Von Hippel-Lindau Syndrome (VHL) Classification**

- I. VHL *without* pheochromocytomas
- II. VHL *with* pheochromocytomas
  - A. Pheochromocytomas, retinal angiomas, and CNS hemangioblastomas
  - B. Pheochromocytomas, retinal angiomas, CNS hemangioblastomas, clear cell RCC, and pancreatic neoplasms and cysts
  - C. Pheochromocytomas only

*Note.* VHL type I and type IIB differ only by the absence (I) or presence (IIB) of pheochromocytoma. Endolymphatic sac tumors and cystadenomas of the epididymis and broad ligament have not been assigned to specific VHL types.

Table III Screening<sup>a</sup> Guidelines for Von Hippel-Lindau Syndrome (VHL)

Test	Age
Ophthalmoscopy yearly	From infancy
Fluorescein angiography	As needed
Urinary catecholamines	From age 2 years, every 1–2 years
Enhanced MRI, brain/spine	From age 11 years, every 2 years; <sup>b</sup> after age 60 years, every 3 years
Abdominal CT w/ and w/o contrast	From age 18 years, every 2 years
Abdominal US	Yearly from age 11 to 18 years when abdominal CT begins
Neurotology examination	Any age if there is hearing loss, tinnitus, or vertigo

<sup>a</sup>Screening tests of specific organ sites are modified and may be increased in frequency at sites where VHL tumors or cysts are detected.

<sup>b</sup>Examine area of internal auditory canals for signs of endolymphatic sac tumor.

When clinical evidence of VHL is unclear, genetic testing (in all but some mosaic cases) should detect a mutation in the VHL gene when there is a risk for VHL. Identified risk indicates a need for age-appropriate baseline clinical screening and a lifetime of periodic monitoring (Table III), as well as a 50% chance for each offspring to inherit the trait and carry the risk for VHL.

Treatment of any one VHL manifestation must be prioritized along with knowledge of the other VHL tumors and cysts present in the patient. In particular, pheochromocytoma must be ruled out before any surgery. In the case of abdominal surgery, preoperatively, one must also consult the neurosurgeons to evaluate for unstable CNS masses.

### Clear Cell Renal Cell Carcinoma: Solid and Cystic Disease

Renal cell carcinoma is detected in approximately 24–45% of individuals with VHL disease. Males and females are affected approximately equally, unlike sporadic RCC, which occurs more often in males.

Multiple bilateral VHL clear cell renal lesions develop, within which a few have scattered granular cells. These tumors tend to be low grade and minimally invasive when they are small and become more aggressive as they enlarge. Only 1 of 66 malignant lesions examined by Poston and colleagues in 1995 was found to have sarcomatoid renal cell carcinoma with clear and granular cells. Although microscopic invasion of the pseudo-capsule was common (52%), no tumor extension through the pseudo-capsule was identified in a study of 161 renal lesions. There is wide variation in the growth rate of solid tumors in VHL, which Choyke and colleagues in 1992 have shown to be 0.2–2.2 cm/year (mean 1.6 cm/year). There is a single case report of metastatic disease occurring in

VHL with a primary renal tumor less than 3 cm in diameter.

As many as 63% of people with VHL have renal cysts. Complex renal cysts in VHL disease have been shown to harbor clear cell RCC in many cases. Although renal cystic disease in VHL may have the appearance of polycystic kidney disease, in VHL, periodic monitoring of renal cystic disease is recommended, as malignancy may be present.

A high prevalence of microscopic lesions in grossly normal renal parenchyma was identified and reported by Walther and colleagues in 1995. Based on extrapolations from grossly normal renal parenchyma removed at the time of partial nephrectomy, up to an estimated 600 microscopic “tumorlets” and 1100 microscopic cysts with clear cell lining may be present in the kidney of a 37-year-old with VHL. In the follow-up study reported in 1996 by Lubensky and colleagues, allelic deletion of the VHL gene was present in microscopic clear cell renal lesions. They identified loss of the wild-type VHL allele and retention of the inherited mutated VHL allele early in the development of renal clear cell lesions.

### Diagnosis

Primary kidney tumors may grow for many years before symptoms manifest. In a screened population, renal tumors are often diagnosed incidentally during imaging studies for another purpose. The classic triad of pain, palpable mass, and hematuria is very rare and when seen is often associated with large renal masses. Presymptomatic screening protocols (Table II) for individuals at risk for VHL are recommended to identify tumors at a stage when they may be managed with good outcome.

Computed tomography (CT) remains the gold standard for diagnosis of renal involvement in VHL. Contrast-enhanced CT in thin sections of 3–5 mm before and after intravenous non-ionic iodinated

contrast medium is essential for best detection and spiral geometry may decrease the chances of missing lesions (Fig. 1). Ultrasound is less sensitive than CT and is not considered reliable for lesions smaller than 2 cm. However, ultrasound is often useful in determining whether a lesion is cystic or solid. Magnetic resonance imaging (MRI) may be used in cases in which CT is contraindicated, for example, when there is impaired renal function. Patients with VHL commonly have normal renal function even when there are multiple tumors and cysts in their kidneys. Renal lesions in patients with VHL are most often considered suspicious for cancer and it is recommended that they be followed if they are not yet large enough to warrant immediate treatment.

### **Treatment**

Nephron-sparing surgery was designed to minimize the risk of metastasis yet preserve kidney function. Walther and colleagues reported in 1999 their 10-year experience with preserving kidney function by using a 3 cm threshold for renal parenchymal-sparing surgery. In 52 patients with VHL who had partial nephrectomies and a median follow-up of 60 months (ranging from 6 to 205 months), there were no metastases and none of these patients needed dialysis or transplantation.

Each patient's treatment may be tailored according to the location and sizes of renal masses on imaging studies. Preoperative evaluation in VHL includes evaluations for pheochromocytoma and

hemangioblastomas of the central nervous systems as both lesions in these locations have the potential to cause significant adverse consequences in the operative and perioperative period if undetected.

Radiofrequency ablation, a newer treatment modality, is being evaluated in VHL patients as an alternative to surgery. Early results are encouraging for this method of tumor ablation, which may have significantly reduced morbidity compared to surgery. However, it is too early to determine the role of this method of treatment for patients with VHL-associated renal cell carcinoma.

Kidney transplantation after bilateral nephrectomy is performed when the option of nephron-sparing surgery is no longer viable. In addition to the usual protocol rules guiding organ transplantation for individuals who have had cancer, the hereditary nature of VHL adds another dimension in selecting a living related donor. It is recommended that potential related donors be screened with genetic testing. Furthermore, the long-term effects of immunosuppression in VHL with regard to tumor development are incompletely understood.

### **Sporadic Clear Cell Renal Carcinoma and VHL Gene Mutations**

Approximately 60% of tumors from patients with sporadic clear cell renal cancers unrelated to VHL have been found to have mutations in the VHL gene. Somatic mutations in both alleles of the VHL



**Figure 1** A CT scan of a 44-year-old female with VHL and severe renal manifestations, including bilateral renal cell carcinomas (arrows) and numerous bilateral renal cysts.

gene have been shown in many clear cell RCCs. This is strong evidence supporting the VHL gene as being important in the development of clear cell carcinoma of the kidney.

### **Phenocopies**

In kindreds with VHL, sporadic clear cell RCC may arise even in the absence of the VHL gene mutation. This has been observed in individuals from two kindreds. One family member with clear cell RCC who was a heavy cigarette smoker was shown to be negative for the VHL germ-line mutation present in her relatives with VHL. These phenocopies apparently represent sporadic cases of renal cancer occurring in families with VHL.

### **Risk Factors in RCC**

Environmental factors, as well as heritable genetic factors, have been shown to be associated with mutations in the VHL gene. It has been widely reported that cigarette smoking may be a risk factor for the development of kidney cancer. In addition, in 1999, Brauch and colleagues reported high cumulative trichloroethylene (TRI) exposure in workers who later developed RCC with a unique mutation pattern found in the VHL gene. In the past, TRI was used in industrial solvents.

## **Pheochromocytoma**

The exclusion of pheochromocytoma prior to any surgery and before the onset of labor and delivery can be a life-saving measure. Pheochromocytomas are one of two types of VHL tumors that can occur before age 10. It is recommended that for children who are at risk for VHL screening should begin at an early age.

VHL has been subclassified based on the tendency to develop pheochromocytomas. Type I VHL is not associated with pheochromocytomas, whereas type II VHL predisposes to pheochromocytomas. Type IIA has a high prevalence of pheochromocytomas. For example, 57% of VHL-affected children and adults had pheochromocytomas, as seen in early studies of a type IIA large multigenerational family with mild CNS hemangioblastomas and retinal angiomas and a missense germ-line mutation in *VHL*. Type IIB, however, predisposes to pheochromocytomas along with the entire spectrum of VHL disease. Atuk *et al.* in 1998 reported a large multigenerational type IIB VHL family with pheochromocytoma having its onset between the ages of 5 and 25 years. Type IIC has been designated as familial pheochromocytoma only,

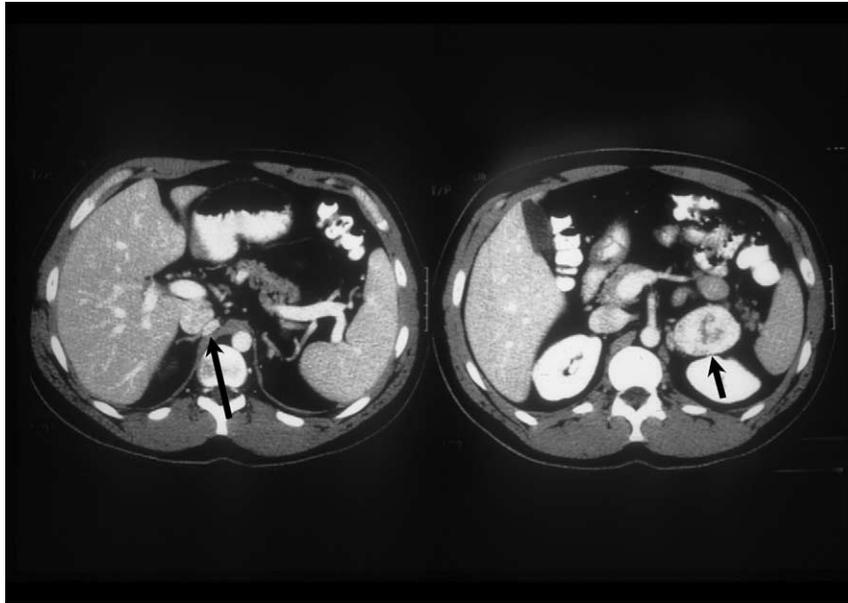
without renal cell carcinoma and hemangioblastomas, in five or more reports since 1995 of kindreds with mutations in the VHL gene.

Pheochromocytomas in VHL arise from chromaffin cells, usually in the adrenal medulla. Less than 2–10% of pheochromocytomas in VHL become metastatic. In 1999, Walther and colleagues reported on 64 patients with VHL pheochromocytomas and noted that missense mutations in the VHL gene tended to be associated with extra-adrenal pheochromocytoma, younger age at presentation, and the only patient in the study with metastases.

### **Diagnosis**

Pheochromocytomas in VHL are frequently multiple and bilateral in the adrenal medulla or may be extra-adrenal (Fig. 2). Ectopic or extra-adrenal pheochromocytomas (paragangliomas) are not uncommon and may be located in the glomus jugulare, carotid body, peri-aortic sites, spleen, kidney, and organ of Zuckerkandl (bifurcation of aorta and femoral arteries). Metastatic disease to nodes and distant organs has been seen with pheochromocytomas in VHL. CT or MRI is helpful in identifying adrenal and peri-adrenal masses, but extra-adrenal sites may require a radionuclide study called metaiodobenzylguanidine (MIBG). Patients with VHL and pheochromocytomas can have CT scans with intravenous non-ionic iodinated contrast without adverse effects. Pheochromocytomas arise from neural crest tissue and produce catecholamines that are stored in neurosecretory granules. Symptoms are variable and may include intermittent or sustained hypertension, palpitations, tachycardia, nervousness, irritability, headaches, episodic sweating, pallor, nausea, and anxiety attacks. Pheochromocytomas may cause life-threatening hypertensive crisis, myocardial infarction, cardiac failure, or metastatic disease. However, VHL-associated pheochromocytomas are often diagnosed by presymptomatic screening of family members. Walther and colleagues reported in 1999 that 35% (13 of 37) of newly diagnosed patients, detected by screening kindreds with the pheochromocytomas, had no symptoms, normal blood pressure, and normal catecholamine test results. Median tumor doubling time was 17 months.

Laboratory tests for pheochromocytoma often provide the foundation for the diagnosis. These functional tests may show activity of a pheochromocytoma when there is no adrenal or extra-adrenal tumor seen on imaging, thus prompting further studies, possibly including MIBG scintigraphy, to locate the site of tumor activity. For many years, 24h urinary measurements of catecholamines have been the practice in testing for functional pheochromocytoma.



**Figure 2** Bilateral pheochromocytomas in a 28-year-old male with VHL. There is a 4 cm pheochromocytoma (short arrow) on the left and a 1 cm pheochromocytoma on the right (long arrow), ectopically located behind the inferior vena cava.

A very sensitive serum test for catecholamine precursors was developed by Eisenhofer and colleagues in 1999. The sensitivity for normetanephrines and metanephrines is 97% and the specificity is 96%, whereas the sensitivities of other biochemical tests ranged from 47 to 74%. Patients with VHL had high plasma concentrations of normetanephrines; this contrasts with multiple endocrine neoplasia type 2 (MEN2) patients, who had high concentrations of metanephrine. Measurements are made of epinephrine, norepinephrine, normetanephrines, and total metanephrines. Twenty-four-hour urinary testing additionally measures dopamine and vanillylmandelic acid. It may also be necessary to perform provocative testing by glucagon stimulation and/or clonidine suppression tests to assess for functional changes that would occur under stress.

#### **Treatment**

Surgical resection is often the most common treatment for pheochromocytoma. Partial adrenalectomies and enucleations are increasingly being recommended to preserve adrenal function. Prior to surgery for pheochromocytoma, systemic treatment may be required for pharmacologic control with a combination of  $\alpha$ - and  $\beta$ -adrenergic blockade. Blockade may be initiated with phenoxybenzamine, an alpha-blocker, followed by beta-blockade with propranolol or methyparatyrosine. Adrenergic blockage may be employed before, during, and even after surgery. Asymptomatic VHL

pheochromocytomas found during screening before catecholamine levels are elevated are sometimes followed at intervals of 6 months to 1 year with imaging and functional tests. Walther and colleagues provided surgical decision guidelines in 1999 that include tumors with abnormal function or size greater than 3.5 cm.

#### **Sporadic Pheochromocytomas**

Brauch *et al.* in 1999 reported their study of germ-line VHL gene mutations and RET gene mutations in 62 patients with pheochromocytomas with no history of hereditary disease. They found 2 patients (3%) with VHL gene mutations and none with mutations in the RET proto-oncogene at exons 10, 11, and 13. Neumann and colleagues in 1993 reported 19.5% of unselected patients with what appeared to be sporadic pheochromocytomas, who actually had VHL as the etiology. Walther *et al.* in 1999 found pheochromocytomas in VHL to be smaller and less functional than sporadic pheochromocytomas.

#### **Pancreatic Neuroendocrine Tumors**

Solid VHL neoplasms of the pancreas are commonly nonfunctional neuroendocrine tumors. Histologically, the tumor may show a trabecular and/or glandular architecture and nests of tumor cells have been demonstrated. Pancreatic and gastrointestinal hormones

were negative by staining in the study of Libutti and colleagues reported in 1998, whereas Lubinski *et al.*, also in 1998, examined 30 tumors and found that fewer than 4 stained positive for pancreatic polypeptide, somatostatin, insulin, and/or glucagon.

A Mayo Clinic study of patients with VHL, enrolled at the clinic during a 10-year period, found neuroendocrine tumors in 17% of patients. An association with pheochromocytomas was also noted. Libutti and colleagues in 1998 studied 17 patients with surgically resected pancreatic neuroendocrine tumors and their ages ranged from 18 to 48 years.

### Diagnosis

CT imaging performed pre- and post-contrast may identify a neuroendocrine tumor of the pancreas as an enhancing mass (Fig. 3). The CT is often obtained during the arterial (early) phase after contrast administration as these tumors enhance early and briefly and then become isodense with the surrounding pancreatic parenchyma. The tumors frequently occur in the pancreatic head, but are also found in the body or tail. They may be multiple and metachronous; thus, monitoring of the remaining pancreas is continued after resection of a tumor. Because these tumors are characteristically non-functional, detection may require periodic scanning of asymptomatic individuals at risk for VHL.

VHL neuroendocrine tumors in some series have been shown to have the potential to metastasize, most often to the liver and regional lymph nodes.

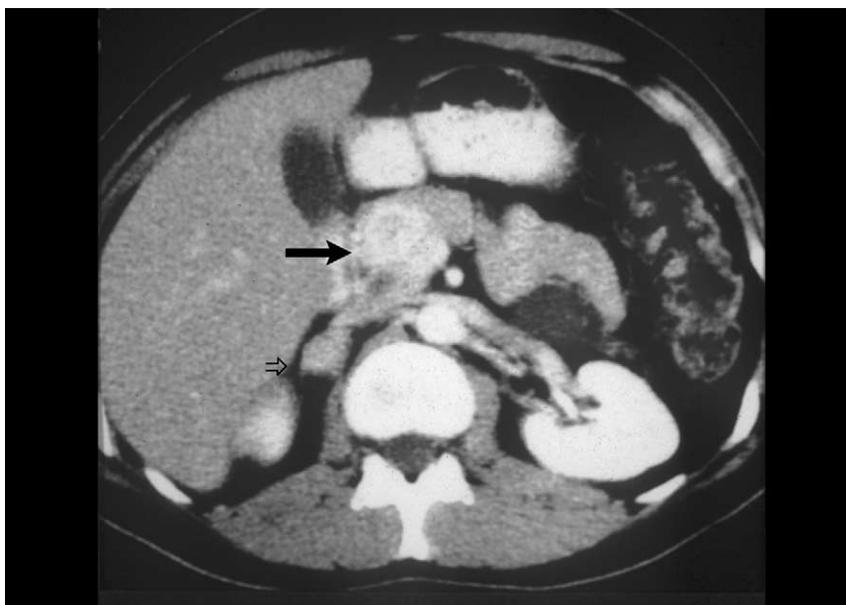
Therefore, it is considered important to carefully evaluate the liver by CT and MRI in patients with VHL and PNETs.

Neuroendocrine tumors may sometimes be difficult to distinguish from multicystic cystadenomas, which are considered to be benign VHL pancreatic masses, which may also show enhancement on CT or MRI. Pancreatic neuroendocrine tumors usually enhance uniformly and intensely, whereas serous or multicystic cystadenomas are more heterogeneous and enhance to a lesser degree.

### Treatment

Surgical resection is often the treatment of choice. The surgical approach is generally dictated by the location and size of the tumor. Resection may be by enucleation, pylorus-preserving pancreaticoduodenectomy (Whipple's procedure), or partial or rarely total pancreatectomy with replacement therapy. Neuroendocrine tumors of the pancreas in VHL may be slow-growing. Libutti and colleagues in 1998 found the size of the primary tumor to be correlated with the presence of metastases. Tumors less than 1 cm in size were monitored with serial scanning at intervals of 12–24 months. Management of larger tumors, 1–3 cm, depended on their location. However, preserving pancreatic functional tissue is often balanced with the known malignant potential.

Metastatic foci in the liver have been treated by therapies such as isolated hepatic perfusion with



**Figure 3** Pancreatic neuroendocrine tumor in a 40-year-old female with VHL. There is a large hyperenhancing mass in the pancreatic head (solid arrow) and a right pheochromocytoma (open arrow). A cyst is present in the pancreatic tail.

Melphalan, hyperthermia, radiofrequency ablation, and wedge resection.

### **Pancreatic Cysts and Cystadenomas**

Benign cysts and cystadenomas are the most common VHL lesions in the pancreas. The frequency of individuals affected varies in different studies; frequencies ranging from 17 to 56% have been reported. Rarely, these lesions may be associated with endocrine or exocrine insufficiency. Some very large benign cysts may require treatment to relieve symptoms due to compression of the stomach outlet, intestine, or bile ducts (Fig. 4).

### **Hemangioblastoma of Brain and Spinal Cord**

A common presenting tumor of VHL, found in 44–72% of patients, is hemangioblastoma of the CNS. This tumor is most often detected using clinical imaging. A mean age of onset of 29 years has been estimated, but affected individuals as young as 11 years and as old as 78 years with CNS hemangioblastoma have been reported.

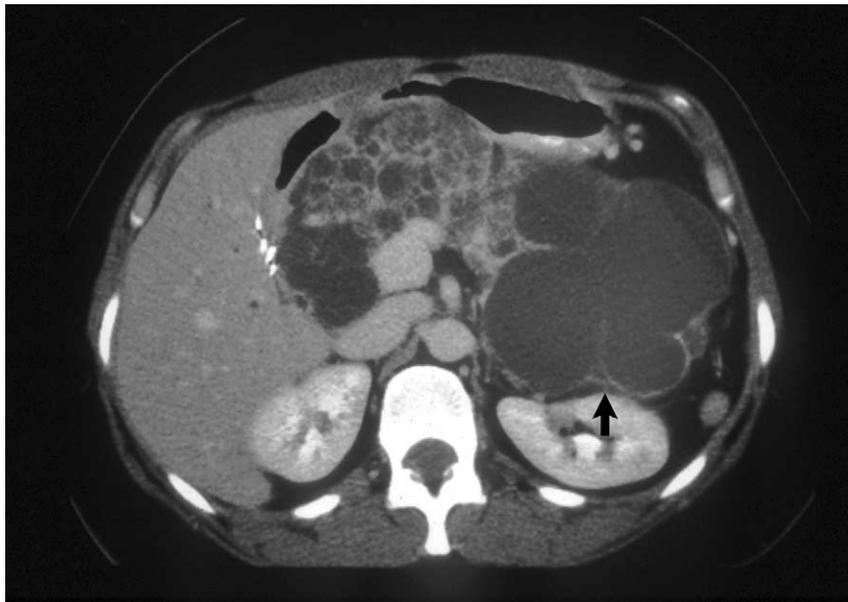
The central nervous system is often affected by multiple tumors, each of which may have an associated cyst. Within the fixed intracranial space, growing tumors or their more rapidly expanding associated cysts impinge

on normal brain structures and may cause obstructive hydrocephalus, increased intracranial pressure, and death, if they are not identified and treated.

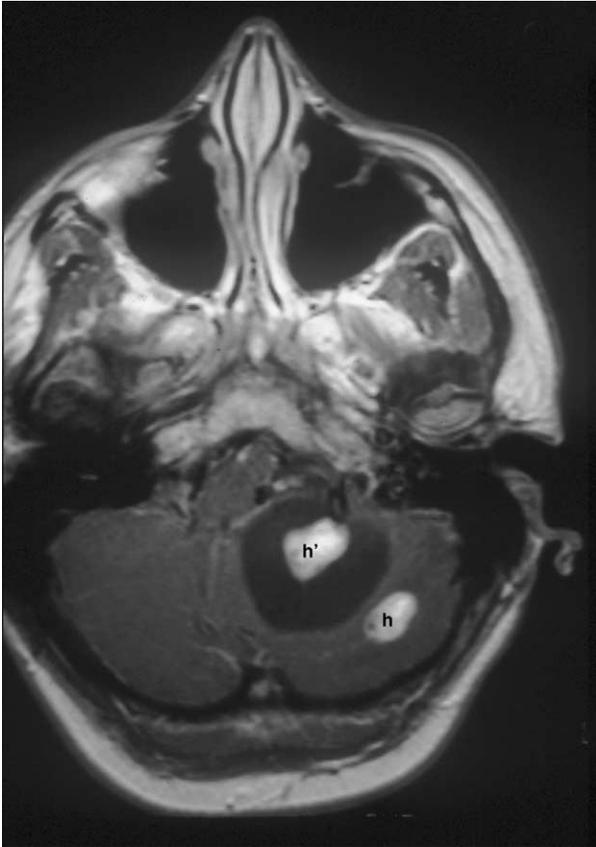
Hemangioblastoma is a vascular tumor with histopathology demonstrating channels lined by cuboidal epithelium, nests of foamy stromal cells, and pericytes. Loss of heterozygosity of the VHL gene has been found in the stromal cell, indicating that it is the likely cell of origin of these tumors. Hemangioblastomas express excess VEGF. Berkman and colleagues demonstrated that these tumors express a large amount of VEGF. Rarely, a patient with hemangioblastoma in the CNS may have polycythemia associated with elevated erythropoietin and the erythropoietin levels decrease after removal of the tumor. Erythrocytosis or secondary polycythemia has been seen even more rarely in VHL.

### **Diagnosis**

MRI with gadolinium demonstrates hemangioblastomas as circumscribed, brightly enhancing, spherical tumors that often arise at several sites in the cerebellum, spinal cord, intrathecal nerve roots, and brainstem. Thus, in screening evaluations, both the brain and spinal cord are imaged. Commonly, the tumors are found in the cerebellum, where they may be silent until they reach a large size. Spinal cord, brainstem, and supratentorial sites of hemangioblastomas are more problematic for treatment. Rarely, tumors occur in the temporal lobe where they may



**Figure 4** Severe pancreatic cystic disease. In this 34-year-old female with VHL, the pancreas is markedly enlarged by innumerable small and large pancreatic cysts including a large multilobulated cyst in the tail of the pancreas (arrow). This patient had poor gastric emptying due to extrinsic compression on the stomach.



**Figure 5** Multiple hemangioblastomas in a 26-year-old female with VHL. One hemangioblastoma (h) is entirely solid and the other (h') is surrounded by a cyst.

cause seizure activity. Also rare are hemangioblastomas that arise in the tuber cinereum, near the pituitary, or in the optic nerve or chiasm, where impingement on the optic nerves or chiasm may produce loss of vision, even though the retina may be free of angiomas. When cysts are associated with the CNS tumors, they may necessitate treatment sooner because of rapidly accumulating fluid enlarging the cyst and causing pressure on normal tissue. In fact, Lindau's original description of the typical CNS lesion was a vascular mural nodule that may be difficult to locate in a large fluid-filled cyst (Fig. 5).

Spinal cord hemangioblastomas arise mainly in the cervical and thoracic regions of the spinal cord and less often in the lumbar region. In the important Melmon and Rosen report of 1964, 80% of intramedullary hemangioblastomas were accompanied by syringomyelia. These cysts in the spinal cord (Fig. 6) can cause pressure on adjacent normal tissue and symptoms such as weakness, sensory loss, urinary disturbance, and loss of balance may result. In addition



**Figure 6** Multiple cervical hemangioblastomas with a syrinx. This 36-year-old male has undergone prior neurosurgical procedures to remove hemangioblastomas. A syrinx (open arrow) and several hemangioblastomas (arrows) are present.

to headaches, symptoms may include vertigo, nausea, vomiting, slurred speech, nystagmus, dysmetria, wide-based gait, ataxia, and motor or sensory deficits.

#### **Treatment**

Surgical resection remains the mainstay of treatment. Because these tumors are benign and do not metastasize, neurosurgical resection is sometimes not considered until there are symptoms or signs that functional compromise may be impending. The timing of surgery is often selected to avoid neurologic deficits. In the surgical treatment of enlarging or symptomatic cysts or syrinxes, neurological surgeons often locate and remove the hemangioblastoma producing the fluid, because if the cyst is only drained and the tumor is left in place, the cyst may recur in a few days. Intraoperative color Doppler ultrasound often is used to aid the neurosurgeon in localizing small mural tumors in the cerebellum and tumors in the spinal cord.

A small fraction of cerebellar hemangioblastomas are treated with high-dose focused radiation (gamma knife surgery, linac, or other stereotactic radiosurgical ablation). In a setting of multiple tiny hemangioblastomas, on rare occasions, selected doses of conventional external-beam radiotherapy targeting the posterior fossa or spinal axis have achieved tumor control for many years.

### Retinal Angiomas

Retinal angiomas (hemangioblastomas) are the commonest VHL tumor, occurring in 59% of individuals with VHL. Approximately 5% of the cases present under age 10 years, with the earliest age of onset being reported in a 1-year-old.

The histology of retinal angiomas appears to be identical to that of hemangioblastomas of the CNS and they are also considered benign. Molecular studies by Chan and colleagues in 1999 were undertaken to determine the cell of origin of hemangioblastomas. They concluded that vacuolated stromal cells represent the true neoplastic cell in retinal angiomas.

### Diagnosis

Direct or indirect ophthalmoscopy may identify most retinal angiomas. Prominence of feeder vessels may be seen leading to and from some tumors. Tonometry periodically may be recommended to detect any associated glaucoma. Screening examinations are often recommended at least yearly. A pediatric ophthalmologist will often be involved in the management early in the life of individuals at risk for VHL.

Multifocal and bilateral (approximately 50%) retinal angiomas may occur. Asymptomatic lesions may be found in the periphery of the retina. However, it is not uncommon to have the tumor on the optic disc, in which case the ophthalmologist might recommend monitoring the lesion, as treatment itself could have the potential to damage the optic nerve head.

Symptoms resulting from untreated retinal angiomas may be the result of retinal detachment, capillary leak with fluid exudates, hemorrhage, macular edema, neovascularization glaucoma, and cataract. Partial or total vision loss occurs in some individuals with VHL retinal disease.

Other causes of vision loss in VHL are hemangioblastomas of the optic nerve or chiasm, increased intracranial pressure from homangioblastoma of the brain resulting in optic nerve atrophy, and postneurosurgical damage to the occipital lobe.

### Treatment

Most clinicians recommend laser photocoagulation for the treatment of retinal tumors or cryotherapy for larger lesions. However, tumors on the optic disc are often monitored without treatment, which itself can be associated with damage. When tumors are diagnosed and treated early, vision loss or blindness may be prevented. Enucleation may be necessary for irreversible glaucoma with severe pain associated with ocular angiomatosis. Vitrectomy may play a role in a few cases. Further investigations are needed for the role of systemic antitumor agents. Various radiotherapy modalities have been applied on a compassionate or an experimental basis in cases of severely affected eyes not responding to the usual treatments.

### Endolymphatic Sac Tumor

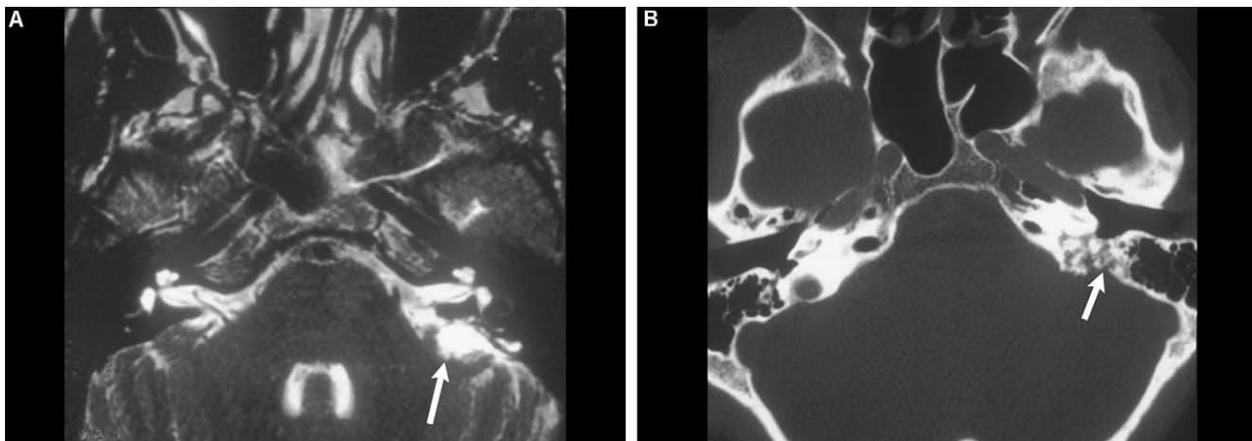
Manski and colleagues reported in 1997 their study of the association of an inner ear tumor with VHL. The ELST was the consensus name given for the previously diverse nomenclature of the tumor. ELSTs were detected in 13 (11%) of 121 individuals with VHL and in none of 253 patients without VHL. Hearing loss occurred at a mean age of 22 years, with the range being 12 to 50 years of age. Additional patients from the collection at the Air Force Institute of Pathology were reviewed by Manski and included a 7-year-old with ELST. Bilateral ELSTs seem to be found exclusively in patients with VHL.

The ELST is characterized by a papillary-cystic adenomatous growth, but lacks generally accepted histologic features of malignancy. However, because of its locally aggressive behavior of eroding the surrounding temporal bone, Heffner *et al.* classified it as a low-grade adenocarcinoma.

### Diagnosis

Common presenting symptoms of ELST are hearing loss, tinnitus, vertigo/disequilibrium, and facial paresis. Sudden onset of complete hearing loss on the side of the tumor was found in 38% of patients with ELST. There is no clear correlation between tumor size and symptoms.

Audiologic assessment is added when there is a suspicion of a tumor. MRI of the brain may reveal a lesion, but the specific studies used for diagnosis are CT and MRI of the internal auditory canals. On CT scans, ELST is seen as an expansile and/or osteolytic lesion centered around the vestibular aqueduct in the posterior petrous bone. On MRI scans, it is characterized by heterogeneous foci of low and high



**Figure 7** Endolymphatic sac tumor. This 32-year-old female with VHL developed sensorineural hearing loss on the left due to an endolymphatic sac tumor (arrow) on T2-weighted sequence MRI (A). A CT scan (B) of the same lesion demonstrates bony erosions due to the tumor (arrow).

intensities in both T1- and T2-weighted sequences (Fig. 7).

If no ELST is identified in patients with symptoms of hearing loss, tinnitus, vertigo or unexplained imbalance, repeated monitoring including audiologic and imaging studies is often advised.

#### **Treatment**

The indication and timing of surgical treatment take into consideration the slow but variable growth rate of ELSTs, preoperative hearing level, and severity of vestibular symptoms. Due to a high incidence of sudden sensorineural hearing loss associated with ELST, an early surgical intervention may be recommended for a patient with serviceable hearing as long as the tumor has not extended into the bony labyrinth. A surgical treatment may be indicated to control recalcitrant vestibular symptoms from ELST.

Surgical treatment of ELST may be curative when the tumor is completely excised via a combined retro-labyrinthine and retrosigmoid approach. Partial resection with or without radiation therapy has been reported as having a high incidence of recurrence. With complete resection, the preoperative level of hearing has been preserved in most cases and the tumor-associated vestibular symptoms can be effectively controlled.

### **Papillary Cystadenoma**

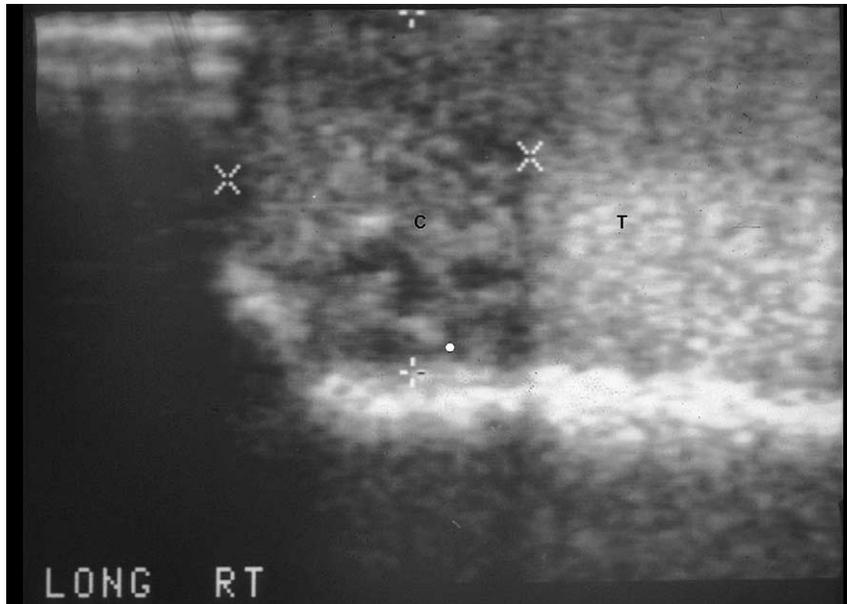
#### **Epididymal Papillary Cystadenomas**

Choyke and colleagues in 1997 reported that 30 (54%) of 57 males with VHL had epididymal abnormalities consistent with epididymal cystadenoma, commonly 15 to 20 mm solid masses in the head of

the epididymis. Onset has been seen in the teenage years. Ultrasounds of the scrotum may be used to distinguish cystadenomas of the epididymis found in males with VHL from cysts of the epididymis found in approximately one-fourth of all males (Fig. 8). The tumors are papillary cystadenomas that may be multiple and bilateral and are often identifiable on physical examination. Because these tumors are benign and most often asymptomatic, treatment is rarely required. One case of VHL with bilateral clear cell papillary cystadenoma of the epididymis that presented as infertility has been reported.

#### **Broad Ligament Papillary Cystadenomas**

In females with VHL, papillary cystadenomas of the broad ligament have been reported in rare cases, but may be unrecognized in many more cases. In 1994, Gaffey and colleagues in their report pointed out the histological similarity of papillary cystadenomas of the middle ear/temporal bone and of the female pelvic adnexa in patients with VHL. They referred to the broad ligament tumors of VHL as adnexal papillary cystadenoma of probable mesonephric origin (APMO). In 2000, Shen *et al.* reported allelic deletion of the VHL gene detected in papillary tumors of the broad ligament, thus confirming it as a VHL tumor. Reports in the literature include descriptions of women with VHL who had diagnoses of broad ligament cystadenomas between the ages of 22 and 46 years. However, in the VHL registry of the authors of this article, one patient was a teenager when she had surgery twice within a few years for resection of papillary cystadenomas of the paratubal region of the broad ligament.



**Figure 8** Cystadenoma of the epididymis. This 28-year-old male with VHL presented with a rock-like mass in the right scrotum. A longitudinal ultrasound of the scrotum reveals a normal testicle (T) and a solid mass in the head of the epididymis, which is typical of a cystadenoma of the epididymis in VHL.

## FUTURE PROSPECTS

The genetics of VHL have relevance to a number of sporadic neoplasms. Of particular importance is the genetic similarity of most clear cell renal cancers. This has led to diverse molecular studies of VHL being initiated although the syndrome itself is a rare disease. Most VHL tumors have been shown to make VEGF and, as a result, VHL has become a model for the study of tumor angiogenesis and the development of anti-angiogenic agents. Investigations of the role of pVHL, elongin B/C, and Cul2 and their role with HIF in oxygen sensing may lead to therapies to restore normal function to important biochemical pathways. This may lead to the development of anti-tumor agents or modalities for use in the prevention or treatment of VHL and sporadic tumors arising due to VHL gene somatic mutations.

## See Also the Following Articles

Adrenal Tumors, Molecular Pathogenesis • Hypertension, Endocrine • Pheochromocytoma

## Further Reading

Chan, C. C., Vortmeyer, A. O., Chew, E. Y., Green, W. R., Matteson, D. M., Shen, D. F., Linehan, W. M., Lubensky, I. A., and Zhuang, Z. (1999). VHL gene deletion and enhanced VEGF

gene expression detected in the stromal cells of retinal angioma. *Arch. Ophthalmol.* **117**, 625–630.

Choyke, P. L., Glenn, G. M., Walther, M. M., Patronas, N. J., Linehan, W. M., and Zbar, B. (1995). Von Hippel-Lindau disease: Genetic, clinical, and imaging features. *Radiology* **194**, 629–642.

Clifford, S. C., Cockman, M. E., Smallwood, A. C., Mole, D. C. R., Woodward, E. R., Maxwell, P. H., Ratcliffe, P. J., and Maher, E. R. (2001). Contrasting effects on HIF-1 $\alpha$  regulation by disease-causing pVHL mutations correlate with patterns of tumorigenesis in von Hippel-Lindau disease. *Hum. Mol. Genet.* **10**, 1029–1038.

Eisenhofer, G., Lenders, J. W. M., Linehan, W. M., Walther, M. M., Goldstein, D. S., and Keiser, H. R. (1999). Plasma normetanephrine and metanephrine for detecting pheochromocytoma in von Hippel-Lindau disease and multiple endocrine neoplasia type 2. *N. Engl. J. Med.* **340**, 1872–1879.

Hoffman, M. A., Ohh, M., Yang, H., Klco, J. M., Ivan, M., and Kaelin, W. G., Jr. (2001). Von Hippel-Lindau protein mutants linked to type 2C VHL disease preserve the ability to down-regulate HIF. *Hum. Mol. Genet.* **10**, 1019–1027.

Latif, F., Tory, K., Gnarr, J., Yao, M., Duh, F.-M., Orcutt, M. L., Stackhouse, T., Kuzmin, I., Modi, W., Geil, L., Schmidt, L., Zhou, F., Li, H., Wei, M. H., Chen, F., Glenn, G., Choyke, P., Walther, M. M., Weng, Y. U., Duan, D. R., Dean, M., Glavac, D., Richards, F. M., Crossey, P. A., Ferguson-Smith, M. A., Le Paslier, D., Chumakov, I., Cohen, D., Chinault, A. C., Maher, E. R., Linehan, W. M., Zbar, B., and Lerman, M. I. (1993). Identification of the von Hippel-Lindau disease tumor suppressor gene. *Science* **260**, 1317–1320.

Libutti, S. K., Choyke, P. L., Alexander, H. R., Glenn, G., Bartlett, D. L., Zbar, B., Lubensky, I., McKee, S. A., Maher, E. R., Linehan, W. M., and Walther, M. M. (2000). Clinical and genetic analysis of patients with pancreatic neuroendocrine tumors associated with von Hippel-Lindau disease. *Surgery* **128**, 1022–1028.

- Lubensky, I. A., Gnarr, J. R., Bertheau, P., Walther, M. M., Linehan, W. M., and Zhuang, Z. (1996). Allelic deletions of the VHL gene detected in multiple microscopic clear cell renal lesions in von Hippel-Lindau disease patients. *Am. J. Pathol.* **149**, 2089–2094.
- Manski, T. H., Heffner, D. K., Glenn, G. M., Patronas, N. J., Pikus, A. T., Katz, D., Lebovics, R., Sledjeski, K., Choyke, P. J., Zbar, B., Linehan, W. M., and Oldfield, E. H. (1997). Endolymphatic sac tumors: A source of morbid hearing loss in von Hippel-Lindau disease. *J. Am. Med. Assoc.* **277**, 1461–1466.
- Ohh, M., Park, C. W., Ivan, M., Hoffman, M. A., Kim, T. Y., Huang, L. E., Pavletich, N., Chau, V., and Kaelin, W. G. (2000). Ubiquitination of hypoxia-inducible factor requires direct binding to the  $\beta$ -domain of the von Hippel-Lindau protein. *Nat. Cell Biol.* **2**, 423–427.
- Sgambati, M. T., Stolle, C., Choyke, P. L., Walther, M. M., Zbar, B., Linehan, W. M., and Glenn, G. M. (2000). Mosaicism in von Hippel-Lindau disease: Lessons from kindreds with germline mutations identified in offspring with mosaic parents. *Am. J. Hum. Genet.* **66**, 84–91.
- Walther, M., Choyke, P., Glenn, G., Lyne, J. C., Rayford, W., Venzon, D., and Linehan, W. M. (1999). Renal cancer in families with hereditary renal cancer: Prospective analysis of a tumor size threshold for renal parenchymal sparing surgery. *J. Urol.* **161**, 1475–1479.
- Walther, M. M., Lubensky, I. A., Venzon, D., Zbar, B., and Linehan, W. M. (1995). Prevalence of microscopic lesions in grossly normal renal parenchyma from patients with von Hippel-Lindau disease, sporadic renal cell carcinoma and no renal disease: Clinical implications. *J. Urol.* **154**, 2010–2014.
- Walther, M. M., Reiter, R., Keiser, H. R., Choyke, P. L., Venzon, D., Hurley, K., Gnarr, J. R., Reynolds, J. C., Glenn, G. M., Zbar, B., and Linehan, W. M. (1999). Clinical and genetic characterization of pheochromocytoma in von Hippel-Lindau families: Comparison with sporadic pheochromocytoma gives insight into natural history of pheochromocytoma. *J. Urol.* **162**, 659–664.



## von Recklinghausen's Disease

see Neurofibromatosis

progresses. Hyperreflexia in the lower extremities with a positive Babinski sign is elicited in all patients. Peripheral neuropathy is manifest in the lower extremities with relative sparing of the hands. Patients show an abnormal wear pattern to their shoes. In most cases, the brain MRI is normal at the time of initial diagnosis, but approximately one-third of AMN patients demonstrate cerebral white matter involvement and probably carry a worse prognosis. Somatosensory-evoked potentials and BAEPs are often abnormal and nerve conduction velocities in the lower extremities may be delayed.

As the neurologic disease progresses over the course of 10–30 years, cerebral involvement becomes more apparent in most patients and signs of subcortical dementia supervene. In addition to the adrenal disease, patients often show diminished testicular function with decreased testosterone levels in combination with elevated luteinizing hormone and follicle-stimulating hormone levels. Nevertheless, many patients have fathered children prior to the onset of neurologic symptoms or early in its course. AMN patients frequently develop premature thinning of their scalp hair.

In addition to CCALD and AMN, there are several less common ALD phenotypes. Some males have adolescent or adult-onset forms of cerebral ALD, with a rapid progression of neurologic disease similar to that seen in CCALD. A small number of X-ALD patients exhibit Addison's disease only and lack neurologic symptoms. By screening families of X-ALD probands, a significant proportion of asymptomatic males who carry the gene, but lack neurologic or adrenal symptoms have been identified. Most of these males are still young and will probably develop AMN or other neurologic variants.

Although males express the full spectrum of symptoms, 20% of female heterozygotes ultimately develop mild spastic paraparesis, similar to AMN males, in the fourth to sixth decades of life. A greater proportion shows subtle signs of neurologic disease, such as hyperreflexia and decreased vibratory sensation in the lower extremities. However, female heterozygotes rarely develop adrenal insufficiency.

A perplexing observation is the lack of consistency in the X-ALD phenotypes within families. Approximately one-half of X-ALD kindreds have patients with CCALD and AMN, even though these patients share the same ALD mutation. This argues strongly for the presence of other genetic or environmental factors that modify the clinical expression of X-ALD. Genetic studies implicate the existence of a modifier gene, which has yet to be identified.

## GENETIC DEFECT

ALD is inherited as an X-linked trait and the gene has been mapped to chromosome Xq28. The gene spans 26 kb and is composed of 10 exons that code for a 745-amino-acid protein (ALDP), which is localized to the peroxisomal membrane. ALDP is a member of the ATP-binding cassette (ABC) family of membrane transport proteins and shows homology to several other ABC proteins that are located within peroxisomal membranes, including ALDR and PMP70. ALDP is a half-transporter that is active when it forms a dimer with itself, ALDR, or other ABC half-transporter proteins. The exact function of ALDP is not known, but the biochemical consequence of its deficiency is an impaired ability to oxidize very-long-chain fatty acids (VLCFAs). The protein has been speculated to act in the intraperoxisomal transport of VLCFA, VLCFA-coenzyme A (CoA) esters, very-long-chain acyl CoA synthetase, or other essential factors for VLCFA oxidation.

More than 200 mutations have been found in the ALD gene, including missense mutations, deletions, insertions, and splicing defects. Most of these mutations result in the absence of immunologically detectable ALDP. It is notable that there is no correlation between the type of mutation and the associated X-ALD phenotype or the severity of the VLCFA oxidative defect.

## BIOCHEMICAL ABNORMALITIES

Patients with X-ALD are deficient in VLCFA metabolism. The biochemical defect is at an early step in peroxisomal  $\beta$ -oxidation in which the free fatty acid form of VLCFA is activated to its acyl-CoA ester. This reaction is catalyzed by a peroxisome-specific very-long-chain acyl-CoA synthetase but, as noted above, it is dependent on functional ALDP, which is missing in X-ALD. As a consequence, VLCFAs are diverted through an alternate pathway, where they are acted on by an analogous synthetase enzyme in the endoplasmic reticulum that provides VLCFA-CoA substrates for biosynthesis of other lipids, including phospholipids, cholesterol esters, and sphingolipids. Fatty acid analyses of the tissue lipids of X-ALD patients reveal abnormal elevations in saturated VLCFAs that are longer than 22 carbons. In the adrenal cortex and brain, cholesterol esters containing saturated VLCFAs (C24:0 and C26:0) are almost undetectable in normal controls, whereas in CCALD they may account for over 50% of the fatty acid composition. VLCFA elevations are seen in other

lipid fractions as well, but to a lesser extent. In plasma from X-ALD patients, C26:0 and C24:0 levels are increased by two- to sixfold and 50%, respectively.

## PATHOGENESIS

The pathogenesis of endocrine and neurologic disease in X-ALD is complex and poorly understood. Pathologic mechanisms must account for the extraordinarily diverse clinical phenotypes and the intrafamilial variation in symptoms. Nevertheless, an excess of saturated VLCFAs is the cardinal biochemical feature of X-ALD. Although the extent of lipid abnormality and the types of VLCFA-containing lipids vary from one tissue to another, the vulnerable organs in X-ALD are those that show the greatest accumulation of VLCFAs in cholesterol esters. In electron micrographs, these cholesterol esters are manifest as lamellar lipid inclusions in adrenal cortical cells, Leydig cells, Schwann cells, and brain macrophages.

The major neuropathologic differences between cerebral forms of X-ALD and pure AMN are the active cerebral inflammatory demyelination observed in the former and the prominent axonal degeneration observed in the latter. The brain in CCALD exhibits decreased myelinated axons and oligodendrocytes, reactive astrocytosis, sudanophilic lipid deposits in macrophages, and a striking perivascular lymphocytic infiltration. Both T and B cell lymphocytes are present. Inflammatory mediators, such as tumor necrosis factor  $\alpha$  and interleukin-1, are produced in regions of active demyelination. Lipid analysis of autopsy brain shows that the fatty acid abnormality precedes demyelination, but the antigenic determinants have not been identified. VLCFA accumulation may initiate in phosphatidylcholine and later spread to other lipids as demyelination proceeds. Actively demyelinating lesions show the highest level of cholesterol esters containing VLCFAs, but these are unlikely to elicit an antigenic response. In addition, myelin proteins have been shown to be acylated with VLCFAs. Irrespective of the mechanism, the inflammatory demyelinating lesions in the brain of ALD are not seen in other lipidoses associated with white matter disease.

In contrast to the cerebral forms of X-ALD, AMN seems to lack an immune component to the neurologic disease. In AMN, the spinal cord is mainly involved with loss of myelinated axons in the ascending and descending tracts. The peripheral nerve is usually less severely affected. The mechanism for axonal degeneration is unclear.

In all forms of X-ALD, VLCFA accumulation in adrenocortical cells is thought to be solely responsible for the adrenal gland dysfunction, but more than one pathogenic mechanism may be in play. Addition of saturated VLCFAs to cultured adrenocortical cells has been shown to blunt adrenocorticotrophic hormone (ACTH) receptor responsiveness, probably by increasing plasma membrane microviscosity. Cholesterol esters containing VLCFAs, which accumulate in the adrenal gland, are poor substrates for hydrolyzing enzymes to provide free cholesterol precursors for steroidogenesis. In time, adrenocortical cells become ballooned and striated from lipid accumulation and then die. In contrast to autoimmune forms of Addison's disease, anti-adrenal antibodies are missing in X-ALD. Whether analogous pathogenic mechanisms occur in Leydig cells and are responsible for testicular dysfunction is not known.

## DIAGNOSIS

The diagnosis of X-ALD may be challenging because of its rare occurrence and striking clinical variation. Identification of X-ALD is often delayed until symptoms involving the nervous system and adrenal gland appear together, a condition that is lacking in a significant proportion of patients. AMN males and symptomatic heterozygote females are frequently misdiagnosed as having multiple sclerosis or familial spastic paraparesis.

Because neurologic symptoms are missing or can develop many years after the onset of adrenal disease, X-ALD should be suspected in any male with isolated primary adrenal insufficiency. In adults, adrenal insufficiency is often caused by other etiologies, but a significant proportion of boys who present with Addison's disease in childhood have X-ALD. In those who initially present with neurologic symptoms, however, adrenal disease might be suspected only after a diagnostic work-up for leukodystrophy has been initiated. In these patients, a 60 min ACTH stimulation test will usually detect adrenal involvement. Measurement of a morning serum cortisol concentration alone is not an adequate screening test, because basal cortisol levels may be maintained in the face of diminished adrenal reserve. In contrast, plasma ACTH is usually elevated before hypocortisolemia develops.

Owing to its wide clinical variation, the diagnosis of X-ALD is critically dependent on laboratory confirmation. The most convenient test is measurement of VLCFAs in plasma, which is noninvasive, is relatively inexpensive, and detects abnormal elevations of

VLCFAs in patients long before symptoms develop, even at the time of birth. Affected patients accumulate C26:0 and C24:0, but have normal C22:0. The C26:0/C22:0 and C24:0/C22:0 ratios are particularly useful for discriminating X-ALD patients from non-ALD controls. In males, there are very few conditions, such as a ketogenic diet, that give rise to false-positive results and false-negative results are uncommon. In circumstances where the plasma VLCFA results are equivocal, fatty acids can be measured in cultured skin fibroblasts grown from a patient. Female carriers for X-ALD tend to show intermediate elevations in plasma VLCFAs, but interpretation of a normal test result is problematic because plasma VLCFAs are elevated in only 85% of carriers. A normal test result, therefore, does not completely eliminate the possibility that the female is carrying the X-ALD gene. DNA testing is more reliable for heterozygote detection if the mutation is known in the family, but the lack of a common mutation in the X-ALD gene hampers the development of simple DNA screening tests for routine diagnosis.

The diagnosis of X-ALD has profound implications for genetic counseling, disease prevention, and potential treatment. Family studies indicate that only approximately 5% of the X-ALD probands represent new mutations. It is therefore important to screen at-risk family members by measuring plasma VLCFA or by DNA analysis. This frequently leads to the identification of asymptomatic males who should be counseled and monitored for the appearance of symptoms. It is not possible to predict which X-ALD phenotype they may develop.

Prenatal diagnosis affords the ability to prevent X-ALD. Affected fetuses can be identified by DNA analysis or by measuring VLCFA content of chorionic villi cells obtained at 8.5 to 10 weeks gestation and in amniocytes obtained during the second trimester.

## MANAGEMENT

The adrenal cortical insufficiency in X-ALD is easily treated with hormone replacement, but it has no effect on the progression of neurologic disease. Hydrocortisone is typically used alone or in combination with fludrocortisone, depending on the severity of mineralocorticoid deficiency. The glucocorticoid dosage should be increased in response to stressful occasions. Patients lacking adrenal insufficiency at the time of diagnosis should be monitored yearly by measuring plasma ACTH in combination with a morning serum cortisol. Elevations in ACTH level

precede a decline in cortisol to alert the physician of impending adrenal disease. As a practical matter, patients are often treated with replacement doses of glucocorticoids once diminished adrenal reserve is discovered. No therapy has been found to prevent or reverse the adrenal insufficiency.

Therapeutic options for the serious neurologic symptoms of X-ALD have emerged over the past decade. Bone marrow transplantation (BMT) has been shown to stabilize, or even reverse, the neurologic disease in CCALD. To receive clinical benefit, however, patients need to be transplanted at the earliest sign of neurologic involvement when only cognitive symptoms are predominant. Even then, all patients do not respond. Given the risks of BMT, the procedure is not recommended for asymptomatic or AMN patients.

Other therapeutic approaches have not proven useful, including attempts to modify the inflammatory reaction in the brain with immunosuppressive drugs, by intravenous administration of gamma globulin or  $\beta$ -interferon. Dietary restriction of saturated VLCFAs has no effect on the disease. Dietary supplementation with monounsaturated fatty acids (Lorenzo's oil) normalizes plasma C26:0 by inhibiting saturated VLCFA synthesis, but has no significant clinical impact once neurologic symptoms develop, probably because fatty acids in the brain are not altered. Studies are ongoing to determine whether Lorenzo's oil can delay the onset of neurologic or adrenal disease in presymptomatic boys.

Pharmacologic approaches to increase the peroxisomal VLCFA-oxidizing activity using 4-phenylbutyrate or statin drugs are under investigation. These drugs act to stimulate VLCFA oxidation by up-regulating peroxisomal ALDR protein, which forms functional ABC protein heterodimers in place of the missing ALDP protein. In this regard, the availability of mouse X-ALD gene knockout animals is essential for future therapeutic investigations.

## See Also the Following Articles

ACTH (Adrenocorticotrophic Hormone) • Adrenal Insufficiency

## Further Reading

- Aubourg, P. (1995). X-Linked adrenoleukodystrophy. In "Handbook of Clinical Neurology: Neurodystrophies and Neurolipidoses" (H. W. Moser, ed.), Vol. 22, No. 66, pp. 447-483. Elsevier, Amsterdam, the Netherlands.
- Kemp, S., Wei, H. M., Lu, J. F., Braiterman, L. T., McGuinness, M. C., Moser, A. B., Watkins, P. A., and Smith, K. D. (1998).

- Gene redundancy and pharmacologic gene therapy: Implications for X-linked adrenoleukodystrophy. *Nat. Med.* **4**, 1261–1268.
- Moser, H. W., Smith, K. D., Watkins, P. A., Powers, J., and Moser, A. B. (2001). X-Linked adrenoleukodystrophy. In "The Metabolic and Molecular Bases of Inherited Disease" (C. R. Scriver, A. L. Beaudet, D. Valle, and W. S. Sly, eds.), 8th ed., pp. 3257–3301. McGraw-Hill, New York.
- Smith, K. D., Kemp, S., Braiterman, L. T., Lu, J.-F., Wei, H.-M., Geraghty, M., Stetten, G., Bergin, J. S., Pevsner, J., and Watkins, P. A. (1999). X-Linked adrenoleukodystrophy: Genes, mutations, and phenotypes. *Neurochem. Res.* **24**, 521–535.
- Van Geel, B. M., Bezman, L., Loes, D. J., Moser, H. W., and Raymond, G. V. (2001). Evolution of phenotypes in adult male patients with X-linked adrenoleukodystrophy. *Ann. Neurol.* **49**, 186–194.



## Zollinger-Ellison Syndrome

see Gastrinomas